TESIS DE LA UNIVERSIDAD DE ZARAGOZA

2012

14

Mariano Llamedo Soria

Signal processing for automatic heartbeat classification and patient adaptation in the electrocardiogram

Departamento
Instituto de Investigación en Ingeniería [I3A]

Director/es
Martínez Cortés, Juan Pablo

ISSN 2254-7606

Prensas de la Universidad Universidad Zaragoza

Universidad Zaragoza
1542

Tesis Doctoral

SIGNAL PROCESSING FOR AUTOMATIC HEARTBEAT CLASSIFICATION AND PATIENT ADAPTATION IN THE ELECTROCARDIOGRAM

Autor

Mariano Llamedo Soria

Director/es

Martínez Cortés, Juan Pablo

UNIVERSIDAD DE ZARAGOZA

Instituto de Investigación en Ingeniería [I3A]

2012

Instituto Universitario de Investigación
en Ingeniería de Aragón
Universidad Zaragoza

PhD Thesis

Signal Processing for Automatic Heartbeat Classification and Patient Adaptation in the Electrocardiogram

Mariano Llamedo Soria

Advisor

Juan Pablo Martínez Cortés, PhD

Zaragoza, June 2012

Abstract

Cardiovascular diseases are currently the biggest single cause of death in developed countries, so the development of better diagnostic methodologies could improve the health of many people. Arrhythmias are related to the sudden cardiac death, one of the challenges for the modern cardiology. On the other hand, the classification of heartbeats on the electrocardiogram (ECG) is an important analysis previous to the study of arrhythmias. The automation of heartbeat classification could improve the diagnostic quality of arrhythmias, specially in Holter or long-term recordings. The objective of this thesis is the study of the methodologies for the classification of heartbeats on the ECG.

First we developed and validated a simple heartbeat classifier based on features selected with the focus on an improved generalization capability. We considered features from the RR interval (distance between two consecutive heartbeats) series, as well as features computed from the ECG samples and from scales of the wavelet transform, at both available leads. The classification performance and generalization were studied using publicly available databases: the MIT-BIH Arrhythmia, the MIT-BIH Supraventricular Arrhythmia and the St. Petersburg Institute of Cardiological Technics (INCART) databases. The Association for the Advancement of Medical Instrumentation (AAMI) recommendations for class labeling and results presentation were followed. A floating feature selection algorithm was used to obtain the best performing and generalizing models in the training and validation sets for different search configurations. The best model found comprehends 8 features, was trained in a partition of the MIT-BIH Arrhythmia, and was evaluated in a completely disjoint partition of the same database. The results obtained were: global accuracy (A) of 93%; for normal beats, sensitivity (S) 95%, positive predictive value (P^+) 98%; for supraventricular beats, S 77%, P^+ 39%; for ventricular beats S 81%, P^+ 87%. In order to test the generalization capability, performance was also evaluated in the INCART, with results comparable to those obtained in the test set. This classifier model has fewer features and performs better than other state of the art methods with results suggesting better generalization capability.

With an automatic classifier developed and validated, we evaluated two improvements. One, to adapt the classifier to ECG recordings of an arbitrary number of leads, or multilead extension. The second improvement was to improve the classifier with a nonlinear multilayer perceptron (MLP). For the multilead extension, we studied the improvement in heartbeat classification achieved by including information from multilead ECG record-

ings in the previously developed and validated classification model. This model includes features from the RR interval series and morphology descriptors for each lead calculated from the wavelet transform. The experiments were carried out in the INCART database, available in Physionet, and the generalization was corroborated in private and public databases. In all databases the AAMI recommendations for class labeling and results presentation were followed. Different strategies to integrate the additional information available in the 12-leads were studied. The best performing strategy consisted in performing principal components analysis to the wavelet transform of the available ECG leads. The performance indices obtained for normal beats were: S 98%, P^+ 93%; for supraventricular beats, S 86%, P^+ 91%; and for ventricular beats S 90%, P^+ 90%. The generalization capability of the chosen strategy was confirmed by applying the classifier to other databases with different number of leads with comparable results. In conclusion, the performance of the reference two-lead classifier was improved by taking into account additional information from the 12-leads. The improvement of the linear classifier classifier by means of a MLP was developed with a methodology similar to the one presented above. The results obtained were: A of 89%; for normal beats, S 90%, P^+ 99%; for supraventricular beats, S 83%, P^+ 34%; for ventricular beats S 87%, P^+ 76%.

Finally we studied an algorithm based on the methodologies previously described, but able to improve its performance by means of expert assistance. We presented a patient-adaptable algorithm for ECG heartbeat classification, based on a previously developed automatic classifier and a clustering algorithm. Both classifier and clustering algorithms include features from the RR interval series and morphology descriptors calculated from the wavelet transform. Integrating the decisions of both classifiers, the presented algorithm can work either automatically or with several degrees of assistance. The algorithm was comprehensively evaluated in several ECG databases for comparison purposes. Even in the fully automatic mode, the algorithm slightly improved the performance figures of the original automatic classifier; just with less than 2 manually annotated heartbeats (MAHB) per recording, the algorithm obtained a mean improvement for all databases of 6.9% in A, of 6.5% in S and of 8.9% in P^+. An assistance of just 12 MAHB per recording resulted in a mean improvement of 13.1% in A, of 13.9% in S and of 36.1% in P^+. For the assisted mode the algorithm outperformed other state-of-the-art classifiers with less expert annotation effort. The results presented in this thesis represent an improvement in the field of automatic and patient-adaptable heartbeats classification on the ECG.

Resumen

Las enfermedades cardiovasculares son en la actualidad la mayor causa de muerte individual en los países desarrollados, por lo tanto cualquier avance en las metodologías para el diagnóstico podrían mejorar la salud de muchas personas. Dentro de las enfermedades cardiovasculares, la muerte súbita cardíaca es una de las causas de muerte más importantes, por su número y por el impacto social que provoca. Sin lugar a duda se trata uno de los grandes desafíos de la cardiología moderna. Hay evidencias para relacionar las arritmias con la muerte súbita cardíaca. Por otro lado, la clasificación de latidos en el electrocardiograma (ECG) es un análisis previo para el estudio de las arritmias. El análisis del ECG proporciona una técnica no invasiva para el estudio de la actividad del corazón en sus distintas condiciones. Particularmente los algoritmos automáticos de clasificación se focalizan en el análisis del ritmo y la morfología del ECG, y específicamente en las variaciones respecto a la normalidad. Justamente, las variaciones en el ritmo, regularidad, lugar de origen y forma de conducción de los impulsos cardíacos, se denominan arritmias. Mientras que algunas arritmias representan una amenaza inminente (Ej. fibrilación ventricular), existen otras más sutiles que pueden ser una amenaza a largo plazo sin el tratamiento adecuado. Es en estos últimos casos, que registros ECG de larga duración requieren una inspección cuidadosa, donde los algoritmos automáticos de clasificación representan una ayuda significativa en el diagnóstico.

En la última década se han desarrollado algunos algoritmos de clasificación de ECG, pero solo unos pocos tienen metodologías y resultados comparables, a pesar de las recomendaciones de la AAMI para facilitar la resolución de estos problemas. De dichos métodos, algunos funcionan de manera completamente automática, mientras que otros pueden aprovechar la asistencia de un experto para mejorar su desempeño. La base de datos utilizada en todos estos trabajos ha sido la MIT-BIH de arritmias. En cuanto a las características utilizadas, los intervalos RR fueron usados por casi todos los grupos. También se utilizaron muestras del complejo QRS diezmado, o transformado mediante polinomios de Hermite, transformada de Fourier o la descomposición wavelet. Otros grupos usaron características que integran la información presente en ambas derivaciones, como el máximo del vectocardiograma del complejo QRS, o el ángulo formado en dicho punto.

El objetivo de esta tesis ha sido estudiar algunas metodologías para la clasificación de latidos en el ECG. En primer lugar se estudiaron metodologías automáticas, con capacidad

para contemplar el análisis de un número arbitrario de derivaciones. Luego se estudió la adaptación al paciente y la posibilidad de incorporar la asistencia de un experto para mejorar el rendimiento del clasificador automático.

En principio se desarrolló y validó un clasificador de latidos sencillo, que utiliza características seleccionadas en base a una buena capacidad de generalización. Se han considerado características de la serie de intervalos RR (distancia entre dos latidos consecutivos), como también otras calculadas a partir de ambas derivaciones de la señal de ECG, y escalas de su transformada wavelet. Tanto el desempeño en la clasificación como la capacidad de generalización han sido evaluados en bases de datos públicas: la MIT-BIH de arritmias, la MIT-BIH de arritmias supraventriculares y la del Instituto de Técnicas Cardiológicas de San Petersburgo (INCART). Se han seguido las recomendaciones de la Asociación para el Avance de la Instrumentación Médica (AAMI) tanto para el etiquetado de clases como para la presentación de los resultados. Para la búsqueda de características se adoptó un algoritmo de búsqueda secuencial flotante, utilizando diferentes criterios de búsqueda, para luego elegir el modelo con mejor rendimiento y capacidad de generalización en los sets de entrenamiento y validación. El mejor modelo encontrado incluye 8 características y ha sido entrenado y evaluado en particiones disjuntas de la MIT-BIH de arritmias. Todas las carácterísticas del modelo corresponden a mediciones de intervalos temporales. Esto puede explicarse debido a que los registros utilizados en los experimentos no siempre contienen las mismas derivaciones, y por lo tanto la capacidad de clasificación de aquellas características basadas en amplitudes se ve seriamente disminuida. Las primeras 4 características del modelo están claramente relacionadas a la evolución del ritmo cardíaco, mientras que las otras cuatro pueden interpretarse como mediciones alternativas de la anchura del complejo QRS, y por lo tanto morfológicas. Como resultado, el modelo obtenido tiene la ventaja evidente de un menor tamaño, lo que redunda tanto en un ahorro computacional como en una mejor estimación de los parámetros del modelo durante el entrenamiento. Como ventaja adicional, este modelo depende exclusivamente de la detección de cada latido, haciendo este clasificador especialmente útil en aquellos casos donde la delineación de las ondas del ECG no puede realizarse de manera confiable. Los resultados obtenidos en el set de evaluación han sido: exactitud global (A) de 93 %; para latidos normales, sensibilidad (S) 95 %, valor predictivo positivo (P^+) 98 %; para latidos supraventriculares, S 77 %, P^+ 39 %; para latidos ventriculares S 81 %, P^+ 87 %. Para comprobar la capacidad de generalización, se evaluó el rendimiento en la INCART obteniéndose resultados comparables a los del set de evaluación. El modelo de clasificación obtenido utiliza menos características, y adicionalmente presentó mejor rendimiento y capacidad de generalización que otros representativos del estado del arte.

Luego se han estudiado dos mejoras para el clasificador desarrollado en el párrafo anterior. La primera fue adaptarlo a registros ECG de un número arbitrario de derivaciones, o extensión multiderivacional. En la segunda mejora se buscó cambiar el clasificador lineal por un perceptrón multicapa no lineal (MLP). Para la extensión multiderivacional

se estudió si conlleva alguna mejora incluir información del ECG multiderivacional en el modelo previamente validado. Dicho modelo incluye características calculadas de la serie de intervalos RR y descriptores morfológicos calculados en la transformada wavelet de cada derivación. Los experimentos se han realizado en la INCART, disponible en Physionet, mientras que la generalización se corroboró en otras bases de datos públicas y privadas. En todas las bases de datos se siguieron las recomendaciones de la AAMI para el etiquetado de clases y presentación de resultados. Se estudiaron varias estrategias para incorporar la información adicional presente en registros de 12 derivaciones. La mejor estrategia consistió en realizar el análisis de componentes principales a la transformada wavelet del ECG. El rendimiento obtenido con dicha estrategia fue para latidos normales: S 98 %, P^+ 93 %; para latidos supraventriculares, S 86 %, P^+ 91 %; y para latidos ventriculares S 90 %, P^+ 90 %. La capacidad de generalización de esta estrategia se comprobó tras evaluarla en otras bases de datos, con diferentes cantidades de derivaciones, obteniendo resultados comparables. En conclusión, se mejoró el rendimiento del clasificador de referencia tras incluir la información disponible en todas las derivaciones disponibles. La mejora del clasificador lineal por medio de un MLP se realizó siguiendo una metodología similar a la descrita más arriba. El rendimiento obtenido fue: A 89 %; para latidos normales: S 90 %, P^+ 99 %; para latidos supraventriculares, S 83 %, P^+ 34 %; y para latidos ventriculares S 87 %, P^+ 76 %.

Finalmente estudiamos un algoritmo de clasificación basado en las metodologías descritas en los anteriores párrafos, pero con la capacidad de mejorar su rendimiento mediante la ayuda de un experto. Se presentó un algoritmo de clasificación de latidos en el ECG adaptable al paciente, basado en el clasificador automático previamente desarrollado y un algoritmo de clustering. Tanto el clasificador automático, como el algoritmo de clustering utilizan características calculadas de la serie de intervalos RR y descriptores de morfología calculados de la transformada wavelet. Integrando las decisiones de ambos clasificadores, este algoritmo puede desempeñarse automáticamente o con varios grados de asistencia. El algoritmo ha sido minuciosamente evaluado en varias bases de datos para facilitar la comparación. Aún en el modo completamente automático, el algoritmo mejora el rendimiento del clasificador automático original; y con menos de 2 latidos anotados manualmente (MAHB) por registro, el algoritmo obtuvo una mejora media para todas las bases de datos del 6.9 % en A, de 6,5 % S y de 8,9 % en P^+. Con una asistencia de solo 12 MAHB por registro resultó en una mejora media de 13,1 % en A , de 13,9 % en S y de 36,1 % en P^+. En el modo asistido, el algoritmo obtuvo un rendimiento superior a otros representativos del estado del arte, con menor asistencia por parte del experto.

Como conclusiones de la tesis, debemos enfatizar la etapa del diseño y análisis minucioso de las características a utilizar. Esta etapa está íntimamente ligada al conocimiento del problema a resolver. Por otro lado, la selección de un subset de características ha resultado muy ventajosa desde el punto de la eficiencia computacional y la capacidad de generalización del modelo obtenido. En último lugar, la utilización de un clasificador

simple o de baja capacidad (por ejemplo funciones discriminantes lineales) asegurará que el modelo de características sea responsable en mayor parte del rendimiento global del sistema.

Con respecto a los sets de datos para la realización de los experimentos, es fundamental contar con un elevado numero de sujetos. Es importante incidir en la importancia de contar con muchos sujetos, y no muchos registros de pocos sujetos, dada la gran variabilidad intersujeto observada. De esto se desprende la necesidad de evaluar la capacidad de generalización del sistema a sujetos no contemplados durante el entrenamiento o desarrollo. Por último resaltaremos la complejidad de comparar el rendimiento de clasificadores en problemas mal balanceados, es decir que las clases no se encuentras igualmente representadas. De las alternativas sugeridas en esta tesis probablemente la más recomendable sea la matriz de confusión, ya que brinda una visión completa del rendimiento del clasificador, a expensas de una alta redundancia.

Finalmente, luego de realizar comparaciones justas con otros trabajos representativos del estado actual de la técnica, concluimos que los resultados presentados en esta tesis representan una mejora en el campo de la clasificación de latidos automática y adaptada al paciente, en la señal de ECG.

Conclusiones

En esta sección se resumen las conclusiones extraídas a lo largo de los capítulos de la tesis. Comenzaremos enfatizando la importancia del diseño de las características y en consecuencia la comprensión del problema fisiológico. En nuestra experiencia, la comprensión pormenorizada del problema permitirá desarrollar características valiosas para la clasificación, y en consecuencia un clasificador con capacidad de generalización. En el momento de la escritura de esta tesis, estamos estudiando la aplicación de los clasificadores denominados *deep belief networks* (DBN) [Hinton et al., 2006], estando aún pendiente su implementación. Este tipo de clasificadores no solo han mejorado el estado de la técnica en otras áreas del reconocimiento de patrones, como el reconocimiento de la escritura y el habla, sino que lo han hecho utilizando directamente las muestras digitalizadas de una señal o los píxeles de una imagen. Simplemente han evitado la etapa del diseño del modelo de características. A pesar de que esto último se contrapone con nuestra primer conclusión, la utilidad de los DBN necesita aún ser corroborada en el campo de la clasificación de latidos. También es probable que otros modelos de características puedan desempeñarse mejor que sólo las muestras digitalizadas del ECG. De cualquier manera, nosotros creemos que los clasificadores del estilo *caja negra* (o cualquier otro no lineal o no paramétrico) no debería ser considerado como primer alternativa a la resolución de un problema de clasificación, sino hacerlo cuando se haya alcanzado un rendimiento de partida con un clasificador más simple.

La importancia de contar con un set de datos *grande* es determinante. En aplicaciones de clasificación de latidos, donde existe una gran variabilidad intersujeto, la definición de *grande* puede ser engañosa. En nuestra experiencia, es más importante contar con sets de datos de muchos sujetos, aunque de corta duración, que registros de larga duración de pocos sujetos, tal vez repetidos. Es necesario aclarar que la aplicación de clasificadores a registros de larga duración no ha sido estudiado minuciosamente en esta tesis, quedando pendiente para mejoras futuras. Este último aspecto refuerza la idea de evaluar un clasificador en tantos sets de datos como sea posible, para tener una mejor estimación de su rendimiento en un contexto real.

En los experimentos de selección de características hemos encontrado dos modelos, tras perseguir diversos criterios de optimización. En la Tabla 3.4 se muestra un modelo con buen rendimiento intersujeto. Como puede verse las características que incluye el modelo son íntegramente mediciones de intervalos. Esto puede explicarse debido a que las bases de datos usadas no incluyen siempre el mismo par de derivaciones de ECG en cada registro. Por lo tanto aquellas características que miden amplitudes se ven muy afectadas por esto. Las características direccionales (como el VCG_ϕ) probablemente también se vean afectadas, a pesar de su conocida utilidad para los cardiólogos [Taylor, 2002]. A diferencia

de estas, los intervalos parecen retener la capacidad de clasificación independientemente de las derivaciones donde se midan. Las primeras cuatro características del modelo están claramente relacionadas a la evolución del ritmo cardíaco, mientras que las otras cuatro podrían interpretarse como mediciones alternativas de la anchura del QRS, y por lo tanto una descripción morfológica del complejo. Estas características no necesitan una detección muy precisa del punto fiducial del complejo QRS, siendo muy adecuadas para registros ECG de mala calidad donde la detección y delineación automática de las ondas del ECG no es confiable o incluso no es posible.

Por otro lado en la Tabla 5.2, se muestra un modelo con buen rendimiento intrasujeto. El modelo incluye también características de ritmo y morfología. Respecto a las características de ritmo, el EMC utiliza adicionalmente P_{RR} y dRR_L, ambas relacionadas con la variación local del intervalo RR. Con respecto a la descripción morfológica, las características S^1_{QRS} y k^1_M podrían interpretarse como una medición alternativa y robusta de la anchura del intervalo QRS; mientras que $r_{\mathrm{QRST}}(k_M)$ describe la similaridad del complejo QRST entre las derivaciones PCA en la escala 3 de la DWT. Esta última medida puede relacionarse con cambios morfológicos y del eje de depolarización del complejo QRST.

Las funciones discriminantes lineales determinadas por el LDC-C han demostrado su utilidad para desarrollar un clasificador con capacidad de generalización. Esto puede explicarse debido a que una función de decisión conservativa, como un hiperplano, es más apropiado para problemas de clasificación complicados o con una gran variabilidad intersujeto. En este tipo de problemas, casi ninguna de las hipótesis impuestas por nuestras decisiones de diseño se cumplen completamente. Sólo para clarificar esto último, según el enfoque propuesto de clasificación automática, nuestro set de entrenamiento debería ser una muestra representativa del universo completo de latidos. Esto no sólo no es factible, sino que podemos afirmar que nuestro set de entrenamiento es distinto a nuestro set de evaluación, tan solo comparando las diferencias de rendimiento entre las tablas 3.2 y 3.3. Con esta evidente limitación, es probable que el clasificador con más capacidad para modelar la información de entrenamiento, en nuestro caso el QDC, es más propenso a fallar más seguido en el set de evaluación. Esta razón probablemente haga que una decisión más conservativa, como el LDC, sea la mejor opción. En la Figura 2.12, las funciones discriminantes producidas por un LDC y un QDC pueden ser comparadas.

Cuando limitamos el problema a un sujeto a la vez, y perseguimos el mejor rendimiento intrasujeto, podemos permitir que el clasificador produzca funciones de decisión no lineales. En nuestro caso hemos usado un clasificador basado en mezcla de Gaussianas, que utiliza el mismo algoritmo EM utilizado para el clustering.

El esquema de selección de características usado resultó una metodología muy conveniente para la reducción de la complejidad del problema de clasificación, y al mismo tiempo para mejorar la capacidad de generalización del modelo obtenido. El algoritmo SFFS fue especialmente útil cuando se utilizaron clasificadores simples y determinísticos, como QDC o LDC, pero para el caso de los no determinísticos, como MLP o mezcla de

Gaussianas, se adoptaron algunas soluciones de compromiso dado que debíamos asegurar (o al menos limitar) la repetibilidad. Esto último debido a que el SFFS necesita reevaluar continuamente búsquedas previas, obteniendo diferentes resultados en el caso que no se asegure la repetibilidad.

La capacidad de generalización de un clasificador es en nuestra opinión, su característica más importante. En el Capítulo 5 mostramos que es posible realizar una evaluación minuciosa del rendimiento y capacidad de generalización de un clasificador exclusivamente en bases de datos públicas y de libre disponibilidad.

La estimación del rendimiento en problemas desbalanceados, como el estudiado en esta tesis, puede ser complicado especialmente cuando se comparan clasificadores. En esta tesis hemos explorado algunas metodologías para tratar con el problema del desbalance. Sin embargo, ninguna de las soluciones sugeridas en los Capítulos 3 y 4, como el cálculo balanceado del rendimiento, asegura la solución del problema. Por este motivo sugerimos siempre que fuera posible la incorporación de la matriz de confusión, ya que clarifica el rendimiento obtenido por un clasificador y asegura la comparabilidad de los resultados. Otro problema referido a la estimación del rendimiento, es cuando se comparan los resultados obtenidos en bases de datos con desbalances diferentes. Para facilitar la interpretación en estos casos, sugerimos una estimación optimísticamente sesgada del rendimiento que representa una cota superior de rendimiento en cada base de datos. De esta manera, se puede utilizar dicha cota como referencia.

Las comparaciones realizadas en los capítulos previos fueron hechas de manera justa de acuerdo a nuestro conocimiento. Los trabajos incluidos en nuestras comparaciones tienen metodologías comparables y son representativos del estado actual de la técnica. En general, como ya fue detallado en los capítulos anteriores, nuestros clasificadores se desempeñaron mejor. En todas las comparaciones realizadas, siempre hemos incluido una descripción detallada de nuestros resultados con la finalidad de facilitar futuras mejoras.

En resumen, los resultados presentados en esta tesis constituyen una mejora en el rendimiento con respecto a otros trabajos publicados y representativos del estado actual de la técnica en el campo de la clasificación automática y adaptada al paciente de latidos.

Contents

Chapter 1

Introduction

1.1 Motivation

The World Health Organization places cardiovascular diseases (CVD) as the first single cause of death globally in the present, and forecasts the same ranking up to 2030 [World Health Organization, 2012]. These diseases affect in a higher degree to low- and middle-income countries, but in the same proportion to women and men. Specifically in Argentina and Spain, more than 30% of the deaths are caused by CVD and is by far, the first single cause of death according to the official agencies [Dirección de Estadísticas e Información en Salud, 2012, Instituto Nacional de Estadística, 2012]. A great part of the deaths caused by CVD occur suddenly, starting with a ventricular fibrillation which leads to a cardiac arrest [Bayés de Luna, 2010]. This situation is known as sudden cardiac death (SCD) and is probably the most important challenge of the modern cardiology. This disease is unusual up to the age of 35, but from there the risk of SCD increases specially during the chronic and acute phases of myocardial infarction, or other cardiopathy related to heart failure. The identification or prediction of SCD has been studied more thoroughly for those risk groups with a previous cardiac condition (cardiac arrest, genetic defects, heart failure, heart attack) than for the people in which SCD is the first manifestation. The importance of the last group is that it represents more than the 50% of people who suffer SCD. However, up to the moment, an exhaustive screening of the population is unfeasible from the technical and economical point of view.

The improvement of cost-effective methodologies for the prediction of SCD received lot of attention from the scientific community in the last decades. It was studied in several works that arrhythmias are responsible of most of the cases of SCD [Bayés de Luna, 2010]. One important advance in the study of arrhythmias was the use of long-term (or Holter) recordings and the software to aid the cardiologist in the detection and diagnostic of abnormalities in the electrocardiogram (ECG). The study of arrhythmias by means of the computerized analysis of the ECG signal, is in the present a cost-effective and well established tool to analyze the heart function. The improvement of the methodologies used

in the study of arrhythmias is likely to aid cardiologists in the diagnostic and screening of SCD. In this thesis we developed and analyzed new algorithms for the classification of ECG heartbeats, which is an important analysis previous to the study of arrhythmias.

1.2 Background

As this thesis is entirely focused on the analysis of the ECG signal, a brief description of its origin is included, as well as the basic concepts of cardiac electrophysiology. We will start with a selection of anatomy and physiology concepts, to subsequently inspect some mechanisms at the cellular level and their manifestation on the ECG. Our objective in the following chapters will be the design of a computer algorithm capable of classifying the concepts explained in this section. This section is based on the books [Bayés de Luna, 2010, Natale and Wazni, 2007, Guyton and Hall, 2006, Sörnmo and Laguna, 2005, Malmivuo and Plonsey, 1995], where the reader is referred for further details and references.

1.2.1 The heart

The heart is an electromechanical pulsatile pump. From the anatomic point of view, as can be seen in Figure 1.1, there are two separate pumps: one at the right that pumps blood through the lungs, and one at the left that pumps blood through the peripheral organs. Each half includes a two-chamber pump composed of an atrium and a ventricle. The atrium pumps blood for the ventricle, and then the ventricles supply the main pumping force either through the pulmonary circulation, by the right ventricle, or through the peripheral circulation by the left ventricle. There are four valves to force the direction of the blood, as is shown in Figure 1.2, two located between the atria and the ventricles, and two between the ventricles and the arteries.

As a periodic electromechanical pump, an electrical impulse is responsible of the mechanical activation of the muscle. Each cycle is initiated by spontaneous generation of an action potential (AP) in the sinus (or sinoatrial in Figure 1.2) node. This node is located in the superior lateral wall of the right atrium near the opening of the superior vena cava. The impulse, or AP, travels through both atria reaching the atrio-ventricular (A-V) bundle, where is delayed about 0.1 seconds. This delay allows the atria to pump blood into the ventricles. After this, the ventricles are filled and ready to be activated. This is done by a special conduction system (SCS), the right and left bundle branches of Purkinje fibers. This system propagates the impulse from the A-V node to the whole ventricular muscle very fast, allowing a synchronized activation and consequently an effective pump of the blood. This cycle is repeated up to the death of the heart.

Now we will try to relate the electrical and mechanical behavior of the heart described above. The activation of the cardiac muscle composed of two phases, contraction and re-

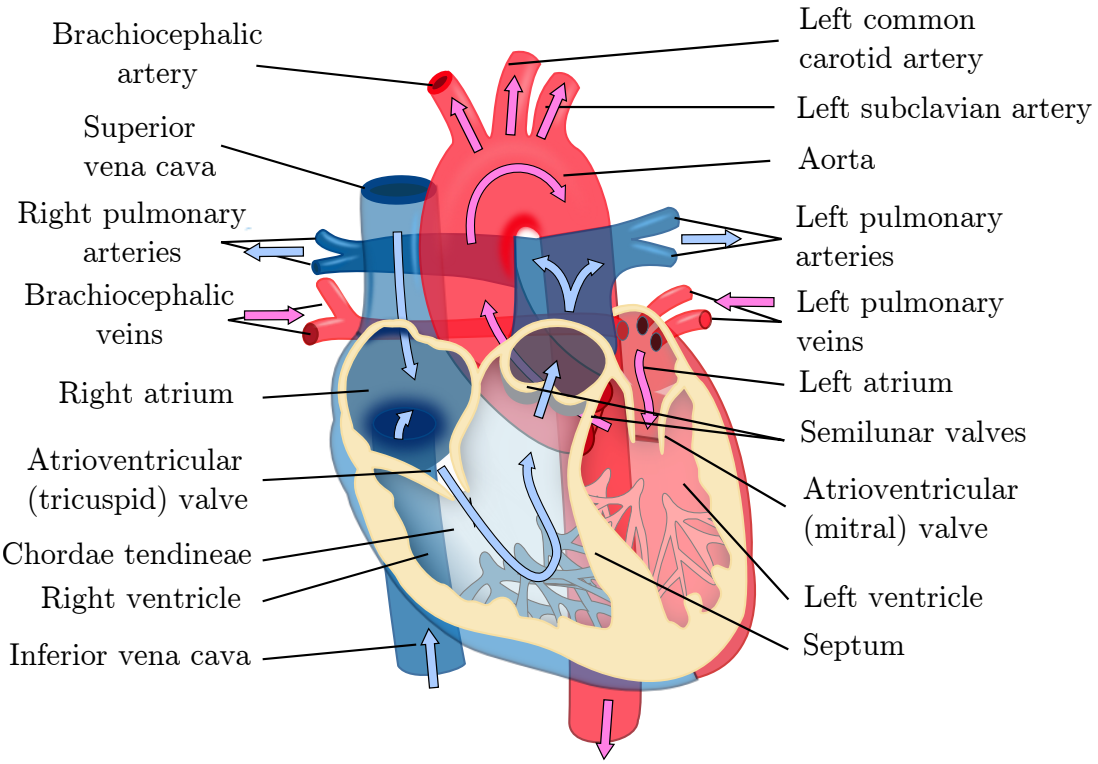


Figure 1.1: Structure of the heart, and course of blood flow through the heart chambers and heart valves. Diagrams based on image http://en.wikipedia ... -en.svg under license CS-BY-SA.

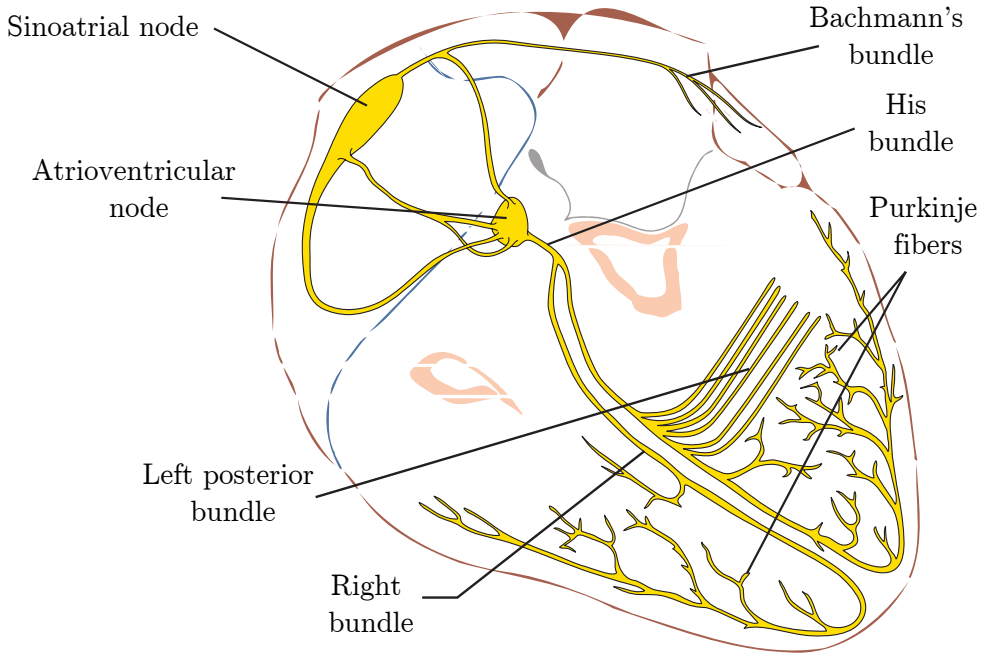


Figure 1.2: Course of the blood flow through the heart, and the electrical conduction system of the heart. Diagrams based on image http://commons.wikimedia ... Heart.svg under license CS-BY-SA.

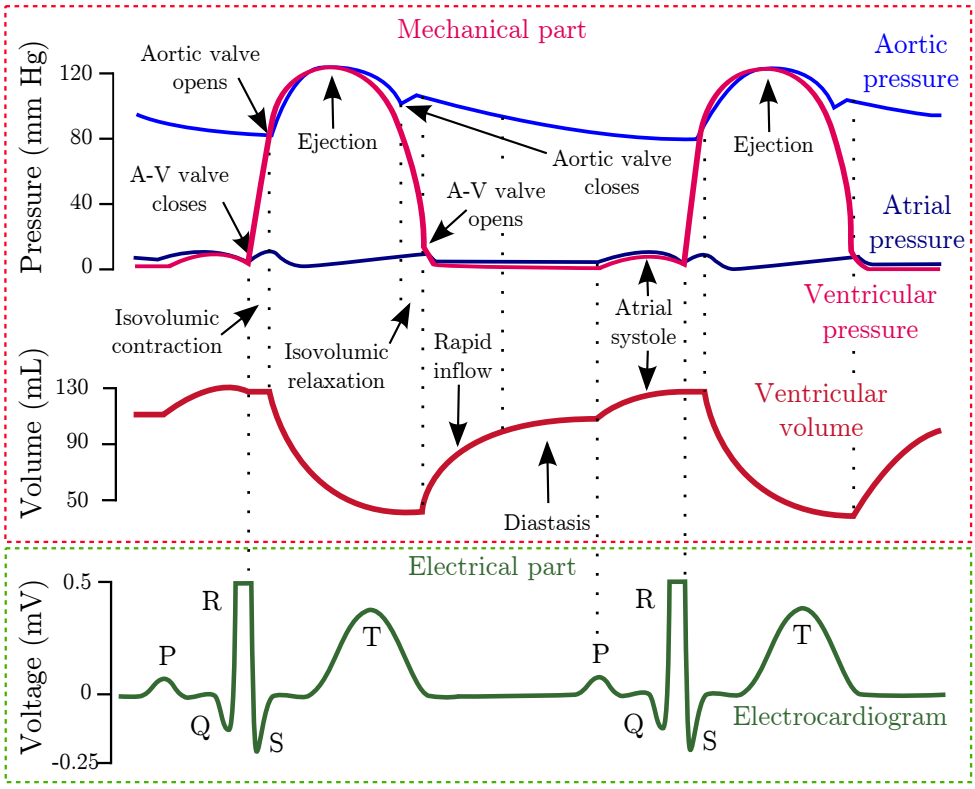


Figure 1.3: Wiggers diagram. Events of the cardiac cycle for left ventricular function, showing changes in left atrial pressure, left ventricular pressure, aortic pressure, ventricular volume, and the electrocardiogram.

laxation, or in electrical terms as depolarization and repolarization. As the heart function produces an electrical field, the voltage generated can be recorded by the electrocardiograph from the surface of the body. The first wave, called with the letter P, is caused by spread of depolarization through the atria. After the electrical activation, follows the atrial contraction which causes a slight rise in the atrial pressure. About 0.16 seconds after the onset of the P wave, the QRS waves appear as a result of electrical depolarization of the ventricles. This initiates the contraction of the ventricles and causes the ventricular pressure to begin rising. Finally, the ventricular T wave in the electrocardiogram represents the stage of repolarization of the ventricles when the ventricular muscle fibers begin to relax. As can be noted in Figure 1.3, the electrical depolarization is preceded by the corresponding mechanical contraction.

1.2.2 From the action potentials to the electrocardiogram

In general heart cells can be grouped in two types: the ones from the SCS and the contractile cells. The first are responsible of the generation of the electrical impulse (rhythmicity) and its conduction to the contractile cells, while the contractile cells are responsible of the pumping or mechanical function. Both cell types are responsible of the electromechanical link. In Figure 1.4 it is showed the waveforms of the voltage, or action potential, and currents measured in the cellular membrane of a contractile cell. Following the depolar-

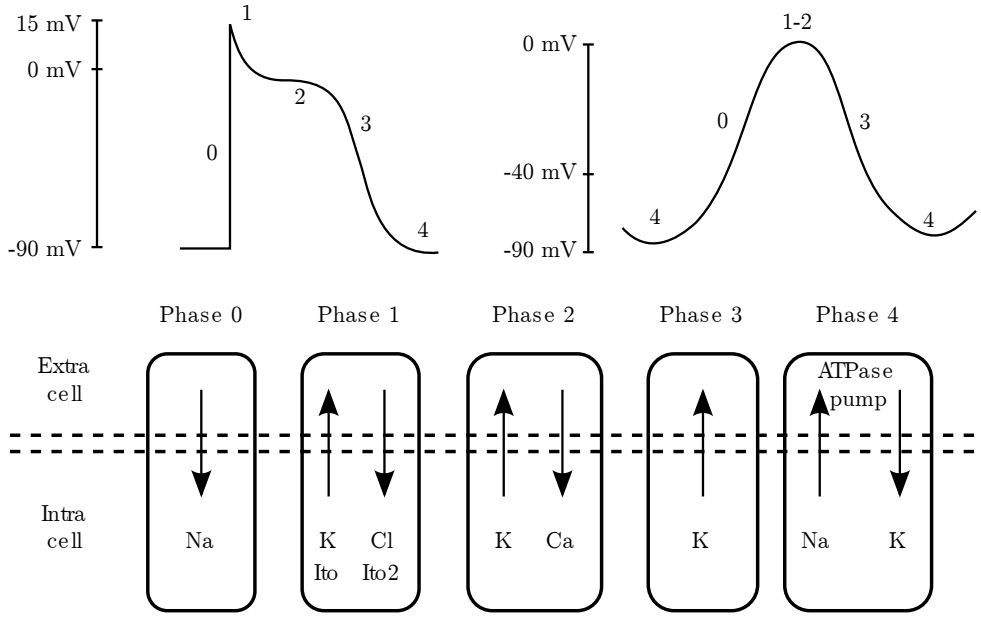


Figure 1.4: Reproduced from [Natale and Wazni, 2007]. Top panel: on left, the action potential in contractile cells, and on the right in SCS cell. Bottom panel: predominant currents during the different phases of Na-channel-dependent action potential.

ization phases in the same Figure, note that when a cell receives depolarizing current, Na channels are activated resulting in a net inward current manifested as phase 0 of the AP. Phase 1 starts with the opening of a rapid outward potassium current. Phase 2 or the plateau phase of the AP is the result of an L-type Ca current that counteracts the outward K currents. With time, L-type Ca channels are inactivated and the plateau subsides. At the same time, the increase in calcium concentration acts as a trigger for release of more Ca stored in the sarcoplasmic reticulum, which in turn provides a contraction signal to the myocyte contractile elements, producing the contraction of the cell. Phase 3 is due to 'delayed rectifier' outward K currents. Phase 4 constitutes a steady, stable, polarized membrane due to voltage-regulated inward rectifiers. Compared to atrial action potential, ventricular AP has a longer duration, a higher phase 2, a shorter phase 3, and more negative phase 4.

On the other hand, the SCS cells have the ability to generate a spontaneous action potential using T-type Ca and K rectifier currents. These currents confer the unstable electrical property of phase 4, causing these cells to develop rhythmic spontaneous slow diastolic depolarization. Once AP reaches –40 mV, L-type Ca channels are activated, generating the slow upstroke of the action potential in these types of cells (phase 0).

There are three types of SCS cells:

1. P cells, found mostly in the sinus node are responsible of automaticity.

2. The Purkinje cells, are found in the His bundle branches and are responsible of the fast transmission of electrical impulses through the ventricles.

3. The transitional cells, with slow conduction velocity, are typically found between

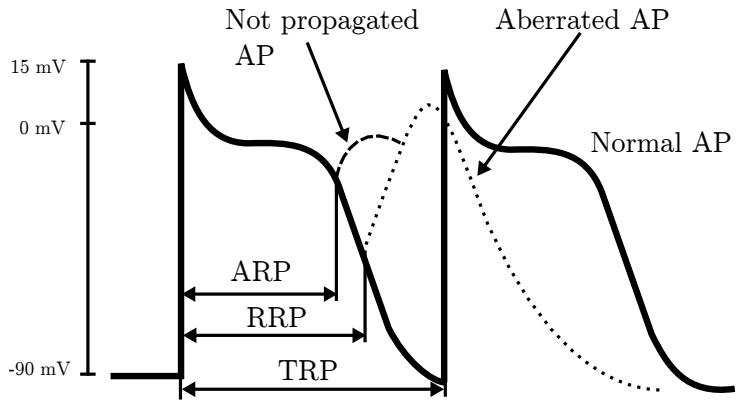


Figure 1.5: Based on Figure 2.20 from [Bayés de Luna, 2010]. Refractory period of ventricular cells. During absolute refractory period (ARP) depolarization is not possible. During the relative refractory period (RRP), an increased activation is necessary to depolarize the cell. After the total refractory period, the cell is able to produce a normal AP upon activation.

the P, Purkinje and contractile cells.

Once any cell is depolarized it takes certain time until it can be normally depolarized again. This time is known as total refractory period (TRP). Also there is a period of time where the cell can not be depolarized, and is known as absolute refractory period (ARP). If the time of arrival of a new activation is greater than ARP, the cell can produce an aberrated AP if the stimulus is big enough. This is known as relative refractory period (RRP). There is a small time window, between RRP and ARP in Figure 1.5, where the cell reacts to an increased activation, but the activation can not be propagated.

Automaticity is an intrinsic property of all myocardial cells. In addition to the sinus node, cells with pacemaking capability in the normal heart are located in some parts of the atria and ventricles. However, the occurrence of spontaneous activity is prevented by the natural hierarchy of pacemaker function, causing these sites to be latent or subsidiary pacemakers. The spontaneous discharge rate of the sinus node normally exceeds that of all other subsidiary pacemakers. Therefore, the impulse initiated by the sinus node depolarizes and keeps the activity of subsidiary pacemaker sites depressed before they can spontaneously reach threshold. However, slowly depolarizing and previously suppressed pacemakers in the atrium, A-V node, or ventricle can become active and assume pacemaker control of the cardiac rhythm if the sinus node pacemaker becomes slow or unable to generate an impulse (e.g., secondary to depressed sinus node automaticity) or if impulses generated by the sinus node are unable to activate the subsidiary pacemaker sites (e.g., sinoatrial exit block, or A-V block). The emergence of subsidiary or latent pacemakers under such circumstances is an appropriate fail-safe mechanism, which ensures that ventricular activation is maintained.

Once introduced the types of AP of the heart cells, it is possible to imagine that the electrical field which produces the ECG in the body surface, results from the integration

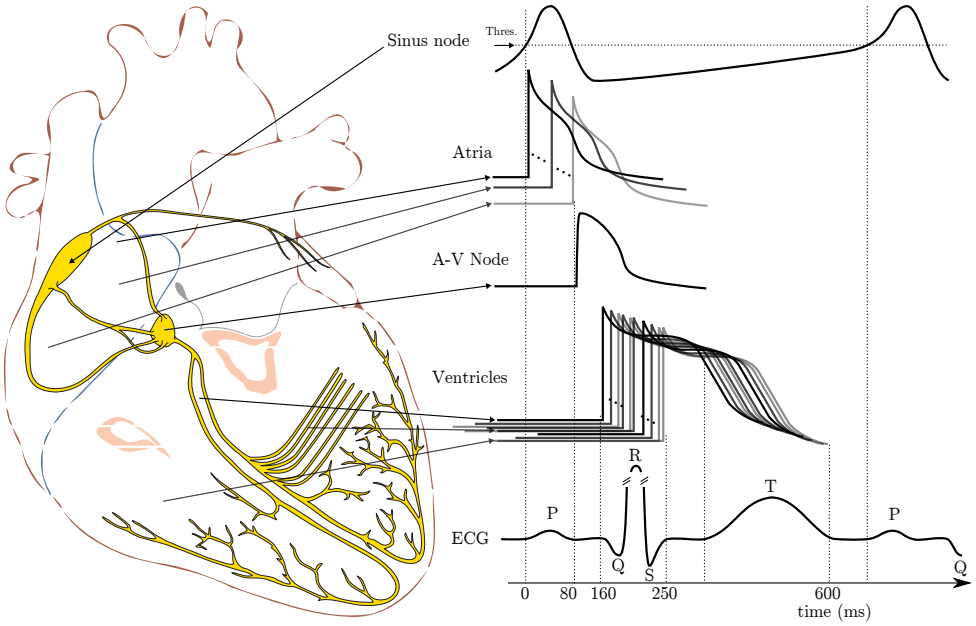


Figure 1.6: The morphology and timing of the action potentials from different regions of the heart and the related cardiac cycle of the ECG as measured on the body surface. Based on Figure 6.2 from [Sörnmo and Laguna, 2005]. Diagrams based on image http://commons.wikimedia ... Heart.svg under license CS-BY-SA.

of the AP of all cells in the heart during a heart cycle. As can be seen in Figure 1.6, the integration of all AP in the atria results in the formation of the P wave of the ECG. The same happens with the ventricles, but in this case the greater amount of mass, and therefore of cells and energy involved, results in a larger ECG amplitude. The tails or terminal parts of the AP, phases 2, 3 and 4 of Figure 1.4, are the responsible of the repolarization waves. Note that in the ECG only the repolarization of the ventricles is visible, and is known as T wave. However, the repolarization of the atria exists, but it is buried by the depolarization of the ventricles. The heart cycle repeats again, thanks to the rhythmic property of the sinus node cells.

Now we will add some details to the cyclic activation mechanism. The cells that constitute the ventricular myocardium are coupled together by gap junctions which, for the normal healthy heart, have a very low resistance. As a consequence, activity in one cell is readily propagated to neighboring cells. It is said that the heart behaves as a syncytium; a propagating wave once initiated continues to propagate uniformly into the region that is still at rest. The activation wavefronts proceed relatively uniformly, from endocardium to epicardium and from apex to base. One way of describing cardiac activation is to plot the sequence of instantaneous depolarization wavefronts. Since these surfaces connect all points in the same temporal phase, the wavefront surfaces are also referred to as isochrones. Such a description is contained in Figure 1.7. After the electric activation of the heart has begun at the sinus node, it spreads along the atrial walls. The resultant vector of the atrial electric activity is illustrated with a thick arrow. After the depolarization has propagated over the atrial walls, it reaches the AV node. The propagation

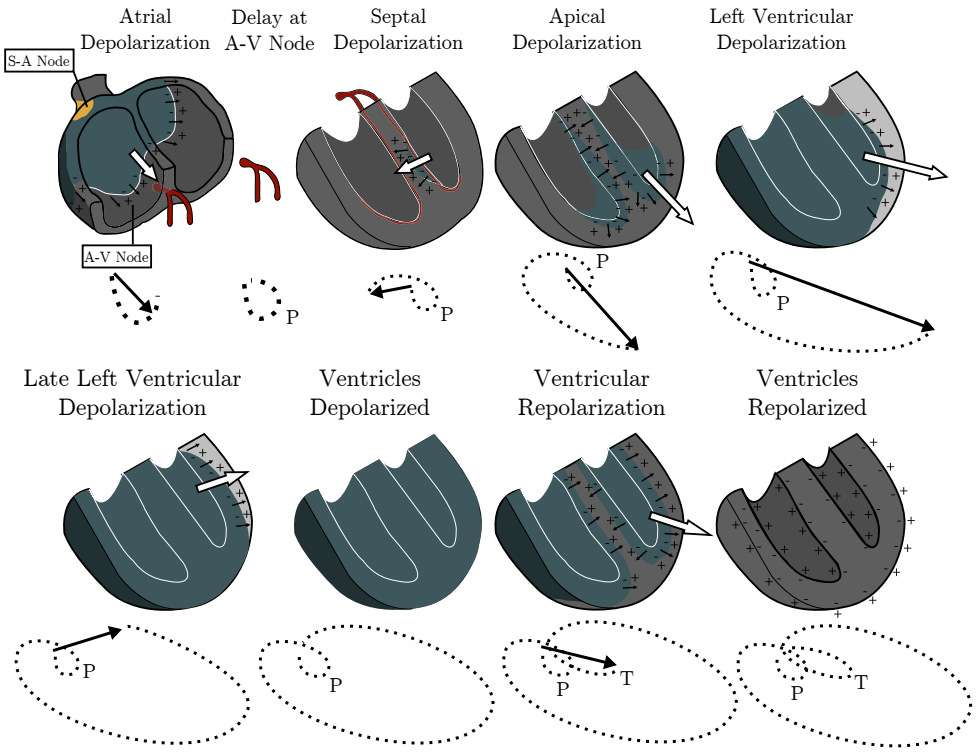


Figure 1.7: The normal sequence of ventricular depolarization. The instantaneous heart vector is shown at four times during the process: 10, 20, 40, and 60 milliseconds. From Massie and Walsh, 1960.

through the AV junction is very slow and involves negligible amount of tissue; it results in a delay in the progress of activation and allows the completion of ventricular filling. Once activation has reached the ventricles, propagation proceeds along the Purkinje fibers to the inner walls of the ventricles. The ventricular depolarization starts first from the left side of the interventricular septum, and therefore, the resultant dipole from this septal activation points to the right. In the next phase, depolarization waves occur on both sides of the septum, and their electric forces cancel. However, early apical activation is also occurring, so the resultant vector points to the apex.

After a while the depolarization front has propagated through the wall of the right ventricle; when it first arrives at the epicardial surface of the right-ventricular free wall, the event is called breakthrough. Because the left ventricular wall is thicker, activation of the left ventricular free wall continues even after depolarization of a large part of the right ventricle. Because there are no compensating electric forces on the right, the resultant vector reaches its maximum in this phase, and it points leftward. The depolarization front continues propagation along the left ventricular wall toward the back. Because its surface area now continuously decreases, the magnitude of the resultant vector also decreases until the whole ventricular muscle is depolarized. The last to depolarize are basal regions of both left and right ventricles. Because there is no longer a propagating activation front, there is no signal either. Ventricular repolarization begins from the outer side of the ventricles and the repolarization front *propagates* inward. This seems paradoxical,

but even though the epicardium is the last to depolarize, its action potential durations are relatively short, and it is the first to recover. Although recovery of one cell does not propagate to neighboring cells, one notices that recovery generally does move from the epicardium toward the endocardium. The inward spread of the repolarization front generates a signal with the same sign as the outward depolarization front, as pointed out in Figure 1.7 (recall that both direction of repolarization and orientation of dipole sources are opposite). Because of the diffuse form of the repolarization, the amplitude of the signal is much smaller than that of the depolarization wave and it lasts longer.

In the previous paragraph we described in detail the electrical activity inside the thorax, now we will focus on how this activity is recorded in the body surface. Augustus Désiré Waller measured the human electrocardiogram in 1887 using Lippmann's capillary electrometer [Waller, 1887]. He selected five electrode locations: the four extremities and the mouth. In this way, it became possible to achieve a sufficiently low contact impedance and thus to maximize the ECG signal. Furthermore, the electrode location is unmistakably defined and the attachment of electrodes facilitated at the limb positions. The five measurement points produce altogether 10 different leads. From these 10 possibilities he selected five designated cardinal leads. Two of these are identical to the Einthoven leads I and III described below. In 1908 Willem Einthoven published a description of the first clinically important ECG measuring system [Einthoven, 1908]. He used the capillary electrometer in his first ECG recordings. His essential contribution to ECG recording technology was the development and application of the string galvanometer, invented by Clément Ader. Its sensitivity greatly exceeded the previously used capillary electrometer. The Einthoven lead system is illustrated in Figure 1.8.

The Einthoven limb leads (standard leads) are defined in the following way:

$$V_{\mathrm{I}} = \Phi_L - \Phi_R$$

$$V_{\mathrm{II}} = \Phi_F - \Phi_R$$

$$V_{\mathrm{III}} = \Phi_F - \Phi_L,$$

where $V_{\mathrm{I,II,III}}$ are the voltages of leads I, II and III and $\Phi_{L,R,F}$ are potentials at the left and right arms and the left foot respectively. According to Kirchhoff's law these lead voltages have the following relationship:

$$V_{\mathrm{I}} + V_{\mathrm{III}} = V_{\mathrm{II}},$$

hence only two of these three leads are independent. The lead vectors associated with Einthoven's lead system are conventionally found based on the assumption that the heart is located in an infinite, homogeneous volume conductor (or at the center of a homogeneous sphere representing the torso). One can show that if the position of the right arm, left arm, and left leg are at the vertices of an equilateral triangle, having the heart located at

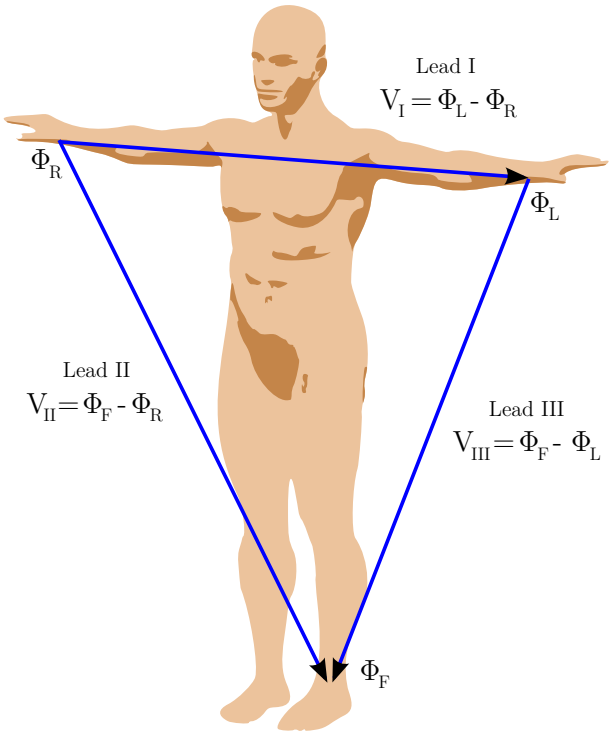


Figure 1.8: Einthoven limb leads and Einthoven triangle. The Einthoven triangle is an approximate description of the lead vectors associated with the limb leads. Diagrams based on image http://commons.wikimedia ... planes.svg under license CS-BY-SA.

its center, then the lead vectors also form an equilateral triangle. A simple model results from assuming that the cardiac sources are represented by a dipole located at the center of a sphere representing the torso, hence at the center of the equilateral triangle. With these assumptions, the voltages measured by the three limb leads are proportional to the projections of the electric heart vector on the sides of the lead vector triangle, as described in Figure 1.8.

Frank Norman Wilson (1890-1952) investigated how electrocardiographic unipolar potentials could be defined. Ideally, those are measured with respect to a remote reference (infinity). But how is one to achieve this in the volume conductor of the size of the human body with electrodes already placed at the extremities? In several articles on the subject, Wilson and colleagues suggested the use of the central terminal as this reference [Wilson et al., 1931]. This was formed by connecting a 5 kΩ resistor from each terminal of the limb leads to a common point called the central terminal, as shown in Figure 1.9. Wilson suggested that unipolar potentials should be measured with respect to this terminal which approximates the potential at infinity. Actually, the Wilson central terminal is not independent of, but rather, is the average of the limb potentials. In clinical practice good reproducibility of the measurement system is vital. Results appear to be quite consistent in clinical applications. Wilson advocated 5 kΩ resistances; these are still widely used, though at present the high-input impedance of the ECG amplifiers would allow much higher resistances.

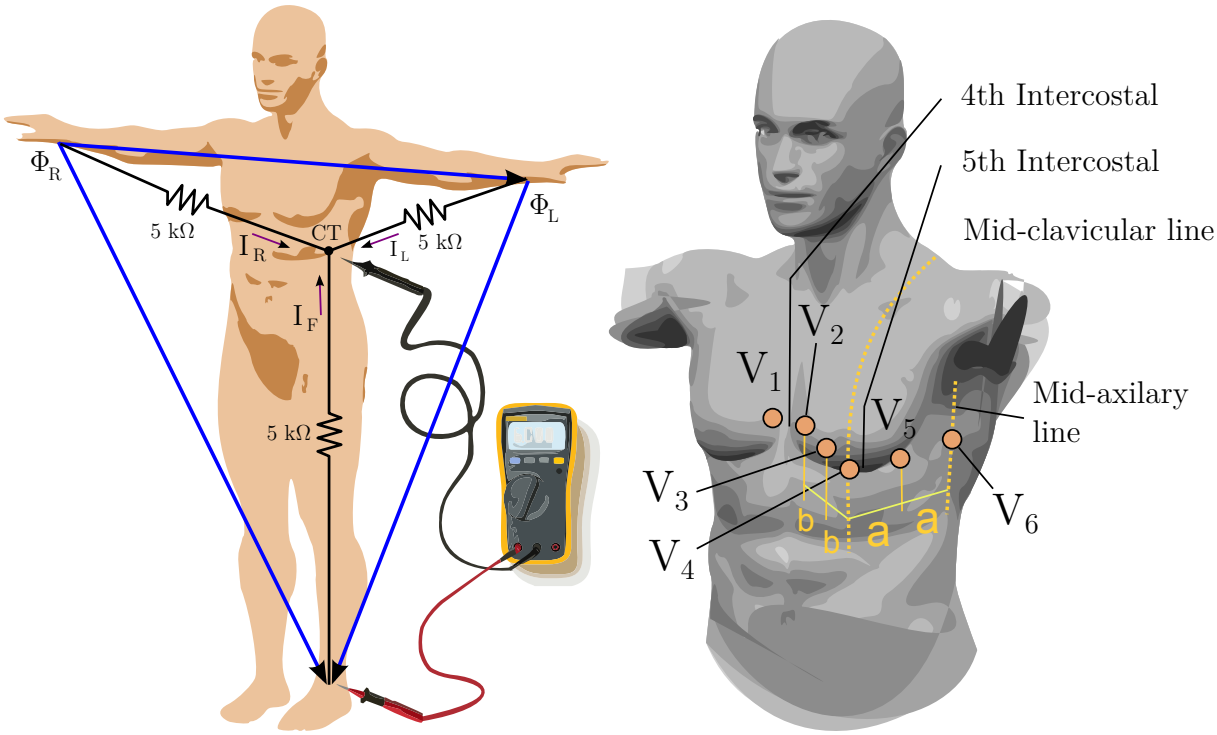


Figure 1.9: Wilson central terminal and precordial leads position on the torso. Diagrams based on image http://commons.wikimedia ... planes.svg under license CS-BY-SA.

Three additional limb leads are obtained by measuring the potential between each limb electrode and the Wilson central terminal. In 1942 E. Goldberger observed that these signals can be augmented by omitting that resistance from the Wilson central terminal, which is connected to the measurement electrode. In this way, the aforementioned three leads may be replaced with a new set of leads that are called augmented leads because of the augmentation of the signal. For measuring the potentials close to the heart, Wilson introduced the precordial leads (chest leads) in 1944. These leads, V1-V6 are located over the left chest as described in Figure 1.9. The points V1 and V2 are located at the fourth intercostal space on the right and left side of the sternum; V4 is located in the fifth intercostal space at the mid-clavicular line; V3 is located between the points V2 and V4; V5 is at the same horizontal level as V4 but on the anterior axillary line; V6 is at the same horizontal level as V4 but at the mid-line. The location of the precordial leads is illustrated in Figure 1.9.

The 12-lead system as described here is the one with the greatest clinical use. There are also some other modifications of the 12-lead system for particular applications. In exercise ECG, the signal is distorted because of muscular activity, respiration, and electrode artifacts due to perspiration and electrode movements. The distortion due to muscular activation can be minimized by placing the electrodes on the shoulders and on the hip instead of the arms and the leg, as suggested by R. E. Mason and I. Likar [Mason and Likar, 1966]. The Mason-Likar modification is the most important modification of the 12-lead system used in exercise ECG. The accurate location for the right arm electrode in

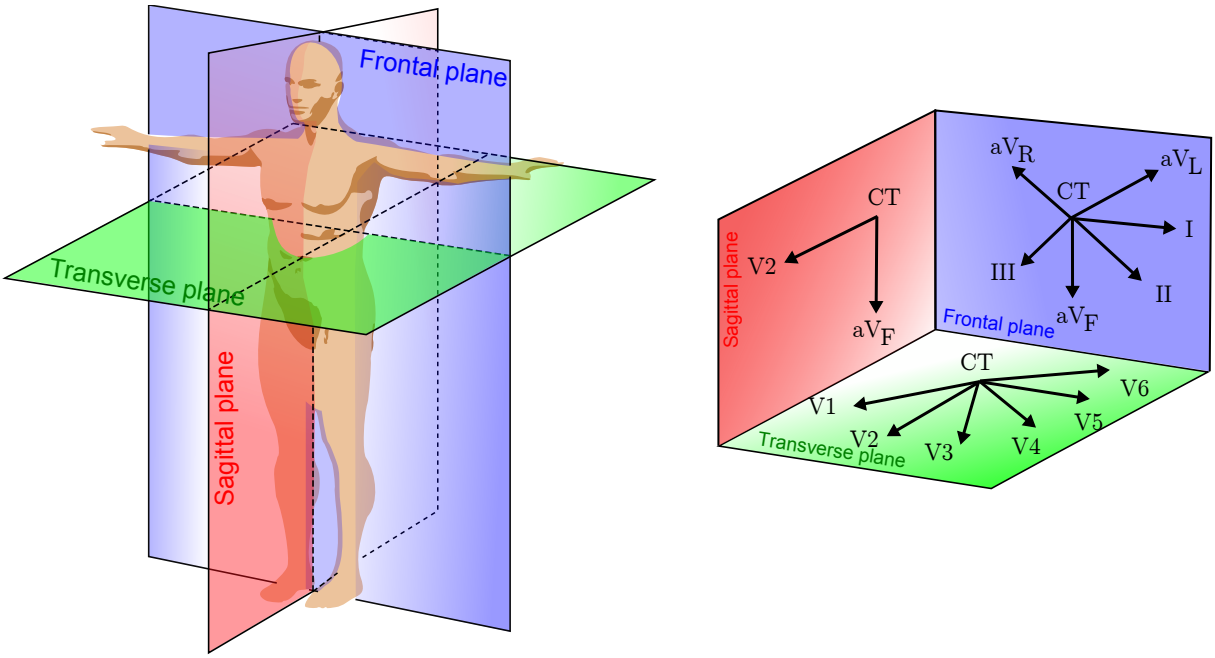


Figure 1.10: The projections of the lead vectors of the 12-lead ECG system in three orthogonal planes when one assumes the volume conductor to be spherical homogeneous and the cardiac source located in the center. Diagrams based on image http://commons.wikimedia ... planes.svg under license CS-BY-SA.

the Mason-Likar modification is a point in the infraclavicular fossa medial to the border of the deltoid muscle and 2 cm below the lower border of the clavicle. The left arm electrode is located similarly on the left side. The left leg electrode is placed at the left iliac crest. The right leg electrode is placed in the region of the right iliac fossa. The precordial leads are located in the Mason-Likar modification in the standard places of the 12-lead system. In ambulatory monitoring of the ECG, as in the Holter recording, the electrodes are also placed on the surface of the thorax instead of the extremities.

Of these 12 leads, the first six are derived from the same three measurement points. Therefore, any two of these six leads include exactly the same information as the other four. However, the precordial leads detect also nondipolar components, which have diagnostic significance because they are located close to the frontal part of the heart. Therefore, the 12-lead ECG system has eight truly independent and four redundant leads. The main reason for recording all 12 leads is that it enhances pattern recognition. This combination of leads gives the clinician an opportunity to compare the projections of the resultant vectors in two orthogonal planes and at different angles.

1.2.3 Arrhythmias

Arrhythmias are defined as any cardiac rhythm other than the normal sinus rhythm. Sinus rhythm originates in the sinus node and subsequently is conducted at appropriate rates through the atria, A-V junction, and the intraventricular specific conduction system. At rest the sinus node discharge cadence tends to be regular, although it presents gen-

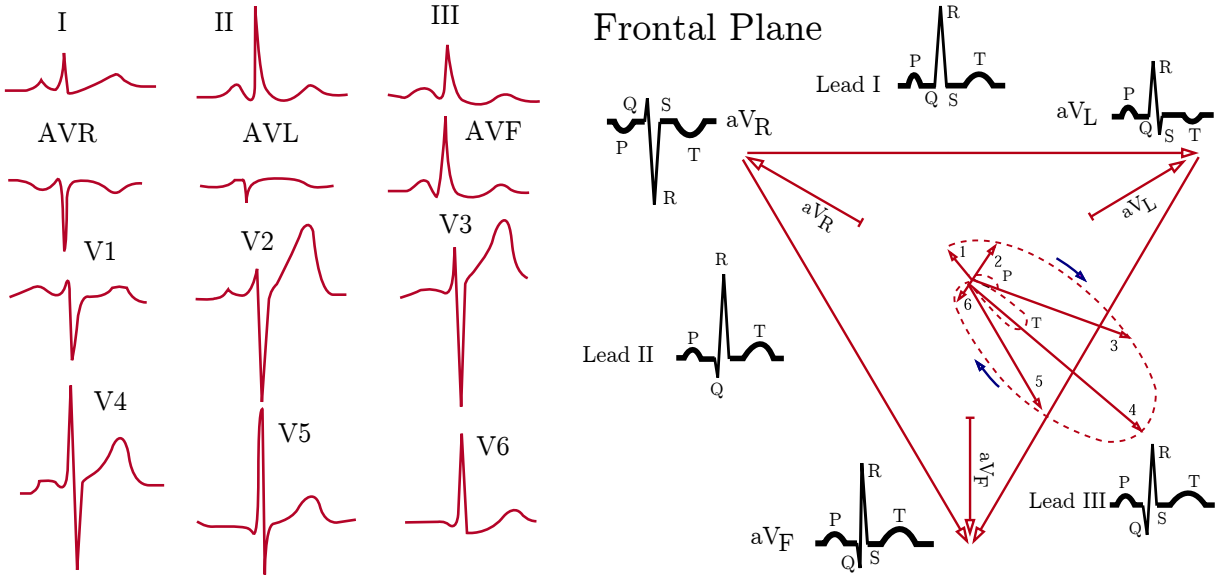


Figure 1.11: Normal Vectocardiogram and the projection to the 12-lead ECG.

erally slight variations. However, under normal conditions and particularly in children, it may present slight to moderate changes dependent on the phases of respiration, with the heart rate increasing with inspiration. In adults at rest the rate of the normal sinus rhythm ranges from 60 to 100 beats per minute (bpm). Thus, sinus rhythms over 100 bpm (sinus tachycardia) and those under 60 bpm (sinus bradycardia) may be considered arrhythmias. However, it should be taken into account that sinus rhythm varies throughout a 24-h period and sinus tachycardia and sinus bradycardia usually are a physiologic response to certain sympathetic (exercise, stress) or vagal (rest, sleep) stimuli. Under such circumstances, the presence of these heart rates should be considered normal. The term arrhythmia does not mean rhythm irregularity, as regular arrhythmias can occur often with absolute stability (flutter, paroxysmal tachycardia, etc.), sometimes presenting heart rates in the normal range. On the other hand, some irregular rhythms should not be considered arrhythmias (mild to moderate irregularity in the sinus discharge, particularly when linked to respiration). Moreover, a diagnosis of arrhythmia in itself does not mean evident pathology. In fact, in healthy subjects, the sporadic presence of certain arrhythmias both active (premature complexes) and passive (escape complexes, certain degree of A-V block, evident sinus arrhythmia, etc.) is frequently observed. There are different ways to classify cardiac arrhythmias:

- According to the site of origin: arrhythmias are divided into supraventricular (including those having their origin in the sinus node, the atria, and the AV junction) and ventricular arrhythmias.

- According to the underlying mechanism: arrhythmias may be explained by: 1) abnormal formation of impulses, which includes increased heart automaticity (extra systolic or parasystolic mechanism) and triggered electrical activity, 2) reentry of different types, and 3) decreased automaticity and/or disturbances of conduction.

- From the clinical point of view: arrhythmias may be paroxysmal, incessant or permanent. In reference to tachyarrhythmias (an example of an active arrhythmia), paroxysmal tachyarrhythmias occur suddenly and usually disappear spontaneously (i.e. A-V junctional reentrant paroxysmal tachycardia). Permanent tachyarrhythmias are always present (i.e. chronic atrial fibrillation), and incessant tachyarrhythmias are characterized by short and repetitive runs of supraventricular or ventricular tachycardia.

- Finally, from an electrocardiographic point of view, arrhythmias may be divided into two different types: active and passive.

  - Active arrhythmias, due to increased automaticity, reentry, or triggered electrical activity (these mechanisms are explained below), generate isolated or repetitive premature complexes on the ECG, which occur before the cadence of the regular sinus rhythm. The isolated premature complexes may be originated in a parasystolic or extrasystolic ectopic focus. The extra systolic mechanism presents a fixed coupling interval, whereas the para systolic presents a varied coupling interval.
    Premature complexes of supraventricular origin are generally followed by a narrow QRS complex, although they may be wide if conducted with aberrancy. The ectopic P wave is often not easily seen as it may be hidden in the preceding T wave. In other cases the premature atrial impulse remains blocked in the AV junction, initiating a pause instead of a premature QRS complex.
    The premature complexes of ventricular origin are not preceded by an ectopic P wave, and the QRS complex is always wide (> 120 ms), unless they originate in the upper part of the intraventricular SCS (ISCS). Premature and repetitive complexes include all types of supraventricular or ventricular tachyarrhythmias (tachycardias, fibrillation, flutter). In active cardiac arrhythmias due to reentrant mechanisms, a unidirectional block exists in some part of the circuit.
  - Passive arrhythmias occur when cardiac stimuli formation and/or conduction are below the range of normality due to a depression of the automatism and/or a stimulus conduction block in the atria, the AV junction, or the ISCS. From an electrocardiographic point of view, many passive cardiac arrhythmias present isolated late complexes (escape complexes) and, if repetitive, slower than expected heart rate (bradyarrhythmia). Even in the absence of bradyarrhythmia, some type of conduction delay or block in some place of the SCS may exist, for example, first-degree or some second-degree sinoatrial or A-V blocks, or atrial or ventricular (bundle branch) blocks. The latter encompasses the aberrant conduction phenomenon. Thus, the electrocardiographic diagnosis of passive cardiac arrhythmia can be made because it may be demonstrated that the ECG

changes are due to a depression of automatism and/or conduction in some part of the SCS, without this manifesting in the ECG as a premature complex, as it does in reentry (see Figure 1.12).

The mechanisms of cardiac arrhythmias are often the results of many factors including fluctuation in intracellular concentration of Ca, after depolarization currents, refractory period shortening or lengthening, autonomic nervous system innervation, repolarization dispersion, and changes in excitability and conduction. For example, bradyarrhythmia is often caused by abnormalities in excitability. This could be caused by dysfunction in the Na channels or by ischemia-induced elevation in extracellular K concentration. Furthermore, inherent or metabolically induced abnormalities in Na channels, Ca channels, or connexin have been shown to play a role in conduction diseases. Mechanisms of tachyarrhythmias can be grouped into three categories: re-entry, triggered activity and automaticity.

- Re-entry is a depolarizing wave traveling through a closed path. There are three prerequisites for re-entry: 1) At least two pathways: slow and fast AV nodal pathways, accessory pathway or the presence of barrier (anatomic: tricuspid valve; pathologic: incisional scars, myocardial infarction, and functional scar). 2) Unidirectional block: This block can be physiologic: caused by a premature complex, or increased heart rate; or pathologic: caused by changes in repolarization gradients. 3) Slow conduction to prevent collision of the head and the tail of the depolarizing wave. In functional re-entry, unidirectional block can be due to dispersion of refractoriness (repolarization) or dispersion of conduction velocity (anisotropic re-entry). See Figure 1.12 for an example of this concept.

- Triggered activities are caused by after depolarization currents. They are classified as early (EAD occurring inside AP: phases 2 and 3) or delayed (DAD: phase 4). These currents can in turn be responsible for both focal and reentrant arrhythmias. The former is caused by eliciting an excitatory response exceeding the activation threshold and the latter can be developed when these currents cause prolongation in action potential which facilitates the development of a unidirectional block due to dispersion of refractoriness.

- Automaticity is driven by spontaneous phase 4 depolarization. Automatic depolarizations in the atria and ventricles are not manifested normally due to overdrive suppression by the faster depolarization caused by the sinus node. However, during excess catecholaminergic states, phase 4 depolarization may exceed sinus node depolarization, causing depolarization to be driven by the abnormal tissue. Ventricular tachycardias during the acute ischemic and reperfusion phases are good examples of

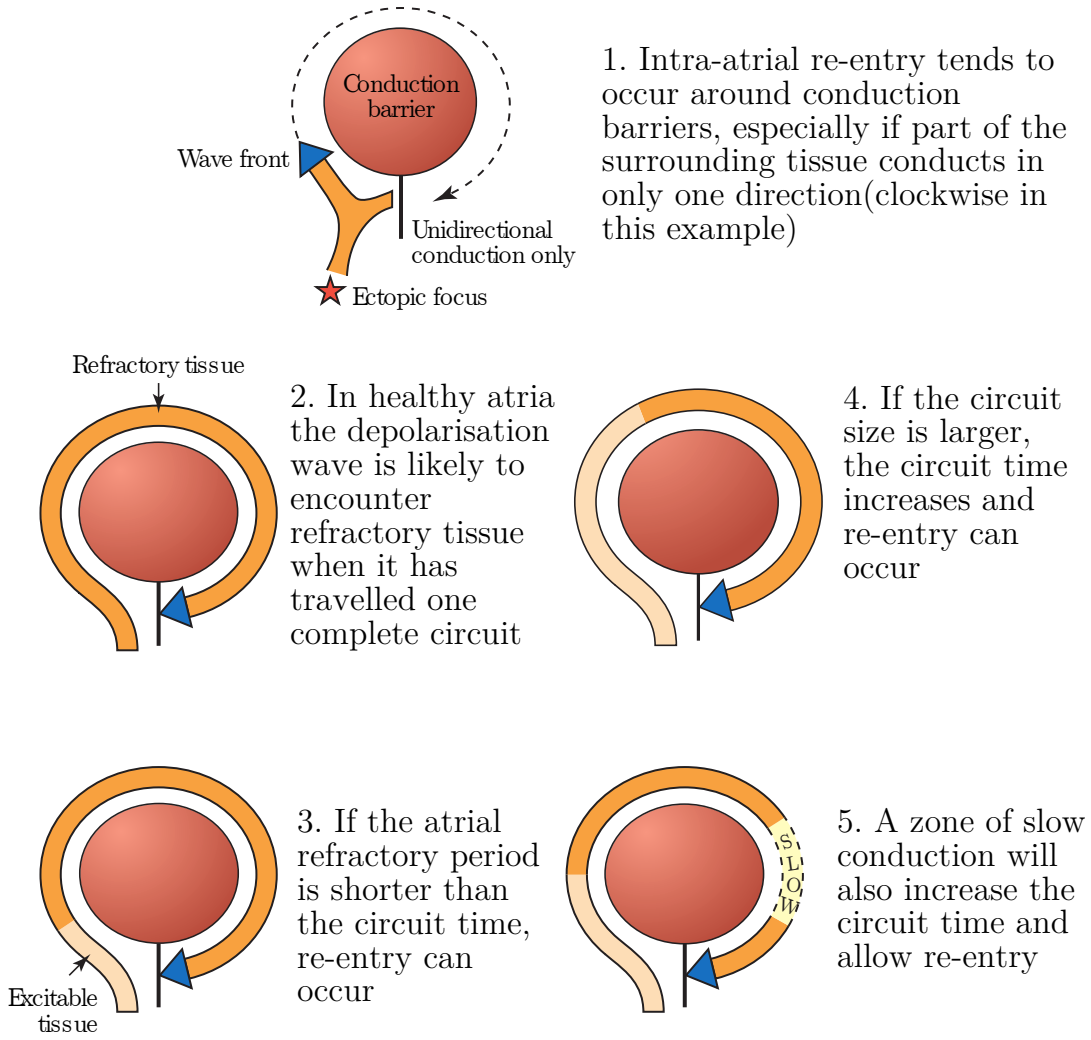


Figure 1.12: Electrical reentry, the mechanism responsible for initiating and maintaining atrial fibrillation. Reproduced from [Grubb and Furniss, 2001].

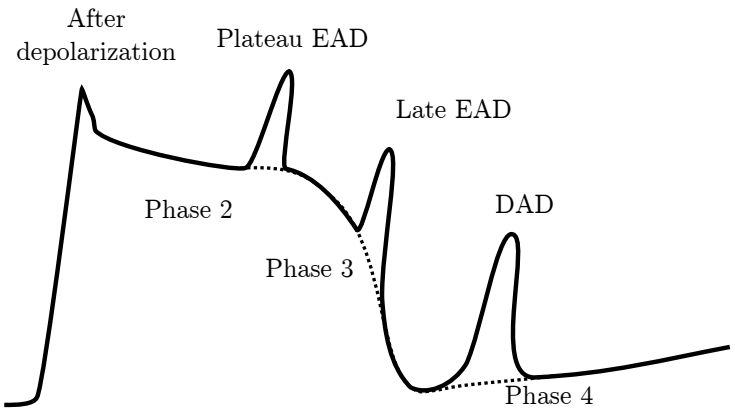


Figure 1.13: Types of after depolarization currents. EAD, early after depolarization; DAD, delayed after depolarization. Reproduced from [Natale and Wazni, 2007].

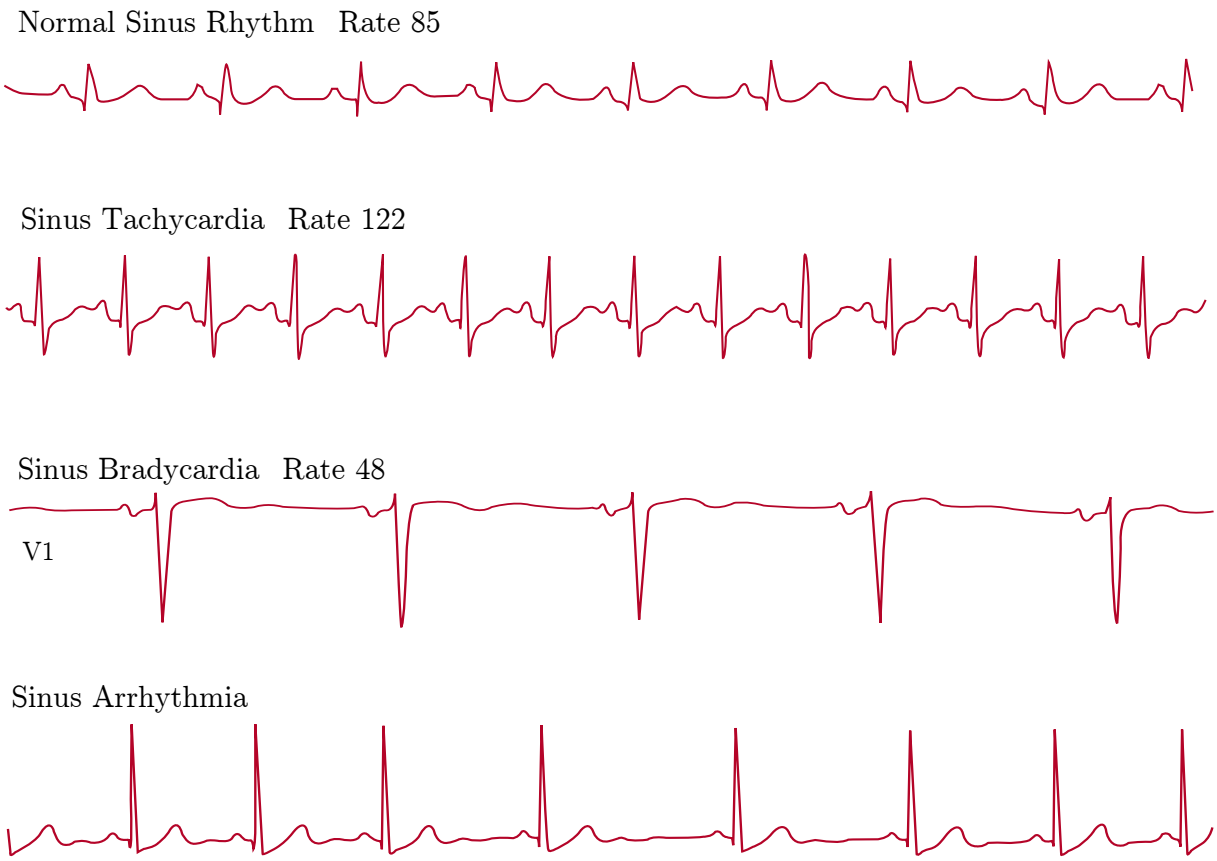


Figure 1.14: Several examples of sinus rhythms.

automaticity. They are often originated from the border zone between normal and ischemic cells.

As described above, the mechanisms that originates arrhythmias are diverse, and therefore the manifestation in the ECG. In the following section we will show the most important mechanisms as they appear in the ECG.

1.2.4 Manifestation of arrhythmias on the ECG

In this subsection several examples of the mechanisms enumerated above are shown in the ECG. Normal sinus rhythm is characterized by a regular cardiac rate with normal QRS complexes whose duration must be less than 120 milliseconds, as can be seen in Figure 1.14. The P-waves are normal in shape, and are synchronized with the QRS complexes. The PR interval must be less than 0.2 seconds. Heart rates may range from 60-100 bpm. There are a number of variant types of sinus rhythm, *sinus arrhythmia* is a normal rhythm in which heart rate varies periodically, usually with the respiratory cycle. There is an acceleration of rate during inspiration, and a slowing of rate during expiration.

Escape beats arise from lower (normally latent) pacemakers outside of the sinus node that fire because of either depressed sinus node function or blocked conduction of sinus impulses. Escape beats may originate at any pacemaker site below the sinus node. If the

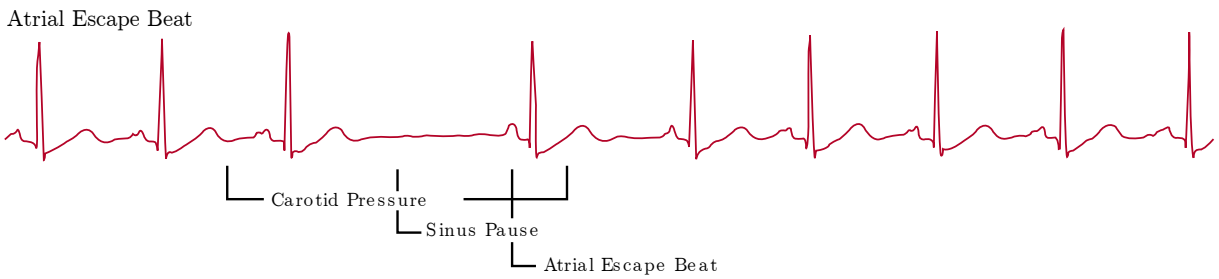


Figure 1.15: Example of an atrial escape beat.

sinus node slows sufficiently (perhaps due to vagal tone), other latent pacemaker sites in the atrium may emerge to establish heart rate. The P-wave resulting from these beats is usually different in shape from the normal, and in many cases is inverted in polarity. This reflects the fact that the beats originate low in the atrium. Such beats are sometimes referred to as *low atrial* or *coronary sinus* beats.

A-V nodal escape beats often terminate prolonged sinus pauses. The QRS complex is normal because the impulse is conducted normally to the ventricles. The P-wave is either not visible at all, or may be found just prior to or immediately following the QRS. In general the P wave is abnormal in shape since it is retrogradely conducted. If the P-wave immediately precedes the QRS complex, the beat is referred to as a *fast conducted beat*. Conversely, if the P-wave follows the QRS, the beat is called a *slow conducted beat*.

Ventricular escape beats protect the heart against asystole in the event of AV block (either fixed or transitory). They are characterized by a wide and usually bizarre QRS complex. The cardiac impulse originates in the ventricular Purkinje system. It is generally conducted with a slow propagation speed (0.5 meter/second) through the myocardium, thus leading to a wide QRS complex (usually greater than 120 ms). Ventricular escape rhythms (idioventricular rhythms) are common in cases of complete heart block, and have rates of about 40 per minute. Ectopic beats could arise from pacemakers outside the sinus node as a result of an abnormal increase in rhythmicity in the ventricular Purkinje system.

Atrial premature beats (APB) are seen frequently in normal individuals and have little clinical significance. They are also seen in heart disease, and when frequent, may be an early sign of atrial irritability which may progress to more serious atrial dysrhythmias. In APBs the QRS complexes are normal since they propagate normally through the ventricles via the conduction system. The P-waves are generally slightly abnormal since they originate from an abnormal focus, and propagate in an abnormal pattern. The impulse generally invades the area of the SA node and resets the sinus pacemaker. APBs occurring quite early following the previous beat may be aberrantly conducted, frequently with a right bundle branch block configuration. Aberrant conduction is particularly likely when the APB follows a long RR interval (the Ashman phenomenon). If an APB is extremely early it may run into refractory tissue in the AV node and be non-conducted.

Ventricular ectopic beats (VPB) originate from somewhere in the ventricles. The QRS complex is wide (greater than 0.12 seconds) and bizarre. VPBs may exhibit fixed coupling

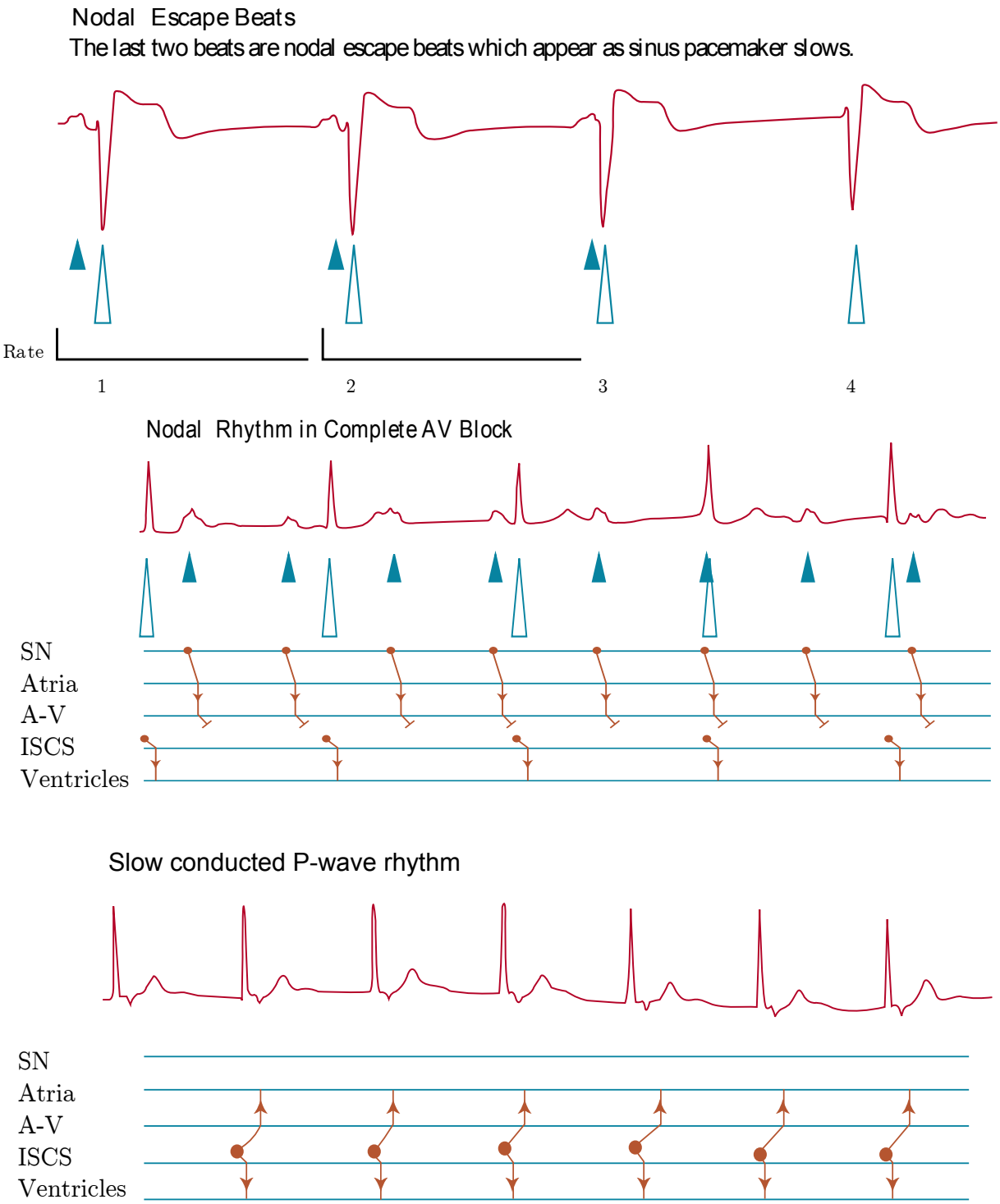


Figure 1.16: Examples of A-V nodal escape beats.

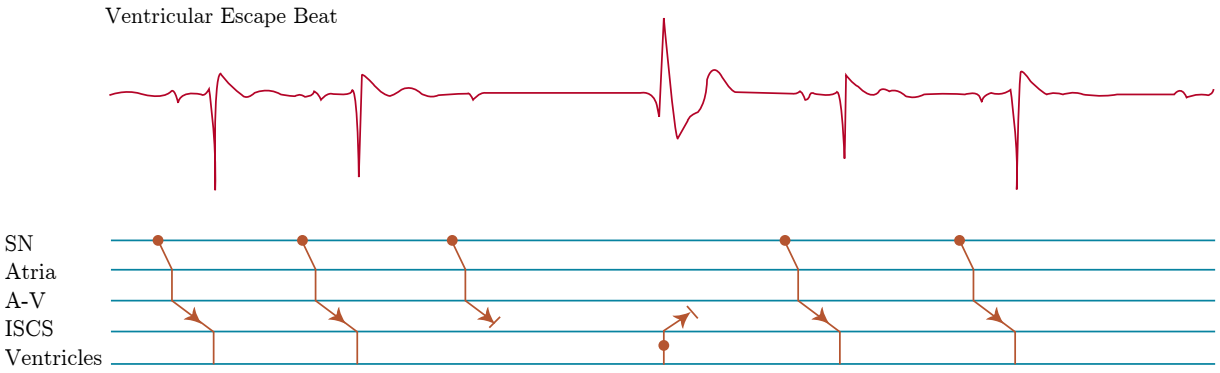


Figure 1.17: Example of a ventricular escape beat.

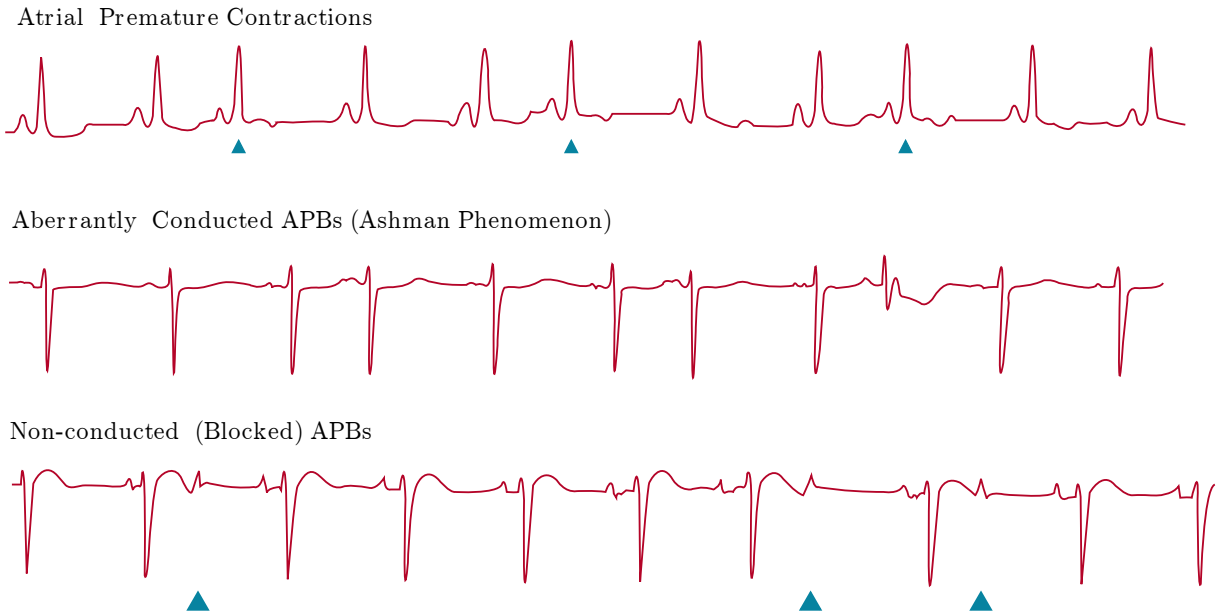


Figure 1.18: Examples of atrial premature beats. The blue triangles indicate the premature beats in the top panel, and the non-conducted beats in the bottom.

to previous normal beats. They may occur early or late in the cycle. The mechanism for PVCs may be reentry or triggered activity as discussed previously. Some VPBs appear to show no fixed coupling to preceding normal beats. If they show a regular rhythm of their own, they may result from a parasystolic focus. Note that some parasystolic depolarizations experience "exit block" and do not result in ventricular excitation. Parasystolic ventricular ectopic beats are usually considered relatively benign. Most VPBs are followed by a pause. The pause is usually compensatory, meaning that the coupling interval to the preceding normal beat plus the pause following the VPB comprise an interval equal to twice the normal R-R interval. An interpolated VPB is one which is sandwiched between two normal QRS complexes which arrive on time with the sinus normal activation.

VPBs are often found in otherwise normal individuals and probably have little significance if they are infrequent. In heart disease, VPBs may be a risk factor for increased incidence of more serious ventricular arrhythmias and sudden death. VPBs may occur singly or in groups and the following ordering of increasing severity of ventricular ectopic activity has been proposed:

1. Occasional: less than 30 per hour VPBs of the same morphology.
2. Frequent: greater than 30 per hour uniform VPBs or bigeminy where every other beat is a VPB
3. Multiform PVCs: different QRS morphologies
4. Couplets: pairs of consecutive VPBs
5. Ventricular Tachycardia: runs of three or more VPBs
6. Ventricular Flutter: rapid ventricular tachycardia with a sinusoidal configuration caused by merging of QRSs and Ts
7. Ventricular Fibrillation chaotic electrical activity without definite QRS complexes

VPBs which occur very early in the cardiac cycle such that they fall on the T-wave of the previous beat are considered particularly dangerous. At the time corresponding to the peak of the T wave, the ventricular myocardium is just beginning to repolarize. Some cells may be in the relatively refractory period, while others may be more fully recovered, and still others quite refractory. The electrical properties of the myocardium are thus quite varied, and conditions favoring reentrant loops are likely. Thus, an extra stimulus in the form of an isolated VPB which is very early-cycle may trigger a repetitive ventricular ectopic rhythm such as ventricular tachycardia or ventricular fibrillation. (The period near the T-wave peak is often referred to as the *vulnerable period*). Proper characterization of ventricular ectopic activity requires long-term (24-hour) ECG monitoring.

The classification of heartbeats on the ECG as can be seen, is an important task for the automatic analysis of arrhythmias. This is the first task performed by a cardiologist

when inspecting a recording, and as shown above, it is a very demanding task. In the next section we will review the state of the art regarding heartbeat classification algorithms.

1.3 Previous works

Many algorithms for ECG heartbeats classification were developed in the last decades. Some of the most relevant before the beginning of this thesis are [Hu et al., 1997, Lagerholm et al., 2000, de Chazal et al., 2004, Inan et al., 2006, Christov et al., 2006, de Chazal and Reilly, 2006], while others were published in the last few years [Llamedo and Martínez, 2007, Jiang and Kong, 2007, Park et al., 2008, Ince et al., 2009]. However, due to the lack of standardization in the development and evaluation criteria, comparison of results across most of these works could not be performed fairly or is impossible. In order to overcome this problem, some methodological aspects in the development and evaluation of heartbeat classifiers were followed in recent works [de Chazal et al., 2004, Jiang and Kong, 2007, Ince et al., 2009, Llamedo and Martínez, 2011a]. The most relevant key-points are:

- Use of public and standard databases, as the ones available in Physionet [Goldberger et al., 2000].
- Fulfillment of AAMI recommendations for class labeling and results presentation [AAMI-EC57, 1998–2008].
- Patient-oriented data division into training and testing sets, as described in [de Chazal et al., 2004].

Another aspect suggested in recent works is the analysis of the capability of the classifier to retain its performance in other databases not considered during the development [Llamedo and Martínez, 2011a]. We refer to this property of a classifier as generalization capability, and its analysis provides a broader idea of the performance achieved. Up to the writing of this thesis, only few of the reviewed works used more than one database either for the development [Watrous and Towell, 1995, Kiranyaz et al., 2011] or for a generalization assessment [Chudácek et al., 2009, Krasteva and Jekova, 2007, Syed et al., 2007].

The AAMI EC57 recommendations [AAMI-EC57, 1998–2008] for class labeling and results presentation are at the present time broadly accepted [de Chazal et al., 2004, Inan et al., 2006, Llamedo and Martínez, 2007, Jiang and Kong, 2007, Park et al., 2008, Ince et al., 2009]. As any classification problem, the goal is to *learn* a function that divides in C regions (or classes) a (hyper) space defined by the features, extracted from the ECG, and then make predictions with this function. In other words, this means assigning a label to an unknown heartbeat as a function of the value of some features. It is not difficult to realize that the lesser the amount of classes (small C), the simpler the partition function. Since cardiologists can group heartbeats into a number of classes that is easily higher than 10, the AAMI EC57 recommendations simplifies the problem into

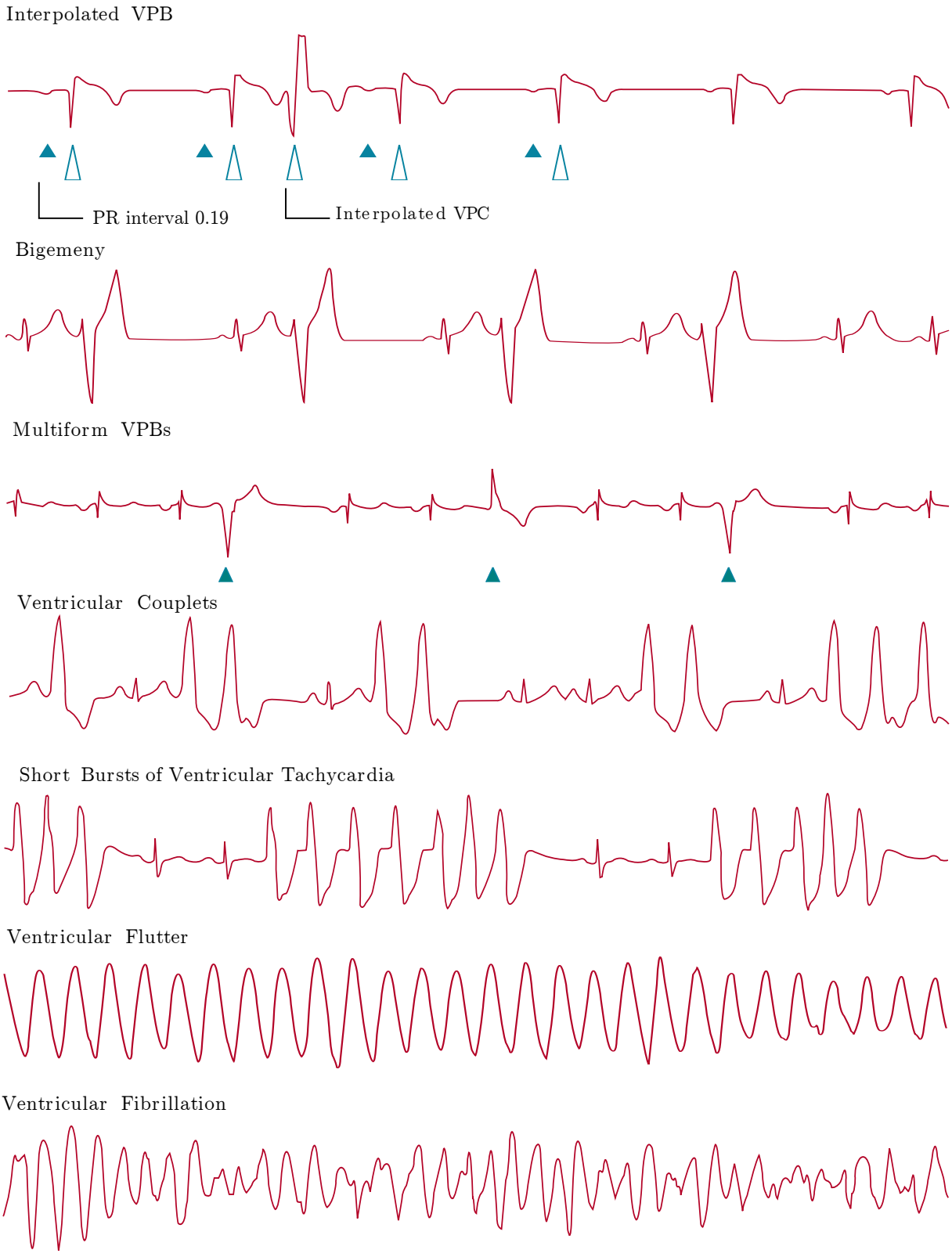


Figure 1.19: Examples of ventricular premature beats.

5 classes. Specifically, the EC57 recommendation [AAMI-EC57, 1998–2008] suggest the supraventricular (S) and ventricular (V) ectopic beats, fusion of normal and ventricular beats (F), a paced beat, a fusion of paced and normal beats or a beat that cannot be classified (Q) and finally a normal or bundle branch block beat (N). It is remarkable that all previous works were interested in discriminating between N and V classes, but only few of these works studied the multiclass classification problem [Lagerholm et al., 2000, de Chazal et al., 2004, Llamedo and Martínez, 2007, Park et al., 2008].

In terms of the data division in some works performed a beat-oriented division, no matter to which subject the heartbeats belongs to, with the inconvenience that sometimes heartbeats from some subjects were included in both the training and testing datasets [Inan et al., 2006, Jiang and Kong, 2007, Ince et al., 2009]. It was shown in [de Chazal et al., 2004] that this approach leads to an optimistic bias of the results, being more advisable a patient-oriented division, which is based on the application scenario where this kind of algorithm would be used.

Regarding to the features used (the classification model), the surrounding RR intervals were used in almost all published works. Other typical choices were the decimated ECG samples (mostly from the QRS complex or T wave) [de Chazal et al., 2004], or transformed by Hermite polynomials [Lagerholm et al., 2000] or wavelet decomposition (WT) [Llamedo and Martínez, 2007]. In [de Chazal et al., 2004], features derived from the delineation of the ECG like the QRS complex and T wave duration, resulted useful for classification. In some works where the dimensionality of the feature-space was an issue, feature transformations like PCA were used to keep the dimension of the model as low as possible [Ince et al., 2009]. The study of the relative importance of each feature within a model to perform a feature selection was not performed in any of the reviewed articles. Some works use features that integrate information present in both leads, like the vectocardiogram (VCG) maximum value (VCG_{max}) and VCG angle (VCG_{angle}) [Christov et al., 2006]. Another multilead strategy can be seen in [de Chazal et al., 2004], where a final decision from several posterior probabilities is calculated from single-lead features. This last approach is not practical for multilead classification because of the need of a different model designed for each set of leads, and the consequent growth in features dimensionality. The room for improvement in the field of heartbeat classification, together with the availability of 3- and 12-lead Holter devices makes necessary the development of algorithms capable of exploiting the increase of recorded information. Recently, moreover, the St. Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database (INCART) became freely available on Physionet [Goldberger et al., 2000], making possible the evaluation of multilead heartbeat classifiers in a comparable way. The generalization of a two-lead classifier to an arbitrary number of leads is one of the contributions of this thesis.

Several classifiers were adopted in the reviewed articles, from simple linear discriminant functions based on the Gaussian assumption of the data [de Chazal et al., 2004, Llamedo

and Martínez, 2007] to others more elaborated, as artificial neural networks (ANN's), self organizing maps (SOM) and learning vector quantization (LVQ) among others [Hu et al., 1997, Lagerholm et al., 2000, Inan et al., 2006, Christov et al., 2006, Jiang and Kong, 2007, Park et al., 2008, Ince et al., 2009].

The database used without exception by all groups was the MIT-BIH arrhythmia database [Moody and Mark, 2001] for training and testing purposes. Up to the moment none of the reviewed articles reported the generalization properties of the proposed algorithms outside the MIT-BIH database.

In the current state-of-the-art, it seems that the automatic classification approach has approximated to a performance upper bound, probably because the train and test datasets do not always have the same probability distribution in the feature space. The patient adaptation technique by means of expert assistance (i.e. manual beat annotation) was reported to be useful in two works to overcome this problem [Hu et al., 1997, de Chazal and Reilly, 2006], at the expense of sacrificing automaticity. Other works also reported better performances than the ones obtained by automatic classifiers, always taking advantage of the expert assistance [Lagerholm et al., 2000, Jiang and Kong, 2007, Ince et al., 2009, Kiranyaz et al., 2011]. One aspect to study when adopting this technique is the efficient use of the assistance, in order to keep the classifier as much automatic as possible. It is interesting to note that some classifiers require from 2 to 5 minutes of manual annotations, which is equivalent to several hundred of expert labeled heartbeats [Hu et al., 1997, de Chazal and Reilly, 2006, Ince et al., 2009, Jiang and Kong, 2007], while [Kiranyaz et al., 2011] requires the annotation of several heartbeats, depending on the number of arrhythmias present. One drawback of several patient-adaptable approaches is that they can not operate without assistance [Lagerholm et al., 2000, Ince et al., 2009, Jiang and Kong, 2007, Kiranyaz et al., 2011]. This is not the case of those developed as an evolution of a previous automatic classifier [Hu et al., 1997, de Chazal and Reilly, 2006].

1.4 Objective

The objective of this thesis is the study of methodologies to improve the classification of heartbeats on the ECG. As a first task we pursued the development of a two-lead automatic classifier. For this, we developed and evaluated a methodology for selecting the most discriminating features, with the best performance and generalization properties, in a multidatabase context according to the following premises:

- Perform fully automatic ECG classification
- Follow AAMI recommendations for class labeling and results presentation
- Use a simple classifier (as linear or quadratic discriminant functions)

- Features should have a physiological meaning, being simple to compute and robust to the typical kind of noise present in the ECG

Then, in a second task, we studied an effective way of accounting for morphologic information present in multilead ECG signals. For that purpose, we compare several multilead classification strategies against the reference two-lead classifier that we developed in [Llamedo and Martínez, 2011a]. We assess the improvement in classification performance as well as the generalization capability to other databases not considered during the development.

Finally, we studied how the classification performance of the previously developed automatic algorithms [Llamedo and Martínez, 2011a, 2012a] can be improved, by implementing a patient-adaptation technique. For that purpose, first we compare several integration strategies in a development dataset, and finally we assess the final performance and generalization capability to other databases not considered during the development. The performance was compared with other state-of-the-art classifiers [de Chazal and Reilly, 2006, Jiang and Kong, 2007, Ince et al., 2009, Kiranyaz et al., 2011, Mar et al., 2011].

1.5 Outline of the Thesis

The thesis is organized as follows:

- In **Chapter 2** we introduce several pattern recognition, signal processing and statistics methodologies used in the development of the thesis. As well as the ECG databases and computing resources used.

- **Chapter 3** includes the development of an automatic ECG heartbeat classifier. In this chapter we give special importance to the generalization achieved by the algorithm. For this purpose we adopt a feature selection algorithm, with an optimization criterion modified to select those features with larger generalization capability. The results of this chapter were the following publications:

  - M. Llamedo and J. P. Martínez. Heartbeat classification using feature selection driven by database generalization criteria. *IEEE Transactions on Biomedical Engineering*, 58:616– 625, 2011.
  - M. Llamedo and J.P. Martínez. Evaluation of an ECG heartbeat classifier designed by generalization-driven feature selection. In Engineering in Medicine and Biology Society. EMBC 2010. Annual International Conference of the IEEE, 2010.
  - M. Llamedo and J.P. Martínez. Analysis of multidomain features for ECG classification. In Computers in Cardiology 2009, volume 36, pages 561 – 564. IEEE Computer Society Press, 2009.

  - M. Llamedo and J.P. Martínez. Clasificación de ECG basada en características de escala, dirección y ritmo. In XXVI Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 09)., 2009.
  - M. Llamedo and J.P. Martínez. An ECG classification model based on multilead wavelet transform features. In Computers in Cardiology 2007, volume 34, pages 105–108. IEEE Computer Society Press, 2007.
  - M. Llamedo and J.P. Martínez. An ECG classification model based on multilead wavelet transform features. In XVI Congreso Argentino de Bioingeniería. San Juan. ISBN 978- 950-605-505-9, pages 531–534, 2007.
  - M. Llamedo, J.P. Martínez, and P Laguna. Un delineador de ECG multiderivacional basado en la transformada wavelet de la señal RMS. In XVI Congreso Argentino de Bioingeniería. San Juan. ISBN 978-950-605-505-9, pages 535–538, 2007.

- **Chapter 4** covers two improvements performed to the automatic classifier developed in the previous chapter. Between them, the adaptation of the algorithm to ECG recordings with an arbitrary amount of leads, and the use of a non-linear classifier, as the multilayer perceptron. The results of this chapter were the following publications:

  - M. Llamedo and J. P. Martínez. Cross-database evaluation of a multilead heartbeat classifier. IEEE Transactions on Information Technology in Biomedicine, Currently under review, with minor revision:–, 2012 expected.
  - T. Mar, S. Zaunseder, J. P. Martínez, M. Llamedo, and R. Poll. Optimization of ECG classification by means of feature selection. Biomedical Engineering, IEEE Transactions on, 58(8):2168 –2177, aug. 2011.
  - M. Llamedo and J.P. Martínez. Analysis of 12-lead classification models for ECG classification. In Computers in Cardiology 2010, volume 37. IEEE Computer Society Press, 2010.

- In **Chapter 5** we used the algorithms developed in the previous chapters to develop a heartbeat classifier with patient-adaptation capability. For this purpose we used a clustering algorithm, with features specially selected to maximize the intrapatient class separation. The resulting algorithm is capable of being used in a fully automatic mode, or with several degrees of expert-assistance. The results of this chapter were the following publications:

  - M. Llamedo and J. P. Martínez. An automatic patient-adapted ECG heartbeat classifier allowing expert assistance. IEEE Transactions on Biomedical Engineering, Currently under review:–, 2012 expected.

  - M. Llamedo and J.P. Martínez. Analysis of a semiautomatic algorithm for ECG heartbeat classification. In Computers in Cardiology 2011, volume 38. IEEE Computer Society Press, 2011.
  - M. Llamedo and J.P. Martínez. Análisis de un algoritmo para la clasificación semiautomática de latidos en ECG. In XXIX Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 11)., 2011.

- Finally, **Chapter 6** presents the most important conclusions drawn in the thesis.

Chapter 2

Materials and Methods

In this chapter we describe the materials and methods used in the following chapters. The materials are mainly two, ECG databases and computing resources. The methods used are many more, but can be grouped in signal processing and classification. The first group contains the methods to calculate the feature vectors from the ECG signal, while the second group contains the methods to classify the feature vectors into heartbeat classes.

2.1 ECG Databases

All experiments performed in this thesis were carried out in several public databases freely available on Physionet [Goldberger et al., 2000], the well known American Heart Association database [American Heart Association], and a database developed at Biosigna GmbH [Fischer et al., 2008]; their relevant details are summarized in Table 2.3. For all databases the AAMI recommendations for class-labeling were adopted. In the next subsection this recommendation is explained. The AAMI Q class (unclassified and paced heartbeats) was discarded since it is marginally represented in all databases. This limitation occurs to a lesser extent with the fusion (F) AAMI class, but instead of discarding the heartbeats of this class, we adopted an alternative labeling scheme already used in [Llamedo and Martínez, 2011a]. It consists in merging the fusion (of normal and ventricular beats) and ventricular classes, as the same ventricular class (V' in Table 2.3). We will refer to this modification as AAMI2 labeling. This labeling does not compromise the comparability with other AAMI compliant works, since according to AAMI recommendation [AAMI-EC57, 1998–2008], errors involving F and Q classes either should not be accounted for the required performance measurements, or are already accounted by considering the V' class. The databases used include different types of ECG recordings: some of them were recorded during routine ambulatory practice, but others were selected to include less common ventricular, junctional or supraventricular arrhythmias, or baseline ST segment displacement or other ECG abnormalities. As a result, we use in this work a dataset

with a broad range of normal and pathological ECG recordings to evaluate the algorithm performance. Further details of each database can be found on Physionet[1] [Goldberger et al., 2000].

We will refer as the *development dataset* to the union of the MITBIH-SUP database and the 22 recordings included in the DS1 subset of MITBIH-AR defined in [de Chazal et al., 2004], while the *evaluation dataset* includes the rest of databases described in Table 2.3. The reason of this division was, first for ensuring a fair comparison with de Chazal *et al.* [de Chazal et al., 2004] results, and second because MIT Arrhythmia databases have heartbeat annotations thoroughly reviewed by several experts, and are therefore more reliable than the rest of databases.

2.1.1 AAMI class labeling recommendations

According to Section 4.2 of [AAMI-EC57, 1998–2008], within annotation files beat labels are defined as follows:

N any beat that does not fall into the S, V, F, or Q categories described below (a normal beat or a bundle branch block beat)

S a supraventricular ectopic beat (SVEB): an atrial or nodal (junctional) premature or escape beat, or an aberrated atrial premature beat

V a ventricular ectopic beat (VEB): a ventricular premature beat, an R-on-T ventricular premature beat, or a ventricular escape beat

F a fusion of a ventricular and a normal beat

Q a paced beat, a fusion of a paced and a normal beat, or a beat that cannot be classified

The description and conversion matrices for the formats used in the databases described below are presented in tables 2.1 and 2.2.

2.1.2 MIT-BIH Arrhythmia Database (MITBIH-AR)

The MIT-BIH Arrhythmia Database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects studied by the BIH Arrhythmia Laboratory between 1975 and 1979 [Moody and Mark, 2001].

The source of the ECGs included in the MIT-BIH Arrhythmia Database is a set of over 4000 long-term Holter recordings that were obtained by the Beth Israel Hospital Arrhythmia Laboratory between 1975 and 1979. Approximately 60% of these recordings were obtained from inpatients. The first group is intended to serve as a representative sample of the variety of waveforms and artifact that an arrhythmia detector might encounter in routine clinical use. A table of random numbers was used to select tapes, and

[1] www.physionet.org

Table 2.1: Original annotation format used in the databases.

MIT format

Sym	Description
N	Normal beat
·	Normal beat
L	Left bundle branch block beat
R	Right bundle branch block beat
A	Atrial premature beat
a	Aberrated atrial premature beat
J	Nodal (junctional) premature beat
S	Supraventricular premature beat
V	Premature ventricular contraction
F	Fusion of ventricular and normal beat
[	Start of ventricular flutter/fibrillation
!	Ventricular flutter wave
]	End of ventricular flutter/fibrillation
e	Atrial escape beat
j	Nodal (junctional) escape beat
E	Ventricular escape beat
/	Paced beat
f	Fusion of paced and normal beat
x	Non-conducted P-wave (blocked APB)
p	Non-conducted P-wave (blocked APB)
Q	Unclassifiable beat
|	Isolated QRS-like artifact
?	Beat not classified during learning
+	Rhythm change
r	R-on-T premature ventricular contraction
s	ST segment change
B	Bundle branch block beat (unspecified)

AHA format

Sym	Description
E	Ventricular Escape
F	Fusion Beat
N	Beat of Non-Ventricular Origin
P	Paced Beat
Q	Questionable Beat - Indeterminate Origin
R	R-on-T Beat
U	Unreadable
V	Premature Ventricular Contraction
[	Beginning of Ventricular Fibrillation or Flutter
]	End of Ventricular Fibrillation or Flutter

Table 2.2: AAMI class conversion matrices for the formats used.

MIT format

Original	AAMI	AAMI2
N	N	N
·	N	N
L	N	N
R	N	N
A	S	S
a	S	S
J	S	S
S	S	S
V	V	V
F	V	V
[	Q	Q
!	Q	Q
]	Q	Q
e	N	N
j	N	N
E	V	V
/	F	V
f	F	V
x	Q	Q
p	Q	Q
Q	F	V
|	Q	Q
?	Q	Q
+	Q	Q
r	V	V
s	Q	Q
B	N	N
n	N	N

AHA format

Original	AAMI	AAMI2
E	V	V
F	F	V
N	N	N
Q	Q	Q
R	V	V
V	V	V

HES format

Original	AAMI	AAMI2
Q	Q	Q
N	N	N
N	N	N
S	S	S
Q	Q	Q
V	V	V
S	S	S

AAMI labels

N	Normal
S	Supraventricular
V	Ventricular
F	Fusion
Q	Unknown

then to select half-hour segments of them. Segments selected in this way were excluded only if neither of the two ECG signals was of adequate quality for analysis by human experts. Records in the second group were chosen to include complex ventricular, junctional, and supraventricular arrhythmias and conduction abnormalities. Several of these records were selected because features of the rhythm, QRS morphology variation, or signal quality may be expected to present significant difficulty to arrhythmia detectors; these records have gained considerable notoriety among database users. The subjects were 25 men aged 32 to 89 years, and 22 women aged 23 to 89 years. (Records 201 and 202 came from the same male subject.)

In most records, one signal is a modified limb lead II (MLII), obtained by placing the electrodes on the chest. While the other is usually a modified lead V1 (occasionally V2 or V5, and in one instance V4); as for the first signal, the electrodes are also placed on the chest. Normal QRS complexes are usually prominent in the first signal. The lead axis for the second signal may be nearly orthogonal to the mean cardiac electrical axis, however (i.e., normal beats are usually biphasic and may be nearly isoelectric). Thus normal beats are frequently difficult to discern in the second signal, although ectopic beats will often be more prominent. A notable exception is record 114, for which the signals were reversed. Since this happens occasionally in clinical practice, arrhythmia detectors should be equipped to deal with this situation. In records 102 and 104, it was not possible to use modified lead II because of surgical dressings on the patients; modified lead V5 was used for the first signal in these records.

The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. An initial set of beat labels was produced by a simple slope-sensitive QRS detector, which marked each detected event as a normal beat. Two identical 150-foot chart recordings were printed for each 30-minute record, with these initial beat labels in the margin. For each record, the two charts were given to two cardiologists, who worked on them independently. The cardiologists added additional beat labels where the detector missed beats, deleted false detections as necessary, and changed the labels for all abnormal beats. They also added rhythm labels, signal quality labels, and comments.

The annotations were transcribed from the paper chart recordings. Once both sets of cardiologists' annotations for a given record had been transcribed and verified, they were automatically compared beat-by-beat, and another chart recording was printed. This chart showed the cardiologists' annotations in the margin, with all discrepancies highlighted. Each discrepancy was reviewed and resolved by consensus. The corrections were transcribed, and the annotations were then analyzed by an auditing program, which checked them for consistency and which located the ten longest and shortest R-R intervals in each record (to identify possible missing or falsely detected beats).

The annotations provided with the database were used for training and testing purposes, following the recommendations and class-labeling of AAMI (See tables 2.1 and 2.2).

We adopted the training ($DS1$) and test ($DS2$) set division scheme used in [de Chazal et al., 2004] for comparative purposes of the results. The division scheme is summarized in Table 2.3.

2.1.3 MIT-BIH Supraventricular Arrhythmia Database (MITBIH-SUP)

The database consists of 78 two-lead recordings of approximately 30 minutes and sampled at 128 Hz. The recordings were chosen to supplement the examples of supraventricular arrhythmias in the MIT-BIH Arrhythmia Database. The annotations of the recordings were performed automatically first, by the Marquette Electronics 8000 Holter scanner and later reviewed and corrected by a medical student [Greenwald, 1990]. These reference annotations are considered a "silver standard" because the granularity of the beat labels only discriminates among normal, ventricular, supraventricular and fusion beats. The original labeling was also adapted to the AAMI recommendations and to the AAMI2 modification. This database will be considered for validation and model selection purposes. The class distribution is shown in Table 2.3.

2.1.4 St. Petersburg Institute of Cardiological Technics (INCART) 12-lead Arrhythmia Database

This database consists of 75 annotated recordings extracted from 32 Holter records. Each record is 30 minutes long and contains 12 standard leads, each sampled at 257 Hz. The reference annotation files contain over 175000 beat annotations in all. The original records were collected from patients undergoing tests for coronary artery disease (17 men and 15 women, aged 18-80; mean age: 58). None of the patients had pacemakers; most had ventricular ectopic beats. In selecting records to be included in the database, preference was given to subjects with ECG's consistent with ischemia, coronary artery disease, conduction abnormalities, and arrhythmias. These diagnoses were confirmed by enzyme assays, coronary angiography, electrophysiological study, and pressure monitoring where necessary. For each record it was included the patient's age, sex, diagnoses, and a summary of features of the ECG.

The annotations were produced by an automatic algorithm and then corrected manually, following the standard PhysioBank beat annotation definitions. The algorithm generally places beat annotations in the middle of the QRS complex (as determined from all 12 leads); the locations have not been manually corrected, however, there may be occasional misaligned annotations as a result. This database will be considered only for testing purposes. More details about the database are shown in Table 2.3.

2.1.5 European ST-T Database (ESTTDB)

This database consists of 90 annotated excerpts of ambulatory ECG recordings from 79 subjects. Myocardial ischemia was diagnosed or suspected for each subject, additional selection criteria were established in order to obtain a representative selection of ECG abnormalities in the database, including baseline ST segment displacement resulting from conditions such as hypertension, ventricular dyskinesia, and effects of medication. Each record is two hours in duration and contains two signals, each sampled at 250 Hz with 12-bit resolution over a nominal 20 millivolt input range.

The European ST-T Database is intended to be used for evaluation of algorithms for analysis of ST and T-wave changes. This database consists of 90 annotated excerpts of ambulatory ECG recordings from 79 subjects. The subjects were 70 men aged 30 to 84, and 8 women aged 55 to 71. Myocardial ischemia was diagnosed or suspected for each subject; additional selection criteria were established in order to obtain a representative selection of ECG abnormalities in the database, including baseline ST segment displacement resulting from conditions such as hypertension, ventricular dyskinesia, and effects of medication. The database includes 367 episodes of ST segment change, and 401 episodes of T-wave change, with durations ranging from 30 seconds to several minutes, and peak displacements ranging from 100 microvolts to more than one millivolt. In addition, 11 episodes of axis shift resulting in apparent ST change, and 10 episodes of axis shift resulting in apparent T-wave change, have been marked. Compact clinical reports document each record. These reports summarize pathology, medications, electrolyte imbalance, and technical information about each recording. Each record is two hours in duration and contains two signals, each sampled at 250 samples per second with 12-bit resolution over a nominal 20 millivolt input range. The sample values were rescaled after digitization with reference to calibration signals in the original analog recordings, in order to obtain a uniform scale of 200 ADC units per millivolt for all signals. The header files include information about the leads used, the patient's age, sex, and medications, the clinical findings, and the recording equipment. A complete description of the database can be found in [Taddei et al., 1992].

2.1.6 The MIT-BIH ST Change Database (MITBIH-ST)

This database includes 28 ECG recordings of varying lengths, most of which were recorded during exercise stress tests and which exhibit transient ST depression. We selected the two-lead recordings, resulting in 18 useful recordings. The recordings were sampled at 360 Hz and 12-bit resolution.

2.1.7 The Long-Term ST Database (LTSTDB)

The Long-Term ST Database contains 86 lengthy ECG recordings of 80 human subjects, chosen to exhibit a variety of events of ST segment changes, including ischemic ST episodes, axis-related non-ischemic ST episodes, episodes of slow ST level drift, and episodes containing mixtures of these phenomena. The database was created to support development and evaluation of algorithms capable of accurate differentiation of ischemic and non-ischemic ST events, as well as basic research into mechanisms and dynamics of myocardial ischemia. Detailed clinical notes and ST deviation trend plots are provided for all 86 records. The entire Long-Term ST Database is also available from its original home page at the Laboratory for Biomedical Computer Systems and Imaging at the University of Ljubljana, Slovenia The individual recordings of the Long-Term ST Database are between 21 and 24 hours in duration, and contain two or three ECG signals. Each ECG signal has been digitized at 250 samples per second with 12-bit resolution over a range of ±10 millivolts. Each record includes a set of meticulously verified ST episode and signal quality annotations, together with additional beat-by-beat QRS annotations and ST level measurements.

Several sources contributed recordings to the Long-Term ST Database:

- Eleven of the recordings included in the Long-Term ST Database are from the initial Long-Term ST Database developed under a joint U.S.-Slovenian research project between 1995 and 1998.

- Ten additional recordings of the Long-Term ST Database are from the collection originally gathered by the Pisa group for the European ST-T Database, which contains two-hour excerpts of some of these same recordings. The original analog recordings were redigitized for the Long-Term ST Database; since the signals have been rescaled as a result, direct comparison of the annotations in the European ST-T Database records with those for the corresponding portions of the Long-Term ST Database records is not possible. The inclusion of these recordings in the Long-Term ST Database allows study of the dynamics of ischemic ST changes over a much longer period in these previously well-studied subjects. Among the samples available here, record s20021 includes the two-hour segment that was previously digitized to produce record e0113 of the European ST-T Database.

- Another 18 of the LTSTDB recordings, those containing recordings with three ECG signals, were contributed to the project by Zymed, Inc.

The annotation of the Long-Term ST Database was performed using SEMIA, a program written by the group in Ljubljana for this purpose. Each recording was reviewed independently by expert annotators using SEMIA at each of the three sites (Ljubljana, Pisa, and Cambridge). Participants met several times annually to obtain the consensus reference annotations. For further details, see [Jager et al., 2003].

2.1.8 American Heart Association (AHA) ECG Database

The American Heart Association (AHA) sponsored the development of the AHA Database for Evaluation of Ventricular Arrhythmia Detectors during the late 1970s and early 1980s at Washington University (St. Louis). The first portions of the AHA Database were released in 1982, and it was completed in 1985. Until recently, the only available portion of the AHA database consisted of 80 two-channel excerpts of analog ambulatory ECG recordings, digitized at 250 Hz per channel with 12-bit resolution over a 10 mV range [American Heart Association]. These recordings, designated as the development set, are divided into eight classes of ten recordings each, according to the highest level of ventricular ectopy present:

- no ventricular ectopy (records 1001 through 1010)
- isolated unifocal PVCs (records 2001 through 2010)
- isolated multifocal PVCs (records 3001 through 3010)
- ventricular bi- and trigeminy (records 4001 through 4010)
- R-on-T PVCs (records 5001 through 5010)
- ventricular couplets (records 6001 through 6010)
- ventricular tachycardia (records 7001 through 7010)
- ventricular flutter/fibrillation (records 8001 through 8010)

The final thirty minutes of each recording are annotated beat-by-beat, although supraventricular ectopic beats are not distinguished from normal sinus beats. At the time the AHA Database was created, a second set of 75 recordings (designated as the test set) was constructed according to the same criteria as the development set (only 5 recordings in the R-on-T PVC class were included in the test set). The test set was intended for evaluations without any possibility that the detectors might have been tuned (optimized) for the test data; for this reason, the test set was unavailable until recently.

2.2 Supercomputing Resources

The development and evaluation of the algorithms presented in this thesis involves a lot of computation. This kind of tasks are unfeasible in ordinary computers, since the time required for an adequate evaluation can easily reach several days. All the results presented in this thesis were calculated using the resources of the *Instituto de Investigación en Ingeniería de Aragón* (I3A). HERMES is the I3A's high throughput computing cluster, and this is a brief overview of its main features:

Table 2.3: Databases and datasets used in this thesis with its class representation.

	Database	Purpose	Length	Leads	Fs (Hz)	N	S	V'		Q	#Rec
								V	F		
Development	*DS*1 of MITBIH Arrhythmia (MITBIH-AR)[1,2]	A	30'	2	360	45673	929	3755	412	6	22
	MITBIH Supr. Arrhythmia (MITBIH-SUP)[1]	A	30'	2	128	162271	12195	9940	23	79	78
Evaluation	*DS*2 of MITBIH Arrhythmia (MITBIH-AR)[1,2]	A	30'	2	360	44053	1833	3202	388	7	22
	American Heart Association (AHA)[3]	A	30'	2	250	319125	0	32745	1266	0	155
	European ST-T (ESTTDB)[1]	ST	2h	2	250	784568	1095	4467	354	0	90
	MITBIH ST Change (MITBIH-ST)[1]	ST	<1h	2	360	46215	798	319	0	0	18
	MITBIH long-term database (MITBIH-LT)[1]	LT	14-22h	2	128	594672	1499	63584	2785	0	7
	Long-Term ST (LTSTDB)[1]	LT/ST	21-24h	2/3	250	6422003	30905	37894		0	86
	Biosigna[4]	A	1h	12	500	286246	1326	2541	0	0	56
	St. Petersburg Inst. of Card. Tech. (INCART)[1]	A	30'	12	257	153651	1959	20005	219	6	75
	Total					8858840	52556	178502	5449	100	609

A: Arrhythmia/Heartbeat classification; ST: Ischemia detection; LT: Long-term detection/classification
Heart beats classes are N: normal, S: supraventricular, V: ventricular, F: fusion, Q: unknown, and V':AAMI2 Ventricular.
[1] [Goldberger et al., 2000]; [2] [de Chazal et al., 2004]; [3] [American Heart Association]; [4] [Fischer et al., 2008]

Dataset	MITBIH-AR recording names
*DS*1	101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230
*DS*2	100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234

- 200 computing nodes.
- 1534 computing cores (Intel & AMD).
- 3.5Tb RAM available as a whole.
- 150Tb high performance storage (based on Lustre).
- Network facilities:
  - 10Gb trunk network
  - 1Gb access network
  - 1 QDR Infiniband Island (16 nodes)
- Operating System: Scientific Linux 5.5
- Queue distribution system: Condor 7.6 [Condor, 2010]

2.3 Signal Processing

2.3.1 ECG preprocessing

The ECG recordings of all databases were first resampled to 360 Hz when necessary, which is the sampling frequency of the MITBIH-AR, and an adequate sampling rate for heartbeat classification. This was performed with a tenth order lowpass FIR filter without observing any notorious distortion (resample function, Signal Processing Toolbox of Matlab, The Mathworks Inc., MA). The next task is to remove the contamination that the ECG may contain, in order to improve as much as possible the signal to noise ratio (SNR). We are interested in removing three types of noise: the baseline wander, the power-line frequency and the out of band high frequency. The baseline wander is an external, low-frequency activity in the ECG which may interfere with the signal analysis, and may result from a variety of noise sources including respiration, body movements, and poor electrode contact. In the following chapters, this kind of noise will be removed with two cancellation methods. One method is based on the estimation of the baseline wander noise by median filtering the ECG, for the subsequent subtraction. The median filter is defined for each sample of a signal x as

$$y[n] = \underset{i=n-w}{\overset{n+w}{\text{med}}} x[i], \tag{2.1}$$

that is, the median of the samples included in a window of $2w + 1$ samples centered at the current sample n; and the med operation is defined as

$$\text{med}\, x[i] = \arg\min_{\mu} \sum_{i=1}^{N} |x[i] - \mu|\,, \tag{2.2}$$

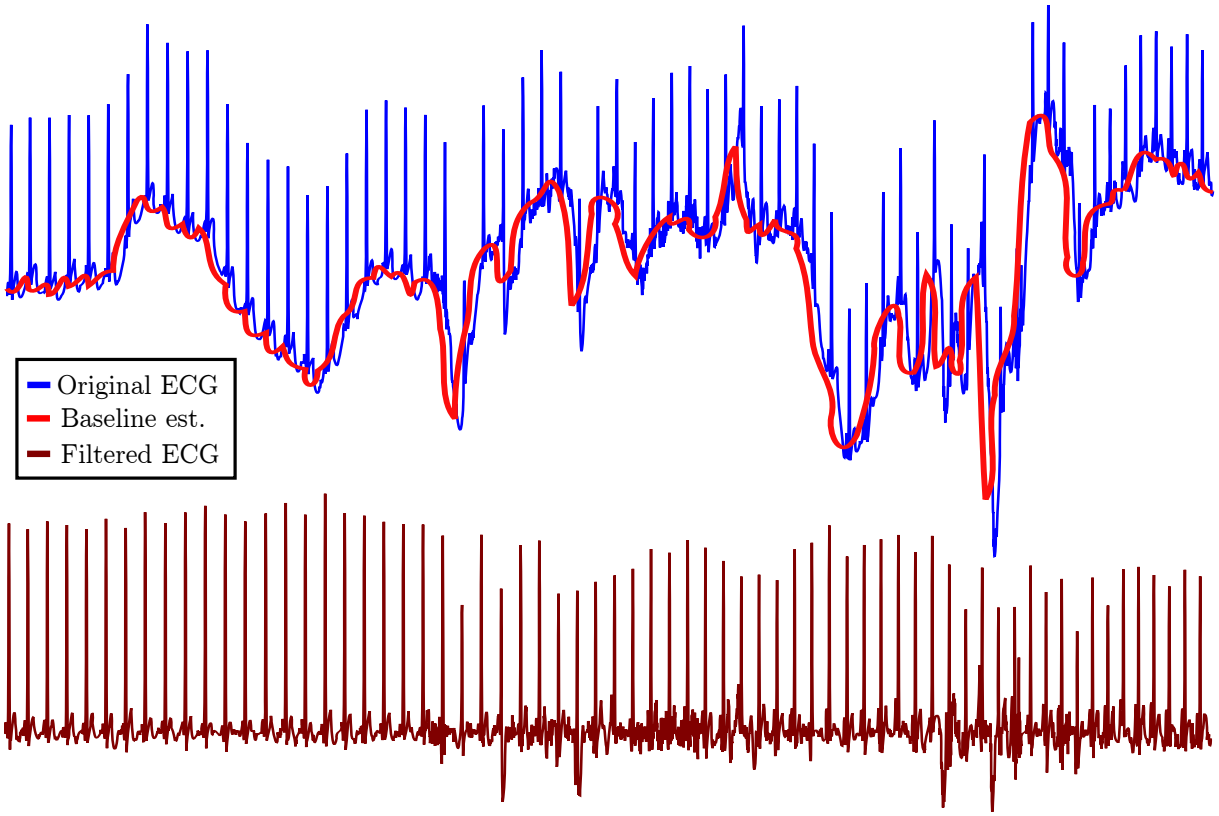


Figure 2.1: Baseline wander removal procedure with median filters.

being μ the median of $x[i]$, a robust estimation of the center location of a group of samples. As can be noted, it is a simple smoothing nonlinear filter, similar to an averaging filter, with only one parameter, which is the window size N. One important difference with the mean filter, is the nonlinear behavior inherited from the med operation (abrupt changes can be noted in Figure 2.1). Each signal was processed with a median filter of 200 ms width to remove QRS complexes and P-waves. The resulting signal was then processed with a median filter of 600 ms width to remove T-waves. The signal resulting from the second filter operation contained the baseline of the ECG signal, which was then subtracted from the original signal to produce the baseline corrected ECG signal. This method for baseline wander removal is simple, but is very demanding in terms of computational cost, and the estimated noise signal can have abrupt changes.

The other method used in this thesis is based on cubic-splines interpolation. In this case, the baseline wander estimation is obtained by fitting a polynomial to several representative samples of the ECG, which are chosen at the PQ interval of successive beats [Sörnmo and Laguna, 2005]. The PQ interval happens when the heartbeat impulse is being delayed at the A-V node (see Figure 1.6), therefore is considered free of electrical activity from the heart. This method is faster (given that the QRS locations were already computed) and avoids the bumps in the estimated signal caused by the median filter. As a disadvantage, the PQ location could be difficult at high heart rates or low SNR recordings.

Unwanted power-line and high-frequency noise was removed from the baseline corrected ECG with low-pass filter, following the same methodology used in [de Chazal et al., 2004]. The filter was a finite impulse response filter with equal ripple in the pass and stop bands. The 3-dB point of the filter was 35 Hz, and the attenuation at the stop-band was 80 dB. The transfer function of this filter is shown in Figure 2.2.

The filtered ECG signals were used in all subsequent processing. No energy or amplitude normalization was done, as we were interested in some amplitude-related features.

2.3.2 Wavelet Transform

Many of the considered features (explained in following sections) were based on the wavelet transform (WT) of the ECG signal. This transformation maps the ECG signal into a time-scale plane (understanding scale as a surrogate of frequency). Therefore, the WT allows the location in time of certain frequency contents. This aspect is of much importance when analyzing ECG signals, since each wave have a particular frequency content.

A *wavelet* $\psi(t)$ is a small wave in the sense of having limited duration, as it is an oscillating function that goes quickly to zero. It has zero mean and finite energy concentrated around a point. A wavelet family is defined from a prototype wavelet $\psi(t)$ by means of dilations (a) and translations (b)

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right),\ a > 0. \tag{2.3}$$

The *continuous wavelet transform* $W_a^y(b)$ (CWT) decomposes a signal $y(t)$ as a combination of functions $\psi_{a,b}$, i. e.

$$W_a^s(b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} y(t)\,\psi_{a,b}^*(t)dt = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} y(t)\,\psi^*\left(\frac{t-b}{a}\right)dt, \tag{2.4}$$

where $\psi^*(t)$ denotes the complex conjugate of $\psi(t)$.

If a real $\psi(t) \in L^2$ is such that its Fourier Transform $\Psi(\omega)$ fulfills the *admissibility condition*

$$\int_0^{+\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty, \tag{2.5}$$

then the CWT satisfies energy conservation [Mallat, 1999]. The CWT presents clear advantages over the *short time Fourier transform* (STFT) for ECG characterization. As a matter of fact, in the case of STFT a sliding time window is applied, obtaining covering units with finite energy, what allows to add *some* time/frequency location to the Fourier transform [Sörnmo and Laguna, 2005]. Each window corresponds to a time/frequency resolution that stands for the entire analysis. Nevertheless, in the case of non stationary signals a fixed resolution is hardly appropriated for the whole signal. The wavelet transform is specially adequate for describing non stationary signals, as is the case of the ECG.

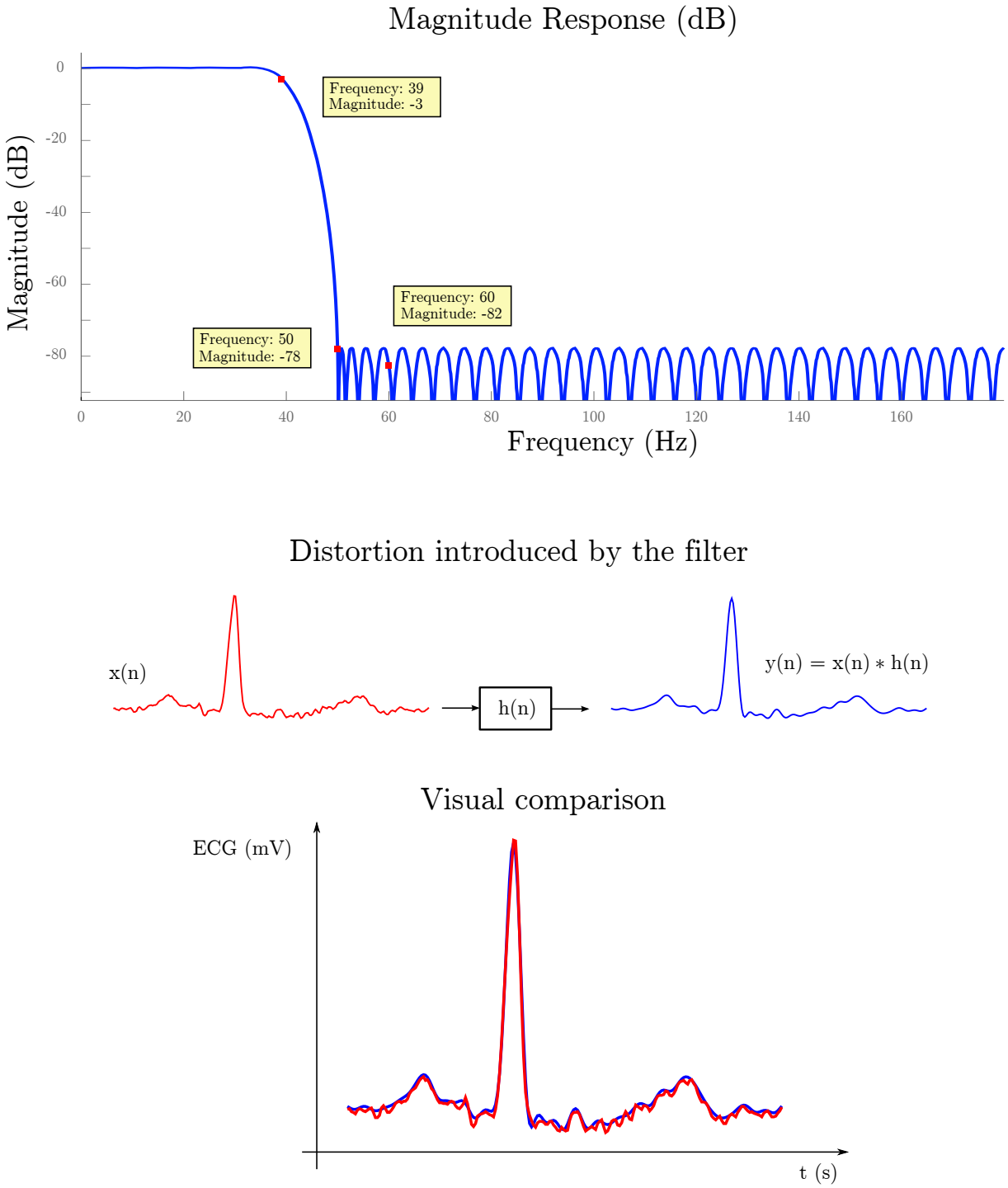


Figure 2.2: Transfer function of the low-pass filter used for ECG preprocessing. Note the adequate attenuation at powerline frequencies and the low distortion introduced in a normal heartbeat.

In fact, with the CWT the lower frequency components of the signal are characterized by the coefficients corresponding to wider $\psi_{a,b}(t)$ resulting from the higher scale factor a, and vice versa. This provides a time/scale domain description of $s(t)$ with higher temporal resolution at high frequencies and higher frequency resolution at low frequencies. For the purpose of automatic computer analysis, a digitized signal $s(k)$ is considered and its *discrete wavelet transform* (DWT) is usually obtained following a dyadic grid on the time-scale plane ($a = 2^m$ and $b = 2^m l$, $m \in N, l \in Z$), denoted as $Wd^s_{2^m}(2^m l)$. With the dyadic grid the DWT is equivalent to an octave filter bank and can be implemented as a cascade of identical cells of low-pass and high-pass FIR filters, with a downsampling operation after each filter, as schematized in Figure 2.3. That is,

$$\begin{aligned} &Wd^s_{2^m}(2^m l) = c_{m-1}(2^{m-1} l) * \bar{g}(2^m l), \\ &c_m(2^m l) = c_{m-1}(2^{m-1} l) * \bar{h}(2^m l), \quad \text{(Mallat's algorithm)} \\ &\bar{g}(k) = g(-k);\ \bar{h}(k) = h(-k), \end{aligned} \tag{2.6}$$

where $g(k)$ and $h(k)$ are the impulse response functions of the associated low-pass and high-pass filters, respectively [Mallat, 1989].

The description in equation (2.6) is time-variant, with more reduced temporal resolution for increasing scales. This drawback can be overcome using the algorithm *à trous* [Cohen and Kovačević, 1996], in which each decimation stage is replaced by an interpolation of the filter impulse responses of the previous scale, as represented in Figure 2.3. In this way a redundant representation of $y(k)$ is obtained, but the temporal resolution is maintained at different scales, that is

$$\begin{aligned} &W^s_{2^m}(k) = c_{m-1}(k) * \bar{g}_m(k), \\ &c_m(k) = c_{m-1}(k) * \bar{h}_m(k), \quad \text{(algorithm } \textit{à trous}\text{)} \\ &\bar{g}_m(k) = g_m(-k);\ \bar{h}_m(k) = h_m(-k). \end{aligned} \tag{2.7}$$

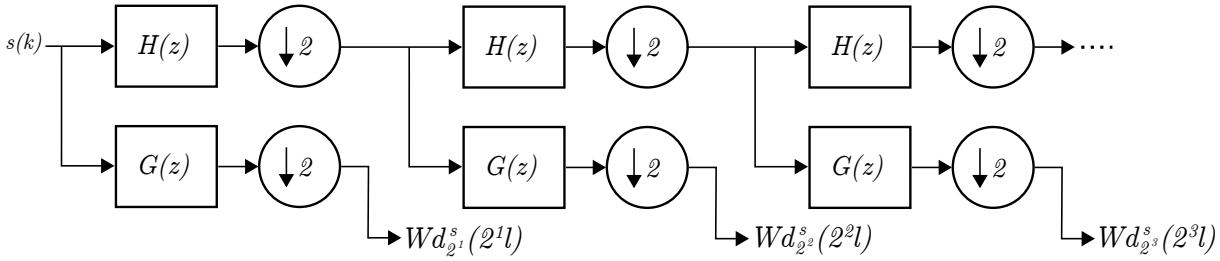
for $g_m(k)$ and $h_m(k)$ denoting the impulse response obtained by inserting $2^m - 1$ zeros between every sample in the impulse response function $g(k)$ and $h(k)$, respectively. Using this algorithm, the equivalent frequency response for scale $a = 2^m$ is

$$Q_m\left(e^{j\omega}\right) = \begin{cases} G\left(e^{j\omega}\right) & m = 1 \\ G\left(e^{j2^{m-1}\omega}\right) \prod_{l=0}^{m-2} H\left(e^{j2^l\omega}\right) & m \geq 2. \end{cases} \tag{2.8}$$

2.3.3 Prototype Wavelet

The frequency response of each scale derives from the prototype wavelet function used, thus it is of extreme importance that its choice attends to the specific frequency content of the signal of interest. It can be shown [Mallat and Zhong, 1992, Sahambi et al., 1998]

a) Mallat's algorithm



b) Algorithm *à trous*

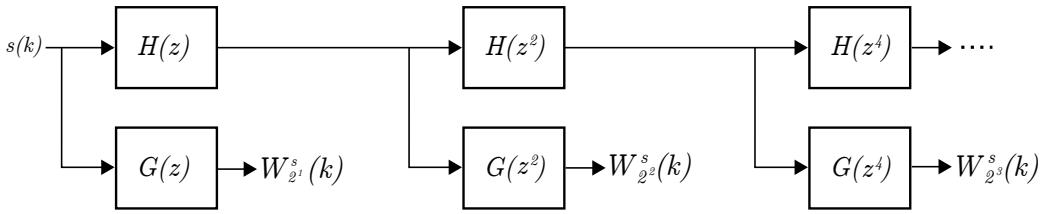


Figure 2.3: Discrete wavelet transform implementation schemes (dyadic grid).

that taking as prototype wavelet $\psi(t)$, the derivative of a smoothing function $\theta(t)$, the CWT of a signal $y(t)$ at scale a is given by

$$W_a^s(b) = -a\left(\frac{d}{db}\right)\int_{-\infty}^{+\infty} s(t)\theta_a(t-b)dt, \tag{2.9}$$

where $\theta_a(t) = \frac{1}{\sqrt{a}}\,\theta\left(\frac{t}{a}\right)$ is the scaled version of the smoothing function.

Therefore, the wavelet transform at scale a is proportional to the derivative of the filtered version of the signal with a smoothing impulse response at scale a.

This property is very convenient for the purpose of modeling ECG waves, which are composed of slopes and local maxima (or minima), as illustrated in Figure 2.9,

- wave peaks in the ECG correspond to zero crossings in the WT,
- slopes in ECG correspond to maxima and minima of the WT.

It was proposed as an example in [Mallat and Zhong, 1992], a wavelet prototype with this property, a quadratic spline that matches the derivative of the convolution of four rectangular pulses. Its Fourier transform is given by

$$\Psi(\omega) = j\omega\left(\frac{\sin(\omega/4)}{\omega/4}\right)^4, \tag{2.10}$$

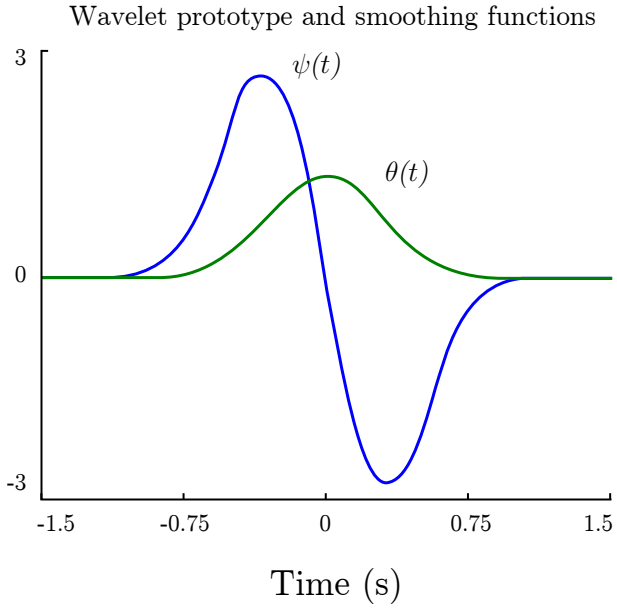


Figure 2.4: Wavelet prototype used in this thesis. A quadratic spline that matches the derivative of the convolution of four rectangular pulses.

and the low-pass and high-pass FIR filters have transfer functions [Li et al., 1995]

$$\begin{aligned} H\left(e^{j\omega}\right) &= e^{j\omega/2}\left(\cos\frac{\omega}{2}\right)^3, \\ G\left(e^{j\omega}\right) &= 4\,j\,e^{j\omega/2}\left(\sin\frac{\omega}{2}\right), \end{aligned} \tag{2.11}$$

with associated impulse responses

$$\begin{aligned} h[k] &= 1/8\cdot\{\delta[k+2]+3\delta[k+1]+3\delta[k]+\delta[k-1]\}, \\ g[k] &= 2\cdot\{\delta[k+1]-\delta[k]\}. \end{aligned} \tag{2.12}$$

This prototype wavelet has been applied to detection and delineation of ECG signals with good results [Li et al., 1995, Martínez et al., 2004]. As the analysis filters in equation (2.12) have linear phase [Li et al., 1995], the outputs of the filters can be realigned in order to present the same delay with respect to the original signal $s(t)$. The equivalent frequency responses $Q_m\left(e^{j\omega}\right)$ for this prototype wavelet using the *algorithme à trous* can be calculated from equations (2.8) and (2.11). The frequency responses are plotted in Figure 2.5 for the first five scales $a = 2^m|_{m=1,2,\ldots,5}$, considering a sampling frequency F_s of 250 Hz. In the rest of the thesis, the wavelet scale $a = 2^m$ will be referred as the m-th scale.

For a value of F_s different from 250 Hz the bands in Figure 2.5 would appear scaled in frequency. It is important to keep the scale fitting to the ECG features on the algorithms independent of F_s: for that purpose a new set of filters, having equivalent frequency responses as close as possible to the ones of Figure 2.5 are constructed for each F_s. The new filters are obtained by resampling adequately the equivalent filter impulse responses

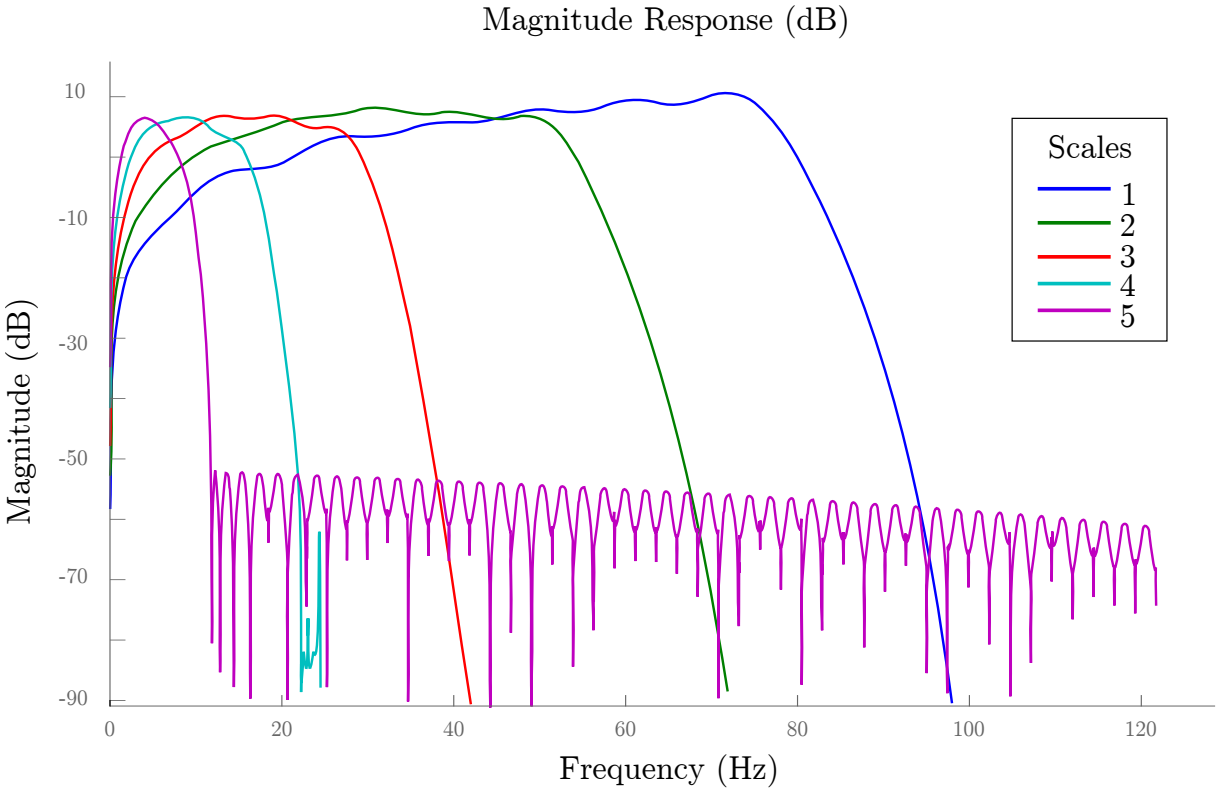


Figure 2.5: Transfer functions of the filter-bank used to calculate the DWT up to the 5th scale.

at 250 Hz. Such a procedure is required to construct a system able to handle equivalently ECG signals with different sampling frequencies. As a final precision of Figure 2.5, the DWT filter-bank response is 250 Hz, but the F_s for this implementation is 360 Hz, which is the sampling rate of the recordings included in the MITBIH-AR.

Following the conclusions of [Martínez et al., 2004], the resulting DWT framework allows an analysis robust to the typical interferences present in routine ECG recordings, so the features derived from the DWT are expected to inherit this desirable property.

2.4 Heartbeat classification

This section includes much of the work carried out in this thesis. In the first two subsections are described the features and classifiers used, which ultimately are the methodologies that will be used for the classification of heartbeats. The other subsections are important during the development of the classification algorithm.

2.4.1 Classification Features

In this subsection all the features that were used in this thesis are defined. However, only a small subset of them are selected to be implemented in the final classification model, as will be described below in section 2.4.6. Many of these features were initially designed to

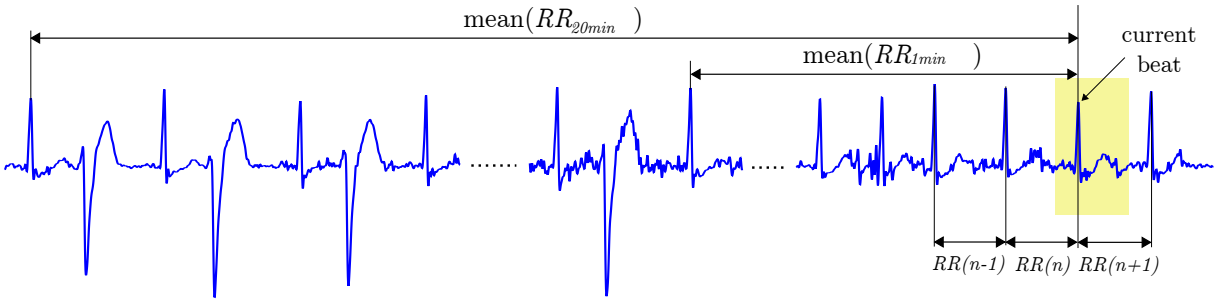


Figure 2.6: Transfer function of the low-pass filter used for ECG preprocessing.

be used in two-lead recordings, as those included in the MITBIH-AR and MITBIH-SUP databases. Later in chapter 4 we present a strategy to adapt these features to recordings with an arbitrary amount of leads.

Following the conclusions of previous works [Hu et al., 1997, de Chazal et al., 2004], we included in our model both rhythm and morphological features. The rhythm features are designed to model the manner in which the heartbeats succeed each other. They are responsible of modeling the sinus node activity, an increased automaticity in any part of the heart or a complete A-V block. As rhythm features we used features from the RR interval (RR) sequence, being

$$RR[i] = Q[i] - Q[i-1] \tag{2.13}$$

the current RR interval the difference between the current ($Q[i]$) and previous ($Q[i-1]$) QRS fiducial points (FP). Then we used $RR[i-1]$, $RR[i]$ and $RR[i+1]$ to describe the local time evolution of the heart rhythm. In order to assess the local variation of the heart rhythm, the feature

$$RR_V[i] = \sum_{j=-1}^{1} |dRR[i-j]|\,, \tag{2.14}$$

being

$$dRR[i] = RR[i] - RR[i-1],$$

characterizes the amount of RR variation in the surrounding heartbeats. The prematurity of a heartbeat, defined as

$$P_{RR}[i] = \frac{RR[i]}{\sum_{k=i-1}^{i+1} RR[k]}, \tag{2.15}$$

measures how anticipated is a heartbeat respect to the previous and next RR interval. We also included estimates of the local and global rhythm by the mean RR interval in the last 1, 5, 10 and 20 minutes (RR_P being $P \in \{1, 5, 10, 20\}$, the interval in minutes of aggregation). Examples of the rhythm features used are shown in Figure 2.6.

The morphological features are designed to model the manner in which heartbeats are conducted through the heart. The features that we used can be grouped in three cate-

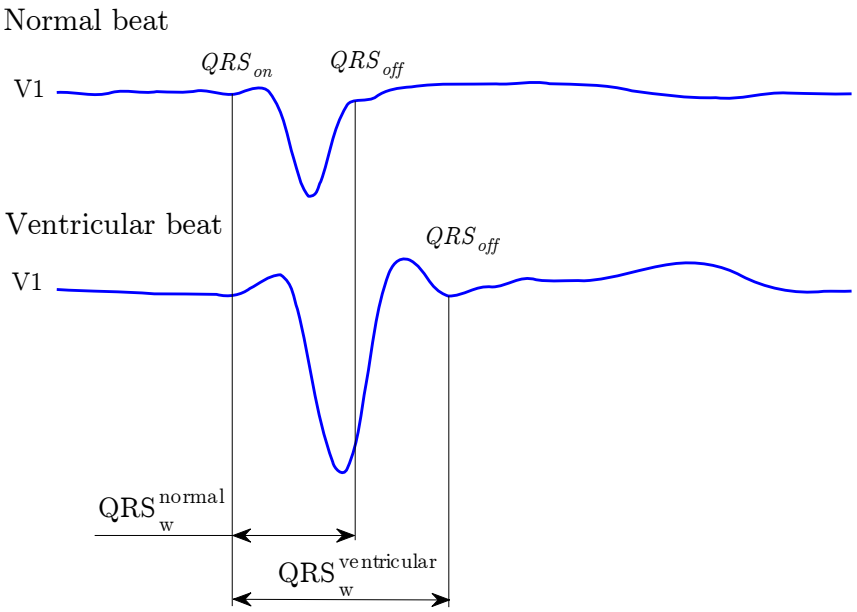


Figure 2.7: QRS width measured for a normal and a ventricular heartbeat.

gories depending on whether they were calculated in the ECG signal, the two-dimensional vectocardiogram (VCG) loop formed by both available leads or in the DWT of the ECG signal.

1) The QRS width is obtained from the delineation of the ECG, as a simple difference

$$QRS_W = QRS_{off} - QRS_{on}$$

of the offset and offset of the QRS complex. The QRS location is assumed to be known, for this purpose we used the annotations included in the databases. Following the QRS complex detection positions, the delineation of each heartbeat was performed with the delineator described in [Martínez et al., 2004].

It is well known that not properly conducted heartbeats, as the ventricular class, have widened QRS complexes as is shown in Figure 2.7. From the clinical point of view, a $QRS_W > 120$ ms is considered to be widened.

2) From the 2-D VCG loop constructed with the two available leads (see Figure 2.8) we calculated two features: the maximal vector of the QRS loop (VCG_M) and the angle of this vector (VCG_ϕ). As any problem in the conduction of the impulse through the heart should change the loop morphology, and therefore this two features, as shown in Figure 2.8 for a normal and a ventricular heartbeat.

3) Regarding the features calculated from the DWT of the ECG, four types can be defined:

3.a) The first type includes 7 features (per lead) that were calculated from peak amplitudes and positions from the fourth scale of the DWT ($W_4^s(k)$), since this scale (between 12.25–22.5 Hz) has good projection of the ECG information. These 7 features are the 2 greatest absolute values of the QRS complex, the 2 greatest absolute values of the T

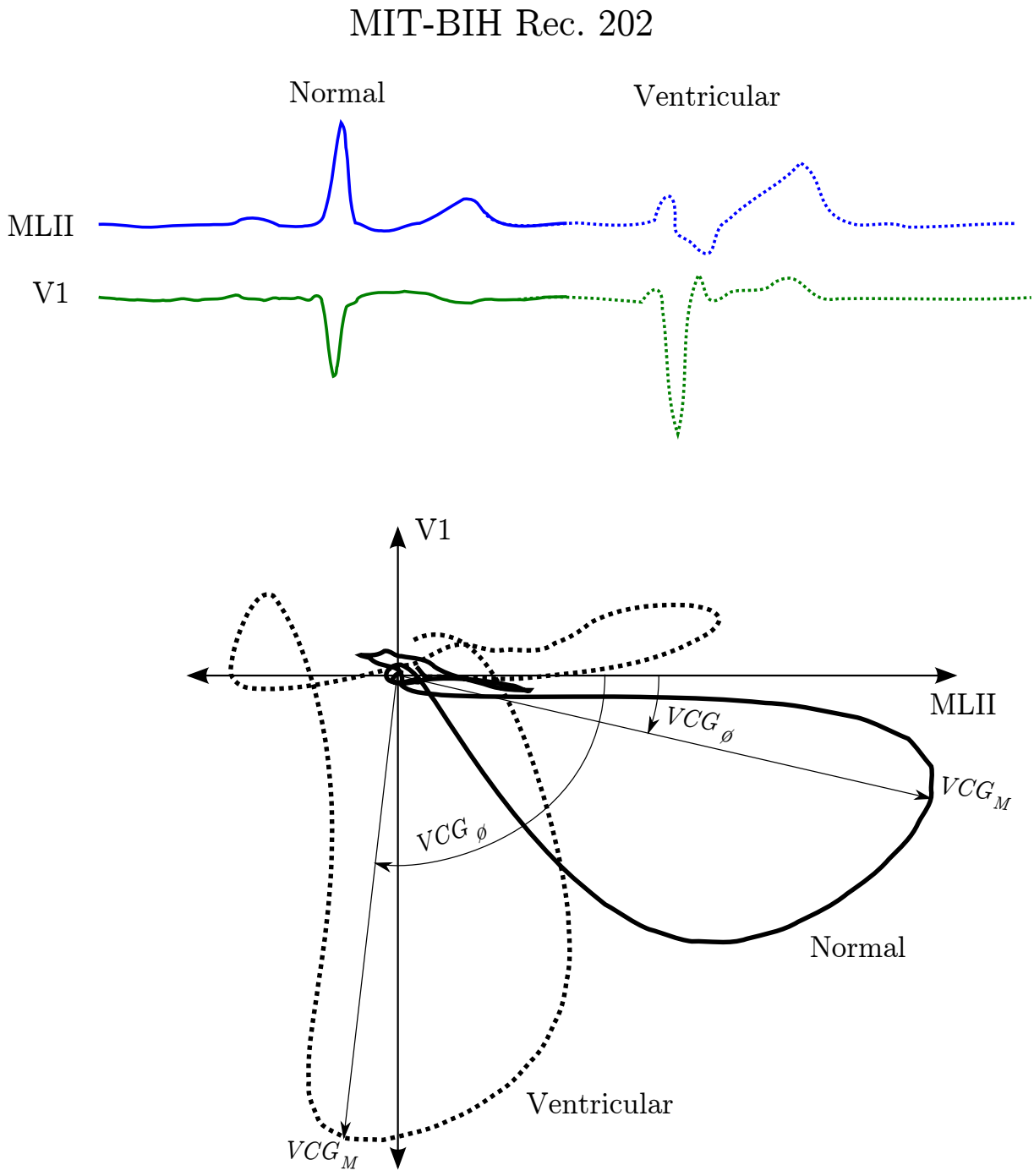


Figure 2.8: Illustration of the features calculated from the VCG loop computed with the two available leads, for a normal (continuous line) and ventricular (dotted line) beats. The maximum value of the loop and the angle at this point are shown.

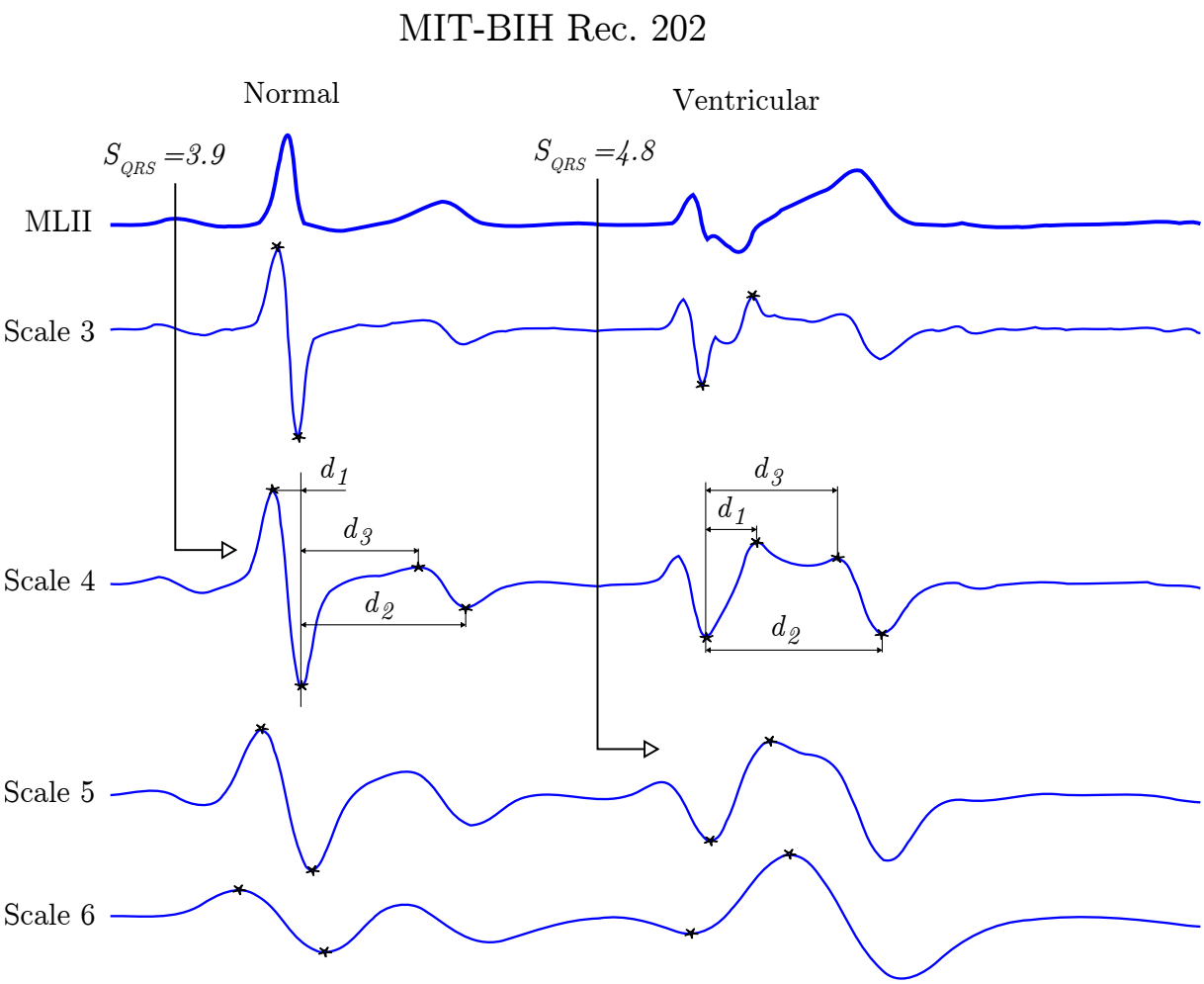


Figure 2.9: Illustration of the features calculated from the wavelet transform for the same normal and ventricular beat in Fig. 2.8. The two most important peaks from the QRS complex and T wave are indicated with an asterisk, and the relative distances (d_i) to the most important peak in the fourth scale. Also the scale where the QRS complex is centered (S^L_{QRS}) is shown for both types of heartbeats used for its calculation (only for one lead).

wave, and their 3 relative positions (to the position of the greatest peak in the heartbeat, see Fig. 2.9).

3.b) The second type is also calculated from the fourth scale of the DWT, but from the autocorrelation sequence

$$r_x^{(4)}(k) = \sum_{i=0}^{N} W_4^x(i) \cdot W_4^x(i-k) \tag{2.16}$$

of the x lead in a window of N samples. This sequence of $2N-1$ samples is symmetric, as can be seen in Figure 2.10. The cross-correlation is very similar

$$r_{xy}^{(4)}(k) = \sum_{i=0}^{N} W_4^x(i) \cdot W_4^y(i-k) \tag{2.17}$$

but involves the two available leads, x and y. In contrast, the cross-correlation sequence

is not symmetric as shown in the bottom of Figure 2.10.

The autocorrelation sequence for both leads ($r_x^{(4)}(k)$ and $r_y^{(4)}(k)$) and the inter-lead cross-correlation sequence ($r_{xy}^{(4)}(k)$) were calculated within a time window which starts 130 ms before the FP and ends 200 ms after. One remarkable aspect is that features calculated from the correlation signals will essentially be synchronized in time, even if the FP is not accurately determined. We calculated for the 3 signals the location and value of the absolute maximum, and for $r_x^{(4)}$ and $r_y^{(4)}$ the location of the first zero-crossing, as shown in Fig. 2.10.

3.c) The third type of feature is the wavelet scale where the QRS complex is centered for each lead. It is known that fast evolving signals (like a normal beat) tend to have their energy located in lower wavelet scales (higher frequency content). The QRS center scale for each lead (S_{QRS}^{Lead}) is calculated as the weighted sum

$$S_{QRS}^{L} = \frac{\sum_{s=1}^{6} A_s^L . s}{\sum_{s=1}^{6} A_s^L} \tag{2.18}$$

where A_s^L is the mean absolute amplitude of the QRS peaks at scale s of the DWT, and lead L

$$A_s^L = \frac{1}{D} \sum_{d=1}^{D} \left| W_s^L s(l_d) \right|, \quad s = 1, 2, \ldots 6 \tag{2.19}$$

being D the number of detected peaks (1 or 2) and l_d the positions of the peaks. In Figure 2.9 the peaks are marked with asterisks, and an arrow points the S_{QRS} feature for a normal and ventricular heartbeats.

3.d) The last type of features computed in the DWT were energies measured at several scales and locations of the heartbeat. Specifically during the occurrence of the P wave, and the QRS complex. For that purpose we defined a window of length 400 ms ending at the QRS_{on}. This window was divided in 4 sections of 100 ms, and the energy of the $W_s^{RMS}(k)$ signal was computed in each section in scales 3 to 5. This amounts 12 features per beat (4 sections $\times$ 3 scales). The same concept was used to study the QRS complex morphology with a window starting at 100 ms before the FP, or the QRS complex location, and ending 120 ms after, for scales 2 to 5, as can be seen in Figure 2.11. In total 16 features per beat.

2.4.2 Discriminant Functions

In this thesis we focused our efforts in developing a classification model which includes features that provide most of the classification power. To encourage this property we used in our experiments very simple classifiers as linear and quadratic discriminant classifiers (or functions). All these classifiers assume that the data is distributed following a Gaussian distribution, which is rarely true. In spite of this evident limitation, these classifiers obtain an acceptable trade-off between the sub-optimal performance achieved, and the simplicity to understand them during operation, and to interpret the effect of their features in the

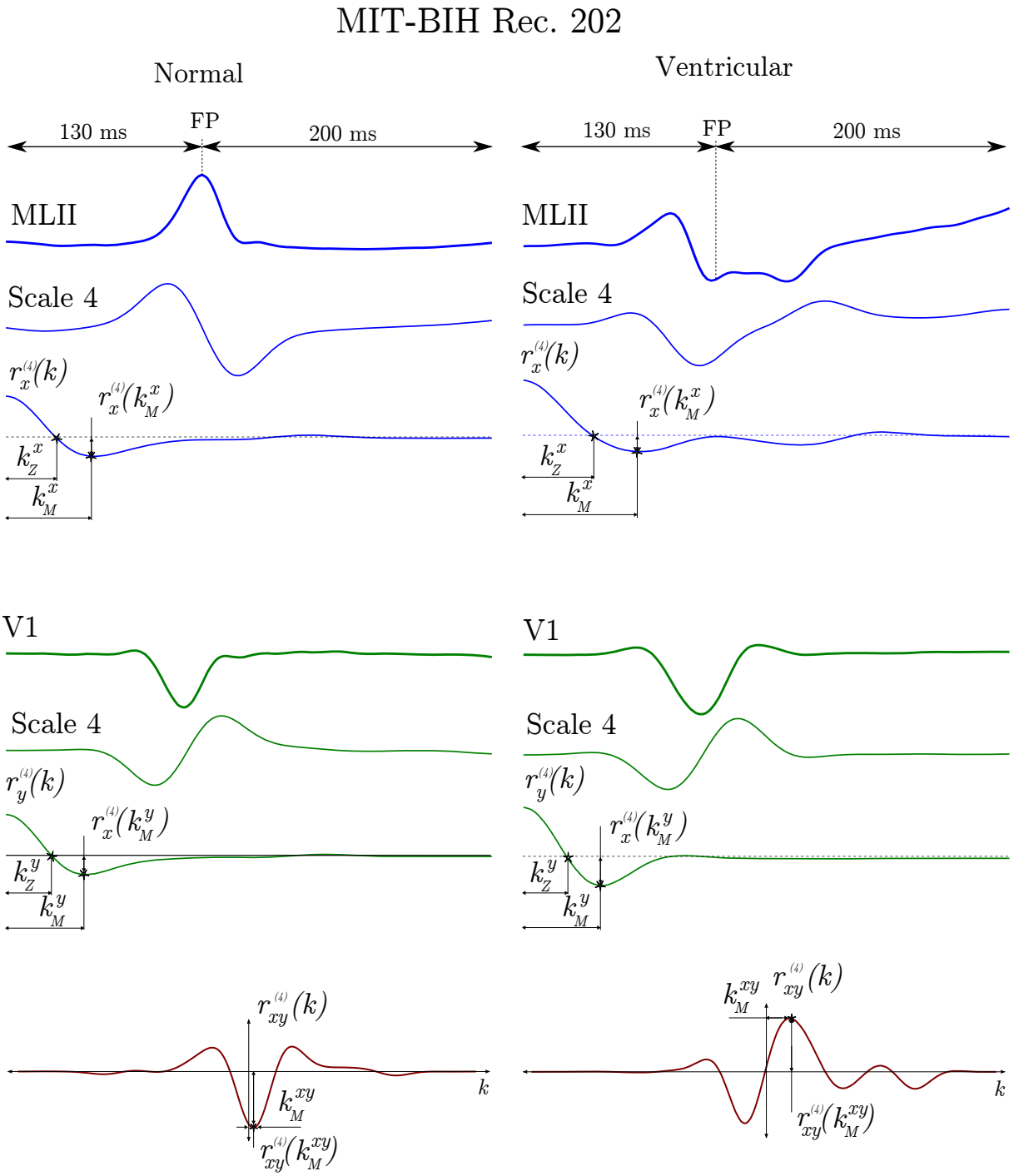


Figure 2.10: Illustration of the features calculated from the wavelet correlation signals for the same normal and ventricular beats. The autocorrelation signal of the QRS complex at scale 4 is shown for both leads (r_x and r_y) as well as the cross-correlation signal (r_{xy}) at the bottom. The zero-crossings and peaks of interest are indicated with an asterisk.

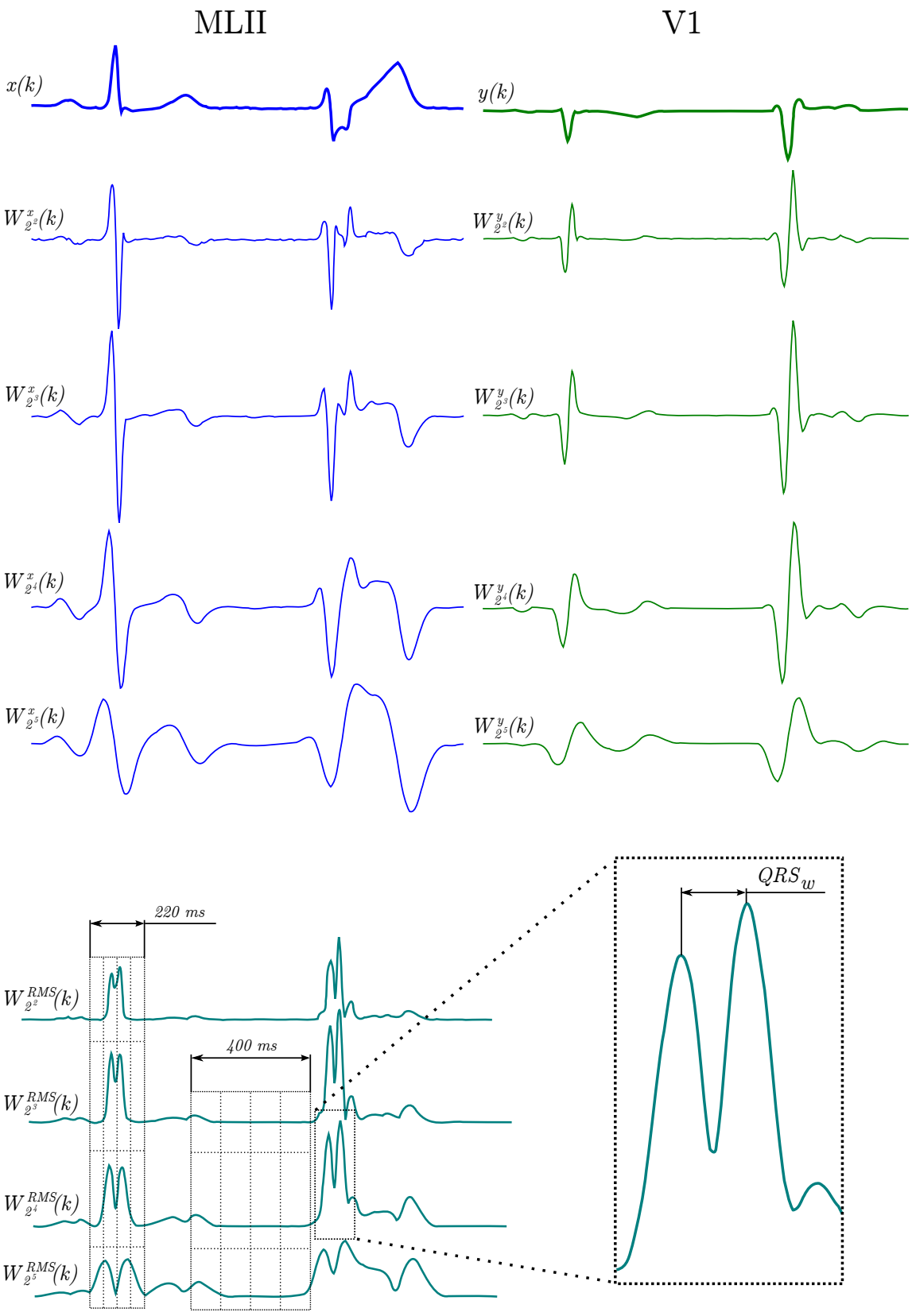


Figure 2.11: Excerpt of record 201 of MIT-BIH database. Normal (N) and ventricular (V) AAMI class heart beats. In the top Figures both ECG leads are shown with their corresponding wavelet decomposition (scales 2-5). The lower panel depicts the RMS composition of both leads wavelet transform ($W_s^{RMS}(k)$). Some features measured in the $W_s^{RMS}(k)$ signal are also shown.

discriminant functions. This later property is not feasible in other classifiers as support vector machines and neural networks for example.

Under the assumption of normally distributed data, the maximum *a posteriori* classification criterion (MAP) leads to quadratic discriminant functions, broadly used for classification purposes [van der Heijden et al., 2005]. In the general case, the quadratic discriminant function of the i-th class and feature vector $\mathbf{x}$, can be written as

$$g_i(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{x} + \boldsymbol{\mu}_{\mathrm{i}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{i}}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\log(|\boldsymbol{\Sigma}_i|) + \log(P(\omega_i)), \tag{2.20}$$

being $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$ and $P(\omega_i)$ the mean vector, covariance matrix and prior probability of the i-th class. The classification rule assigns $\mathbf{x}$ to the class i which results in the maximum posterior probability

$$P(i|\mathbf{x}) = \frac{e^{g_i(\mathbf{x})}}{\sum_{j=1}^{C} e^{g_j(\mathbf{x})}}. \tag{2.21}$$

As can be seen in (2.21), the denominator is equal for all classes and finding the maximum posterior probability is equivalent to find the maximum $g_i(\mathbf{x})$. The values of $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ were computed from the training data with the sample mean and covariance matrix expressions

$$\boldsymbol{\mu}_i = \frac{1}{M_i}\sum_{m=1}^{M_i}\mathbf{x}_m \tag{2.22}$$

$$\boldsymbol{\Sigma}_i = \frac{1}{M_i - 1}\sum_{m=1}^{M_i}(\mathbf{x}_m - \boldsymbol{\mu}_i)\cdot(\mathbf{x}_m - \boldsymbol{\mu}_i)^{\mathrm{T}} \tag{2.23}$$

being M_i the number of examples ($\mathbf{x}_m$) of the i-th class. The values for the prior probabilities $P(\omega_i)$ were considered the same for all classes. In the case that the covariance matrix $\boldsymbol{\Sigma}$ is considered to be the same for all classes ($\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}, \forall i \neq j$), the quadratic discriminant classifier (QDC) becomes linear in $\mathbf{x}$ leading to the linear discriminant classifier (LDC)

$$g_i(\mathbf{x}) = \boldsymbol{\mu}_{\mathrm{i}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \log(P(\omega_i)), \tag{2.24}$$

where $\boldsymbol{\Sigma}$ can be estimated as the weighted sample covariance

$$\boldsymbol{\Sigma} = \frac{1}{\sum_{i=1}^{C} w_i}\sum_{i=1}^{C}\frac{w_i\sum_{m=1}^{M_i}(\mathbf{x}_m - \boldsymbol{\mu}_i)\cdot(\mathbf{x}_m - \boldsymbol{\mu}_i)^{\mathrm{T}}}{M_i}, \tag{2.25}$$

being C the total amount of classes and w_i the class-weighting coefficients. This class-weighting possibility is of much interest due to the heavy imbalance of the class-sizes inherent to this application, where the normal class is in general one order of magnitude (at least) more represented than other classes. We refer as LDC to the linear classifier where $w_i = w_j, \forall i \neq j$, any other weight scheme will be referred as compensated linear

classifier (LDC-C). The weights used in the following chapters were $w_S = 10$; w_V=10 and $w_N = 1$. In this thesis, all classification tasks were performed using and modifying the PRtools toolbox [Duin et al., 2008] for Matlab (The Mathworks Inc., MA).

For the classifiers described above, the procedure of estimating the parameters $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$ or $\boldsymbol{\Sigma}$ is known as *training*. Then the process of using these parameters in equations (2.20) or (2.24) for every unseen or unlabeled feature vector $\mathbf{x}$, is known as *testing*. The results of equations (2.20) and (2.24) are related to *a posteriori* probabilities, and as said above, the classification rule assign the class i which results in the maximum $g_i(\mathbf{x})$.

In Figure 2.12 the discriminant functions for three classifiers trained in the same data are shown. In the top panels, the difference between the LDC and LDC-C is shown. For the LDC-C the two less-represented classes (S and V classes) has $w_i = 10$. Note the different orientation of the covariance matrices, represented as ellipses, estimated in both cases. In the lower panels the QDC is shown with equal priors and with a prior weights scheme similar to the LDC-C explained before, the two less-represented classes (S and V classes) has $P(\omega_{S,V}) = 10 \cdot P(\omega_N)$. Note in this last case how the red area decrease as the normal class is less probable *a priori*.

2.4.3 Domain Handling for some Features

As the features to include in our model belong to diverse domains, like $\mathbb{R}$, $\mathbb{R}^+$and S^2 (angular or directional domain) we have to transform or deal with them in order to perform classification tasks. In our case, we assume that each feature is normally distributed and therefore valid in the $\mathbb{R}$ domain. According to this, all rhythm and morphological features defined in $\mathbb{R}^+$ are first being transformed to the $\mathbb{R}$ domain by a (natural) logarithm operation. In contrast, circular (or S^2) features require an special treatment that will be briefly described, the interested reader is referred to [Bahlmann, 2006] for details. First considering the directional feature ϑ as the argument of a complex number of unitary module $e^{j\vartheta}$ (being $j^2 = -1$), the expectation of this modified feature defines the mean direction and directional variance [Bahlmann, 2006], counterparts of the regular mean and variance.

$$E[e^{j\vartheta}] = \rho_\vartheta e^{j\mu_\vartheta^c} \tag{2.26}$$

$$\mu_\vartheta^c = arg(E[e^{j\vartheta}]) \tag{2.27}$$

$$V_\vartheta^c = 1 - \left\| E[e^{j\vartheta}] \right\| = 1 - \rho_\vartheta \tag{2.28}$$

Where ρ_ϑ is also known as the resultant length. Then for a multivariate model, where F is the set of indexes of the directional features, the maximum likelihood estimates are

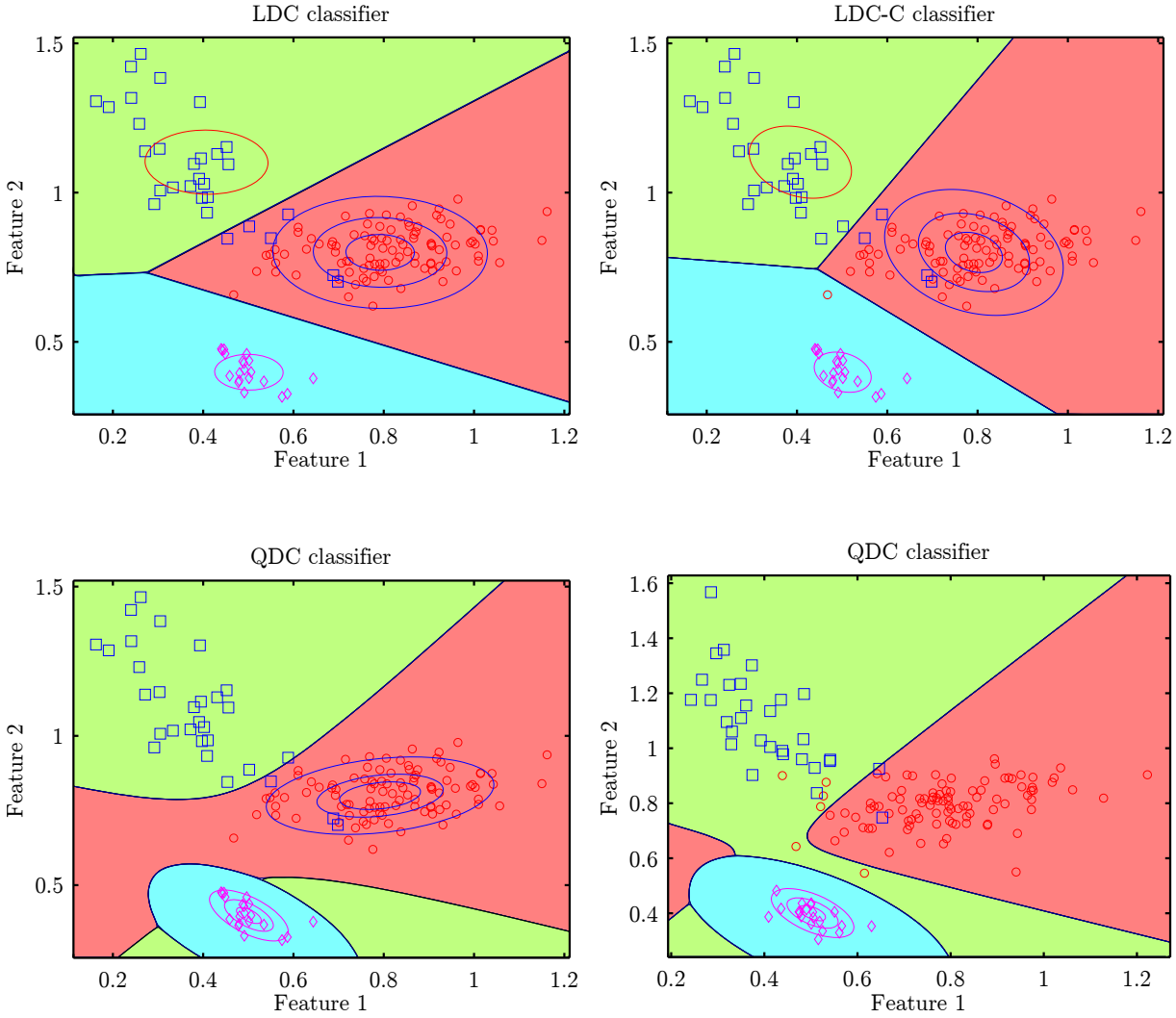


Figure 2.12: The three classifiers, and their discriminant functions, for a three imbalanced-classes problem. Note the different regions caused by the classifiers trained over the same data. The imbalance is evidenced in the difference between the LDC and the LDC-C. See in the lower panels the effect of the class priors.

$$(\hat{\mu}_x)_f = \begin{cases} \frac{1}{M}\sum_{m=1}^{M} x_f(m) & \text{if } f \notin F \\ \arg\left(\frac{1}{M}\sum_{m=1}^{M} e^{\left(jx_f(m)\right)}\right) & \text{else} \end{cases} \tag{2.29}$$

$$\hat{\Sigma}_x = \frac{1}{M-1}\sum_{m=1}^{M} \mathbf{x}^{'}(m).\mathbf{x}^{'}(m)^{\mathrm{T}} \tag{2.30}$$

being

$$\mathbf{x}^{'}(m) = \left[x_1^{'}(m)\ldots x_F^{'}(m)\right]^T \tag{2.31}$$

$$x_f^{'}(m) = \begin{cases} x_f(m) - (\hat{\mu}_x)_f & \text{if } f \notin F \\ \left(x_f(m) - (\hat{\mu}_x)_f\right)_{mod\ 2\pi} & \text{else} \end{cases} \tag{2.32}$$

As it can be noted from equation (2.29), $\hat{\Sigma}_x$ can be easily calculated from the directional mean $(\hat{\mu}_x)_f$ and the raw data. Some examples illustrate the use of the equations described in the last two subsections. For this purpose we will use two Gaussian bivariate datasets, both with 2 classes which can be easily separated by a linear function. The first dataset has its two features directionals, and will be referred as the *wrapped dataset*. We will talk about the other dataset later in this subsection. The wrapped dataset is shown in Figure 2.13 and has the following design parameters:

$$\mu_1 = \begin{pmatrix} -\frac{3}{5}\pi & \frac{3}{5}\pi \end{pmatrix}$$

$$\mu_2 = \begin{pmatrix} \frac{3}{5}\pi & -\frac{3}{5}\pi \end{pmatrix}$$

$$\Sigma_1 = \Sigma_2 = \begin{pmatrix} -\frac{2}{3}\pi & 2 \\ 2 & -\frac{2}{3}\pi \end{pmatrix}$$

This is a simple 2-class example, very illustrative of two features that linearly separates the classes, but not alone in their own. The typical example of *good* (in a class separation sense) correlation between features. In the top panel the dataset is plotted as if both features were linear, but showing a dashed box evidencing where the data should be confined in case of directional features. Also the underlying linear distribution is showed by equiprobabilistic contours. In the lower panel, the same dataset is showed, but considering both features as directional. Note that in this case the tails of the Gaussian that fall outside the domain of the feature, in this case$(-\pi : \pi]$, are translated $\pm 2\pi$ to fall into the directional domain again. The resulting distribution, from a linear point of view, is clearly not Gaussian in the wrapped space, as it can be seen by the two evident modes present for both classes.

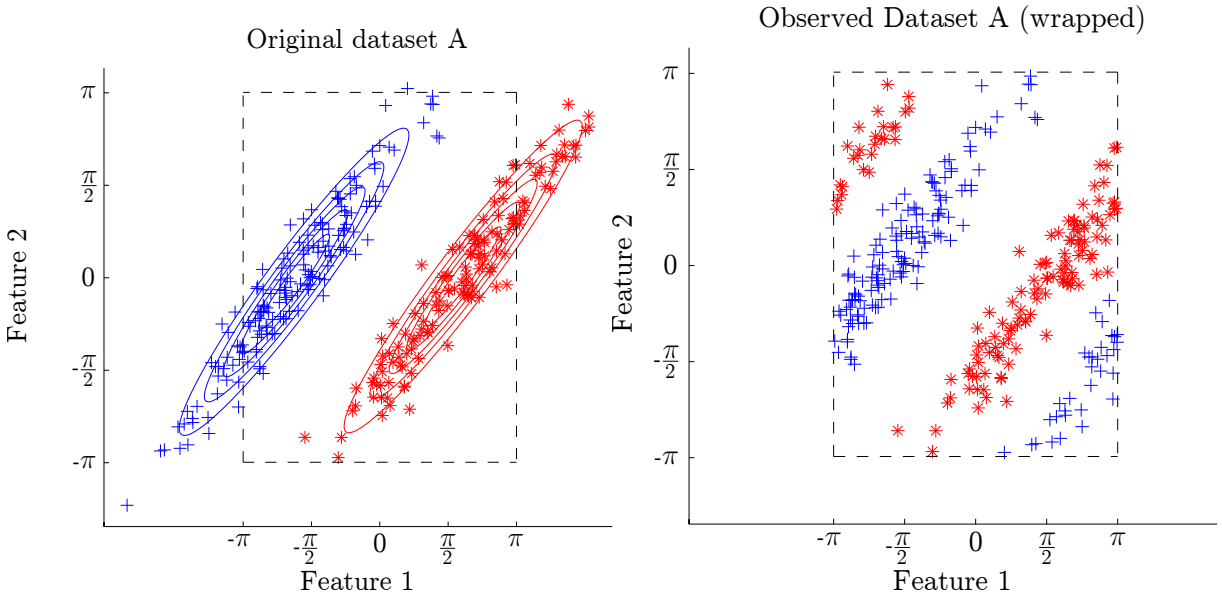


Figure 2.13: Example of wrapped dataset A showing the original dataset before and after considering feature 1 and 2 as directional in the $(-\pi : \pi]$, domain.

The other dataset has the first feature directional, and the second linear, and is defined as the *semi-wrapped* or mixed dataset. This dataset has also two classes with the following parameters:

$$\mu_1 = \begin{pmatrix} -\frac{3}{5}\pi & 0 \end{pmatrix}$$

$$\mu_2 = \begin{pmatrix} \frac{3}{5}\pi & 0 \end{pmatrix}$$

$$\Sigma_1 = \Sigma_2 = \begin{pmatrix} -\frac{2}{3}\pi & 2 \\ 2 & -\frac{2}{3}\pi \end{pmatrix}$$

The same observations described above are valid on the first feature, as can be seen in Figure 2.14.

At this point should be clear that the equations (2.22) and (2.23) are not valid in the context of directional features. Finally in Figures 2.15 and 2.16, equations (2.29) and (2.29) are used with our example datasets to evidence the parameterization results considering both features directionals (directional model), one directional and the other linear (semi-wrapped or mixed) and both linear features (linear model).

2.4.4 Outlier Removal

The classification performance proposed strongly depends on the parameter estimation of the multidimensional Gaussians in the training datasets. The parameter estimation (or training) process can be disrupted mainly by the lack of Gaussianity of the data, and in second degree by the presence of outliers. The presence of outliers in the training

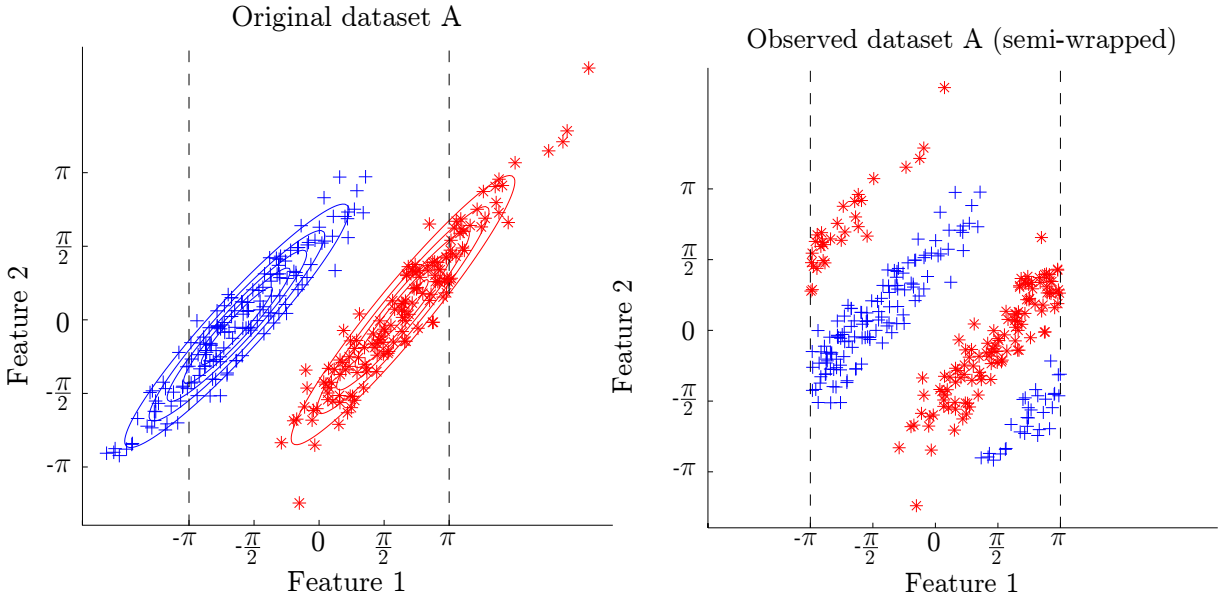


Figure 2.14: Example of semi-wrapped dataset A showing the original dataset before and after considering feature 1 as directional in the $(-\pi : \pi]$, domain.

data can be addressed by the removal of these atypical examples prior to the parameter estimation process. In this work the outliers removal is performed by the algorithm suggested in [Filzmoser et al., 2008], which is a projection pursuit method based on the robust estimation of the translation, scale and kurtosis of the distribution. The translation is estimated by the median, defined for K dimensional data as

$$\underset{i}{\text{med}}\, \mathbf{x}_i = \underset{\boldsymbol{\mu}\in\mathbb{R}^K}{\arg\min} \sum_{i=1}^{N} \|\mathbf{x}_i - \boldsymbol{\mu}\| \tag{2.33}$$

while the scale is estimated by the median absolute deviation (MAD) calculated as

$$\text{MAD}(\mathbf{X}) = 1.4826 \cdot \underset{j}{\text{med}} \left| x_j - \underset{i}{\text{med}}\, \mathbf{x}_i \right| \tag{2.34}$$

finally the kurtosis estimation is performed with the expression

$$\text{KUR}(\mathbf{X}) = \left| \frac{1}{N} \sum_{j=1}^{N} \frac{\left(\mathbf{x}_j - \text{med}_i\, \mathbf{x}_i\right)^4}{\text{MAD}(\mathbf{X})^4} - 3 \right|. \tag{2.35}$$

The presence of a small number of outliers will make the tails of a distribution heavier, increasing the kurtosis coefficient; while a large number of outliers give raise to other modes in the distribution, decreasing the kurtosis coefficient. In a first phase, the algorithm search for outliers in the directions where the kurtosis of the data is large or small to find location outliers. Then, in a second phase, the directions of large variance are explored to address scatter outliers. For both phases each example in the distribution gets one weight, which are finally combined in a final decision weight. Based on the final weight, the data is sorted and the 5% of the most outlying examples are discarded as

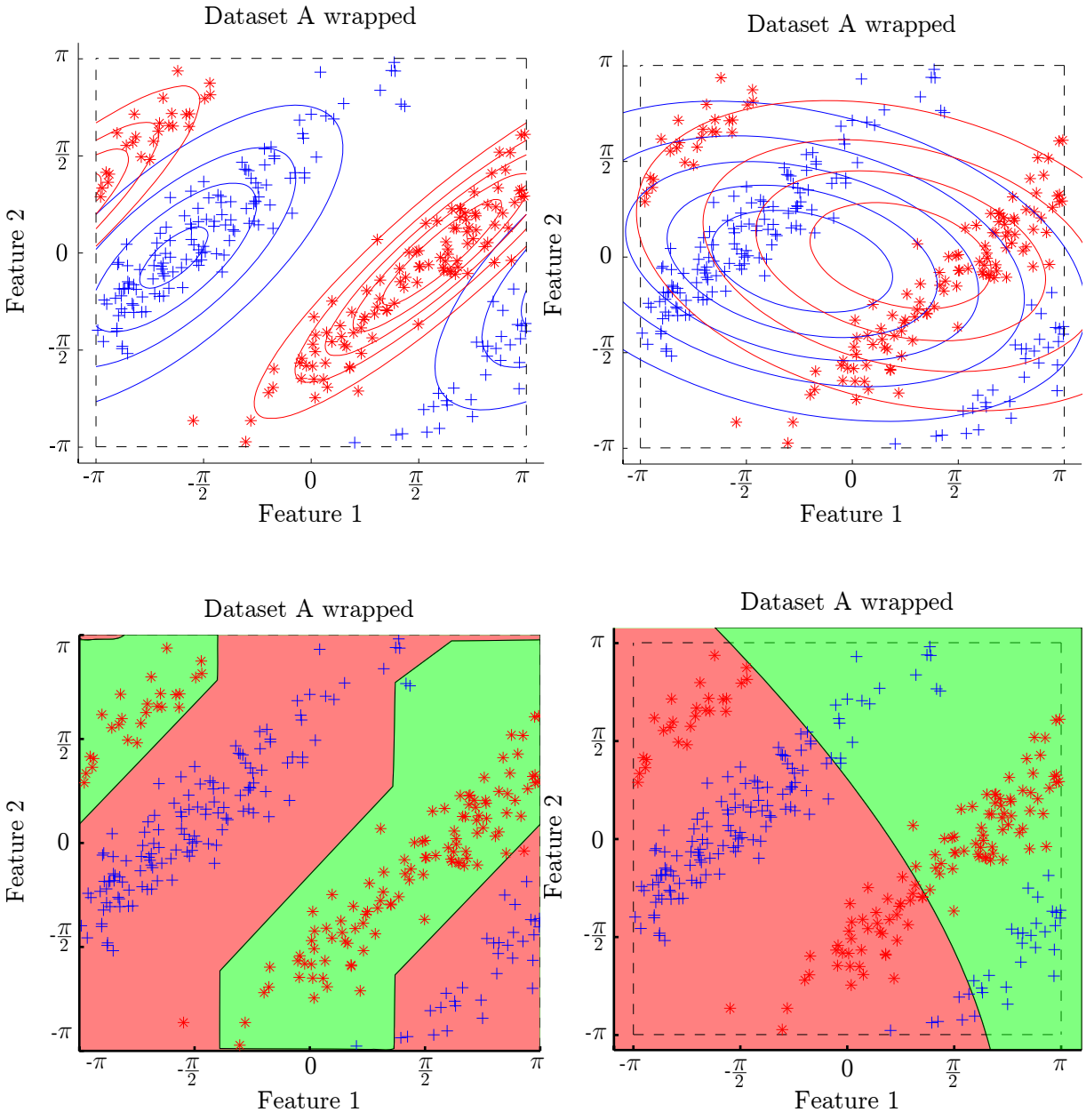


Figure 2.15: In the top panels, *wrapped* dataset using the approximated wrapped Gaussian distribution and the linear Gaussian distribution respectively. In the bottom panels the decision region for both classes and both distributions is showed. Region filled with red color is for Class 1, corresponding to the blue crosses. Note the huge difference on classification performance.

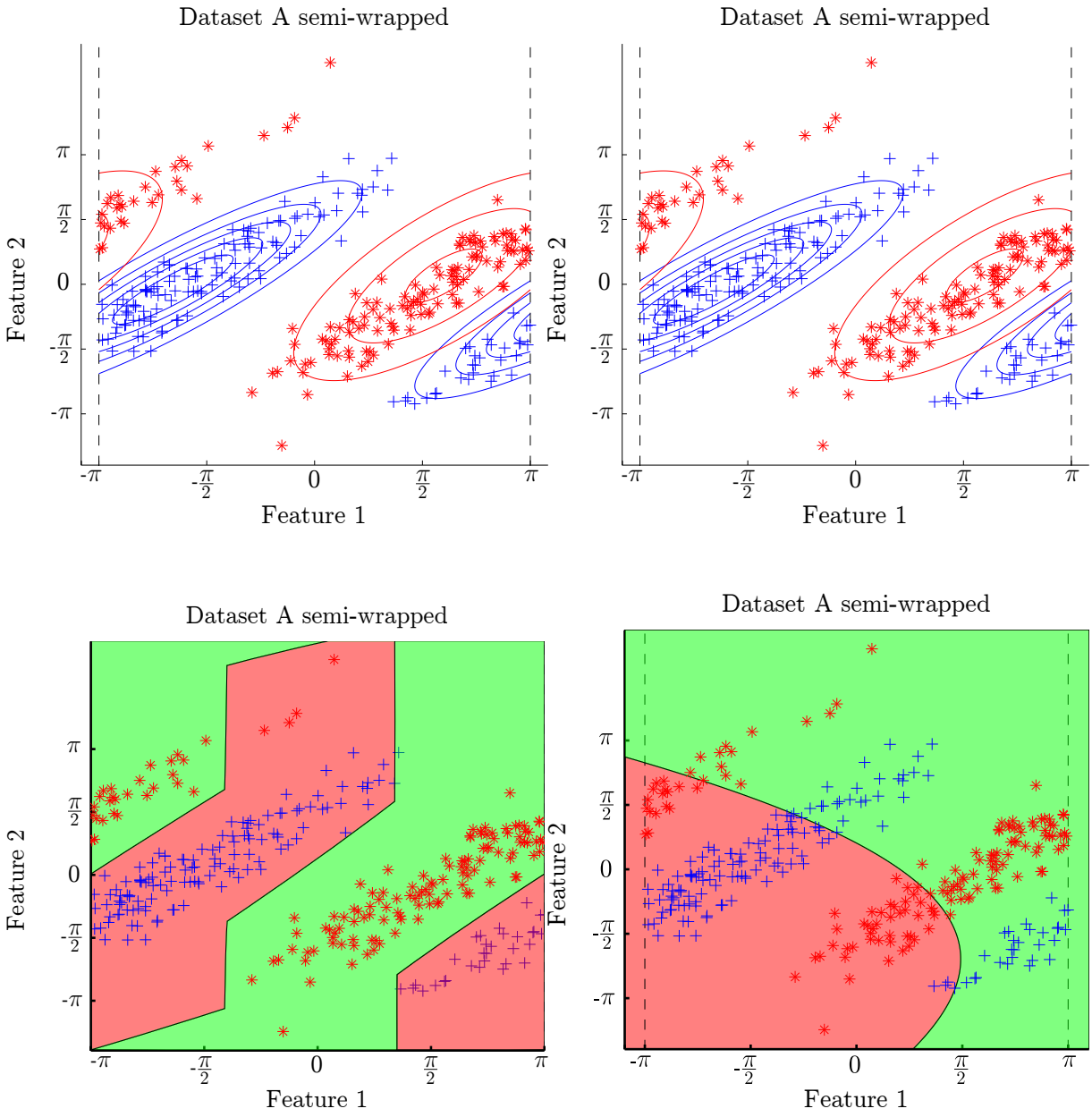


Figure 2.16: In the top panels, *semi-wrapped* dataset using the approximated wrapped Gaussian distribution and the linear Gaussian distribution respectively. In the bottom panels the decision region for both classes and both distributions is showed. Region filled with red color is for Class 1, corresponding to the blue crosses. Note the huge difference on classification performance.

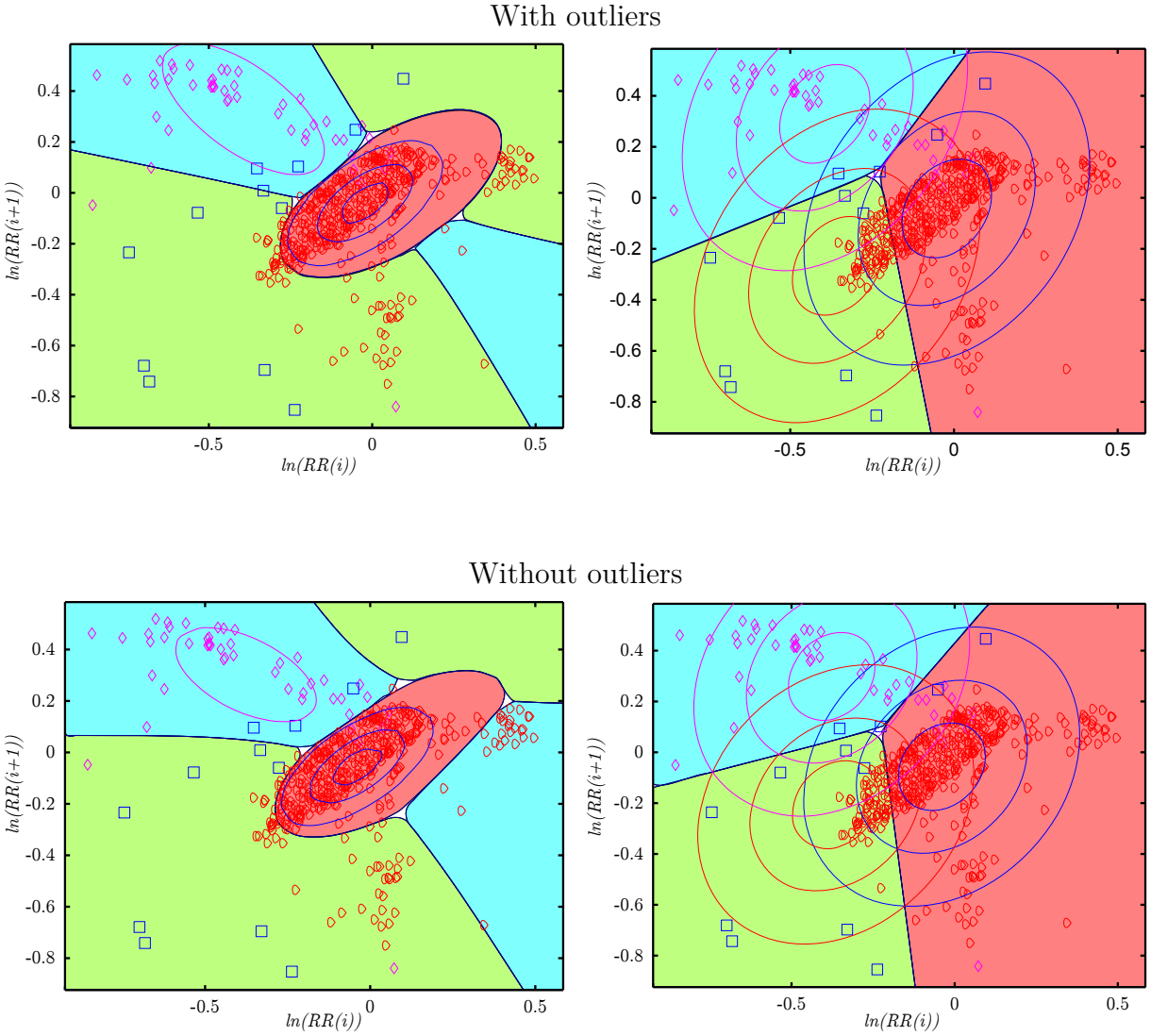


Figure 2.17: Example of the outliers removal when estimating the parameters for both types of classifiers used. In both examples the features used are $RR[i-1]$ and $RR[i]$. Note the different shape of the classification regions as a consequence of the the outliers removal in the lower panels.

outliers. With this assumption of slightly contaminated data, we set an operating point for the trade-off between discarding useful data and allowing the presence of outliers in the parameter estimation process. In Figure 2.17 the use of the algorithm for outliers removal is evidenced in the classification regions, as can be further seen for the QDC classifier.

2.4.5 Performance evaluation

We use three approaches to evaluate the performance of a classification experiment. The first consists in estimating the parameters of the classifier in a training dataset, then with these parameters, evaluate their predictive performance in a disjoint test dataset. The data division should be performed by subject, which means that feature-vectors (or

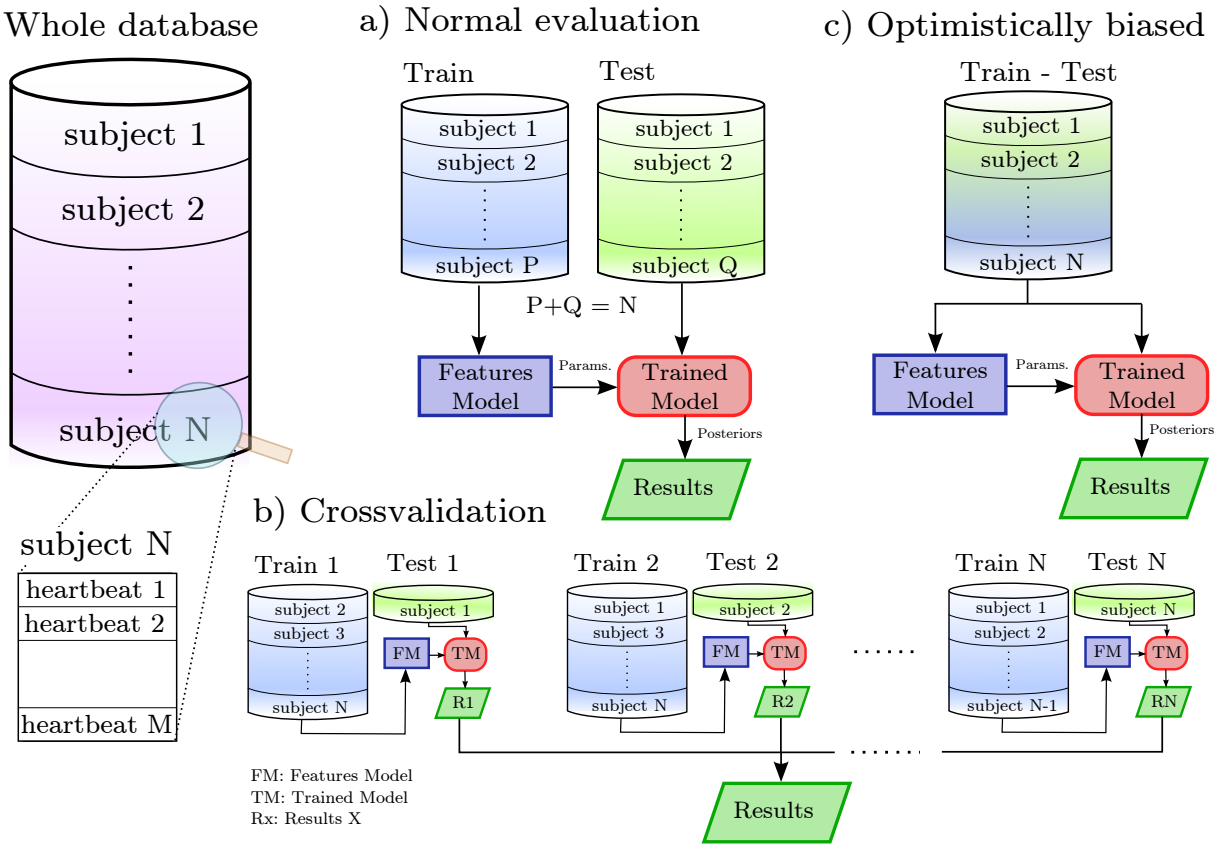


Figure 2.18: Scheme showing the three data division methods used.

heartbeats) of the same subject can not be included in both the training and testing sets. In [de Chazal et al., 2004] it was shown that the heartbeat oriented data division leads to an optimistic bias in the performance estimation calculated in the test dataset.

The second approach used is known as cross-validation, and is adopted when there are not many examples to build the train and test datasets. It consists in dividing a dataset into k disjoint folds (we used 10 folds), and use them as test sets, obtaining therefore k performance measures, one for each fold. The division is again performed by patients, to avoid the presence of heartbeats of the same patient in both training and test datasets. Note that each cross-validation fold implies training in $^{k-1}/_{k}$ of the database patients, and testing in the remaining $^{1}/_{k}$. The resulting performance is the mean of the k performances.

In the last approach, although training and testing in the same dataset leads to an optimistically biased performance estimation, this measure could be useful to have an idea of how good can perform a given classifier in the most favorable situation. This optimistically biased performance serves as an upper bound, and represents the performance of the classifier if the probability distributions of examples in both training and test datasets were identical. In Figure 2.18 a scheme shows the data division methods used.

Now we focus in how the performance is evaluated for any of the approaches described above. According to Section 4.3 of [AAMI-EC57, 1998–2008], the end product of a beat-by-beat comparison between the algorithm and reference labels is a matrix in which each element is a count of the number of beat label pairs of the appropriate type:

Reference label \ Algorithm label	N	s	v	f	q
N	Nn	Ns	Nv	Nf	Nq
S	Sn	Ss	Sv	Sf	Sq
V	Vn	Vs	Vv	Vf	Vq
F	Fn	Fs	Fv	Ff	Fq
Q	Qn	Qs	Qv	Qf	Qq

Moreover, from Subsection 2.4.2, we showed that the outcome of the LDC, LDC-C or QDC classifiers (eq. 2.20 and 2.24), were the posterior probabilities of the classes for each of the heartbeats presented to a trained classifier (see also Figure 2.18). As said before, for each heartbeat the classification rule selects the class with greater posterior. Therefore, given the ground truth or true classes for each example, the number of possible outcomes for a multiclass classification problem are $C \times C$. The amount of events for each of the possible outcomes are represented in a square matrix of dimension C, known as confusion matrix. This is the same matrix with AAMI classes, but for any multiclass problem.

$$\begin{array}{c} \\ \text{True classes} \end{array} \begin{array}{c} \text{Estimated classes} \\ \begin{array}{ccccccc} 1 & & \dots & & i & \dots & C \end{array} \\ \begin{array}{c} 1 \\ \vdots \\ i \\ \vdots \\ C \end{array} \begin{pmatrix} n_{11}^T & \dots & n_{1i}^F & \dots & n_{1C}^F \\ \vdots & \ddots & \vdots & & \\ n_{i1}^F & \dots & n_{ii}^T & \dots & n_{iC}^F \\ \vdots & & \vdots & \ddots & \\ n_{C1}^F & \dots & n_{Ci}^F & \dots & n_{CC}^T \end{pmatrix} \begin{array}{c} N_1 \\ \vdots \\ N_i \\ \vdots \\ N_C \end{array} \\ \begin{array}{ccccccc} P_1 & \dots & P_i & \dots & P_C \end{array} \quad N_T. \end{array}$$

For the i-th class n_{ii}^T is the number of correctly classified examples and n_{ij}^F is the number of examples of class i classified as class j; N_i is the total number of examples for class i, P_i is the number of examples classified as class i and N_T is the total number of examples in the dataset.

$$N_i = n_{ii}^T + \sum_{m \neq i} n_{im}^F$$

$$P_i = n_{ii}^T + \sum_{m \neq i} n_{mi}^F$$

$$N_T = \sum_{i=1}^{C} N_i = \sum_{i=1}^{C} P_i$$

The performance is calculated from this matrix, for each class, in terms of the class sensitivity (S_i) and class positive predictive value (P_i^+)

$$S_i = \frac{n_{ii}^T}{N_i} \tag{2.36}$$

$$P_i^+ = \frac{n_{ii}^T}{P_i}, \tag{2.37}$$

or globally, with the global accuracy (A), sensitivity (S) and positive predictive value (P^+)

$$A = \frac{1}{N_T}\sum_{i=1}^{C} n_{ii}^T = \sum_{i=1}^{C}\frac{N_i}{N_T}S_i \tag{2.38}$$

$$S = \frac{1}{C}\sum_{i=1}^{C} S_i \tag{2.39}$$

$$P^+ = \frac{1}{C}\sum_{i=1}^{C} P_i^+ \tag{2.40}$$

as suggested in [AAMI-EC57, 1998–2008]. From these equations it is clear that any imbalance in the class representation affects P^+, P_i^+ and A calculation, but not S and S_i. Although the AAMI recommendation does not suggest any measure to deal with the strong class size imbalance (see Table 2.3), we considered weighting the classes previous to the calculation of P_i^+ and A in order not to neglect the performance of the less represented classes. The balancing approach used in this work consists in multiplying each row of the confusion matrix by a constant such that the sum of each row N_i is equal for all classes, or $N_i = N_j, \forall i \neq j$. This is equivalent to repeat examples of the less represented classes, in order to balance the class presence.

2.4.6 Model Selection and Dimensionality Reduction

The amount of features included in a classification model should be limited, mainly, because the training set is rarely able to fully describe the class distributions. Then, one of the greatest challenges during the training of a classifier is to avoid overfitting. Overfitting happens when the trained classifier is adapted in excess to the training examples. It is somehow the opposite of generalization in classification. In section 2.4.2 we presented our first strategy to limit the complexity of our classifier, by using linear or quadratic discriminant functions. In this section we present our approach to control the complexity of the features model. It is well known that low dimensional models generalize better to examples not presented during the training phase, resulting in a more robust and realistic classifier. There are two approaches typically used to achieve this, *feature transformation* (FT) or *feature selection* (FS). As can be anticipated, both have pros and cons. FT can be defined as a mapping

$$\mathbf{y} = \mathbf{W}(\mathbf{z}) \tag{2.41}$$

of the input space of dimension N into another of dimension D, given that $D \ll N$. One of the advantages of FT is that take into account all the features in the input space, which can also be thought as a disadvantage in case that some features are useless

for classification. There are many criteria to the design of the mapping function W, which imposes assumptions over the data [van der Heijden et al., 2005]. The same design requirements in term of training data, overfitting and generalization should be taken into account for the correct development of W. This means that the design of W is as important as the training process itself.

On the other hand, the wrapper approach consists in reducing the feature space by selecting the D most important features given a criterion function $J(\cdot)$. We can comment similar pros and cons of FT for FS. However, a property of FS is that the feature model selected is tailored for a specific classifier. If the classifier changes, there is no guarantee that the same feature model maximizes the criterion $J(\cdot)$. In general, the FS problem consists in evaluating $J(\cdot)$ across several feature models. The exhaustive and optimal evaluation across all model combinations requires $N!$ evaluations, which is unfeasible even for moderate N. However, many algorithms exist to overcome this, obtaining a suboptimal model. In this thesis a sequential floating feature selection algorithm (SFFS) was used [Pudil et al., 1994] to obtain the smallest and best performing model.

The SFFS algorithm can be briefly explained as the combination of two simpler steps, a sequential forward selection (SFS) algorithm followed by a sequential backwards selection (SBS) algorithm. Following Figure 2.19, the SFFS iterates for all model sizes, starting from a model of one feature, and registering all the best performances found for each model size. Each iteration starts with an SFS step, and from a model size greater than two features after each SFS step, an SBS step is repeated until the performance of the model found is not greater than the registered for this smaller model size. This way the algorithms goes forward and backwards (like floating) searching at each step for the path of maximum performance. The algorithm ends when the specified greater model size is reached. The result of the algorithm is the model found with maximum performance. The interested reader is referred to [Pudil et al., 1994] for a detailed description and to [Duin et al., 2008] for an implementation of the SFFS algorithm. Two examples of the results obtained from this algorithm are shown in Figure 3.2.

The performance metrics used by the FS algorithm were a weighted class Se and PP, calculated as

$$J_S = \frac{\sum_{i=1}^{C} \pi_i . S_i}{\sum_{i=1}^{C} \pi_i} \tag{2.42}$$

$$J_{P^+} = \frac{\sum_{i=1}^{C} \pi_i . P_i^+}{\sum_{i=1}^{C} \pi_i} \tag{2.43}$$

with C classes and being S_i and P_i^+ the class sensitivities and positive predictive defined in the previous subsection. The class weights π_i allows the possibility of directing the search to specific class performances, specially in the case of the P^+ which is directly influenced by the class size unbalance. Then, as a result of using the J_S criterion, the

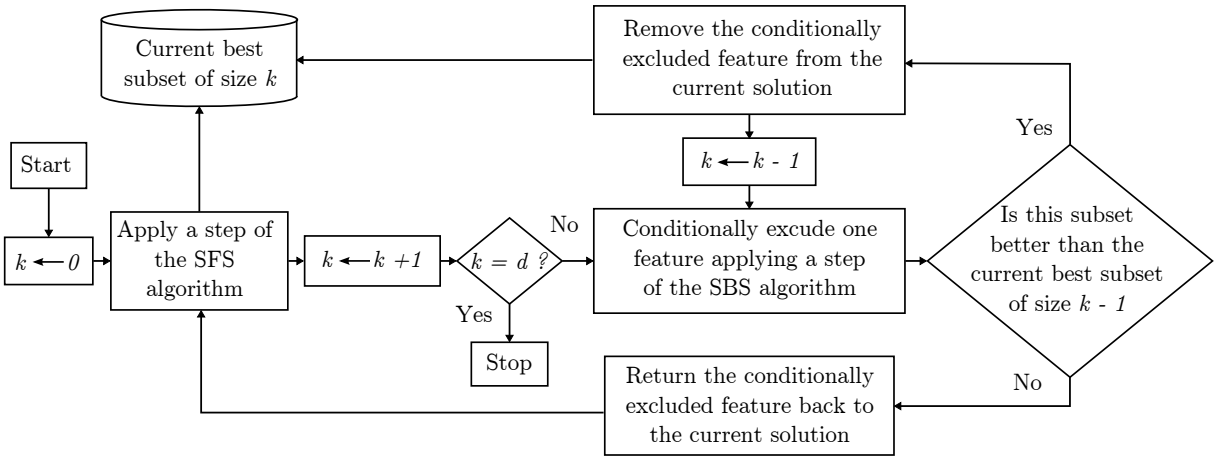


Figure 2.19: Flow diagram of the sequential floating feature selection (SFFS) algorithm used for the feature selection among d features.

SFFS is likely to find sensitive feature models, in the other hand, the J_{P+}criterion will yield specific feature models.

Chapter 3

Automatic ECG Heartbeat Classification

3.1 Introduction

In this chapter we describe the development of a simple heartbeat classifier, based on the methodologies described in the previous chapter, that will be enhanced in the following chapters. In section 1.3 we described other approaches published by other groups, but in our opinion the work of de Chazal *et al.* [de Chazal et al., 2004] is the most representative of the state of the art, and will be used as a reference. The main novelties with respect to [de Chazal et al., 2004] are: 1) the morphological features used are based on the wavelet transform (Section 2.4.1), 2) the use of a feature selection algorithm to select those features with generalization capability (Section 2.4.6) and finally 3) the evaluation of the generalization of the classifier outside the MITBIH Arrhythmia database (MITBIH-AR). The results of this chapter were published in [Llamedo and Martínez, 2011a].

In summary, the objective pursued in this chapter is to develop and evaluate a heartbeat classification algorithm according to the following conditions:

- Perform automatic ECG classification.
- Follow AAMI [AAMI-EC57, 1998–2008] recommendations for class labeling and results presentation.
- Use a simple classifier (as linear or quadratic discriminant functions) to ensure that the classification performance is due to the features selected.
- The features used should have a physiological interpretation, being simple to compute and robust to the typical kind of noise present in the ECG.
- Preference for features that can be used for ECG delineation, as those from the discrete wavelet transform (DWT), since the heartbeat classifier is intended to be used after a DWT based ECG delineator [Martínez et al., 2004].

Table 3.1: Class distribution of the databases used and division of the MITBIH-AR database into training ($DS1$) and testing ($DS2$) sets. Recordings with paced beats were excluded.

		MITBIH-AR					
Dataset	Purpose	N	S	V	F	Q	#Rec
$DS1$	train	45784	940	3783	413	8	22
$DS2$	test	44188	1835	3218	388	7	22
Totals		89972	2775	7001	801	15	44

Heartbeat classes are N: normal, S: supraventricular, V: ventricular and F: fusion

Dataset	MITBIH-AR recordings
$DS1$	101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230
$DS2$	100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234

		Other Databases					
Database	Purpose	N	S	V	F	Q	#Rec
MITBIH-SUP	validation	161902	12083	9897	193	78	78
INCART	test	153517	1958	19991	219	5	75

Heartbeat classes are N: normal, S: supraventricular, V: ventricular and F: fusion

- Use a multidatabase validation approach for feature selection to ensure better generalization properties of the selected feature set.

3.2 Methodology

3.2.1 ECG Databases

In this chapter we used the well-known MITBIH-AR [Moody and Mark, 2001] for training and testing purposes. Additionally, the MITBIH Supraventricular Arrhythmia database (MITBIH-SUP) [Mark et al., 1990] and the St. Petersburg Institute of Cardiological Technics (INCART) database were used for evaluation and testing purposes, in order to assess the generalization achieved by the classification models developed in the MITBIH-AR. All databases are freely available on Physionet [Goldberger et al., 2000] and their details were summarized in Table 3.1 and Section 2.1.

3.2.2 ECG preprocessing

The ECG recordings of the MITBIH-SUP and INCART databases were first resampled to 360 Hz, which is the sampling frequency of the MITBIH-AR. This was performed with a tenth order lowpass FIR filter without observing any notorious distortion (resample

function, Signal Processing Toolbox of Matlab, The Mathworks Inc., MA). All recordings in all databases were first preprocessed to remove artifacts as described in [de Chazal et al., 2004] and Section 2.3.1. No energy or amplitude normalization was done, as we were interested in some amplitude-related features to describe morphology.

As our objective is the evaluation of a heartbeat classifier, the QRS location is assumed to be known and we use the annotations included in the databases. Following the QRS complex detection positions, the delineation of each heartbeat was performed with the delineator described in [Martínez et al., 2004]. The result of delineation as well as the discrete wavelet transform (DWT) of the ECG signals (which are intermediate signals for the delineator) were used to calculate most of the features used in the following subsections.

3.2.3 Features and Classifiers

Following the conclusions of previous works [Hu et al., 1997, de Chazal et al., 2004], we included in our model both rhythm and morphological features. As rhythm features we used features from the RR interval sequence, including local and global estimations of the heart-rate and its variations. Remember that the RR interval is the distance between two consecutive heartbeats. Regarding to the morphologic features, we included several types depending where the features were calculated. In general, these features intended to model the waveforms of the ECG in several domains. In total we used 39 features for the experiments of feature selection. The details of the features used can be found in Section 2.4.1.

Since some of the features belong to different domains, a special treatment must be taken for these features. The methods described in Section 2.4.3 were used for features representing angles, and a natural logarithm transformation for those features that belong to $\mathbb{R}^+$, as intervals. This is shown in Table 3.4.

It is well known that low dimensional models generalize better to examples not presented during the training phase, resulting in a more robust and realistic classifier [van der Heijden et al., 2005]. Other problems of high dimensional feature models are the increased risk of overfitting to the training dataset, and the higher computational cost. The overfitting appears when there are not enough training examples to estimate the classifier parameters. The higher computational cost is the result of calculating more features to perform the same classification task. In order to obtain a small and well performing model, a sequential floating feature selection algorithm (SFFS) was used [Pudil et al., 1994] (Section 2.4.6). This feature was repeated for the three classifiers described in Section 2.4.2, the linear discriminant classifier (LDC), the LDC with class compensation (LDC-C) and the quadratic discriminant classifier (QDC).

3.2.4 Experiment Setup

The experiment can be described as a sequence of three steps:

1) In the first step we search for the best performing model, from the 60 available features, in the training ($DS1$ of MITBIH-AR) and validation (MITBIH-SUP) sets (Fig. 3.1a). For each iteration of the SFFS algorithm, the current model was trained in $DS1$ of MITBIH-AR and its performance was evaluated in the MITBIH-SUP database. As the data divisions in both databases do not share any recording, the features selected should retain the generalization properties. Several parameter configurations were studied for the SFFS algorithm, like the effect of the type of classifier (LDC, LDC-C and QDC) and the optimization criterion (J_S or J_{P^+}) for the search. The weight compensation used in the experiments for the LDC classifier is $w_N = 1$, $w_S = 10$ and $w_V = 10$. The same weights were also studied for the J_{P^+} and J_S criterion, $\pi_N = 1$, $\pi_S = 10$ and $\pi_V = 10$. At the end of this step we have an optimal feature set for each parameter configuration. In Figure 3.2 the evolution of the SFFS algorithm for a particular set of parameters is shown.

2) The second step (Fig. 3.1b) is the selection of the best performing model, among the best models obtained in the previous step for each parameter configurations. For that purpose, we compare the global results (A, S, P^+) obtained in the union set of $DS1$ of MITBIH-AR dataset and the MITBIH-SUP database, using a recording-based k-fold cross validation with $k = 10$ recordings.

3) Finally the performance of the selected model is evaluated in $DS2$ for comparison with [de Chazal et al., 2004], as shown in Fig. 3.1c. Additionally the performance in the INCART database is compared to that obtained in $DS2$ to assess how the model behaves in completely different databases.

The results presented in this chapter are compared to the classifier developed in [de Chazal et al., 2004] (reference classifier in the rest of this chapter), being this, to our knowledge, the best performing fully automatic multiclass classifier (AAMI compliant) reviewed in the literature. In order to perform a fair comparison, some methodological aspects were maintained as similar as possible. The implementation of the classifier suggested in [de Chazal et al., 2004] was contrasted with the reported results obtaining comparable results. With this implementation, we could evaluate the generalization capability of the reference classifier in the MITBIH-SUP database, since this experiment was not performed in [de Chazal et al., 2004]. In those situations where the experiments were already performed in [de Chazal et al., 2004], the reported results were used.

All experiments described in this chapter will focus to achieve automatic classification between the three AAMI2 classes (N, S and V'), since the fusion class is poorly represented in the databases used. The restrictions imposed by the recording-oriented division of the data, and the fact that only a few recordings concentrate the majority of the examples of the fusion heartbeats, makes unfeasible to perform the feature selection using the original

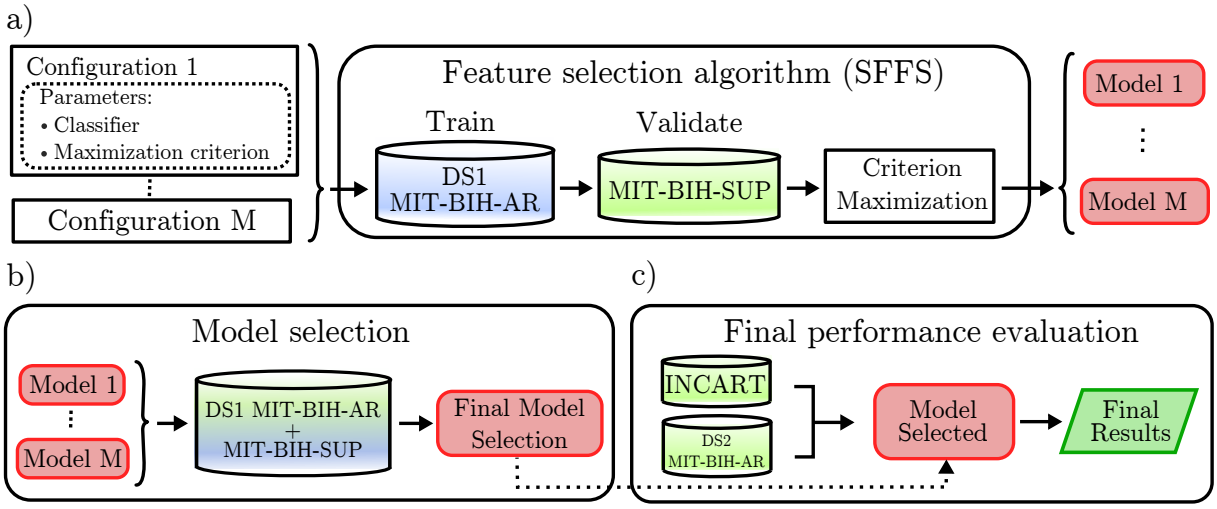


Figure 3.1: Block diagram describing the experiments performed in this chapter. In panel a) the feature selection algorithm is summarized, indicating the train and validation dataset division, as well as the different parameters of the algorithm. In panel b) is shown the methodology to obtain the best performing model among the different searches performed. Finally in panel c), the best performing model is selected for the final performance evaluation in the test datasets.

AAMI labeling. Despite this limitation for the model selection, the model obtained for the three AAMI2 classes was also retrained and evaluated classifying the four AAMI classes to show its utility.

3.3 Results

The main results for the experiments described in the previous section are summarized in tables 3.2 and 3.3. Table 3.2 shows the results of the best models obtained for the different parameter configurations during the model selection. The best performing of these models was an 8 feature model trained in the $DS1$ of the MITBIH-AR. The 8 features that the model comprehends are listed in Table 3.4. The classifier used was an LDC-C, using equal prior probabilities $P(\omega_i)$. The optimization criterion used in the SFFS was J_{P+} with equal weights π_i.

In Figure 3.2 one of the feature selections performed, and the floating evolution of the search algorithm are shown. In this example, the best model is found at the beginning. It can be seen in Table 3.2 that most of the best performing models only included around 10 features.

The performance of the selected model in the test set ($DS2$) is compared with the reported by de Chazal *et al.* [de Chazal et al., 2004] in Table 3.3. The model found in this chapter achieves better performance for the three classes. Table 3.7 presents the performance by recording in the test set, following the recommendations of the AAMI [AAMI-EC57, 1998–2008] for results presentation.

Table 3.2: Summary of the best performing models found with the SFFS algorithm separating all AAMI2 classes.

			Model Evaluation								
	Opt.		Normal		Suprav.		Ventr.		Total		
Classifier	Crit.	# Features	S	P^+	S	P^+	S	P^+	A	S	P^+
LDC-C	J_{P+}	**8**	**93**	**98**	**78**	**40**	**68**	**70**	**91**	**80**	**70**
LDC	J_{P+}	10	92	98	57	38	77	50	89	75	62
QDC	J_{P+}	7	80	98	7	12	89	22	77	59	44
LDC-C	J_S	10	90	98	77	33	68	74	88	79	68
LDC	J_S	10	92	98	74	37	70	67	89	78	67
QDC	J_S	9	87	98	43	32	80	33	84	70	55
[de Chazal et al., 2004]		48	87	98	57	30	63	36	84	69	55

In the left the parameters used for the SFFS to find the model, and in the right the performance estimated for the dataset composed by $DS1$ of the MITBIH-AR and the whole MITBIH-SUP database (with cross-validation of k=10 recordings). The best performing model (in bold) is selected for the final performance evaluation. The results are expressed in percentages.

The performance of the selected model with the four AAMI classes (N, S, V, F) is reported in Table 3.5. The model found achieves a performance slightly lower than the reference, but it must be noted that the selected model was optimized for the three AAMI2 classes (N, S, V').

Finally the performance of the model found in the INCART database is presented in Table 3.6. The performance obtained in this database is comparable with that obtained in $DS2$ for all classes.

3.4 Discussion and Conclusions

In this chapter we have presented a methodology to develop a simple and robust heartbeat classification system. We evaluated it focusing in the generalization capability. In order to do this, we take into consideration the MITBIH-SUP [Mark et al., 1990] and the INCART databases in addition to the widely-used MITBIH-AR, all freely available in Physionet [Goldberger et al., 2000]. Although these databases are bigger than the original MITBIH-AR, the fusion class defined in the AAMI standard [AAMI-EC57, 1998–2008] is not so well represented as the other classes. This limitation is overcome by adopting the alternative labeling AAMI2 proposed in this chapter. The AAMI2 labeling make sense from a physiological point of view, since the AAMI fusion class comprehends those heartbeats which results from the simultaneous occurrence of normal and ventricular heartbeats.

From the results obtained for the model selection presented in Table 3.2, several models that outperform the reference classifier [de Chazal et al., 2004] were achieved. The best model found consists of 8 features: $\ln(RR[i])$, $\ln(RR[i+1])$, $\ln(RR_1)$, $\ln(RR_{20})$, k_Z^x, k_Z^y,

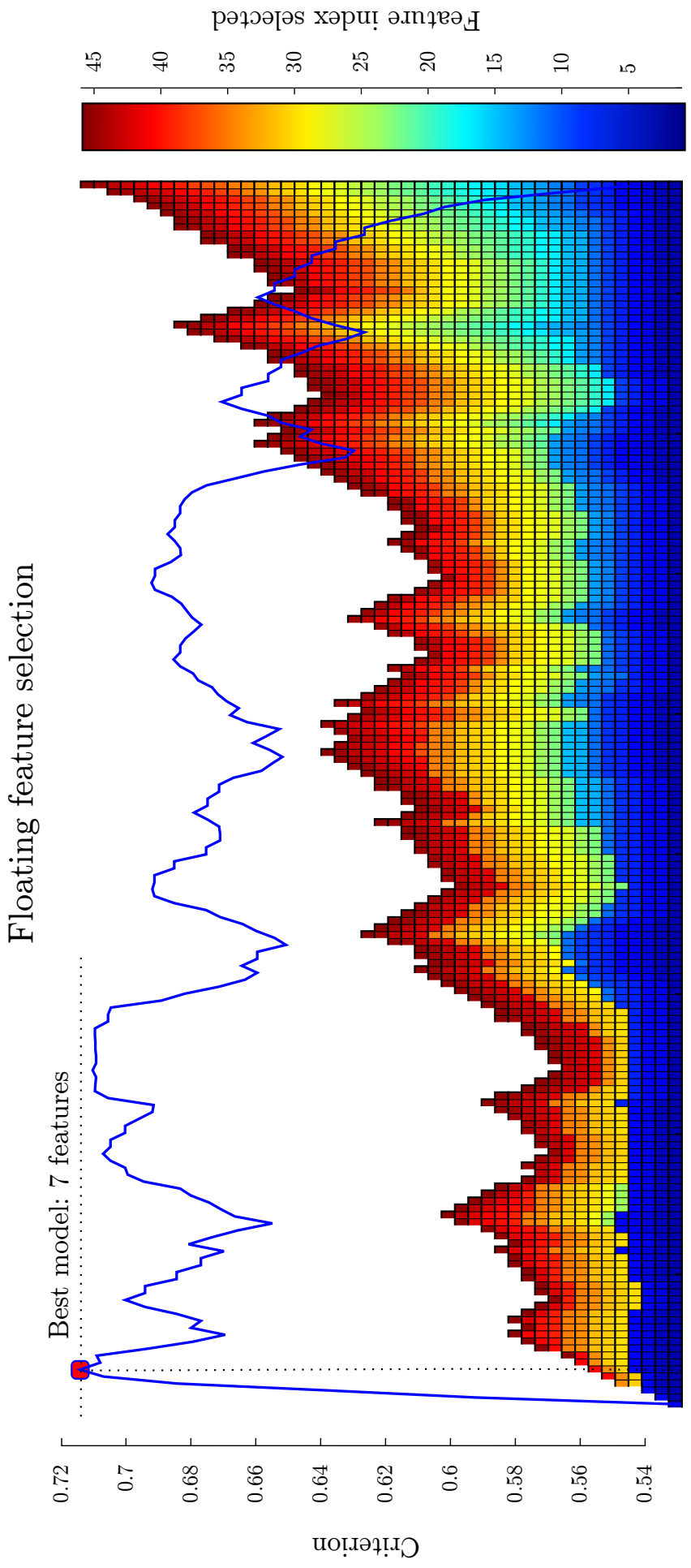


Figure 3.2: Example of a particular SFFS search. The left axis represents the criterion value, and is matched with the blue signal. The colored axis on the right indicates the index of a feature, and the abscissas the iterations of the algorithm. For this example we restricted the search for 45 features, so the first feature in the pool has color dark blue, and the last dark red. The height of the bar is related to the feature model size, at iteration 1 we have a model with only 1 feature, and at iteration 176, a model with all the 45 features. The best model in this search was found soon, at iteration 6 with only 7 features. A summary of all the searches performed is shown in Table 3.2.

Table 3.3: Performance comparison between the model selected in Table 3.2 and the reference classifier [de Chazal et al., 2004] separating all AAMI2 classes in $DS2$ of MITBIH-AR.

[de Chazal et al., 2004]

		Algorithm			
		n	s	v'	Total
Truth	N	40718	1863	1677	44258
	S	307	1361	169	1837
	V'	235	845	2529	3609
	Total	41260	4069	4375	49704

Model selected in Table 3.2

		Algorithm			
		n	s	v'	Total
Truth	N	41950	2002	236	44188
	S	216	1422	197	1835
	V'	473	222	2911	3606
	Total	42639	3646	3344	49629

Performance calculation mode	Classifier	Automatic	# Features	Normal S	Normal P^+	Suprav. S	Suprav. P^+	Ventr. S	Ventr. P^+	Total A	Total S	Total P^+
Imbalanced	1	yes	8	95	98	77	39	81	87	93	84	75
	2	yes	48	92	99	74	33	70	58	90	79	63
	3	no	48	94	99	88	47	95	82	94	92	76
Balanced	1	yes	8	95	79	77	88	81	88	84	84	85
	2	yes	48	92	80	74	73	70	84	79	79	79

1: [Llamedo and Martínez, 2011a] 2: [de Chazal et al., 2004] 3: [de Chazal and Reilly, 2006]

Both models were trained in $DS1$ of the same database. First the confusion matrices for both models are shown, and below the class and total performances are summarized. The performances are expressed in percentages for both, balanced and imbalanced class presence in the dataset.

Table 3.4: Features used in the model selected in Table 3.2 for the final performance evaluation.

Feature	**Description**
$\ln(RR[i])$	Current RR interval[1]
$\ln(RR[i+1])$	Next RR interval[1]
$\ln(RR_1)$	Average RR interval in the last minute[1]
$\ln(RR_{20})$	Average RR interval in the last 20 minutes[1]
k_Z^x	Zero-cross position of the WT autocorrelation signal in lead 1[2]
k_Z^y	Zero-cross position of the WT autocorrelation signal in lead 2[2]
k_M^x	Maximum position of the WT autocorrelation signal in lead 1[2]
k_M^y	Maximum position of the WT autocorrelation signal in lead 2[2]

[1] See Figure 2.6 [2] See Figure 2.10

Table 3.5: Performance comparison between the model selected in Table 3.2 and the reference classifier [de Chazal et al., 2004] separating all AAMI classes in $DS2$ of MITBIH-AR.

de Chazal *et al.* [de Chazal et al., 2004]

Truth \ Algorithm	f	n	s	v	Total
F	347	33	1	7	388
N	3509	38444	1904	303	44160
S	16	173	1395	252	1836
V	176	117	321	2504	3118
Total	4048	38767	3621	3066	49502

Model selected in Table 3.2

Truth \ Algorithm	f	n	s	v	Total
F	370	11	2	5	388
N	8031	34270	1807	80	44188
S	28	124	1403	280	1835
V	321	46	182	2669	3218
Total	8750	34451	3394	3034	49629

Performance calculation mode	Classifier	Fusion S	Fusion P^+	Normal S	Normal P^+	Suprav. S	Suprav. P^+	Ventr. S	Ventr. P^+	Total A	Total S	Total P^+
Imbalanced	1	95	4	78	99	76	41	83	88	78	83	58
	2	89	9	87	99	75	39	80	81	86	83	57
Balanced	1	95	76	78	88	76	88	83	83	83	83	84
	2	89	86	87	80	76	84	80	83	83	83	83

1: [Llamedo and Martínez, 2011a] 2: [de Chazal et al., 2004]

Both models were trained in $DS1$ of the same database. First the confusion matrices for both models are shown, and below the classes and total performances are summarized. The performances are expressed in percentages for both, balanced and imbalanced class presence in the dataset.

Table 3.6: Confusion matrix as a result of separating all AAMI2 classes in the INCART database.

		Algorithm			
		n	s	v'	Total
Truth	N	140983	10576	1958	153517
	S	84	1660	214	1958
	V'	644	3007	16559	20210
	Total	141711	15243	18731	175685

Performance		Normal		Suprav.		Ventr.		Total		
calculation mode	Dataset	S	P^+	S	P^+	S	P^+	A	S	P^+
Imbalanced	$DS2$ MITBIH-AR	95	98	77	39	81	87	93	84	75
	INCART	92	99	85	11	82	88	91	86	66
Balanced	$DS2$ MITBIH-AR	95	79	77	88	81	88	84	84	85
	INCART	92	92	85	80	82	87	86	86	86
By recording	$DS2$ MITBIH-AR	95	83	61	73	75	86	94	80	82
	INCART	93	90	64	66	71	86	91	79	85

The model used is the selected from Table 3.2, trained in $DS1$ of the MITBIH-AR database. The performances are expressed in percentages, and grouped by performance calculation mode for easy comparison with the results obtained from $DS2$ of MITBIH-AR (Table 3.3).

k_M^x and k_M^y; which are described in Table 3.4. As it can be noted, the selected features are computed without exception from time interval measurements. This could be explained given that the used databases do not always include the same pair of ECG leads in each recording. Therefore the classification performances of features which are calculated from amplitudes are heavily degraded. The directional features (like the VCG_ϕ) were also probably affected by this fact, even if the clinical importance of this kind of features is well-known by cardiologists [Taylor, 2002]. In contrast, intervals seem to retain the classification ability with independence of the pair of leads chosen. The first four features in the model are clearly connected to the evolution of heart rhythm, while the other four can be understood as surrogate measurements of the QRS width, and therefore the QRS morphology. As a result, the model found has the evident advantage of a lower size, which results in a computational saving and lower error in the parameter estimation during the training phase. In addition, it only relies on the QRS fiducial point detection, making the classifier model robust to degraded signals where the delineation of the ECG waves is not reliable.

It is worth noting that the performance achieved by the reference classifier [de Chazal et al., 2004] in the union of training and validation dataset (Table 3.2) is lower for all classes than the obtained in the final performance reported in Table 3.3. The same phenomenon happens with the suggested model in a smaller degree, with the exception of the supraventricular performance. This phenomenon was also reported in [de Chazal et al.,

2004], obtaining better performance in the test set than in the training set. These results suggest that $DS2$ dataset may not be a good data sample to measure the actual performance of a classifier. To avoid this bias in the actual performance, it may be convenient in future works that the final performance estimation is computed applying other methodologies or redefining the test dataset. One reason that could be biasing the results in $DS2$ is the different amount of examples by recording for the supraventricular class. As can be seen in Table 3.7, recordings 232 and 222 concentrate the majority of the examples for the supraventricular class, which means that failing in these recordings impacts considerably to the S class performance. For this reason, the average performances presented in Table 3.7 could also be of importance since each recording or subject is equally weighted in the average.

The results presented in [de Chazal and Reilly, 2006], where the automatic classifier of [de Chazal et al., 2004] is assisted by a local expert to improve its performance, are also compared in Table 3.3. This suggests that a similar approach of combining the knowledge of a local expert with our model, could also lead to a comparable improvement in the baseline performance.

An additional assessment of the suggested model classifying the four AAMI (N, S, V and F) classes is presented in Table 3.5. The results verify the validity of the model achieving slightly lower performance than the results presented in [de Chazal et al., 2004]. It must be noted that the model presented in this work was optimized for the AAMI2 labeling (N, S and V'), and the classifier is mainly misclassifying normal heartbeats as fusion, as shown in Table 3.5.

The results in Table 3.6 suggest that the selected features have good generalization capability when evaluating the performance in heartbeats not considered during the development phase, as the ones from the INCART database. The imbalanced performance is comparable for all classes except the supraventricular where a decrease in the P^+ occurred. This could be explained by an increased class imbalance in the INCART database which is about 75-to-1, while in MITBIH-AR is 22-to-1 approximately. This is confirmed by the balanced results (equivalent to a class balance of 1-to-1) in the same table, where the performance figures are very similar. The validity of the generalization capability of the proposed model, is somehow restricted to the available data, and should be corroborated in future works by including new databases in the analysis or other methodologies. Despite this limitation, the degree of generalization of the suggested model is expected to be better than models obtained considering only the MITBIH-AR database.

One limitation of the presented approach is the Gaussian assumption of the data imposed by the classifier, since many features were observed not to fulfill this requirement. Despite this evident limitation, the linear decision regions in the feature space defined by the LDC-C allowed us to select those features which inherently provide better classification performance. Considering the proposed classifier and feature model as a reference for future improvements, the effect of the lack of Gaussianity can be mitigated using more

Table 3.7: Detailed results grouping by recording (or subject).

	Number of beats			Normal		Supraventr.		Ventricular		Totals		
Rec	N	S	V'	S	P^+	S	P^+	S	P^+	A	S	P^+
100	2235	33	1	100%	77%	70%	100%	100%	100%	100%	90%	92%
103	2079	2	0	99%	50%	0%	0%	–	–	99%	50%	25%
105	2524	0	41	97%	100%	–	–	51%	94%	96%	74%	97%
111	2120	0	1	99%	100%	–	–	100%	99%	99%	100%	100%
113	1785	6	0	99%	100%	100%	99%	–	–	99%	100%	100%
117	1531	1	0	100%	100%	100%	100%	–	–	100%	100%	100%
121	1858	1	1	99%	100%	100%	99%	100%	100%	99%	100%	100%
123	1512	0	3	100%	100%	–	–	0%	0%	100%	50%	50%
200	1733	30	826	96%	58%	27%	81%	92%	91%	94%	72%	77%
202	2059	55	20	67%	87%	93%	56%	50%	87%	68%	70%	77%
210	2419	22	205	94%	86%	91%	81%	69%	87%	92%	85%	85%
212	2745	0	0	100%	100%	–	–	–	–	100%	100%	100%
213	2637	28	582	100%	63%	46%	100%	44%	47%	89%	63%	70%
214	2000	0	257	97%	100%	–	–	94%	98%	97%	96%	99%
219	2080	7	65	86%	46%	0%	0%	82%	100%	86%	56%	49%
221	2029	0	396	93%	99%	–	–	99%	100%	94%	96%	100%
222	2271	208	0	72%	92%	89%	76%	–	–	73%	81%	84%
228	1685	3	362	100%	60%	33%	84%	93%	100%	99%	75%	81%
231	1565	1	2	98%	49%	0%	0%	50%	100%	97%	49%	50%
232	397	1381	0	100%	90%	78%	100%	–	–	83%	89%	95%
233	2227	7	841	100%	92%	71%	90%	83%	74%	95%	85%	85%
234	2697	50	3	100%	78%	72%	100%	100%	100%	99%	91%	93%
Average	44188	1835	3606	95%	83%	61%	73%	75%	86%	94%	80%	82%
Gross				95%	98%	77%	39%	81%	87%	93%	84%	75%

For the model selected in Table 3.2 separating all AAMI2 classes in $DS2$ of MITBIH-AR, following AAMI recommended performance measures. Average statistics gives each subject equal weight. Gross statistics weight each heartbeat equal.

complex classifiers, like ANN's or mixture of Gaussians. These classifiers allow more complex decision regions in the feature space, retaining details of the training data which may improve the classification performance.

Despite the improved results presented in this work, there is still room for improvement in the field since the S and P^+ for the supraventricular class are of 77% and 39%, and for the ventricular class (though better) are of 81% and 87%. These results suggest that other features, classifiers or meta-classifier strategies (like local expert assistance) should be developed in order to improve the performance, specially in the supraventricular class.

3.A Detailed Results

In this section we present the performances by recording that the AAMI EC57 standard suggests for performance comparison.

Table 3.8: Detailed results grouped by recordings in the INCART database

	Number of beats			Normal		Supraventr.		Ventricular		Totals		
Rec	N	S	V	S	P^+	S	P^+	S	P^+	A	S	P^+
I01	2410	0	344	95%	100%	–	–	100%	95%	95%	98%	98%
I02	2442	0	229	94%	100%	–	–	44%	96%	90%	69%	98%
I03	2322	2	125	80%	61%	0%	0%	99%	59%	81%	60%	40%
I04	2267	16	138	64%	55%	56%	55%	62%	77%	64%	61%	62%
I05	1517	0	256	70%	74%	–	–	27%	48%	64%	49%	61%
I06	2434	48	9	100%	84%	81%	100%	100%	100%	99%	94%	95%
I07	2637	65	1	100%	94%	94%	100%	100%	100%	100%	98%	98%
I08	1775	2	350	94%	100%	100%	65%	53%	100%	87%	82%	88%
I09	2953	0	41	80%	97%	–	–	88%	100%	81%	84%	99%
I10	3596	0	83	95%	96%	–	–	96%	96%	95%	96%	96%
I11	2099	0	4	95%	100%	–	–	75%	100%	95%	85%	100%
I12	2797	1	8	76%	100%	100%	50%	25%	100%	76%	67%	83%
I13	1791	0	230	87%	100%	–	–	3%	100%	77%	45%	100%
I14	1799	0	64	94%	100%	–	–	0%	0%	91%	47%	50%
I15	2629	0	3	96%	100%	–	–	100%	100%	96%	98%	100%
I16	1518	0	2	100%	100%	–	–	0%	0%	100%	50%	50%
I17	1643	0	27	98%	100%	–	–	0%	0%	96%	49%	50%
I18	2660	0	422	76%	90%	–	–	89%	100%	77%	83%	95%
I19	1212	0	849	95%	100%	–	–	97%	100%	95%	96%	100%
I20	2358	179	111	86%	94%	99%	70%	67%	100%	86%	84%	88%
I21	2070	104	8	95%	98%	98%	77%	75%	100%	95%	89%	92%
I22	2814	124	186	76%	93%	98%	69%	76%	100%	77%	83%	87%
I23	2190	0	13	100%	100%	–	–	46%	100%	99%	73%	100%
I24	2562	0	6	99%	100%	–	–	83%	100%	99%	91%	100%
I25	1702	2	5	94%	100%	100%	49%	0%	0%	94%	65%	50%
I26	1496	7	4	99%	57%	100%	99%	25%	100%	99%	75%	85%
I27	1883	0	719	99%	100%	–	–	100%	100%	99%	100%	100%
I28	1710	0	4	96%	100%	–	–	50%	100%	96%	73%	100%
I29	1833	0	783	82%	100%	–	–	57%	99%	75%	70%	100%
I30	1703	0	755	95%	100%	–	–	51%	100%	82%	73%	100%
I31	1843	0	1363	92%	100%	–	–	79%	98%	86%	86%	99%
I32	1559	0	57	97%	100%	–	–	0%	0%	94%	49%	50%
I33	1244	589	1	98%	97%	87%	98%	100%	92%	95%	95%	96%
I34	1426	536	0	97%	97%	96%	97%	–	–	97%	97%	97%
I35	3198	0	475	72%	86%	–	–	79%	100%	73%	76%	93%
I36	3445	0	462	81%	86%	–	–	87%	99%	81%	84%	93%
I37	2007	0	452	98%	99%	–	–	99%	100%	98%	99%	100%
I38	2153	0	542	100%	99%	–	–	90%	100%	98%	95%	100%
I39	1459	0	313	100%	100%	–	–	96%	100%	99%	98%	100%
I40	2566	6	92	82%	83%	67%	95%	98%	75%	83%	82%	84%
I41	1622	5	1	99%	62%	40%	28%	0%	0%	99%	46%	30%
I42	1544	0	1561	94%	100%	–	–	85%	100%	90%	90%	100%
I43	1084	0	1121	98%	100%	–	–	94%	100%	96%	96%	100%
I44	1801	8	683	100%	89%	88%	99%	99%	100%	100%	96%	96%

continues on the next page

For the model selected in Table 3.2 separating all AAMI2 classes, following AAMI recommended performance measures.

concluded from previous page

	Number of beats			Normal		Supraventr.		Ventricular		Totals		
Rec	N	S	V	S	P^+	S	P^+	S	P^+	A	S	P^+
I45	1434	0	491	99%	100%	–	–	99%	100%	99%	99%	100%
I46	2230	1	425	86%	97%	100%	87%	96%	100%	88%	94%	95%
I47	1857	1	92	93%	48%	0%	0%	47%	100%	91%	47%	49%
I48	2117	2	236	99%	100%	100%	99%	100%	100%	99%	100%	100%
I49	2117	0	27	89%	100%	–	–	100%	100%	89%	95%	100%
I50	2992	0	4	95%	100%	–	–	100%	100%	95%	98%	100%
I51	1968	3	802	73%	52%	33%	55%	100%	100%	81%	69%	69%
I52	1608	0	137	100%	100%	–	–	100%	100%	100%	100%	100%
I53	1149	0	1109	99%	100%	–	–	100%	100%	100%	100%	100%
I54	2338	1	22	95%	91%	0%	0%	77%	43%	94%	57%	45%
I55	2145	1	17	100%	100%	0%	0%	94%	48%	100%	65%	49%
I56	1669	26	7	97%	93%	31%	64%	86%	58%	96%	71%	72%
I57	2839	3	24	91%	69%	67%	89%	92%	100%	91%	83%	86%
I58	2310	0	12	99%	92%	–	–	75%	100%	99%	87%	96%
I59	2064	0	81	77%	100%	–	–	5%	99%	75%	41%	100%
I60	2472	0	0	87%	100%	–	–	–	–	87%	87%	100%
I61	1450	1	0	100%	100%	100%	100%	–	–	100%	100%	100%
I62	1451	9	807	96%	71%	67%	60%	43%	80%	77%	69%	70%
I63	1844	1	146	100%	74%	100%	86%	49%	100%	96%	83%	87%
I64	1883	0	26	100%	72%	–	–	31%	100%	99%	66%	86%
I65	2271	5	386	86%	59%	40%	61%	89%	100%	86%	72%	73%
I66	2136	1	200	95%	49%	0%	0%	56%	100%	92%	50%	50%
I67	2435	5	532	93%	54%	20%	73%	99%	100%	94%	71%	76%
I68	2479	2	161	100%	67%	50%	100%	100%	100%	100%	83%	89%
I69	1997	1	168	100%	99%	0%	0%	99%	50%	100%	66%	50%
I70	1537	126	0	100%	100%	1%	100%	–	–	92%	51%	100%
I71	1631	35	0	90%	97%	94%	91%	–	–	90%	92%	94%
I72	1872	8	386	99%	86%	88%	94%	92%	100%	98%	93%	93%
I73	1888	32	70	100%	96%	97%	72%	61%	100%	98%	86%	89%
I74	2079	0	322	100%	79%	–	–	73%	100%	96%	87%	90%
I75	1482	0	618	100%	95%	–	–	95%	100%	98%	98%	98%
Average	153517	1958	20210	93%	90%	64%	66%	71%	86%	91%	79%	85%
Gross				92%	99%	85%	11%	82%	88%	91%	86%	66%

For the model selected in Table 3.2 separating all AAMI2 classes, following AAMI recommended performance measures. Average statistics gives each subject equal weight. Gross statistics weight each heartbeat equal.

Chapter 4

Extensions to the Automatic Classifier

4.1 Introduction

In this chapter we study two improvements to the original classifier developed in the previous chapter. The first one allows the classification of recordings of an arbitrary number of leads, while the second one explores the utility of more complex classifiers, as neural networks.

4.2 Multilead classification

The room for improvement in the field of heartbeat classification, together with the availability of 3- and 12-lead Holter devices makes necessary the development of algorithms capable of exploiting the increase of recorded information. Moreover, the St. Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database (INCART) has become recently freely available on Physionet [Goldberger et al., 2000], making possible the evaluation of multilead heartbeat classifiers in a comparable way.

The objective of this study is to find an effective way to include morphologic information present in multilead ECG signals. For that purpose, we compare several multilead classification strategies against the reference two-lead classifier that we developed in [Llamedo and Martínez, 2011a]. We assess the improvement in classification performance as well as the generalization capability to other databases not considered during the development. The main novelty presented in this work is the generalization of the model developed in [Llamedo and Martínez, 2011a] to an arbitrary number of leads.

4.2.1 Material and methods

In this study we used the well-known MITBIH Arrhythmia database (MITBIH-AR) [Moody and Mark, 2001] and other public databases already described in Section 2.1.

Table 4.1: Databases used in this work. Heart beats classes are N: normal, S: supraventricular, V: ventricular, F: fusion, and Q: unknown.

Database	N	S	V	F	Q	#Rec
INCART	153651	1959	20005	219	6	75
Biosigna	286246	1326	2541	0	0	56
MITBIH-AR	90089	2779	7007	802	15	44
MITBIH-SUP	162271	12195	9940	23	79	78
Totals	692257	18259	39493	1044	100	253

All public databases are available on Physionet [Goldberger et al., 2000] and their details were summarized in Table 4.1. For all databases the AAMI recommendations for class-labeling were adopted (Section 4.2 in [AAMI-EC57, 1998–2008]).

Additionally, we used a private database called Biosigna. This database was developed at Biosigna GmbH, and consists of 56 recordings containing a broad set of pathologies. Each recording is one hour length, sampled at 500 Hz with an amplitude resolution of 12-bit over a range of 10mV. The recordings were manually annotated by experienced annotators. More detailed information about this database can be found in [Fischer et al., 2008].

For the preprocessing of the ECG recordings, we used the same methodologies described in Section 2.3.1 of the previous chapter.

We follow the results obtained in Chapter 3 where we developed a heartbeat classifier with good generalization capability, using rhythm and morphological features together with a linear classifier compensated for class-imbalance, called LDC-C. This classifier has the possibility of weighting the class contribution to the covariance matrix (equation (2.25)) during the training, to deal with the class imbalance seen in Table 4.1. The equations of the LDC-C were described in Section 2.4.2.

The features obtained from the sequential floating feature selection (SFFS) in [Llamedo and Martínez, 2011a] are shown in Table 4.2. As the rhythm features of the model do not depend on the number of available leads, the first four features in Table 4.2 remain the same. Therefore we will focus the analysis on those features describing heartbeat morphology, which are the ones that can be improved by the addition of new leads.

The morphology features used in the model are the first zero-crossing (k_Z^L) and maximum position (k_M^L) of the autocorrelation sequence of the ECG DWT at scale 4 for each lead L (see Section 2.4.1 for details). Both features were calculated in four sets of leads to study the most suitable way of integrating the information from all leads:

1. The first strategy consists of just including k_Z^L and k_M^L from all available ECG leads and is referred as 12L (or 3L when only 3 leads are available), resulting in two morphology features per lead.

2. The second strategy computes k_Z^L and k_M^L from the three vectocardiogram (VCG)

Table 4.2: Features used in the model obtained in [Llamedo and Martínez, 2011a] only for two-lead recordings.

Feature	Description
$\ln(RR[i])$	Current RR interval[1]
$\ln(RR[i+1])$	Next RR interval[1]
$\ln(RR_1)$	Average RR interval in the last minute[1]
$\ln(RR_{20})$	Average RR interval in the last 20 minutes[1]
k_Z^x	Zero-cross position of the WT autocorrelation signal in lead 1[2]
k_Z^y	Zero-cross position of the WT autocorrelation signal in lead 2[2]
k_M^x	Maximum position of the WT autocorrelation signal in lead 1[2]
k_M^y	Maximum position of the WT autocorrelation signal in lead 2[2]

[1] See Figure 2.6 [2] See Figure 2.10

leads X, Y and Z, transformed from 12L by the Dower matrix. This strategy can only be performed in 12-leads recordings.

3. In the third strategy, referred as ECG-PCA, we apply principal component analysis (PCA) to the available ECG leads, then the morphology features are computed from the two most important components.

4. Finally for the fourth strategy, called WT-PCA, the PCA is applied not to the ECG, but to the fourth scale of the DWT ($W_4 s(k)$), and the two morphological features k_Z^L and k_M^L were calculated from the principal components.

The last two strategies are the result of projecting all ECG leads or its fourth scale DWT ($W_4^s(k)$), into the two most important basis of a principal component analysis (PCA). The PCA consists of finding an orthogonal linear transformation such that the first component of the transformation comes to lie in the direction of the greatest variance of the data. For our multilead ECG signal $\mathbf{S}$, with each lead of N samples as a column resulting in a $N \times L$ matrix, the transformation is defined by

$$\mathbf{R} = \mathbf{S} \cdot \mathbf{P}, \tag{4.1}$$

where $\mathbf{P}$ is the matrix which defines the linear transformation and $\mathbf{R}$ is the transformed ECG. The matrix $\mathbf{P}$ is the result of computing the eigenvectors of the sample covariance matrix $\mathbf{C}$, calculated as described in the next subsection,

$$\mathbf{P}^{-1}\mathbf{C}\mathbf{P} = \mathbf{Q}, \tag{4.2}$$

being

$$\mathbf{C} = \mathbf{S}^T\mathbf{S}. \tag{4.3}$$

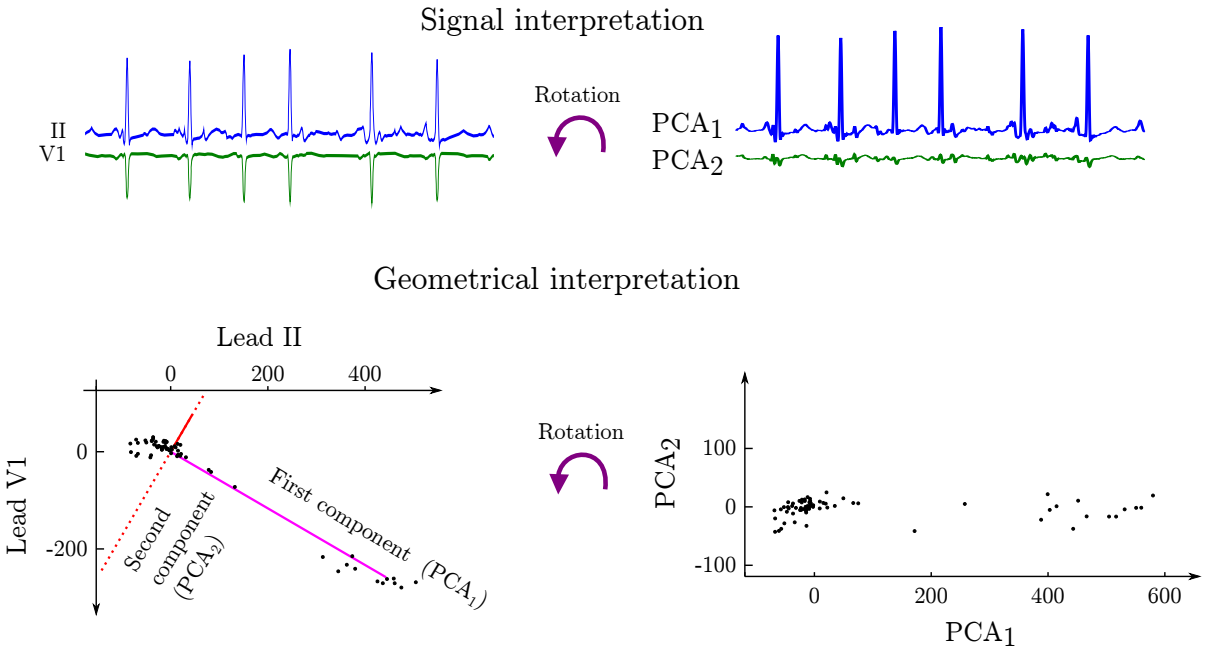


Figure 4.1: Toy example where a two-lead ECG excerpt is also interpreted geometrically, and PCA transformation is performed. Note the rotation involved in the PCA transformation.

The associated diagonal matrix of eigenvalues $\mathbf{Q}$, is related to the importance of each eigenvector of $\mathbf{P}$ in the sense of the variance of the original data $\mathbf{S}$. That is, if we extract only the first two components of $\mathbf{R}$, we are retaining a fraction of the original QRS complex variance which is typically above the 90% for normal sinus rhythm heartbeats. In the rest of this work we will refer to perform PCA to calculate the two most important projections of a multilead signal.

In the toy example presented in Figure 4.1, a two-lead ECG excerpt is interpreted as a multilead signal (typical interpretation), and as a bivariate collection of unrelated samples (geometrical interpretation). The latter interpretation neglects the time relation between each sample of the sequence. This interpretation is useful to visualize the main directions of variation of the signal, and the rotation involved in a PCA transformation. It is useful to think the columns of the $\mathbf{P}$ matrix, as weights of a linear transformation. Each element of the column vectors, are related to the contribution of each lead, as can be seen in the center of Figures 4.2 and 4.3. The variance shown in the right part of the figures, is the ratio between the first two eigenvalues of $\mathbf{Q}$ and the sum of all eigenvalues. Note the similar weight patterns for the normal and supraventricular classes, which are both similarly conducted through the ventricles. These patterns depend on the heartbeat morphology as can be seen for the ventricular and fusion classes in Figure 4.3.

The PCA is performed for each heartbeat in a 160 ms window centered at the QRS complex detection sample, or fiducial point (PCA window in Figure 4.4). Then, the multilead signal is projected into the PCA basis in a wider window starting 130 ms before and ending 200 ms after the QRS fiducial point. As it is known from previous works

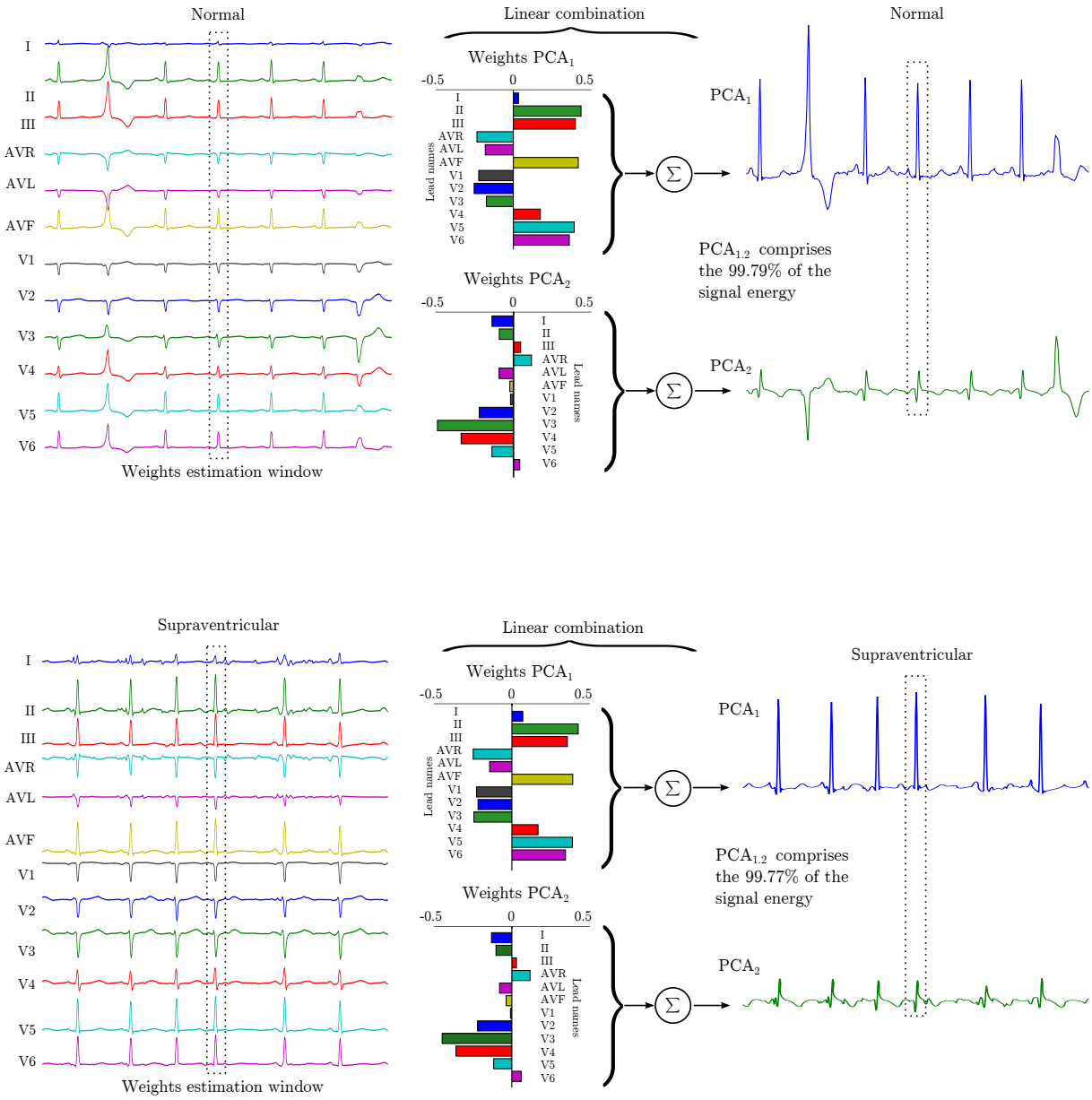


Figure 4.2: Examples of using PCA in 12-lead ECG recordings which includes normal and supraventricular heartbeats. Only the first two components, $PCA_{1,2}$, are retained. The PCA weights (or basis) are calculated in the QRS complex region limited by a dotted box. Note the similar weight patterns for the normal and supraventricular classes. These patterns depends on the heartbeat morphology as can be seen for the ventricular and fusion classes in Figure 4.3.

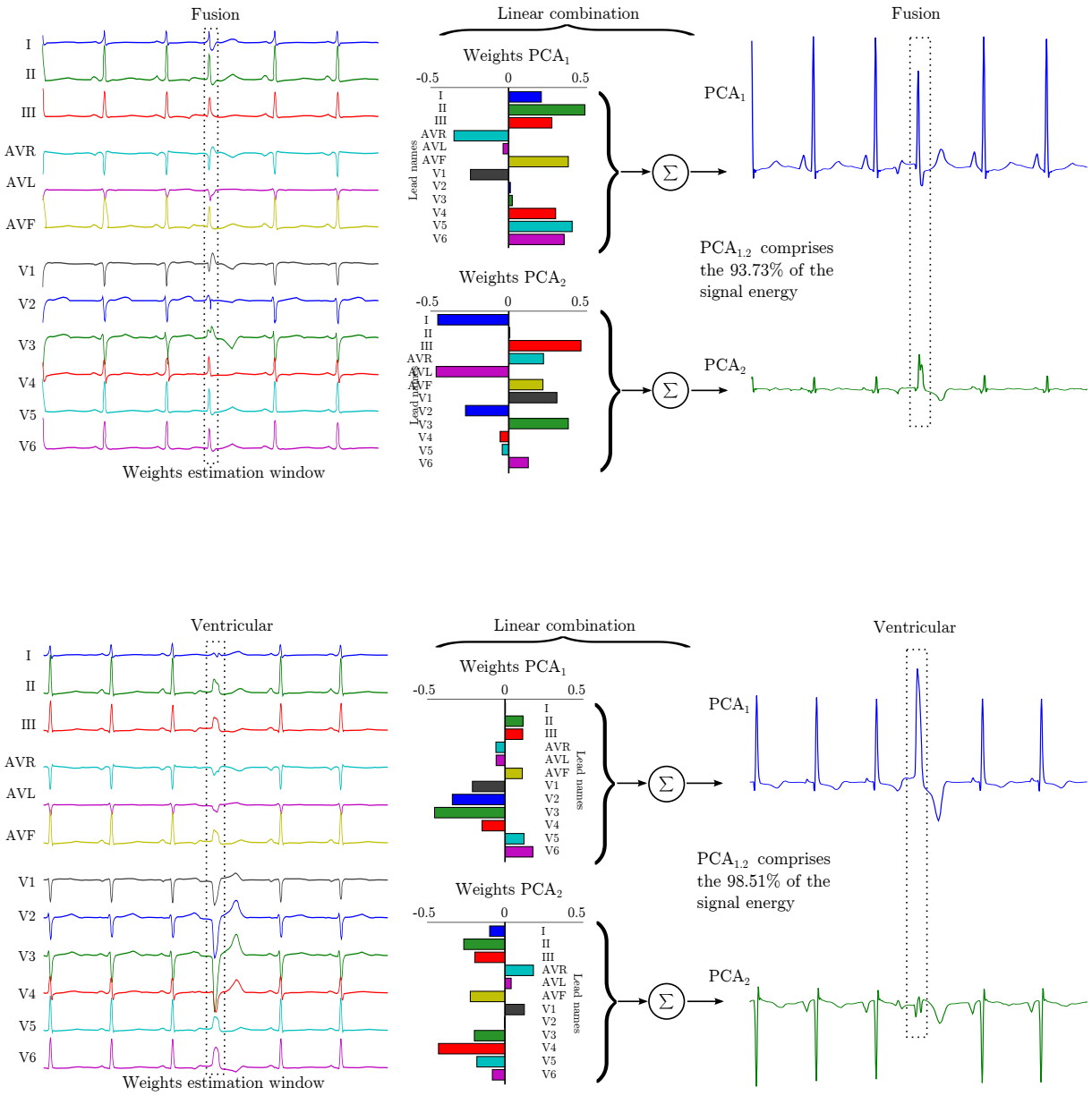


Figure 4.3: Examples of using PCA in 12-lead ECG recordings which includes fusion and ventricular heartbeats. Only the first two components, $PCA_{1,2}$, are retained. The PCA weights (or basis) are calculated in the QRS complex region limited by a dotted box.

that the first two PCA of the ECG retain most of the signal energy [Acar and Köymen, 1999], one or two principal components are selected to compute in them the morphological features. Note that the main difference between ECG-PCA and WT-PCA strategies, as it can be seen in Figure 4.4, is where the PCA is calculated: in the first case, PCA is applied to the ECG signal, while in the second case, PCA is applied to the $W_4^s(k)$ signal.

Then for strategies 12/3L, VCG, ECG-PCA we calculated $W_4^s(k)$ for each ECG lead. Remember that for WT-PCA, the WT was calculated previous to the PCA. After that, the autocorrelation sequence of $W_4^s(k)$ is calculated for each strategy, obtaining $r_L(k)$, where the final step consists in detecting the first zero-crossing (k_Z^L), and the position of the first minimum (k_M^L), as shown in Figure 4.4.

4.2.1.1 Robust Covariance Matrix Computation

As shown above, the PCA computation relies heavily on the calculation of the covariance matrix $\mathbf{C}$. We already discussed in Section 2.4.4 the importance of performing robust estimations, since our signals $\mathbf{S}$ can be corrupted with noise. For this reason the estimation of the $\mathbf{C}$ matrix was performed with the minimum covariance determinant estimator (MCD) presented in [Rousseeuw, 1984] and improved in [Rousseeuw and Van Driessen, 1999]. This is given by the subset of h observations with smallest covariance determinant. The MCD location estimate is then the mean of those h points, and the MCD scatter estimate is their covariance matrix. The default value of h is roughly $0.75 \cdot n$, where n is the total number of observations. Based on the raw estimates, weights are assigned to the observations such that outliers get zero weight. The reweighted MCD estimator is then given by the mean and covariance matrix of the cases with non-zero weight. To compute the MCD estimator, the *FastMCD* algorithm is used [Rousseeuw and Van Driessen, 1999]. The MCD method is intended for continuous variables, and assumes that the number of observations n is at least 5 times the number of variables p.

The MCD is a robust method in the sense that the estimates are not unduly influenced by outliers in the data, even if there are many outliers. Due to the MCD's robustness, we can detect outliers by their large robust distances. The latter are defined like the usual Mahalanobis distance [van der Heijden et al., 2005]

$$d_{\boldsymbol{C}}^{\mathrm{Mah}}(\mathbf{x}, \boldsymbol{\mu}) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{C}^{-1} (\mathbf{x} - \boldsymbol{\mu})}, \tag{4.4}$$

but using the MCD location estimate $\boldsymbol{\mu}_k$, and scatter matrix $\boldsymbol{C}_k$

$$d_{\boldsymbol{C}}^{\mathrm{MCD}}(\mathbf{x}, \boldsymbol{\mu}_k) = \sqrt{(\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{C}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)}, \tag{4.5}$$

instead of the nonrobust sample mean $\boldsymbol{\mu}$, Eq. (2.22) and covariance matrix $\boldsymbol{C}$, Eq. (2.23). The matrix $\boldsymbol{C}_k$ has the smaller $\det(\boldsymbol{C}_k)$ among the K data subsets explored iteratively as described in [Rousseeuw and Van Driessen, 1999]. The *FastMCD* algorithm uses several

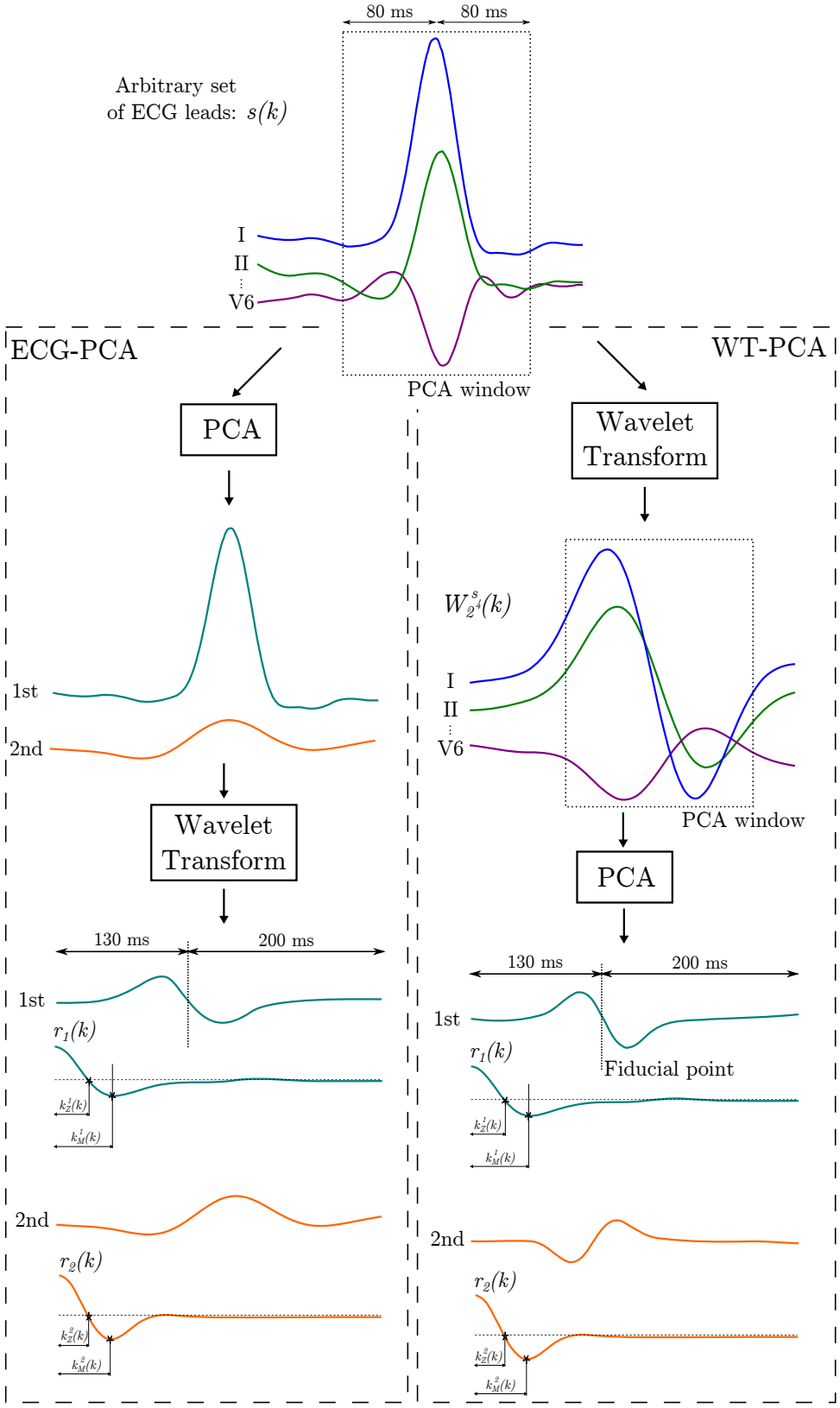


Figure 4.4: Illustration of the features calculated from the wavelet correlation signals. The autocorrelation sequence of the QRS complex at scale 4 is shown for both 12L-PCA and WT-PCA strategy. The calculated features, zero-crossings and peaks of the autocorrelation sequence, are indicated with an asterisk.

Table 4.3: Performance comparison between the different strategies separating AAMI2 classes (N, S, V') in INCART.

				Normal		Suprav.		Ventr.		Total		
Set of leads	Comments	Leads	# Features	S	P^+	S	P^+	S	P^+	A	S	P^+
12L	best single lead	III	6	98	92	86	87	83	88	89	89	89
	Ref. model[1]	II-V1	8	97	93	87	87	84	89	90	90	90
	all leads	all	28	97	93	86	89	86	89	90	90	90
VCG	best lead	Y	6	98	93	83	83	81	85	87	87	87
		X Y Z	10	98	93	82	84	83	85	87	87	87
ECG-PCA		1	6	98	93	87	87	86	90	90	90	90
		1-2	8	98	93	82	87	86	86	89	89	88
WT-PCA		1	6	99	93	86	90	89	90	91	91	91
	Selected model	**1-2**	**8**	**98**	**93**	**86**	**91**	**90**	**90**	**92**	**92**	**91**

The selected strategy is in bold.
[1][Llamedo and Martínez, 2011a]

Table 4.4: Performance comparison between the different strategies separating AAMI2 classes (N, S, V') in three pseudo-orthogonal leads from INCART.

				Normal		Suprav.		Ventr.		Total		
Set of leads	Comments	Leads	# Features	S	P^+	S	P^+	S	P^+	A	S	P^+
3L	best lead	V2	6	98	91	88	82	77	89	87	87	88
	all leads	all	10	98	93	86	86	84	88	89	89	89
ECG-PCA		1-2	8	98	93	87	87	86	91	90	90	90
WT-PCA		**1-2**	**8**	**98**	**93**	**87**	**89**	**88**	**91**	**91**	**91**	**91**

The selected strategy is in bold.

time-saving techniques which make it available as a routine tool to analyze data sets with large n, and to detect deviating substructures in them. For the experiments performed in this thesis, we used the implementation of *FastMCD* available in LIBRA: the Matlab Library for Robust Analysis [Verboven and Hubert, 2005].

4.2.2 Results

As a first experiment we compared the classification performance of the different multilead strategies (12L, VCG, ECG-PCA, WT-PCA) in the INCART database. For each of the strategies we tested the performance in different subsets of leads. Table 4.3 shows class and global classification performances using a k-fold ($k = 10$) cross-validation approach. In Table 4.4, we show the results of a similar experiment in the same database, but considering that only three pseudo-orthogonal leads were available (AvF, V2, V5). In both cases the best model resulted the one with features computed from the first two leads of the WT-PCA set. Therefore, this multilead strategy is considered for the next experiment.

In the last experiment we validated the generalization capability of the classification

Table 4.5: Performance for all databases where the generalization of the WT-PCA strategy was studied.

				Normal		Suprav.		Ventr.		Total		
	Tested in	Obs.	Trained in	S	P^+	S	P^+	S	P^+	A	S	P^+
12 leads	INCART	*biased*	*INCART*	*99*	*94*	*89*	*92*	*91*	*92*	*93*	*93*	*93*
			crossval.	98	93	86	91	90	90	92	92	91
			Biosigna	99	91	86	92	89	90	91	91	91
	Biosigna	*biased*	*Biosigna*	*98*	*86*	*92*	*86*	*77*	*97*	*89*	*89*	*90*
			crossval.	98	85	91	85	75	97	88	88	89
			INCART	97	86	89	84	77	96	88	88	89
3 leads	INCART	*biased*	*INCART*	*98*	*94*	*92*	*91*	*89*	*95*	*93*	*93*	*93*
			crossval.	98	93	87	89	88	91	91	91	91
			Biosigna	99	91	84	90	86	88	90	90	90
	Biosigna	*biased*	*Biosigna*	*98*	*85*	*93*	*84*	*73*	*97*	*88*	*88*	*89*
			crossval.	98	84	92	84	73	97	87	87	88
			INCART	96	86	89	84	76	92	87	87	87
2 leads	MITBIH-AR	*biased*	*MITBIH-AR*	*95*	*82*	*83*	*87*	*78*	*87*	*85*	*85*	*85*
			INCART	97	74	75	83	71	89	81	81	82
			crossval.	94	77	70	82	75	80	79	79	80
			MITBIH-SUP	91	68	67	70	65	93	75	75	77
			Biosigna	97	58	40	71	68	89	68	68	72
	MITBIH-SUP	*biased*	*MITBIH-SUP*	*94*	*83*	*78*	*75*	*75*	*90*	*82*	*82*	*83*
			crossval.	93	80	74	74	74	88	80	80	81
			Biosigna	97	66	49	77	76	85	74	74	76
			MITBIH-AR	94	77	45	75	81	69	73	73	73
			INCART	96	75	44	74	77	69	72	72	73
	INCART	*biased*	*INCART*	*98*	*93*	*91*	*88*	*84*	*93*	*91*	*91*	*91*
			MITBIH-AR	93	93	88	84	85	89	89	89	89
			Biosigna	99	89	84	88	83	88	89	89	88
			crossval.	98	91	82	83	79	84	86	86	86
			MITBIH-SUP	88	95	88	76	81	88	86	86	86
	Biosigna	*biased*	*Biosigna*	*98*	*84*	*91*	*85*	*74*	*97*	*88*	*88*	*88*
			crossval.	98	84	90	83	72	96	87	87	88
			MITBIH-SUP	85	91	94	69	68	96	82	82	85
			INCART	97	86	68	81	74	73	80	80	80
			MITBIH-AR	94	87	52	77	80	65	76	76	76

For the biased case train and test distributions are forced to be the same, this being the best performance achievable for this model.

model to other databases for different number of leads. For all databases available for a given number of leads, we assessed the performance using all possible pairs of different databases as train and test sets. We also evaluated the crossvalidated performance within each database. To have an upper-bound reference, we additionally assessed the performance of the model when trained and tested in the same database. This optimistically biased performance serves as an upper bound, and represents the performance of the model if the distributions of the examples in both training and test datasets were identical. These results, grouped by test database, are presented in Table 4.5 for databases with 12, 3 and 2 leads. Results show that the reference model extended with the selected WT-PCA multilead strategy presents good generalization properties for 3 and 12 leads, while certain degradation is observed when using only two leads.

4.2.3 Discussion and conclusions

In this work we have adapted and improved a two-lead heartbeat classifier by including the additional morphology information present in multilead recordings, like those of INCART database. We followed the concept of the morphology features assessed in [Llamedo and Martínez, 2011a], but calculating these features in sets of leads obtained by following different lead transformation strategies. The simplest strategies consisted in computing the features k_Z^L and k_M^L in all available leads (12L/3L), and in the derived orthogonal leads (VCG). The other two strategies apply PCA to the ECG or its WT previously to the morphology feature computation. The results suggest that strategies using PCA performed better. Moreover the WT-PCA strategy obtains the best improvements with respect to the two-lead classifier obtained in [Llamedo and Martínez, 2011a], either in recordings of 12 or 3 leads. This can be explained because in WT-PCA, the PCA is calculated in $W_4^s(k)$ where most of the noise and other components not related to the QRS have been filtered out [Martínez et al., 2004], and therefore PCA provides a better representation of the multilead evolution of the QRS complex. It must be remembered that although both PCA and WT are linear transformations, the eigenvector calculation is not linear, and therefore sets ECG-PCA and WT-PCA differ in how the multilead signal is projected into the principal components. Another important improvement achieved with the WT-PCA strategy is the robustness against lead misplacement or recordings with undocumented leads. Tables 4.3 and 4.4 show that only the WT-PCA strategy showed the largest observed performance improvement with respect to the two-lead reference model developed in [Llamedo and Martínez, 2011a].

Results in Table 4.5 confirm the generalization capability of the model using the selected WT-PCA strategy to the rest of the studied databases. For the case of 12-lead recordings, as those included in the INCART and Biosigna databases, results show very good generalization for both databases since the performance is slightly lower than the biased performance when training in the other database. In the same table, almost the

same figures can be seen for the case of 3-lead recordings. The last results presented in Table 4.5 are for databases of two leads. It can be noted that the inter-database dispersion in the performance increased, probably because of the heterogeneity of the databases considered.

The results suggest that databases INCART and MITBIH-AR share similar distributions in the feature space, since both obtained the maximal reciprocal performance. Other interesting aspect regarding the MITBIH databases is the lower performance obtained even for the biased case. This fact evidences the diversity of patients and ECG contamination (noise, lead disconnection and misplacement) included in these databases; and therefore the limitation of our classifier to model the data and achieve higher performances. Certainly the biased performance can be thought as a metric of how difficult to classify is a database by a given classification model, the closer to 100%, the easier. This last result reinforces the importance of evaluating the performance of a classifier in several databases.

One advantage of the proposed approach is that it can be used for an arbitrary number of leads, because after the PCA we only retain the two most important components for the morphological feature calculation. These components are calculated specifically for the QRS complex, and in the $W_4^s(k)$ signal (with a band between 11.25 and 22.5 Hz), typically where the ECG presents high SNR. However in case of a large-scale artifact during the QRS complex (as a lead disconnection), the PCA calculation would be corrupted, being this the main limitation found for this approach. This problem is addressed by the robust MCD algorithm to compute the covariance matrix.

The performance improvement with respect to [Llamedo and Martínez, 2011a] is however moderate, probably because the automatic classification approach is close to the performance limit achievable with the current classification model. The worst aspect of performance remains classification of supraventricular ectopic beats, where further study is needed. Regarding the ventricular class, techniques of patient adaptation, as described in [de Chazal and Reilly, 2006], will be presented in Chapter 5.

These results represent an improvement in performance with respect of the previous two-lead classification model, concluding that the adequate addition of multilead information allows the performance improvement of a heartbeat classifier.

4.3 Neural network classifier

In this section, the objective is to improve the classification method used. It is not difficult to understand that a simple classifier as the LDC, which is capable of dividing the hyperspace with hyperplanes, is a suboptimal solution for a complex problem such as the classification of heartbeats. The design of an automatic classifier based on neural networks shares the same methodology described in Chapter 3, but were used in a different set of features. The results presented in this section were published in [Mar et al., 2011],

in collaboration with the Institut für Biomedizinische Technik, at the Dresden University of Technology, Germany.

4.3.1 Feature Sets

We used a set of 71 features, divided (as shown in Table 4.6) into the following categories:

- Temporal features, which already proved, in different studies, to be the most relevant [Lannoy, 2010]. This category includes heart-rate features, which were the only features computed just once for both channels, and features obtained from the segmentation information yielded by *ecgpuwave* [Laguna et al., 1994].

- Morphological features, also previously assessed as being of great relevance, made the bulk of the feature set. Direct samples from the ECG signal, and computed measurements such as area, power or extrema, were included.

- Statistical features completed the feature set, including different order moment-based indexes and histogram variance.

Unlike in temporal features, which were acquired from time domain signals exclusively, the DWT of the ECG signal was used to obtain some of the features belonging to the morphological and statistical categories. DWT features were based on the same heartbeat intervals as the features obtained from the time domain, but using the scales 2, 3, 4 and 5 of the DWT ECG signal. Details about the DWT can be found in Section 2.3.2.

Finally, all features were individually normalized, by computing the necessary scaling to make the features from DS1 signals be mean 0 and variance 1, and normalizing the corresponding features from DS2 with the obtained scaling factors.

4.3.2 Feature Selection

The SFFS procedure is the same described in Section 2.4.6, but the optimization criterion was different. We propose a new performance measure which tackles specifically the problem of providing in a single value complete information about how *good* an ECG classification has been. To this end, this new index was chosen to be a combination of two values defined in Section 2.4.5: the j index, which specifically evaluates the discrimination of the most important ECG arrhythmias (S and V beats),

$$j = S_S + S_V + P_S^+ + P_V^+ \tag{4.6}$$

and the *Kappa* (κ) index, which globally evaluates the confusion matrix [Cohen, 1960]. This index, despite having been proposed as an evaluation coefficient several decades ago, and its potential convenience, had never been applied before in the heartbeat classification.

Table 4.6: Features distributed by categories.

Features	Time Domain	DWT Domain*
Temporal	Previous RR, current RR, average RR, average RR of the last 10 beats, QRS duration, T-Wave duration, P-Wave flag.	
Morphological	Downsampled (10 samples) QRS, downsampled (9 samples) T-Wave, QRS area, QRS power, QRS max, QRS min, QRS Max-Min ratio, peak width at 70% Max, peak slope, beat area, beat power, beat max, beat min, beat Max-Min ratio.	Max(3,4,5), Min(3,4,5), difference between Max and Min (3,4,5), distance (in samples) between Max and Min (3,4), power(2,3,4,5), power ratio(3-2,4-3,5-4).
Statistical	QRS variance, QRS skewness, QRS kurtosis, QRS histogram (20 slots) variance, beat mean, beat variance, beat skewness, beat kurtosis, beat histogram (20 slots) variance.	Mean (3,4), standard deviation(3,4), skewness (3,4).

* Numbers between parenthesis represent the scales at which the feature was obtained.

From its definition,

$$\kappa = \frac{\sum_{i=1}^{C} n_{ii}^{T} - \sum_{i=1}^{C} D_i}{N_T - \sum_{i=1}^{C} D_i} \tag{4.7}$$

where

$$D_i = \frac{(N_i P_i)}{N_T} \tag{4.8}$$

is known as the weighted detections, it can be seen that it evaluates the global quality of the classification: like the multiway accuracy, it also represents a complete evaluation of the confusion matrix (in a single value and weighting each beat equally), but it is much less influenced by the class imbalance.

The resulting combined index, which we named $j\kappa$ index ($I_{j\kappa}$), takes into account the misclassification and the imbalance present between all the considered classes, thanks to the included κ index, and at the same time emphasizes the discrimination of the most important arrhythmias (S and V), thanks to the j index (I_j).

$$I_{j\kappa} = w_1\kappa + w_2 I_j \tag{4.9}$$

As j takes values in the 0-4 range and κ in the 0-1 range, w_1 was set to 1/2 and w_2 to 1/8, so that both factors influence equally the overall result. Consequently $I_{j\kappa}$ takes values between 0 and 1, where 1 indicates perfect classification.

Table 4.7: DS1 Division Scheme for MLP Evaluation

Dataset	N	S	V	F	Total
Eval1 DS1	23379	442	1936	40	25797
Eval2 DS1	22283	501	1842	373	24999
Total (DS1)	45662	943	3778	413	50796

Eval1 DS1 comprises data from the recordings 109, 114, 118, 119, 124, 201, 203, 205, 215, 220 and 223
Eval2 DS1 comprises data from the recordings 101, 106, 108, 112, 115, 116, 122, 207, 208, 209 and 230.

4.3.3 Multi-Layer Perceptron

The multilayer perceptron (MLP) belongs to the class of supervised learning networks, on which the discriminative power is gained through a preliminary learning phase, where labeled examples are presented to the network. The most common training strategy, also used in the present study, is the *backpropagation* (BP) algorithm [Rumelhart et al., 1986]. It works by computing the error between the returned and the known, desired output, employing it to adjust the MLP weights. Although the training process requires a rather long time, the implementation and execution of a trained MLP is very simple, making this paradigm very suited too for classification on ambulatory settings. On the other hand, its characteristics make this paradigm very inadequate to guide the feature selection (FS) process: The random initialization makes MLPs' results not constant, which renders the FS procedure unreliable if only one MLP is evaluated for each tested subset. The unreliability could be overcome by training many MLPs for each tested subset, and performing statistical analysis to obtain a result that would lead to the next step in the FS process. Unfortunately, due to the many subsets tested by the SFFS procedure, plus the relatively long time that training each MLP requires, the time and resources that a reliable MLP-SFFS procedure would take are beyond our computing capabilities.

Therefore, in the present study the MLP paradigm was only applied to classify ECG arrhythmias with the feature set obtained from the SFFS. The values of the different parameters governing the MLP were determined by applying *2*-fold cross-validation on DS1, training with one fold the MLP parameterized with the desired combination and evaluating with the remaining one, and vice versa, averaging the results. Again, this sub-division was performed inter-patiently in such a way that all heartbeat classes were similarly represented in each of the folds, as shown on Table 4.7. MLPs with a single hidden layer of 25 neurons were used, trained with a learn rate of 0.25 and a momentum of 0.03 to avoid getting stuck into local minima. The number of training cycles was chosen to be the one for which the mean results from the 2-fold evaluation began to get worse, i.e. when symptoms of over-learning appeared.

4.3.4 Classifier Combination

As mentioned above, information from both ECG leads was considered throughout the whole study. Except for the heart-rate ones, all features were obtained separately for each lead, and the two resulting feature sets applied independently to perform classification. The posterior probabilities obtained after classification with each feature set were then combined using the Bayesian product integration scheme [de Chazal et al., 2004]

$$P(i|\mathbf{x}) = \frac{\prod_{l=1}^{L} P_l(i|\mathbf{x})}{\sum_{c=1}^{C} \prod_{l=1}^{L} P_l(c|\mathbf{x})}, \tag{4.10}$$

for $L = 2$ leads and C classes. Finally, each heartbeat was labeled with the class with higher posterior probability after the combination.

4.3.5 Results

In the present study aiming at reducing the classifiers complexity as much as possible, but without making its performance worse, we selected as the most suited for our purposes the classifier with the smallest number of features which achieved at least the performance obtained with the original feature set on DS1. The reduced subset accomplishing this criterion contains 9 features which includes Previous RR, current RR, RR average, beat min, beat max, QRS max-min ratio, peak slope, max-min Difference on DWT scale 3.

After evaluating the performance of the SFFS procedure with the matched classifier on DS1, the original feature subset and the selected from the SFFS were tested on DS2 to carry out the final evaluation of the classifier model. Additionally, these subsets were also tested on DS2 with the MLP classifier, in order to analyze the generalizing capability of the SFFS procedure in the case where the criterion function and the classifier paradigm do not match. At the same time, this analysis also tackles the suitability of the MLP for heartbeat classification, in direct comparison to the LDC classifier. In Table 4.8, complete classification description is displayed in the form of the confusion matrices for the results obtained by applying the reduced feature set with either classifier paradigm. These matrices provide insight into the beat-by-beat performance and ease future comparison attempts by other authors. For the rest of studied feature sets, results both for LDA and MLP classification are given through the considered performance measures on Table 4.8.

4.3.6 Discussion and conclusions

In addition to the analysis of the SFFS process itself, it is also interesting to identify the most relevant features obtained. Inspecting the selected subset we can observe the *previous RR*, *current RR* and *RR average* features ($RR[i-1]$, $RR[i]$ and RR_A in Figure 2.6) were present. This indicates the uttermost importance that heart-rate features have on ECG classification. This fact was also corroborated the model found in Chapter 3 and

Table 4.8: Classifier Performances on DS2 Obtained for the most Relevant Studied Classifier Models

LDC		Algorithm				
		n	s	v	f	Total
Truth	N	37384	2726	691	3260	44061
	S	60	1517	237	16	1830
	V	45	225	2782	156	3208
	F	137	1	50	200	388
	Total	37626	4469	3760	3632	49487

MLP		Algorithm				
		n	s	v	f	Total
Truth	N	39497	2778	771	1015	44061
	S	122	1523	93	92	1830
	V	104	235	2783	86	3208
	F	125	6	20	237	388
	Total	39848	4542	3667	1430	49487

	Normal		Suprav.		Ventr.		Fusion		Total					
Classifier	S	P^+	S	P^+	S	P^+	S	P^+	A	S	P^+	$\boldsymbol{j}$	κ	$I_{j\kappa}$
LDC	85	99	83	34	87	74	52	6	85	77	53	2.78	0.51	0.60
MLP	90	99	83	34	87	76	61	17	89	80	57	2.79	0.60	0.65

Table 4.9: Relevant Indices for AAMI standard and Inter-Patient Division Conform Studies, Including Present Study's ones.

			Total		
Classifier	A	S	$\boldsymbol{j}$	κ	$j\kappa$
LDC [Mar et al., 2011]	85	77	2.78	0.51	0.60
MLP [Mar et al., 2011]	89	80	2.79	0.60	0.65
LDC [de Chazal et al., 2004]	86	83	2.764	0.532	0.612
LDC* [de Chazal and Reilly, 2006]	94	88	3.234	0.754	0.781
LDC [Llamedo and Martínez, 2007]	80	80	2.372	0.421	0.507
SVM [Park et al., 2008]	86	76	–	–	–
SVM [Lannoy, 2010]	–	83	–	–	–
LDC** [Llamedo and Martínez, 2011a]	78	83	2.887	0.412	0.567

* Patient adapting: Requires expert intervention.

** Feature set optimized for classification of N, S and V' classes, where V' class included V and F beats.

presented in Table 4.2. Regarding the results for the feature set obtained, which contained a much larger number of features, including many statistical ones, it is remarkable to observe that the 6 non-heart-rate features are all morphological features too, and even more noteworthy, that 5 of them represent extrema. This fact suggests that extrema values possess the highest discriminative power among all morphological features.

In spite of the large number of studies in which the MLP classifier paradigm has been applied for ECG classification, none among them could be found in which the results were evaluated in conformance with the AAMI standard and inter-patient dataset distribution. Yet, results show that, when applying reduced feature sets, the MLP can clearly outperform LDC in the task of heartbeat classification. Comparing the results of both paradigms, an improvement in the range of the 4% can be observed in the *global accuracy* (A) and in the *global sensitivity* (S). Nevertheless, it should be noted that these numbers are just orientative of the possible improvement, as, due to their random initialization, successive evaluations of the MLP with the same feature set could yield different results.

Comparing the achieved performances with those of previous studies, it provides further insight on the suitability of the proposed techniques. As mentioned, this comparison can only be objectively done among those studies following the same constrains. Thus, the performances obtained in the present study on both classifier paradigms have been compared with the results of the other AAMI conforming studies which followed the inter-patient division scheme (see Table 4.9).

However, the ones achieved with the MLP outperform all previous non-adapting proposed methods. These results evince that the non-linear classification capabilities of this type of MLP are extremely suitable to perform heartbeat classification, which is intrinsically non-linear too. Moreover, they also show, in the context of this study, a greater generalization capability of the MLP when compared to algorithmic methods such as LDC, suggesting that they may be a more appropriate tool for ECG heartbeat classification.

4.A Detailed Results

In this section we present the confusion matrices for ease the comparison of future works. The summarized performances presented in the previous sections were calculated from these numbers, according to the methodologies described in Section 2.4.5.

Table 4.10: Confusion matrices of the results presented in Table 4.5 for 12-leads databases.

biased INCART

Truth \ Algorithm	n	s	v	Total
N	151649	1553	449	153651
S	69	1741	149	1959
V	481	1361	18382	20224
Total	152199	4655	18980	175834

biased Biosigna

Truth \ Algorithm	n	s	v	Total
N	280874	5031	341	286246
S	77	1222	27	1326
V	260	337	1944	2541
Total	281211	6590	2312	290113

crossval INCART

Truth \ Algorithm	n	s	v	Total
N	151077	2055	519	153651
S	91	1685	183	1959
V	526	1674	18024	20224
Total	151694	5414	18726	175834

crossval Biosigna

Truth \ Algorithm	n	s	v	Total
N	280555	5206	485	286246
S	83	1212	31	1326
V	266	357	1918	2541
Total	280904	6775	2434	290113

Biosigna - INCART

Truth \ Algorithm	n	s	v	Total
N	151488	1399	764	153651
S	88	1685	186	1959
V	828	1267	18129	20224
Total	152404	4351	19079	175834

INCART - Biosigna

Truth \ Algorithm	n	s	v	Total
N	280163	6586	805	287554
S	109	1189	37	1335
V	213	366	1993	2572
Total	280485	8141	2835	291461

Table 4.11: Confusion matrices of the results presented in Table 4.5 for 3-leads databases.

biased INCART

Truth \ Algorithm	n	s	v	Total
N	150937	1857	857	153651
S	70	1811	78	1959
V	565	1606	18053	20224
Total	151572	5274	18988	175834

biased Biosigna

Truth \ Algorithm	n	s	v	Total
N	280840	5176	230	286246
S	74	1227	25	1326
V	289	391	1861	2541
Total	281203	6794	2116	290113

crossval INCART

Truth \ Algorithm	n	s	v	Total
N	150307	2278	1066	153651
S	88	1705	166	1959
V	557	1841	17826	20224
Total	150952	5824	19058	175834

crossval Biosigna

Truth \ Algorithm	n	s	v	Total
N	279365	5567	1314	286246
S	83	1219	24	1326
V	298	400	1843	2541
Total	279746	7186	3181	290113

Biosigna - INCART

Truth \ Algorithm	n	s	v	Total
N	151516	1393	742	153651
S	89	1653	217	1959
V	1114	1778	17332	20224
Total	152719	4824	18291	175834

INCART - Biosigna

Truth \ Algorithm	n	s	v	Total
N	275435	5401	5410	286246
S	84	1181	61	1326
V	231	389	1921	2541
Total	275750	6971	7392	290113

Table 4.12: Confusion matrices of the results presented in Table 4.5 for 2-leads databases.

biased MITBIH-AR

Truth \ Algorithm	n	s	v	Total
N	85189	3809	1091	90089
S	181	2305	293	2779
V	1087	663	6059	7809
Total	86457	6777	7443	100677

Biosigna - MITBIH-AR

Truth \ Algorithm	n	s	v	Total
N	86959	1594	1536	90089
S	1479	1109	191	2779
V	1366	1147	5296	7809
Total	89804	3850	7023	100677

Biosigna - MITBIH-SUP

Truth \ Algorithm	n	s	v	Total
N	157104	1740	3427	162271
S	4848	5918	1429	12195
V	1093	1327	7543	9963
Total	163045	8985	12399	184429

crossval MITBIH-AR

Truth \ Algorithm	n	s	v	Total
N	84903	3838	1348	90089
S	365	1935	479	2779
V	1144	843	5822	7809
Total	86412	6616	7649	100677

MITBIH-SUP - MITBIH-AR

Truth \ Algorithm	n	s	v	Total
N	82076	7093	920	90089
S	810	1870	99	2779
V	1048	1657	5104	7809
Total	83934	10620	6123	100677

INCART - MITBIH-SUP

Truth \ Algorithm	n	s	v	Total
N	156237	1935	4099	162271
S	2863	5332	4000	12195
V	860	1393	7710	9963
Total	159960	8660	15809	184429

INCART - MITBIH-AR

Truth \ Algorithm	n	s	v	Total
N	87249	1993	847	90089
S	493	2075	211	2779
V	1230	1014	5565	7809
Total	88972	5082	6623	100677

biased MITBIH-SUP

Truth \ Algorithm	n	s	v	Total
N	152814	8623	834	162271
S	1743	9466	986	12195
V	540	1984	7439	9963
Total	155097	20073	9259	184429

MITBIH-AR - MITBIH-SUP

Truth \ Algorithm	n	s	v	Total
N	152165	4887	5219	162271
S	2614	5475	4106	12195
V	713	1198	8052	9963
Total	155492	11560	17377	184429

Table 4.13: Confusion matrices of the results presented in Table 4.5 for 2-leads databases.

crossval MITBIH-SUP

Truth \ Algorithm	n	s	v	Total
N	151390	9278	1603	162271
S	2126	9010	1059	12195
V	610	2019	7334	9963
Total	154126	20307	9996	184429

INCART - Biosigna

Truth \ Algorithm	n	s	v	Total
N	278776	3292	4178	286246
S	74	907	345	1326
V	266	385	1890	2541
Total	279116	4584	6413	290113

biased INCART

Truth \ Algorithm	n	s	v	Total
N	150838	2084	729	153651
S	58	1780	121	1959
V	953	2276	16995	20224
Total	151849	6140	17845	175834

biased Biosigna

Truth \ Algorithm	n	s	v	Total
N	280772	5251	223	286246
S	90	1204	32	1326
V	295	375	1871	2541
Total	281157	6830	2126	290113

MITBIH-AR - Biosigna

Truth \ Algorithm	n	s	v	Total
N	269190	12528	4528	286246
S	71	695	560	1326
V	226	276	2039	2541
Total	269487	13499	7127	290113

Biosigna - INCART

Truth \ Algorithm	n	s	v	Total
N	151875	1423	353	153651
S	97	1650	212	1959
V	1367	2165	16692	20224
Total	153339	5238	17257	175834

crossval Biosigna

Truth \ Algorithm	n	s	v	Total
N	279915	6081	250	286246
S	94	1196	36	1326
V	308	399	1834	2541
Total	280317	7676	2120	290113

MITBIH-SUP - Biosigna

Truth \ Algorithm	n	s	v	Total
N	243485	42445	316	286246
S	44	1249	33	1326
V	132	688	1721	2541
Total	243661	44382	2070	290113

MITBIH-AR - INCART

Truth \ Algorithm	n	s	v	Total
N	142971	9666	1014	153651
S	52	1718	189	1959
V	784	2196	17244	20224
Total	143807	13580	18447	175834

Table 4.14: Confusion matrices of the results presented in Table 4.5 for 2-leads databases.

MITBIH-SUP - INCART

Truth \ Algorithm	n	s	v	Total
N	135717	17613	321	153651
S	28	1716	215	1959
V	653	3249	16322	20224
Total	136398	22578	16858	175834

crossval INCART

Truth \ Algorithm	n	s	v	Total
N	150278	2783	590	153651
S	72	1610	277	1959
V	1109	3138	15977	20224
Total	151459	7531	16844	175834

Chapter 5

Patient-Adapted ECG Heartbeat Classification

5.1 Introduction

In the current state-of-the-art, it seems that the automatic classification approach has approximated to a performance upper bound, probably because the huge interpatient variability makes impossible that the probability distribution learned in a train set be representative of that found in a test set and during the normal operation of the classifier. The patient adaptation technique by means of expert assistance (i.e. manual heartbeat annotation) was reported to be useful in two works to overcome this problem [Hu et al., 1997, de Chazal and Reilly, 2006], at the expense of sacrificing automaticity. Other works also reported better performances than the ones obtained by automatic classifiers, always taking advantage of the expert assistance [Lagerholm et al., 2000, Jiang and Kong, 2007, Ince et al., 2009, Kiranyaz et al., 2011]. One aspect to study when adopting this technique is the efficient use of the assistance, in order to keep the classifier as much automatic as possible. It is interesting to note that some assisted classifiers require from 2 to 5 minutes of manual annotations, which is equivalent to several hundred of expert labeled heartbeats [Hu et al., 1997, de Chazal and Reilly, 2006, Ince et al., 2009, Jiang and Kong, 2007], while [Kiranyaz et al., 2011] requires the annotation of a number of heartbeats, depending on the amount of beat classes present. One drawback of several patient-adaptable approaches is that they can not operate without assistance [Lagerholm et al., 2000, Ince et al., 2009, Jiang and Kong, 2007, Kiranyaz et al., 2011]. This is not the case of those developed as an evolution of a previous automatic classifier [Hu et al., 1997, de Chazal and Reilly, 2006]. In those cases a fully automatic and an assisted mode can be used.

In this chapter we propose an expert assistance approach pursuing two objectives: first, to be able to perform automatic classification adapted to the patient's heartbeat characteristics, and second, if assistance is available, to take advantage of it efficiently. For this purpose we suggest the integration of a well-known clustering algorithm based

on mixture of Gaussians [van der Heijden et al., 2005], with the linear classifier (LDC) presented in [Llamedo and Martínez, 2011a]. In this solution clustering is responsible of retaining all patient-specific data ordering, while the automatic classifier performs the cluster labeling, and can be assisted by an expert, as will be described in detail in the following sections.

The objective of this chapter is to study how the classification performance of a previously developed multilead algorithm [Llamedo and Martínez, 2011a, 2012a] can be improved, by implementing a patient-adaptation technique based on clustering. Our working hypothesis is that within a given recording, classes are clearly separated and beats of the same class tend to be grouped in one or more homogeneous clusters. In other words, after clustering the beats from a recording, all beats grouped in the same cluster would likely belong to the same class. For that purpose, first we search for an appropriate set of features for intra-recording clustering, and compare several integration strategies in a development dataset, to finally assess the final performance and generalization capability to other databases not considered during the development. The performance will be compared with other state-of-the-art classifiers [de Chazal and Reilly, 2006, Jiang and Kong, 2007, Ince et al., 2009, Kiranyaz et al., 2011, Mar et al., 2011].

5.2 Methodology

5.2.1 ECG databases

All experiments performed in this chapter were carried out in several public databases available on Physionet [Goldberger et al., 2000], and the well known American Heart Association database [American Heart Association]; their relevant details are summarized in Table 2.3. For all databases the AAMI recommendations for class-labeling were adopted (Section 4.2 in [AAMI-EC57, 1998–2008]). The details of the databases used in this chapter can be found in Section 2.1. The databases used include different types of ECG recordings: some of them were recorded during routine ambulatory practice, but others were selected to include less common ventricular, junctional or supraventricular arrhythmias, or baseline ST segment displacement or other ECG abnormalities. As a result, we use in this chapter a dataset with a broad range of normal and pathological ECG recordings to evaluate the algorithm performance. Moreover, the different length of recordings will evidence the ability of the algorithm to handle the nonstationarities present in the ECG.

We will refer as the *development dataset* to the union of the MITBIH-SUP database and the 22 recordings included in the DS1 subset of MITBIH-AR defined in [de Chazal et al., 2004], while the "evaluation dataset" includes the rest of databases described in Table 2.3.

5.2.2 Heartbeats classification

Following the scheme presented in Figure 5.1, the patient-adaptable algorithm includes a linear discriminant classifier (LDC) and an expectation-maximization clustering algorithm (EMC). Both LDC and EMC work independently and each performs a preliminary classification/clustering task in different feature spaces. The LDC was developed and trained as described in [Llamedo and Martínez, 2011a], while the EMC development will be described later. Finally, the heartbeat and cluster labels provided by the LDC and EMC respectively, are integrated with a voting scheme into a final heartbeat label. Three modes of operation are proposed, depending on the degree of expert assistance available in the application scenario: 1) automatic, 2) slightly assisted and 3) assisted. The algorithm performs the following procedures: a) cluster and centroid identification, b) LDC automatic classification and c) expert assistance.

For the *automatic mode*, in each record, K clusters and centroids are identified, corresponding to groups of similar heartbeats, while at the same time, the LDC computes the labels for each heartbeat. Then for each cluster, the algorithm tests if any label obtains a qualified majority, meaning that the most represented label exceeds the α percent of the cluster population. In case this label exists, it is assigned to the whole cluster, superseding the LDC labels. If the qualified majority is not reached, the uncertainty is considered to be too high to change the labels, an thus the LDC labels remain unchanged.

The *slightly assisted mode* is similar to the automatic, with the exception that in case of not finding a class with qualified majority, expert assistance is required to label the cluster centroid and propagate it to the whole cluster, ignoring LDC labels. The procedure of expert assistance is simulated by inspecting the true labels provided with each database.

Finally in the *assisted mode*, K clusters and centroids are identified, then the expert is required to label each centroid. The algorithm concludes assigning these labels to the rest of heartbeats in each cluster.

To better understand the three modes of operation, a toy example can be found in the center of Figure 5.1. The LDC by itself makes 4 errors in cluster 1 and 3 in cluster 2. For the automatic mode, clusters 2 and 3 have majority of V and N classes respectively. Then, votation propagates centroid labels to the rest of examples within both clusters, occurring 1 mistake in cluster 2. In cluster 1 there is no qualified majority for $\alpha = 50\%$, so the LDC labels remain unchanged and 5 mistakes happen. The automatic mode made 1 mistake less than the LDC. For the slightly-assisted mode, only cluster 1 would be modified, by propagating the true label of the centroid S, resulting 4 errors for this cluster. Finally for the assisted mode, the result is the same as in the previous mode, four errors in cluster 1, one error in cluster 2 and no errors in cluster 3. In summary, 7 errors for the LDC, 6 for the automatic mode and 5 for the slightly and assisted modes. As it was shown, the algorithms rely heavily in the ability of the EMC to cluster the heartbeats adequately.

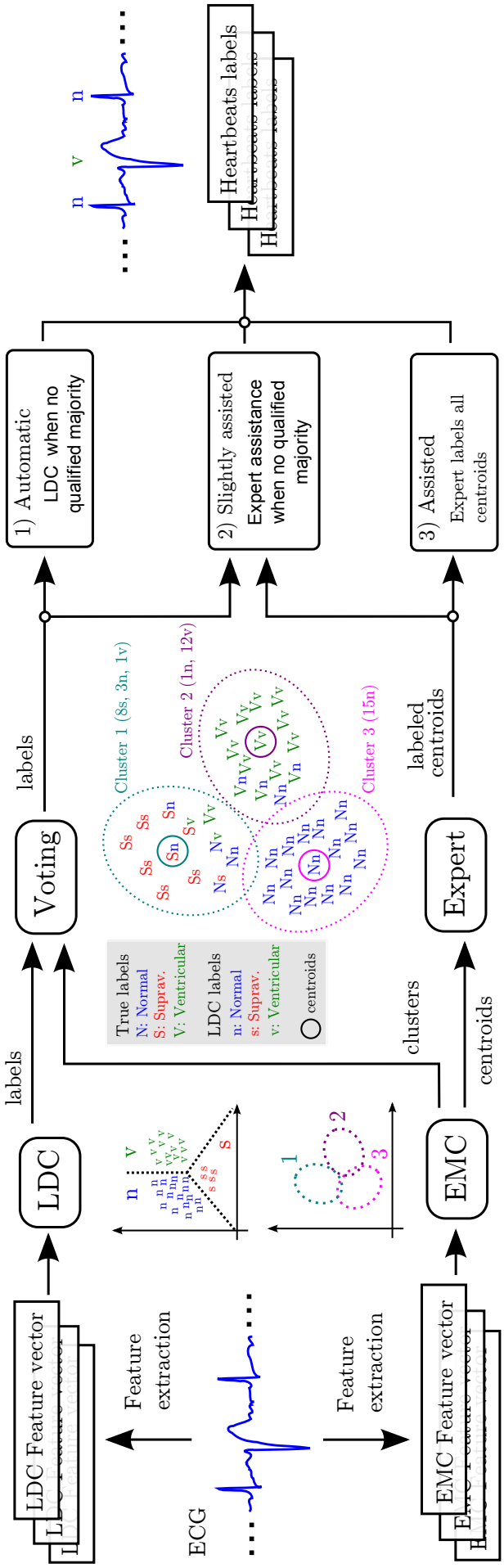


Figure 5.1: Overview of the proposed algorithm. There is a graphical description in the center of the scheme about the task carried out by each block. The toy-example in the middle is also commented in the text to better understand the three modes of operation.

Table 5.1: feature model used by the automatic classifier for recordings of 2 or more leads .

Feature	Description
$\ln(RR[i])$	Current RR interval[1]
$\ln(RR[i+1])$	Next RR interval[1]
$\ln(RR_1)$	Average RR interval in the last minute[1]
$\ln(RR_{20})$	Average RR interval in the last 20 minutes[1]
k_Z^1	Zero-cross position of the WT autocorrelation signal in lead 1[2]
k_Z^2	Zero-cross position of the WT autocorrelation signal in lead 2[2]
k_M^1	Maximum position of the WT autocorrelation signal in lead 1[2]
k_M^2	Maximum position of the WT autocorrelation signal in lead 2[2]

[1] See Figure 2.6 [2] See Figure 2.10

5.2.3 Automatic classifier

We follow a scheme similar to the one in [Llamedo and Martínez, 2011a, 2012a], where we developed a multilead heartbeat classifier with good generalization capability. We used a linear classifier compensated for the class-imbalance, while as feature model we adopted rhythm and morphological features computed in a multilead manner. Regarding to the classifier used, we found that linear discriminant functions were suitable for the heartbeat classification task in terms of performance and generalization capability. The details and equations of this classification model can be found in Section 2.4.2.

The features used by the automatic classifier are described in Table 5.1. The morphology features k_Z^L and k_M^L for lead L, are calculated in the two principal ECG leads after integrating the multilead information with a principal component analysis (PCA). In Chapter 4 it was shown that WT-PCA strategy was a good strategy to include multilead morphology information. Therefore these features account for a multilead morphological description of the QRS complex. For a detailed description of the features and the multilead strategy used see Chapters 3 and 4.

5.2.4 Clustering algorithm

The EMC algorithm used in this chapter is based on the mixture of Gaussians model [van der Heijden et al., 2005]. It consists of estimating the parameters of a density function

$$p(\mathbf{x}|\Psi) = \sum_{k=1}^{K} \pi_k \cdot f(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_\mathrm{k}) \tag{5.1}$$

$$f(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_\mathrm{k}) = \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{(2\pi)^m |\boldsymbol{\Sigma}_k|}} \exp^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)}, \tag{5.2}$$

where the m-dimensional vector $\mathbf{x}$ is modeled by K Gaussians with mixing coefficients π_k, in order to retain a more realistic structure of the data. The parameter set $\Psi = \{\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k | k = 1, \ldots, K\}$ is estimated by maximum likelihood criterion. We maximize the log likelihood

$$L(X|\Psi) = \ln \prod_{n=1}^{N} p(\mathbf{x}_n|\Psi), \tag{5.3}$$

for the N heartbeats in each recording named $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. Since there is not a closed form solution for Ψ by maximizing $L(X|\Psi)$, the well-known expectation-maximization algorithm (EM) is used to obtain the estimation equations of the parameters Ψ, which are the mixing coefficient for each cluster

$$\hat{\pi}_k = \frac{1}{N} \sum_{m=1}^{N} \hat{\beta}_{m,k}, \tag{5.4}$$

the cluster mean

$$\hat{\boldsymbol{\mu}}_k = \frac{1}{N\hat{\pi}_k} \sum_{m=1}^{N} \hat{\beta}_{m,k} \mathbf{x}_m \tag{5.5}$$

and cluster covariance matrix

$$\hat{\boldsymbol{\Sigma}}_k = \frac{1}{N\hat{\pi}_k} \sum_{m=1}^{N} \hat{\beta}_{m,k} (\mathbf{x}_m - \hat{\boldsymbol{\mu}}_k) \cdot (\mathbf{x}_m - \hat{\boldsymbol{\mu}}_k)^{\mathrm{T}}. \tag{5.6}$$

Where $\hat{\beta}_{m,k}$ is known as the ownership variable, which indicates the probability of sample $\mathbf{x}_m$ to have been generated by the k-th component

$$\hat{\beta}_{m,k} = \frac{\hat{\pi}_k \cdot f(\mathbf{x}_m; \hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_{\mathrm{k}})}{\sum_{j=1}^{K} \hat{\pi}_j \cdot f(\mathbf{x}_m; \hat{\boldsymbol{\mu}}_j, \hat{\boldsymbol{\Sigma}}_{\mathrm{j}})}. \tag{5.7}$$

The EM algorithm iteratively computes the weight, location and dispersion for each of the K clusters (Eq. (5.4)-(5.6) respectively), until $\hat{\beta}_{m,k}$ does not change significantly, which is equivalent to obtaining stable clusters. The interested reader is referred to [van der Heijden et al., 2005, Duin et al., 2008] for details, equations and the implementation used in this chapter.

The EM algorithm guarantees the convergence, at least, to a local optimum. The mathematical demonstration of this can be found in [van der Heijden et al., 2005]. However, in Figure 5.2 an example of this is shown. The toy example is build from three Gaussian distributions with different parameters. Even for the case where the algorithm tried to find less components ($K = 2$), the algorithm converges. For the rest of the cases, the redundancy of the components is notorious. This leads to one of the most critical aspects of this kind of clustering algorithms, how to estimate the complexity of the data to cluster. As there is not a reliable methodology to answer this, one recommendation

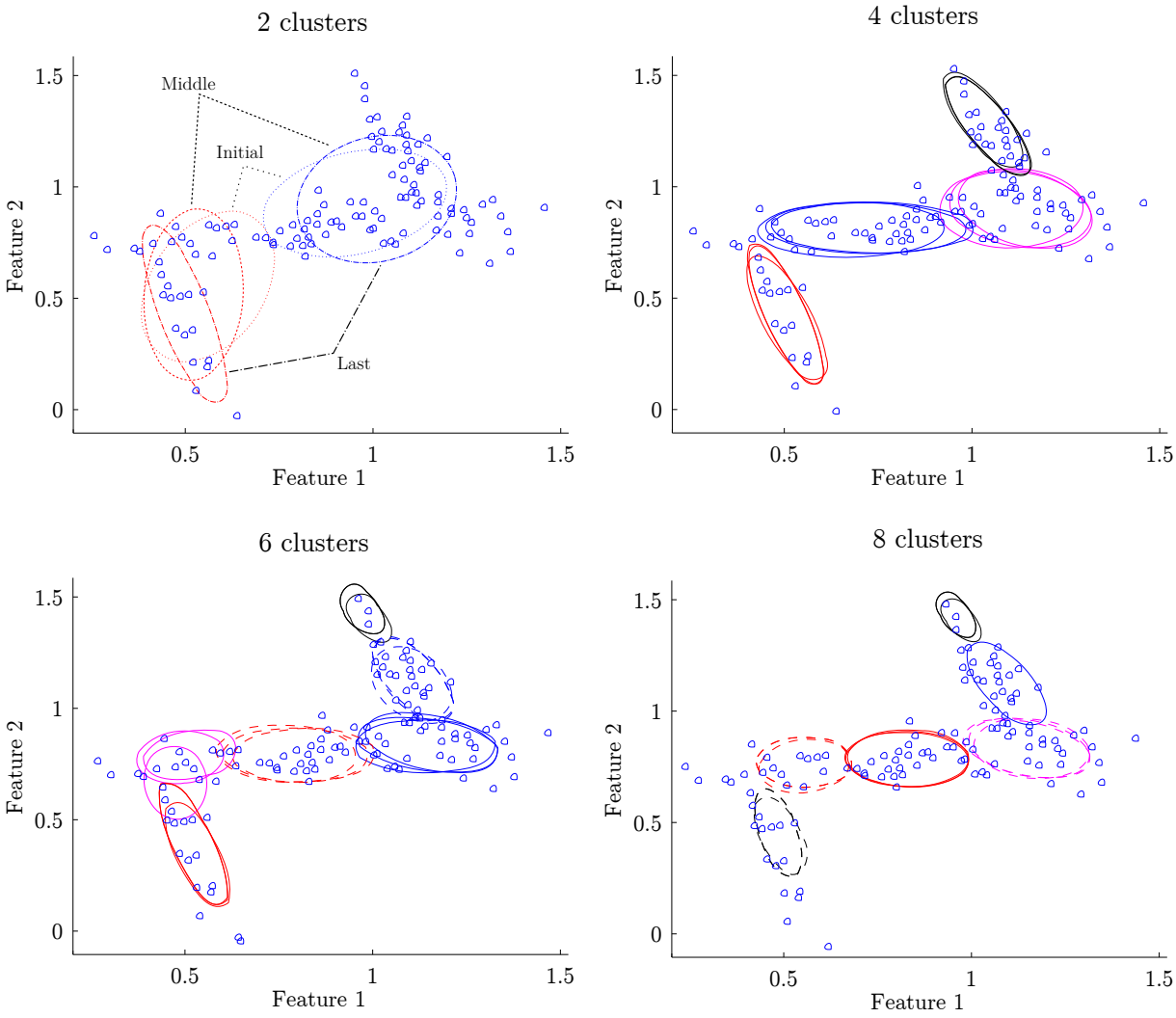


Figure 5.2: Toy example of a non-Gaussian distribution with several amount of clusters to be found. The elliptic equiprobable contour shows the estimated Gaussian distribution component at each step. Three situations of the EM algorithm are shown: initial, middle and last iteration. Note how the components are adapted to the data.

is to follow the *a priori* knowledge of the problem. As will be explained below in the results section, as our classification problem involves 3 classes, we will allow between 3 and 4 clusters per class, which means 9 or 12 clusters in total. An example with real data is shown in Figure 5.3. In this example two clusters are shown where the EMC grouped normal and ventricular heartbeats. In the top-right of the same figure, among the more distant heartbeats, some misclassified examples can be seen. Note the presence in the morphology details of some widened QRS complexes.

5.2.5 Feature selection for clustering

Regarding the feature model used with the EMC, we followed the same feature selection procedure described in Section 2.4.6, by means of a sequential floating feature selection algorithm (SFFS) [Pudil et al., 1994, van der Heijden et al., 2005]. The complete pool

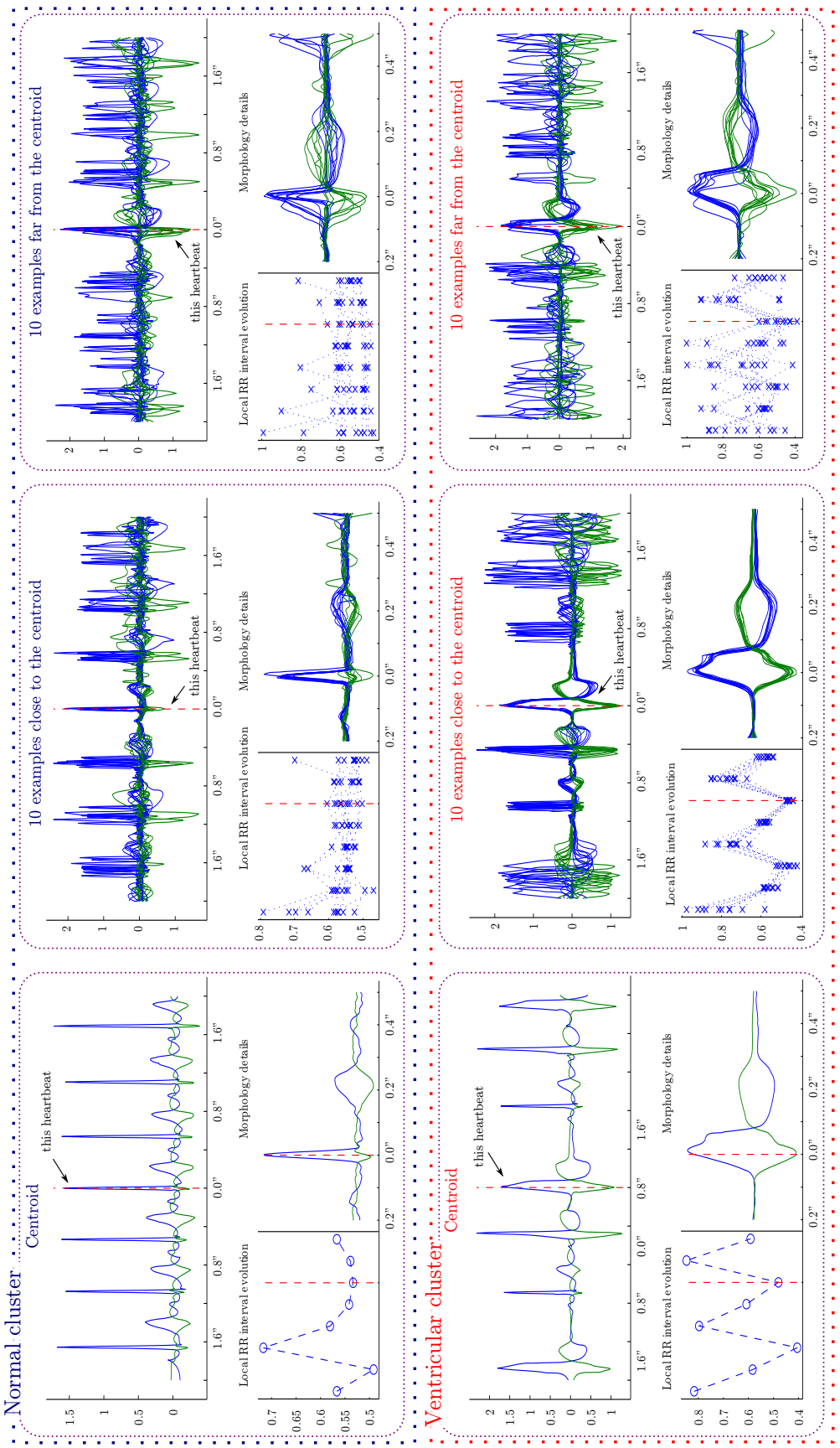


Figure 5.3: Clustering algorithm applied to the recording 208 of MITBIH-AR database. Only two clusters are shown for simplicity. In the top panel one with normal heartbeats, and below one with ventricular heartbeats. Within each panel, some heartbeats sampled from each cluster, from left to right, the centroid heartbeat and the 10 closer and farther heartbeats to the centroid. The red dashed lines indicates the heartbeat position. The rhythm evolution and the morphology details are also shown below. Note some bad-clustered heartbeats in the farther examples.

Table 5.2: Features used with the EMC algorithm.

Feature	Description
$\ln(RR[i])$	Current RR interval[1]
$\ln(RR[i-1])$	Previous RR interval[1]
$\ln(P_{RR})$	Prematurity of the heartbeat[1]
$\ln(dRR_L)$	Local RR interval variation[1]
$\ln(RR_{20})$	Mean RR interval within the last 20 minutes[1]
$\ln(S^1_{\text{QRS}})$	QRS mean wavelet scale in the first principal component[2]
k^1_M	First minimum position of the WT autocorrelation sequence in the 1st principal component[3]
$r_{\text{QRST}}(k_M)$	Value of the first maximum in the QRST complex crosscorrelation sequence between WT scale 3 of the[3] first two principal components

[1] See Figure 2.6 [2] See Figure 2.9 [3] See Figure 2.10

of features consisted of 61 features, described in Section 2.4.1. For the case of clustering, instead of looking for features with generalization capability or interpatient separability, we looked for those with high intrapatient separability. This criterion was achieved by modifying the SFFS' optimization criterion explained in Section 2.4.6, in order to find a feature model that provides as much intrapatient class separability as possible, facilitating the clusters identification. The first modification consisted in evaluating our clustering algorithm in a patient by patient fashion since this is how this algorithm will be used in practice. The second is that the performance will be evaluated in an optimistically biased fashion, described in Section 2.4.5, assuming that we know *a priori* the true labels of the heartbeats. The feature selection experiments were carried out in a dataset formed by the union of MITBIH-SUP with DS1 subset of MITBIH-AR [de Chazal et al., 2004]. As the SFFS performs thousands of model evaluations, this task is very demanding in processing power, specially for the random and iterative nature of the EMC. For this reason we replaced the EMC, only for the feature selection task, for a classifier based on mixture of Gaussians (MoG), which uses the same algorithm used for cluster discovery. The classifier based on MoG (MoGC) models each AAMI class with K Gaussian distributions, in contrast with the LDC-C that models each class with a mean vector and a pooled covariance matrix for all classes (Eq. (2.22) and (2.23) respectively). The MoGC uses during training the EM algorithm for the estimation of the Gaussian components. This modification results, first, in moving through deterministic paths through the performance surface evaluated with the SFFS, and second in easing the EM iteration since the heartbeat labels are known *a priori*. As a result, a model of 8 features was obtained. This model also includes a description of the rhythm and morphology of heartbeats as shown in Table 5.2.

Among the rhythm features used in the model, some of them have already been used in previous works. The prematurity of a heartbeat

$$P_{RR}[i] = \frac{RR[i]}{\sum_{k=i-1}^{i+1} RR[k]}, \tag{5.8}$$

measures how anticipated is a heartbeat respect to the previous and next RR interval. The local RR interval variation is defined as $dRR_L[i] = \sum_{k=i-1}^{i+1} |dRR[k]|$, where $dRR[i] = RR[i] - RR[i-1]$. One of the morphology related features is the wavelet scale where the QRS complex is mostly projected. It is known that fast evolving signals, as a normal heartbeat, tend to be projected in lower wavelet scales or contains higher frequency components. The QRS center scale for each lead (S_{QRS}^{Lead}) is calculated as the weighted sum

$$S_{QRS}^{L} = \frac{\sum_{s=1}^{6} A_s^L.s}{\sum_{s=1}^{6} A_s^L} \tag{5.9}$$

where A_s^L is the mean absolute amplitude of the QRS peaks at scale s of the DWT, and lead L

$$A_s^L = \frac{1}{D} \sum_{d=1}^{D} \left| W_s^L s(l_d) \right|, \quad s = 1, 2, \ldots 6 \tag{5.10}$$

being D the number of detected peaks (1 or 2) and l_d the positions of the peaks. The last morphology feature used is the maximum of the autocorrelation sequence of the ECG WT at scale 3 ($r_{\mathrm{QRST}}(k_M)$), which describes the QRST complex similarity between PCA leads at scale 3 of the WT. This feature is related to changes in the multilead morphology and the depolarization axis of the QRST complex. See Figure 2.10 for details about the calculation of all the morphology features used.

5.2.6 Performance evaluation

The performance is calculated from the confusion matrix after performing a classification experiment, in terms of the class sensitivity (S_i), class positive predictive value (P_i^+), global accuracy (A), global sensitivity (S) and global positive predictive value (P^+) as suggested in [AAMI-EC57, 1998–2008] and described in [Llamedo and Martínez, 2011a, 2012a]. As the initialization of the EMC is random, the results of the clustering algorithm are not deterministic. Then each experiment is repeated 30 times to evaluate the mean and standard deviation of the performance estimates. The amount of expert assistance required in the patient-adaptable modes of operation will be also accounted for each experiment.

5.3 Results

We performed two experiments, in the first one we studied the values of the algorithm parameters, that will be used in the second to evaluate its performance. The objective of the first experiment was to set up the number of clusters (K) and the qualified majority percentage used in votations (α), both parameters used in automatic and slightly

Table 5.3: Performance obtained in the development dataset for the election of K and α parameters.

			Normal		Supraventricular		Ventricular		Total		
Operation mode	α	K	S	P^+	S	P^+	S	P^+	A	S	P^+
Slightly assisted	50	9	96 ± 0	97 ± 0	58 ± 2	53 ± 1	79 ± 1	67 ± 1	93 ± 0	77 ± 0	72 ± 0
		12	96 ± 0	97 ± 0	56 ± 2	51 ± 1	78 ± 1	66 ± 1	92 ± 0	77 ± 1	72 ± 0
	75	**9**	**97±0**	**98±0**	**60±1**	**61±1**	**84±1**	**71±1**	**94±0**	**80±1**	**77±1**
		12	97 ± 0	98 ± 0	61 ± 1	60 ± 1	84 ± 1	71 ± 1	94 ± 0	80 ± 0	76 ± 0
Automatic	0	9	96 ± 0	97 ± 0	52 ± 3	49 ± 1	78 ± 1	65 ± 1	92 ± 0	75 ± 1	71 ± 1
		12	95 ± 0	97 ± 0	50 ± 2	49 ± 1	77 ± 1	63 ± 1	92 ± 0	74 ± 1	70 ± 1
	50	**9**	**96±0**	**97±0**	**53±2**	**51±2**	**77±1**	**65±1**	**92±0**	**75±1**	**71±0**
		12	95 ± 0	97 ± 0	52 ± 1	48 ± 1	77 ± 1	64 ± 0	92 ± 0	75 ± 0	70 ± 0
	75	9	95 ± 0	97 ± 0	50 ± 1	47 ± 0	77 ± 0	61 ± 0	92 ± 0	74 ± 0	68 ± 0
		12	95 ± 0	97 ± 0	50 ± 1	47 ± 0	77 ± 0	61 ± 0	92 ± 0	74 ± 0	68 ± 0

K: number of clusters; α: majority threshold (in percent)

assisted modes of operation. These parameters were assessed in the development dataset (MITBIH-SUP and DS1 subset of MITBIH-AR), and then used for the final performance evaluation in the remaining datasets. Table 5.3 shows the results of this experiment for two values of the evaluated parameters. As a result of this experiment we adopted $K = 9$ and $\alpha = 50\%$ for the automatic mode, and $K = 9$ and $\alpha = 75\%$ for the slightly assisted mode.

The final evaluation of the algorithm was performed in a broad set of databases in order to obtain a realistic estimation of its performance, as done in [Chudácek et al., 2009, Llamedo and Martínez, 2012a]. The three modes of operation were evaluated for each database with the parameter values obtained in the first experiment. The results of this experiment are presented in Tables 5.4 and 5.5 grouped by dataset. Comparison with the most relevant algorithms found in the literature are presented separately in Table 5.4. In Table 5.5 the performance obtained for all databases are presented. For each database we present the performance of our previous classifier [Llamedo and Martínez, 2011a] at the bottom for comparison, and a biased performance estimation on top as an upper bound. This biased performance is obtained when a quadratic classifier [Llamedo and Martínez, 2011a, van der Heijden et al., 2005, Duin et al., 2008] and the feature model presented in Table 5.2 are trained and tested in the same patient, for each patient in a database. This optimistically biased performance serves as an upper bound, and represents the performance of the model if it could be re-trained for each patient. From the results presented in Table 5.4, the proposed algorithm outperforms almost all reviewed algorithms, except the algorithms of Jiang [Jiang and Kong, 2007] and Ince [Ince et al., 2009], both in a small subset of MITBIH-AR, and the algorithm of Kiranyaz [Kiranyaz et al., 2011] in the MITBIH-LT. Finally the results showed in Table 5.5 evidence that the algorithm improves the baseline performances obtained by the LDC.

Table 5.4: Performance comparison with reference algorithms.

Dataset	Observation	#MAHB	Normal		Supraventricular		Ventricular		Total		
			S	P^+	S	P^+	S	P^+	A	S	P^+
MITBIH-AR (DS2)$^\bullet$	[de Chazal and Reilly, 2006]	500	94	99	88	47	95	82	94	92	76
	[Llamedo and Martínez, 2012b]	12	100±0	99±0	92±1	90±3	92±1	97±1	99±0	94±1	96±1
	[de Chazal et al., 2004]	0	87	99	76	39	87	43	87	83	60
	[Mar et al., 2011]	0	90	99	83	34	87	61	89	87	65
	[Llamedo and Martínez, 2012b]	0	93±0	99±0	77±0	39±0	82±0	70±0	92±0	84±0	69±0
MITBIH-AR (DS1-DS2)$^\bullet$	[Ince et al., 2009]	$\approx 300^\dagger$	98	98	64	54	85	87	96	82	80
	[Llamedo and Martínez, 2012b]	12	100±0	99±0	89±2	88±3	90±1	97±0	98±0	93±1	95±1
MITBIH-AR (24 rec.)$^\triangledown$	[Jiang and Kong, 2007]	$\approx 300^\dagger$	99	96	51	68	85	93	95	78	86
	[Llamedo and Martínez, 2012b]	12	99±0	98±0	91±1	88±2	90±1	96±1	98±0	93±1	94±1
MITBIH-AR (11 rec.)$^\circ$	[Hu et al., 1997]	$\approx 300^\dagger$	–	–	–	–	79	76	–	–	–
	[de Chazal and Reilly, 2006]	500	–	–	76	39	76	91	–	–	–
	[Jiang and Kong, 2007]	$\approx 300^\dagger$	–	–	75	79	94	96	–	–	–
	[Ince et al., 2009]	$\approx 300^\dagger$	–	–	82	63	90	92	–	–	–
	[Llamedo and Martínez, 2012b]	12	99±0	99±0	89±2	88±3	93±1	97±1	98±0	85±1	85±2
MITBIH-LT	[Kiranyaz et al., 2011]	$\leq 900^\star$	99	100	40	17	98	99	99	79	72
	[Llamedo and Martínez, 2012b]	20	99±1	99±0	16±16	0±0	94±0	91±4	98±0	70±5	74±9

$^\circ$ Comparison against results presented in Table II of [Ince et al., 2009].
$^\triangledown$24 recordings from 200 to 234 used in [Jiang and Kong, 2007].
$^\bullet$ DS1 and DS2 datasets defined in [de Chazal et al., 2004].
† Heartbeats in the first 5 min of each recording.
* Heartbeats in approx. 15 min of each recording.
#MAHB: manually annotated heartbeats per recording.

Table 5.5: Generalization evaluation of the proposed algorithm for all databases.

Dataset	Observation	Ref.	#MAHB	Normal S	Normal P^+	Supraventricular S	Supraventricular P^+	Ventricular S	Ventricular P^+	Total A	Total S	Total P^+
AHA	*biased*		*all*	*100*	*100*	–	–	*99*	*97*	*100*	*100*	*98*
	Assisted		12	100±0	100±0	–	–	98±0	98±0	100±0	99±0	99±0
			9	100±0	100±0	–	–	97±0	98±0	99±0	98±0	99±0
	Slightly assisted		1.2	96±0	100±0	–	–	86±0	82±0	95±0	91±0	60±0
	Automatic	[1]	0	93±0	99±0	–	–	75±1	71±1	92±0	85±0	57±0
		[2]	0	91	99	–	–	75	66	89	83	55
ESTTDB	*biased*		*all*	*100*	*100*	*98*	*41*	*97*	*83*	*100*	*98*	*74*
	Assisted		12	100±0	100±0	42±2	75±5	91±2	90±1	100±0	77±1	88±3
			9	100±0	100±0	37±4	65±13	88±1	81±3	100±0	75±2	82±4
	Slightly assisted		1	97±0	100±0	43±3	2±0	81±1	43±3	96±0	73±1	48±1
	Automatic	[1]	0	92±0	100±0	28±12	1±0	77±1	32±4	92±0	66±4	44±1
		[2]	0	89	100	47	1	78	31	89	72	44
INCART	*biased*		*all*	*99*	*100*	*98*	*68*	*99*	*98*	*99*	*99*	*88*
	Assisted		12	100±0	100±0	85±2	91±2	98±0	98±0	99±0	94±1	96±1
			9	100±0	99±0	79±3	90±2	98±0	98±1	99±0	92±1	96±1
	Slightly assisted		1.1	96±0	99±0	74±3	18±1	93±1	97±1	95±0	88±1	72±0
	Automatic	[1]	0	89±0	99±0	74±4	8±0	88±0	96±1	89±0	84±1	68±0
		[2]	0	88	100	76	7	84	94	87	83	67
LTSTDB	*biased*		*all*	*99*	*100*	*96*	*49*	*98*	*84*	*99*	*98*	*78*
	Assisted		15	100±0	100±0	51±0	58±0	82±0	85±0	99±0	78±0	81±0
			9	100±0	100±0	52±1	58±5	78±2	84±3	99±0	77±1	80±3
	Slightly assisted		1.2	99±0	100±0	44±1	36±5	771	59±4	99±0	73±1	65±0
	Automatic	[1]	0	96±0	100±0	50±2	8±0	66±1	36±2	96±0	71±0	48±1
		[2]	0	92	100	57	4	70	29	91	73	44

#MAHB: manually annotated heartbeats per recording. [1]this work [2][Llamedo and Martínez, 2012a]

Continues on the next page

Concludes from the previous page

Dataset	Observation	Ref.	#MAHB	Normal		Supraventricular		Ventricular		Total		
				S	P^+	S	P^+	S	P^+	A	S	P^+
MITBIH-AR	*biased*		*all*	*98*	*100*	*98*	*83*	*98*	*93*	*98*	*98*	*92*
	Assisted		12	100±0	99±0	89±2	88±3	90±1	97±0	98±0	93±1	95±1
			9	99±0	99±0	86±1	85±2	88±1	96±1	98±0	91±1	93±1
	Slightly assisted		1	98±0	99±0	79±3	59±3	83±2	92±1	97±0	87±1	83±1
	Automatic	[1]	0	96±0	98±0	76±2	43±2	80±2	82±3	94±0	84±1	74±1
		[2]	0	95	98	76	35	76	77	93	82	70
MITBIH-LT	*biased*		*all*	*97*	*100*	*95*	*11*	*92*	*91*	*97*	*95*	*67*
	Assisted		20	99±1	99±0	16±16	0±0	94±0	91±4	98±0	70±5	74±9
			9	99±0	99±0	1±1	0±0	89±6	92±2	98±0	63±2	64±1
	Slightly assisted		2.6	95±1	98±0	0±0	0±0	83±1	65±3	94±0	59±0	54±1
	Automatic	[1]	0	88±0	95±0	2±0	0±0	37±2	44±1	83±0	42±1	46±0
		[2]	0	86	95	33	1	38	44	81	52	47
MITBIH-ST	*biased*		*all*	*99*	*100*	*98*	*89*	*100*	*98*	*99*	*99*	*96*
	Assisted		12	100±0	100±0	89±1	87±2	95±1	99±0	100±0	94±0	95±1
			9	100±0	100±0	83±5	80±6	94±2	99±0	99±0	93±2	93±2
	Slightly assisted		1.4	71±1	99±0	76±6	5±0	93±3	25±1	71±1	79±3	43±0
	Automatic	[1]	0	67±1	99±0	57±6	3±0	43±3	11±1	66±1	56±3	38±0
		[2]	0	65	99	62	3	53	9	65	60	37
MITBIH-SUP	*biased*		*all*	*97*	*100*	*95*	*68*	*94*	*90*	*96*	*95*	*86*
	Assisted		12	99±0	98±0	74±1	79±1	88±1	90±1	96±0	87±0	89±0
			9	98±0	98±0	73±2	77±3	87±2	90±1	96±0	86±1	89±1
	Slightly assisted		1.3	96±0	98±0	62±1	61±1	85±1	65±1	93±0	81±1	74±1
	Automatic	[1]	0	94±0	97±0	47±3	50±1	82±0	54±1	91±0	75±1	67±1
		[2]	0	93	98	51	46	79	49	90	74	64

#MAHB: manually annotated heartbeats per recording. [1]this work [2][Llamedo and Martínez, 2012a]

Table 5.6: Results for the datasets most used in other works, with the AAMI labeling.

			Normal		Supraventr.		Ventricular		Total
Dataset	Observation	#MAHB	S	P^+	S	P^+	S	P^+	A
MITBIH-AR	Assisted	12	100±0	99±0	89±2	88±3	91±1	97±1	98±0
		9	99±0	99±0	86±1	85±2	89±2	97±1	98±0
	Slightly assisted	1	98±0	99±0	78±3	62±3	86±2	93±1	96±0
	Automatic	0	96±0	98±0	74±4	42±2	84±1	88±1	94±0
DS2	Assisted	12	100±0	99±0	92±1	90±3	93±1	97±1	98±0
		9	99±0	99±0	92±1	86±3	91±3	96±1	98±0
	Slightly assisted	1	97±0	99±0	83±4	58±5	91±3	90±2	96±0
	Automatic	0	95±0	99±0	79±2	46±2	89±1	87±1	94±0

#MAHB: manually annotated heartbeats per recording.

5.4 Discussion and Conclusions

In this chapter we presented a versatile ECG heartbeat classification algorithm suitable for a broad range of scenarios, from automatic or unassisted to fully assisted mode. The automatic part of the algorithm relies on a previously developed automatic classifier with proven generalization capability [Llamedo and Martínez, 2011a], referred as LDC in Figure 5.1. The main limitation of the LDC is the inability to handle large inter-patient rhythm and morphology variations. Many works overcame this limitation with the assistance of an expert [Hu et al., 1997, Lagerholm et al., 2000, de Chazal and Reilly, 2006, Jiang and Kong, 2007, Ince et al., 2009, Kiranyaz et al., 2011]. The approach to handle assistance presented in this chapter is based on a clustering algorithm, responsible of retaining most of the patient specific characteristics of the heartbeats (EMC in Figure 5.1). For this reason the feature model used with the clustering algorithm pursues the maximum intrapatient class separability. This approach is different to the one used in the development of the LDC feature set in [Llamedo and Martínez, 2011a], which pursued the maximization of a generalization criterion. As a result, Tables 5.1 and 5.2 show the different feature models used in the algorithm. As it can be seen, both feature models use rhythm and morphology features for heartbeats representation. Regarding to the rhythm features, the EMC has the addition of features P_{RR} and dRR_L, which are related to the local RR interval variation. As for the morphology description, features S^1_{QRS} and k^1_M may together represent a robust surrogate of the QRS width; while feature $r_{\mathrm{QRST}}(k_M)$ describes the QRST complex similarity between PCA leads at scale 3 of the WT. This measure is related to morphologic and depolarization-axis changes in the QRST complex.

To operate in automatic and slightly assisted modes, the EMC algorithm uses K as the number of clusters to model, and the voting scheme uses α as a threshold to assume that a whole cluster belongs to a class. As can be seen in Table 5.3, the performance intervals are comparable for the selected configurations but a mild improvement can be observed for the highlighted configurations. These configurations will be used for the final

evaluation and comparison of the algorithm.

The reference performances used for comparison purposes were in all cases AAMI [AAMI-EC57, 1998–2008] compliant. This means that the AAMI2 alternative labeling used is equivalent when calculating the S and V' class-performances, presented in tables 5.4 and 3.3. The reader interested in this aspect is referred to section A.3.5.2 of [AAMI-EC57, 1998–2008]. In Table 5.6 are shown the results obtained in DS2 of MITBIH-AR and MITBIH-AR, when considering the five AAMI classes. It can be seen that both AAMI and AAMI2 labeling yield the same results for the N, S and V classes.

The performance comparisons presented in Table 5.4 evidence the usefulness of the proposed algorithm. Without expert assistance the proposed algorithm performs better than a recent algorithm of Mar *et al.* [Mar et al., 2011] and the automatic version of de Chazal *et al.* [de Chazal et al., 2004]. Furthermore, when assistance is available our algorithm outperforms the reviewed algorithms [Hu et al., 1997, de Chazal and Reilly, 2006, Jiang and Kong, 2007, Ince et al., 2009] in the MITBIH-AR, with the following clarification. The algorithms of Jiang *et al.* [Jiang and Kong, 2007], Ince *et al.* [Ince et al., 2009] and de Chazal *et al.* [de Chazal and Reilly, 2006] outperform our algorithm in a subset of MITBIH-AR, but using more expert assistance. However the same algorithms perform worse in bigger subsets of the same database, as can be seen in Table 5.4. This fact reinforces the importance of evaluating arrhythmia classifiers in a wide range of databases, to have a complete idea of its performance. Finally, the algorithm presented by Kiranyaz [Kiranyaz et al., 2011] performed better than our algorithm in the MITBIH-LT, but with an increased effort in assistance of 15 minutes, or approximately 900 manually annotated heartbeats (MAHB) per recording, respect to the 20 MAHB required by this algorithm. However the differences in performance are moderate, and considering that the algorithm presented in [Kiranyaz et al., 2011] was specifically developed for long-term recordings. It is worth remarking that the MITBIH-LT presents the bigger class imbalance among the studied databases, showing some limitations of the EMC to detect scarcely represented classes, as the supraventricular.

An interesting aspect of the proposed algorithm is the improvement achieved in the amount of expert assistance required, 42 times less annotation effort than the algorithms of Hu [Hu et al., 1997] and de Chazal [de Chazal and Reilly, 2006], 25 times less than Jiang [Jiang and Kong, 2007] and Ince [Ince et al., 2009], and 45 times less than Kiranyaz [Kiranyaz et al., 2011].

Regarding to the comparison with the previous automatic multilead classifier, the fully automatic mode of the patient-adapted algorithm presented in this chapter achieved performance figures higher than those obtained in [Llamedo and Martínez, 2011a] for all databases except in the two databases including long-term recordings (MITBIH-LT and MITBIH-ST), where the performance was slightly lower. The decrease in MITBIH-LT and ST shows a limitation of the clustering features to adequately account for changes in long recordings. Moreover, as both databases include a small number of recordings (7

and 18, respectively), a particular recording could have an exaggerated influence on the whole database performance. However, for the automatic mode the mean improvement across databases with respect to the LDC is of 2.3% in A, of -2% in S and of 1.8% in P^+. With a small degree of assistance, 1-2 MAHB per recording in the slightly assisted mode, we obtained a mean improvement of 6.9% in A, of 6.5% in S and of 8.9% in P^+. Furthermore, an assistance of just 12 MAHB per recording results in a mean improvement of 13.1% in A, of 13.9% in S and of 36.1% in P^+. The important improvement in performance achieved in assisted mode shows that our working hypothesis, separate and homogeneous clusters within a recording, is corroborated in most cases. The algorithm showed robust dealing with different types of noise present throughout the evaluated databases, as a result of using robust features. In addition, the results presented in Table 3.3 are consistent, as more assistance is translated into larger performance improvement. This experiment evidences that the algorithm can handle properly the different degrees of assistance provided by an expert. Note that the development dataset, which includes DS1 of MITBIH-AR and MITBIH-SUP, is included in the results presented in Table 3.3. These results are optimistically biased and should be considered only as an additional description of the algorithm performance.

In slightly assisted mode, it is worth noting that the intra-cluster class-heterogeneity in the recordings analyzed is proportional to the assistance required. Remember that the algorithm ask for assistance in those clusters where a qualified majority is not reached. According to Table 3.3, from 11 to 16% of the clusters did not reach a qualified majority. For the particular case of the MITBIH-LT this figure raised to the 29%. This increase is reasonable since the nonstationarities of the ECG signal, and thus the cluster heterogeneity, are more evident in long-term recordings.

The algorithm's computational efficiency was not analyzed in detail, however it takes around 25 seconds to classify a MITBIH-AR recording (30 minutes of two-lead ECG) in a desktop PC (Intel Core2 E8500 CPU). The measurement was performed in a freely available implementation of the algorithm in Matlab [Llamedo, 2012]. Although the execution time is not excessively large, there is room for improvement with an optimized implementation.

From the evaluation of the algorithm some limitations were found and need to be addressed in future improvements. The first one, is the inability of the clustering algorithm to find marginally represented classes. This problem slightly affects the global performance since the less represented classes have a mild effect in a database-aggregated performance estimates. However, in certain applications the misclassification of this kind of infrequently arrhythmias could limit the usefulness of algorithms based on clustering techniques. Other limitations are related to the feature model used by the EMC, presented in Table 5.2. In certain recordings where the classes are reasonably represented to be clustered, the EMC fails to recognize the clusters probably due to the inability of the feature model to separate the classes. This problem is also evidenced in the biased

evaluation performance, showed on top for each database in Table 5.5. Theoretically if the features and classifier used could adequately model the data, the biased performance should be perfect for all classes (S_i and P_i^+ 100%). Since this is not true, it can be concluded that the presented model still has limitations. This could be improved with the development of better features or a most sophisticated classifier. The last limitation found during the evaluation appeared in long-term recordings. In these recordings the evident nonstationarities in the feature space made the algorithm performance to decrease considerably. For this reason, the assistance provided to the algorithm was increased for long-term recordings. Strategies to deal with nonstationarities, as the proposed in [Kiranyaz et al., 2011], will be studied in the future.

The results presented in this chapter represent a performance improvement with respect to the published works in the field of automatic and patient-adaptable heartbeats classification. These results show that the performance of an automatic classifier can be improved with an efficient handling of the expert assistance. The authors freely distribute a Matlab implementation of the algorithm for academic use, presented in Appendix A.

5.A Detailed Results

In this section we present the confusion matrices for ease the comparison of future works. The summarized performances presented in the previous sections were calculated from these numbers, according to the methodologies described in Section 2.4.5.

Table 5.7: Confusion matrices of the results presented in Table 5.5 for AHA database.

biased

Truth \ Algorithm	n	s	v	Total
N	265807	–	1175	266982
S	–	–	–	–
V	177	–	33833	34010
Total	265984	–	35008	300992

Slightly assisted

Truth \ Algorithm	n	s	v	u	Total
N	305655±423	5548±361	6515±251	18±10	317735
S	–	–	–	–	–
V	1398±164	3399±284	29157±280	33±18	33987
U	171±21	7±5	118±25	276±27	572
Total	307224±478	8954±429	35789±365	327±44	352294

Assisted 12 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	316325±1054	–	642±112	31±20	317735
S	–	–	–	–	–
V	809±80	–	32885±334	61±40	33987
U	131±35	–	93±14	348±40	572
Total	317266±1068	–	33619±364	440±83	352294

Automatic

Truth \ Algorithm	n	s	v	u	Total
N	296739±736	10589±363	10406±647	–	317735
S	–	–	–	–	–
V	1660±131	6653±231	25674±250	–	33987
U	201±23	65±11	306±24	–	572
Total	298600±752	17308±469	36387±818	–	352294

Assisted 9 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	316865±151	–	846±149	24±12	317735
S	–	–	–	–	–
V	975±140	–	32965±146	47±19	33987
U	145±26	–	114±22	314±28	572
Total	317984±267	–	33925±265	385±41	352294

Automatic [Llamedo and Martínez, 2012a]

Truth \ Algorithm	n	s	v	u	Total
N	289116	15746	12873	–	317735
S	–	–	–	–	–
V	1616	7003	25368	–	33987
U	253	77	242	–	572
Total	290985	22826	38483	–	352294

Table 5.8: Confusion matrices of the results presented in Table 5.5 for ESTTDB database.

biased

Truth \ Algorithm	n	s	v	Total
N	534551	1486	979	537016
S	12	1056	8	1076
V	96	49	4673	4818
Total	534659	2591	5660	542910

Slightly assisted

Truth \ Algorithm	n	s	v	u	Total
N	759079±2090	20811±1122	4748±1184	0	784638
S	619±113	429±110	47±14	0	1095
V	585±61	359±111	3875±104	2±5	4821
U	8±3	2±2	1±1	1±1	11
Total	760291±2066	21601±1081	8671±1272	3±6	790565

Assisted 12 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	783911±394	177±102	532±300	18±27	784638
S	565±53	478±53	52±2	0	1095
V	418±127	18±9	4375±132	11±3	4821
U	10±1	0	0	1±1	11
Total	784904±509	673±149	4959±375	30±40	790565

Automatic

Truth \ Algorithm	n	s	v	u	Total
N	724671±1956	52226±2271	7742±1147	–	784638
S	687±173	289±147	119±34	–	1095
V	608±95	493±107	3720±106	–	4821
U	7±2	1±1	1±1	–	11
Total	725973±1810	53011±2289	11581±1155	–	790565

Assisted 9 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	781204±7044	199±147	908±347	0	784638
S	632±87	413±94	50±11	0	1095
V	583±106	61±71	4175±120	2±5	4821
U	9±2	1±2	1±1	1±1	11
Total	782428±7048	673±295	5133±411	3±6	790565

Automatic [Llamedo and Martínez, 2012a]

Truth \ Algorithm	n	s	v	u	Total
N	700021	76527	8090	–	784638
S	433	517	145	–	1095
V	297	748	3776	–	4821
U	3	8	0	–	11
Total	700754	77800	12011	–	790565

Table 5.9: Confusion matrices of the results presented in Table 5.5 for INCART database.

biased

Truth \ Algorithm	n	s	v	Total
N	146861	886	458	148205
S	31	1917	10	1958
V	167	25	20030	20222
Total	147059	2828	20498	170385

Slightly assisted

Truth \ Algorithm	n	s	v	u	Total
N	147498±904	5787±918	418±99	–	153703
S	313±85	1445±89	203±52	–	1960
V	512±139	944±85	18774±166	–	20230
U	3±1	0	3±1	–	6
Total	148325±925	8176±938	19398±281	–	175899

Assisted 12 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	153088±231	255±238	359±78	2±5	153703
S	266±56	1667±55	27±20	0	1960
V	405±85	20±18	19805±89	0	20230
U	3±1	0	3±1	0±1	6
Total	153761±278	1941±265	20195±144	2±6	175899

Automatic

Truth \ Algorithm	n	s	v	u	Total
N	137391±936	15876±1052	436±241	–	153703
S	200±62	1464±105	296±85	–	1960
V	623±76	1806±206	17800±200	–	20230
U	3±1	1±1	3±1	–	6
Total	138217±931	19147±1082	18535±383	–	175899

Assisted 9 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	153046±169	184±133	472±114	2±5	153703
S	359±88	1550±91	51±50	0	1960
V	496±114	16±13	19718±118	0	20230
U	2±1	0	3±1	0±1	6
Total	153903±266	1750±144	20244±244	2±6	175899

Automatic [Llamedo and Martínez, 2012a]

Truth \ Algorithm	n	s	v	u	Total
N	135069	18013	621	–	153703
S	65	1496	399	–	1960
V	607	2561	17062	–	20230
U	3	0	3	–	6
Total	135744	22070	18085	–	175899

Table 5.10: Confusion matrices of the results presented in Table 5.5 for LTSTDB database.

biased

Truth \ Algorithm	n	s	v	Total
N	7651109	63997	12862	7727968
S	2071	63461	644	66176
V	437	894	72174	73505
Total	7653617	128352	85680	7867649

Slightly assisted

Truth \ Algorithm	n	s	v	u	Total
N	8692959±1939	33376±8330	41710±6390	–	8768045
S	20152±448	19881±573	5588±125	–	45620
V	17462±480	2426±239	67070±719	–	86958
U	371±2	8±2	91±4	–	469
Total	8730943±2869	55691±8661	114458±5792	–	8901092

Assisted 12 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	8746770±123	13167±787	5698±910	–	8768045
S	18748±2845	24273±2032	2600±813	–	45620
V	14777±1051	6150±3415	66031±2364	–	86958
U	367±18	15±7	87±11	–	469
Total	8783073±1653	43605±603	74415±2256	–	8901092

Automatic

Truth \ Algorithm	n	s	v	u	Total
N	8412810±54249	239484±9806	96661±5438	–	8768045
S	18640±968	22590±712	4371±253	–	45620
V	13328±2069	16229±1542	57388±527	–	86958
U	315±19	126±10	27±8	–	469
Total	8445092±54044	278429±8976	158446±5156	–	8901092

Assisted 9 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	8743249±6069	14798±3169	9999±2901	–	8768045
S	18225±301	23718±285	3678±585	–	45620
V	16059±1896	2643±28	68257±1924	–	86958
U	353±20	10±0	107±20	–	469
Total	8777885±8285	41168±2856	82039±5429	–	8901092

Automatic [Llamedo and Martínez, 2012a]

Truth \ Algorithm	n	s	v	u	Total
N	8048986	576535	142524	–	8768045
S	10563	25940	9117	–	45620
V	8663	17024	61271	–	86958
U	264	152	53	–	469
Total	8068476	619651	212965	–	8901092

Table 5.11: Confusion matrices of the results presented in Table 5.5 for MITBIH-AR database.

biased

Truth \ Algorithm	n	s	v	Total
N	67839	519	604	68962
S	38	2727	8	2773
V	123	32	7649	7804
Total	68000	3278	8261	79539

Slightly assisted

Truth \ Algorithm	n	s	v	u	Total
N	88554±163	1286±184	287±91	0±1	90127
S	307±59	2225±144	248±118	0	2780
V	995±185	269±48	6547±188	0	7811
U	12±2	0	2±1	0±1	15
Total	89868±301	3780±202	7084±268	1±2	100733

Assisted 12 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	89681±121	226±101	219±70	1±1	90127
S	277±67	2477±70	26±12	0	2780
V	651±148	105±45	7055±136	0	7811
U	12±1	0	2±1	1±1	15
Total	90621±248	2807±185	7303±180	2±2	100733

Automatic

Truth \ Algorithm	n	s	v	u	Total
N	86716±350	2375±256	1036±358	–	90127
S	303±72	2117±74	360±28	–	2780
V	1112±212	409±58	6290±212	–	7811
U	12±2	1±1	3±1	–	15
Total	88142±444	4901±233	7689±472	–	100733

Assisted 9 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	89555±122	318±103	254±75	0±1	90127
S	362±51	2358±52	34±13	0	2780
V	785±181	103±36	6923±189	0	7811
U	12±1	0	2±1	0±1	15
Total	90714±246	2806±122	7212±226	1±2	100733

Automatic [Llamedo and Martínez, 2012a]

Truth \ Algorithm	n	s	v	u	Total
N	85465	3220	1442	–	90127
S	284	2113	383	–	2780
V	1061	783	5967	–	7811
U	10	0	5	–	15
Total	86820	6116	7797	–	100733

Table 5.12: Confusion matrices of the results presented in Table 5.5 for DS2 of MITBIH-AR database.

biased

Truth \ Algorithm	n	s	v	Total
N	31871	288	168	32327
S	19	1812	1	1832
V	42	26	3536	3604
Total	31932	2126	3705	37763

Slightly assisted

Truth \ Algorithm	n	s	v	u	Total
N	43007±217	1106±237	146±97	–	44259
S	74±28	1585±119	178±120	–	1837
V	324±132	77±31	3208±127	–	3609
U	6±1	0	1±1	–	7
Total	43411±291	2768±235	3534±222	–	49712

Assisted 12 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	44032±96	140±83	88±37	–	44259
S	133±33	1688±34	16±9	–	1837
V	246±71	66±41	3298±73	–	3609
U	6±1	0	1±1	–	7
Total	44416±131	1893±122	3402±84	–	49712

Automatic

Truth \ Algorithm	n	s	v	u	Total
N	41902±335	1562±225	795±354	–	44259
S	7483±29	1476±21	277±26	–	1837
V	478±177	83±32	3048±173	–	3609
U	5±1	0	2±1	–	7
Total	42469±421	3121±218	4122±421	–	49712

Assisted 9 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	43942±109	208±91	110±57	–	44259
S	134±35	1684±36	20±10	–	1837
V	289±110	69±33	3251±105	–	3609
U	6±1	0	1±1	–	7
Total	44369±190	1961±101	3382±148	–	49712

Automatic [Llamedo and Martínez, 2012a]

Truth \ Algorithm	n	s	v	u	Total
N	41209	2045	1005	–	44259
S	141	1413	283	–	1837
V	434	211	2964	–	3609
U	4	0	3	–	7
Total	41788	3699	4255	–	49712

Table 5.13: Confusion matrices of the results presented in Table 5.5 for MITBIH-LT database.

biased

Truth \ Algorithm	n	s	v	Total
N	579074	9217	6381	594672
S	56	1431	12	1499
V	2489	2872	61008	66369
Total	581619	13520	67401	662540

Slightly assisted

Truth \ Algorithm	n	s	v	u	Total
N	570716±4038	–	29516±4038	–	600232
S	1287±31	–	213±31	–	1500
V	11197±991	–	55806±991	–	67003
U	–	–	–	–	–
Total	583200±4998	–	85535±4998	–	668735

Assisted 20 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	593776±3263	148±149	6309±3114	–	600232
S	1029±291	242±244	230±48	–	1500
V	4118±303	0	62885±303	–	67003
U	–	–	–	–	–
Total	598923±3252	390±393	69423±2859	–	668735

Automatic

Truth \ Algorithm	n	s	v	u	Total
N	529319±1917	39867±2631	31047±714	–	600232
S	1313±36	24±3	163±33	–	1500
V	26086±877	16092±252	24826±1129	–	67003
U	–	–	–	–	–
Total	556718±1076	55983±2886	56035±1810	–	668735

Assisted 9 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	594826±1502	121±122	5286±1624	–	600232
S	1288±30	11±11	202±19	–	1500
V	5055±1711	2061±2078	59888±3789	–	67003
U	–	–	–	–	–
Total	601168±3183	2192±2210	65376±5394	–	668735

Automatic [Llamedo and Martínez, 2012a]

Truth \ Algorithm	n	s	v	u	Total
N	513227	55080	31925	–	600232
S	873	493	134	–	1500
V	26256	15288	25459	–	67003
U	–	–	–	–	–
Total	540356	70861	57518	–	668735

Table 5.14: Confusion matrices of the results presented in Table 5.5 for MITBIH-ST database.

biased

Truth \ Algorithm	n	s	v	Total
N	13772	100	5	13877
S	17	779	1	797
V	0	0	314	314
Total	13789	879	320	14988

Slightly assisted

Truth \ Algorithm	n	s	v	u	Total
N	33065±794	12323±690	839±171	–	46228
S	194±68	604±68	0±1	–	798
V	13±7	25±362	281±34	–	319
U	–	–	–	–	–
Total	33272±817	12952±706	1121±194	–	47345

Assisted 12 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	46118±46	107±46	3±2	–	46228
S	96±21	702±21	0±1	–	798
V	13±6	4±2	302±6	–	319
U	–	–	–	–	–
Total	46227±54	813±52	306±7	–	47345

Automatic

Truth \ Algorithm	n	s	v	u	Total
N	30781±341	14351±366	1096±229	–	46228
S	316±72	470±67	12±11	–	798
V	9±5	164±30	146±30	–	319
U	–	–	–	–	–
Total	31106±320	14984±342	1254±222	–	47345

Assisted 9 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	46076±75	149±75	3±1	–	46228
S	128±44	669±44	0±1	–	798
V	16±8	4±2	299±8	–	319
U	–	–	–	–	–
Total	46220±103	822±101	303±8	–	47345

Automatic [Llamedo and Martínez, 2012a]

Truth \ Algorithm	n	s	v	u	Total
N	30168	14339	1721	–	46228
S	266	491	41	–	798
V	1	149	169	–	319
U	–	–	–	–	–
Total	30435	14979	1931	–	47345

Table 5.15: Confusion matrices of the results presented in Table 5.5 for MITBIH-SUP database.

biased

Truth \ Algorithm	n	s	v	Total
N	156674	4810	787	162271
S	359	11558	278	12195
V	105	515	9343	9963
Total	157138	16883	10408	184429

Slightly assisted

Truth \ Algorithm	n	s	v	u	Total
N	155136±485	4000±452	3202±165	2±6	162340
S	3311±268	7530±228	1355±155	2±4	12198
V	595±170	930±169	8439±209	3±5	9966
U	23±7	27±5	28±6	2±2	79
Total	159065±785	12486±663	13024±450	8±15	184583

Assisted 12 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	159991±300	1869±327	455±132	25±37	162340
S	2597±223	9081±206	507±106	12±20	12198
V	606±108	600±103	8754±112	7±11	9966
U	33±7	11±3	26±5	6±7	79
Total	163230±458	11561±459	9742±253	50±56	184583

Automatic

Truth \ Algorithm	n	s	v	u	Total
N	153420±415	4496±301	4424±312	–	162340
S	3993±266	5751±420	2454±304	–	12198
V	518±71	1305±55	8143±76	–	9966
U	20±5	27±3	32±4	–	79
Total	157951±513	11579±550	15053±326	–	184583

Assisted 9 MAHB

Truth \ Algorithm	n	s	v	u	Total
N	159842±510	2008±517	468±132	22±32	162340
S	2820±321	8871±313	498±113	9±15	12198
V	687±213	620±162	8656±231	4±6	9966
U	38±7	10±4	26±7	5±5	79
Total	163386±876	11509±853	9648±439	40±53	184583

Automatic [Llamedo and Martínez, 2012a]

Truth \ Algorithm	n	s	v	u	Total
N	151246	6040	5054	–	162340
S	2735	6185	3278	–	12198
V	704	1365	7897	–	9966
U	12	37	30	–	79
Total	154697	13627	16259	–	184583

Table 5.16: Confusion matrices of the results presented in Table 5.6 for MITBIH-AR database.

Assisted 12 MAHB

Truth \ Algorithm	n	s	v	f	u	Total
N	89681±121	226±101	174±67	45±30	1±1	90127
S	277±67	2477±70	26±12	0	0	2780
V	417±133	96±42	6373±148	121±71	0	7008
F	234±48	9±5	159±100	402±116	0	803
U	12±1	0	2±1	1±1	1±1	15
Total	90621±248	2807±185	6734±215	569±184	2±2	100733

Assisted 9 MAHB

Truth \ Algorithm	n	s	v	f	u	Total
N	89555±122	318±103	202±68	51±41	0±1	90127
S	362±51	2358±52	33±13	0	0	2780
V	540±157	93±34	6178±185	197±105	0	7008
F	245±71	11±5	146±110	402±86	0	803
U	12±1	0	1±1	1±1	0±1	15
Total	90714±246	2806±122	6561±250	651±190	1±2	100733

Slightly assisted

Truth \ Algorithm	n	s	v	f	u	Total
N	88602±206	1251±206	202±68	44±35	28±39	90127
S	311±57	2222±138	246±112	0	1±2	2780
V	707±158	111±41	5995±189	195±106	1±1	7008
F	290±82	11±5	145±110	358±130	0	803
U	12±1	0	2±1	1±1	0±1	15
Total	89922±338	3594±225	6589±324	597±236	30±41	100733

Automatic

Truth \ Algorithm	n	s	v	f	u	Total
N	86958±277	2359±248	344±119	374±160	91±59	90127
S	273±65	2056±121	449±99	0±1	2±6	2780
V	652±111	398±60	5868±108	86±56	4±2	7008
F	511±49	6±3	95±60	191±63	0	803
U	10±2	1±1	3±1	0	1±1	15
Total	88405±275	4819±268	6760±207	651±156	95±59	100733

Table 5.17: Confusion matrices of the results presented in Table 5.6 for DS2 of MITBIH-AR database.

Assisted 12 MAHB

Truth \ Algorithm	n	s	v	f	u	Total
N	44032±96	140±83	84±36	4±8	–	44259
S	133±33	1688±34	16±9	0	–	1837
V	139±53	58±38	2971±72	53±46	–	3221
F	107±43	8±5	61±52	213±59	–	388
U	6±1	0	1±1	0	–	7
Total	44416±131	1893±122	3133±107	270±93	–	49712

Assisted 9 MAHB

Truth \ Algorithm	n	s	v	f	u	Total
N	43942±109	208±91	106±57	4±13	–	44259
S	134±35	1684±37	20±10	0	–	1837
V	178±106	60±32	2902±109	81±57	–	3221
F	111±41	10±4	39±36	229±30	–	388
U	6±1	0	1±1	0	–	7
Total	44369±190	1961±101	3068±152	314±54	–	49712

Slightly assisted

Truth \ Algorithm	n	s	v	f	u	Total
N	43049±262	1072±275	114±55	0±1	24±38	44259
S	79±28	1585±116	173±115	0	1±2	1837
V	172±105	65±32	2903±109	81±57	–	3221
F	153±69	10±5	39±36	187±98	–	388
U	6±1	0	1±1	0	–	7
Total	43459±333	2731±272	3229±218	268±138	24±39	49712

Automatic

Truth \ Algorithm	n	s	v	f	u	Total
N	42186±298	1562±226	103±65	356±148	53±49	44259
S	61±17	1409±97	364±98	0±1	2±6	1837
V	224±73	94±31	2878±53	22±26	3±2	3221
F	317±40	0	50±39	21±24	0	388
U	5±1	0	2±1	0	0±1	7
Total	42792±293	3065±232	3397±122	399±149	58±50	49712

Chapter 6

Conclusions and Future Work

In this chapter we summarize the main contributions of the thesis and the conclusions of previous chapters. We will also enumerate some of the future works to be done as a continuation of the thesis.

6.1 Summary

In the previous chapters we suggested and analyzed a framework for the development of an ECG heartbeat classifier. We started from scratch designing the heartbeat features. We were interested in using features that had a physiological meaning, were simple to compute and robust to the noise present in the ECG. We explored several features of different nature and domains, at the same time that the methodology to put them together in the same classification model. To keep the complexity of the model as low as possible, without sacrificing performance, we adopted a floating feature selection algorithm. As our knowledge about the classification problem was limited at the beginning, we decided to adopt a simple classifier to be able to easily understand and interpret results and any issue during the development. Our first objective was developing an automatic classifier with generalization capability. To achieve this goal, we used a feature selection algorithm with a modified criterion to obtain a model with these properties. The model evaluation yielded good results in two databases [Llamedo and Martínez, 2011a], and comprehends the contents of Chapter 3.

With the availability of multilead databases, we suggested a strategy that extended the model developed in [Llamedo and Martínez, 2011a] . As a result we found a useful solution at the signal level, based on principal components analysis (PCA), that allows our two-lead model to handle signals of an arbitrary number of leads. Another important improvement achieved with the WT-PCA strategy is the robustness against lead misplacement or recordings with undocumented leads. With this strategy we slightly improved the performance of our classifier in the data used in [Llamedo and Martínez, 2011a], and confirmed the generalization of the model when applied to unseen 12-leads recordings.

These results were published in [Llamedo and Martínez, 2012a] and the methodology is explained in Section 4.2. In Section 4.3 the framework used in Chapter 3 was used to evaluate the improvement of a non-linear classifier, as the MLP, and these results were also published in [Mar et al., 2011].

Other aspect that we were interested in, was the adaptation to the patient under analysis. This technique was reported to be useful for improving the performance in problems with high interpatient variability. For this purpose we designed an adaptation algorithm that uses the classifier developed in Chapters 3 and Section 4.2, together with a clustering algorithm (see Figure 5.1). The feature model of the clustering algorithm was also obtained from the same feature selection algorithm used in Chapter 3, but with a modified criterion that maximized intrapatient class separability. Finally, a semiautomatic algorithm that integrates the methodologies developed in [Llamedo and Martínez, 2011a, 2012a] was evaluated in a set of eight public databases, described in Section 2.1, and the results were published in [Llamedo and Martínez, 2012b].

6.2 Conclusions

In this section the conclusions drawn through the chapters of this thesis are summarized. We start emphasizing the importance of the feature design, and therefore the understanding of the problem. In our experience, without a proper understanding of the problem it is impossible to design adequate features, and then a classifier with generalization capability. At the moment of writing this thesis, we are studying deep belief networks (DBN) [Hinton et al., 2006] classifiers and have pending its implementation. This kind of classifiers not only improved the state-of-the-art in other pattern recognition tasks, as handwriting or speech recognition, but they did it directly from the digitized signal samples or pixels. They simply skip the feature design step in the development of a classifier. Despite being in contradiction with of our first conclusion, DBN classifiers need further corroboration in the field of heartbeat classification. Probably other feature models can achieve further improvement that the samples by itself. Anyhow, we believe that classifiers of the *black box* style (or any other nonlinear or nonparametric) should be avoided as a first attempt to solve a classification problem, and used when a baseline performance was achieved with a simple classifier.

The importance of having a *large* dataset to perform the experiments is determinant. In applications as the classification of heartbeats, where there exist large interpatient variations, the definition of *large* can be misleading. In our experience, it is more important having databases with many subjects, than a small number of subjects, maybe repeated, of long-term recordings. However, the classification of long-term recordings is an aspect not deeply studied in this thesis and requires further work. This fact reinforces the importance of evaluating a classifier in as many databases as possible, to have a better estimation of its real performance.

In our feature selection experiments we found two models, pursuing different optimization criteria. In Table 3.4 a model with good intersubject performance is shown. As can be noted, the selected features are computed without exception from time interval measurements. This could be explained given that the used databases do not always include the same pair of ECG leads in each recording. Therefore the classification performances of features which are calculated from amplitudes are heavily degraded. The directional features (like the VCG_ϕ) were also probably affected by this fact, even if the clinical importance of this kind of features is well-known by cardiologists [Taylor, 2002]. In contrast, the intervals seem to retain the classification ability with independence of the pair of leads chosen. The first four features in the model are clearly connected to the evolution of heart rhythm, while the other four can be understood as surrogate measurements of the QRS width, and therefore a description of the QRS morphology. In addition, it only relies on a coarse QRS fiducial point detection, making the classifier model robust to degraded signals where the delineation of the ECG waves is not reliable or feasible.

On the other hand, in Table 5.2 a model with good intrasubject performance is shown. The model also includes rhythm and morphology features. Regarding to the rhythm features, the EMC has the addition of features P_{RR} and dRR_L, which are related to the local RR interval variation. As for the morphology description, features S^1_{QRS} and k^1_M may together represent a robust surrogate of the QRS width; while feature $r_{\mathrm{QRST}}(k_M)$ describes the QRST complex similarity between PCA leads at scale 3 of the DWT. This measure can be related to morphologic and depolarization-axis changes in the QRST complex.

The LDC-C linear decision functions showed to be useful in the development of a classifier with generalization capability. This can be explained because a conservative decision function, as a hyperplane, is more appropriate for difficult classification problems, or with large interpatient variability. In this kind of problems, almost none of the hypothesis imposed by our design decisions are completely fulfilled. Just to clarify this last sentence, we must say that in our initial problem we should have a training dataset which resembles the whole universe of heartbeats, this is not only unfeasible, but in our limited case we can say that is even quite different from the test set, just by simple inspection of the performance figures (compare results in Table 3.2 with 3.3). With this evident limitation, the classifier with more capacity to model the training data, in our case the QDC, is likely to fail more often in the test set. This limitation is probably the reason that makes a conservative classifier, as the LDC, the better choice. In Figure 2.12, the discriminant functions produced by the LDC and QDC can be compared.

When we limit the problem to one subject at a time, and we pursue the best intrapatient performance, we allow the classifier to produce non-linear decision functions. In our case we used a mixture of Gaussian classifier, which uses the same EM algorithm used by the clustering algorithm.

The feature selection scheme used resulted in a convenient methodology to reduce

the complexity of the classification task and to improve the generalization of the model obtained. The SFFS was specifically useful when using simple and deterministic classifiers (QDC or LDC), but for the case of non-deterministic classifiers (MLP, mixture of Gaussians) some workarounds were adopted, since we must ensure (or at least limit) the repeatability.

The generalization capability of a classifier is, in our opinion, one of the most important characteristics of a heartbeat classifier. In Chapter 5 we showed that is possible to perform a thorough evaluation of a classifier's performance, exclusively in publicly available databases.

The performance estimation for class-imbalanced problems, as the studied in this thesis, can be misleading specially when comparing different classifiers. In this thesis we explored different methodologies for dealing with the class-imbalance problem. However, none of the solutions suggested in Chapters 3 and 4, as the balanced performance, guarantee the resolution of the problem. Despite of this problem, the confusion matrix can always provide clarification at expense of redundancy, and in our opinion, should always be included to ensure comparability. Another problem is when comparing performances across databases with different class imbalances. To help interpretation we suggested the comparison against the *biased* performance estimation, which provides an upper-bound reference.

The comparisons performed in the previous chapters were performed in a fair manner up to our best knowledge. The works included in our comparisons have comparable methodologies and were representative of the state of the art. In general, as were already detailed in the previous chapters, our classifiers performed better than others. For all the comparisons, we always provided a detailed description of our performances in order to allow future improvements.

Summing up, the results presented in this thesis represent a performance improvement with respect to the published works in the field of automatic and patient-adaptable heartbeats classification.

6.3 Future work

Some of the possible future works that we identified during the development of this thesis were:

1. *More complex classifiers.* One possible improvement to the classification performance is the implementation of complex classifiers, as support vector machines (SVM) [Ivanciuc, 2007] or DBN, instead of the LDC-C used. In Section 4.3 we pointed out some of the limitations of the LDC-C, and showed how in this context the substitution by the MLP performed better. Following this reasoning, we think that may be possible to achieve some performance improvement by replacing the

LDC-C in the framework presented in Chapters 3 and 4.

2. *Improved model for supraventricular heartbeats.* Throughout the chapters, the weakest point of all the classifiers presented was the supraventricular class performance. From Chapter 5 it is clear to us that both feature models, LDC-C and EMC, still have limitations to discriminate the supraventricular class. New features and classification strategies should be explored to improve this aspect, for example classification models that take into account the sequence of heartbeats, as hidden Markov models.

3. *Other physiological signals.* Signals as the blood pressure or the plethysmographic signal are of much interest for heartbeat classification. The main reason is because in our model, we only take into account the electrical activity of the heart. Any signal that includes the mechanical counterpart can allow us to make a better characterization of the heart function. Therefore, the strategies to include them in the classification model would probably lead to a performance improvement.

4. *Arbitrary length of recordings.* One aspect not considered in this thesis were the long-term recordings included in some databases. Specially in this kind of recordings the nonstationarities of the ECG signal are evidenced. Although the performance figures achieved by our classifier were good for these recordings, we believe that improvements in performance and/or expert assistance required (amount of MAHB) can be achieved if we deal with the length of the recordings. Some solutions were suggested in the work of Kiranyaz *et al.* [Kiranyaz et al., 2011].

5. *Other classification strategies.* One promising classification alternative to the framework presented in the first chapters can be the SVM semi-supervised learning methods [Joachims, 1999], which can take advantage of unlabeled data to improve the training of the classifier. This feature can be important in order to improve the performance in the less represented classes, as the supraventricular class. Note that this kind of heartbeats are not easy to detect even for experienced cardiologists, therefore it is hard to find databases with reliable and high quality heartbeats annotations as the MITBIH-AR.

6. *Extrapolation of the semiautomatic framework to ECG detection/delineation.* The semiautomatic framework presented in Chapter 5 can be useful for similar ECG processing tasks, such as the QRS detection and ECG delineation. The possibility of interacting (e.g. correcting errors) with the algorithm when processing long-term recordings could be a useful feature. This would avoid the necessity of changing thresholds or recording for cases not considered during the development of the algorithm.

Scientific Contributions

This thesis produced the following contributions to peer-reviewed journals and conference proceedings:

Journal articles

M. Llamedo and J.P. Martínez. Heartbeat classification using feature selection driven by database generalization criteria. *IEEE Transactions on Biomedical Engineering*, 58(3): 616–625, march 2011. ISSN 0018-9294. doi: 10.1109/TBME.2010.2068048.

M. Llamedo and J. P. Martínez. Cross-database evaluation of a multilead heartbeat classifier. *IEEE Transactions on Information Technology in Biomedicine*, In press:–, 2012a.

M. Llamedo and J. P. Martínez. An automatic patient-adapted ECG heartbeat classifier allowing expert assistance. *IEEE Transactions on Biomedical Engineering*, Currently under review, with major revision:–, 2012b.

T. Mar, S. Zaunseder, J. P. Martínez, M. Llamedo, and R. Poll. Optimization of ECG classification by means of feature selection. *Biomedical Engineering, IEEE Transactions on*, 58(8):2168 –2177, aug. 2011. ISSN 0018-9294. doi: 10.1109/TBME.2011.2113395.

Conference proceedings

J Bolea, E Pueyo, Almeida R, M Sotaquira, M Llamedo, J.P. Martínez, P Laguna, and E.G. Caiani. Microgravedad simulada mediante head-down-bed-resting y su influencia en la dinámica del QT/RR. In *XXIX Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 11).*, 2011.

M. Llamedo and J.P. Martínez. Analysis of a semiautomatic algorithm for ECG heartbeat classification. In *Computers in Cardiology 2011*, volume 38. IEEE Computer Society Press, 2011a.

M. Llamedo and J.P. Martínez. Análisis de un algoritmo para la clasificación semiautomática de latidos en ECG. In *XXIX Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 11).*, 2011b.

M. Llamedo and J.P. Martínez. Analysis of 12-lead classification models for ECG classification. In *Computers in Cardiology 2010*, volume 37. IEEE Computer Society Press, 2010a.

M. Llamedo and J.P. Martínez. Evaluation of an ECG heartbeat classifier designed by generalization-driven feature selection. In *Engineering in Medicine and Biology Society. EMBC 2010. Annual International Conference of the IEEE*, 2010b.

M. Llamedo and J.P. Martínez. Analysis of multidomain features for ECG classification. In *Computers in Cardiology 2009*, volume 36, pages 561 – 564. IEEE Computer Society Press, 2009a.

M. Llamedo and J.P. Martínez. Clasificación de ECG basada en características de escala, dirección y ritmo. In *XXVI Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 09).*, 2009b.

M. Llamedo and J.P. Martínez. An ECG classification model based on multilead wavelet transform features. In *Computers in Cardiology 2007*, volume 34, pages 105–108. IEEE Computer Society Press, 2007a.

M. Llamedo and J.P. Martínez. An ECG classification model based on multilead wavelet transform features. In *XVI Congreso Argentino de Bioingeniería. San Juan. ISBN 978-950-605-505-9*, pages 531–534, 2007b.

Appendix A

Matlab Implementation

A.1 Introduction

The *Argentinian-Aragonese heartbeat classifier* (*a2hbc*[1]) [Llamedo, 2012] is a Matlab script developed for research purposes during the development of this thesis. It is just a heartbeat classifier and its graphical user interface (GUI) to be operated. The main objective of this software is to ease the performance comparison against other (hope better) heartbeat classifiers. It can also be used for classifying unlabeled ECG recordings as well.

A.2 Features

The main features of *a2hbc* are:

- *Validated performance.* The performance was thoroughly evaluated in [Llamedo and Martínez, 2012b].
- *More common ECG formats accepted* (MIT, ISHNE, AHA and HES)
- *Open source.* Documented and easily customizable.
- *Multiprocessing ready.* It is ready to run in both a desktop PC or a high performance cluster.
- *User interface.* The algorithm has a simple graphical user interface (GUI) to ease the labeling of heartbeat clusters.

A.3 Installation and Usage

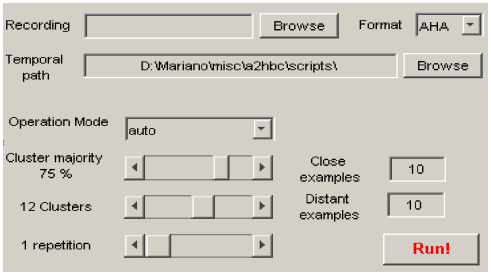
The *a2hbc* can be downloaded from here, just unpack it, read the documentation, try some of the examples included and you are done. Several examples are included in the

[1] http://code.google.com/p/a2hbc/

examples.m script. In this appendix we will show some examples to show the main features of *a2hbc*. There are also some recordings included to follow the examples. It can be started simply by executing the main script

```
>>a2hbc
```

In this case the *control panel* is displayed to gather some information from the user:

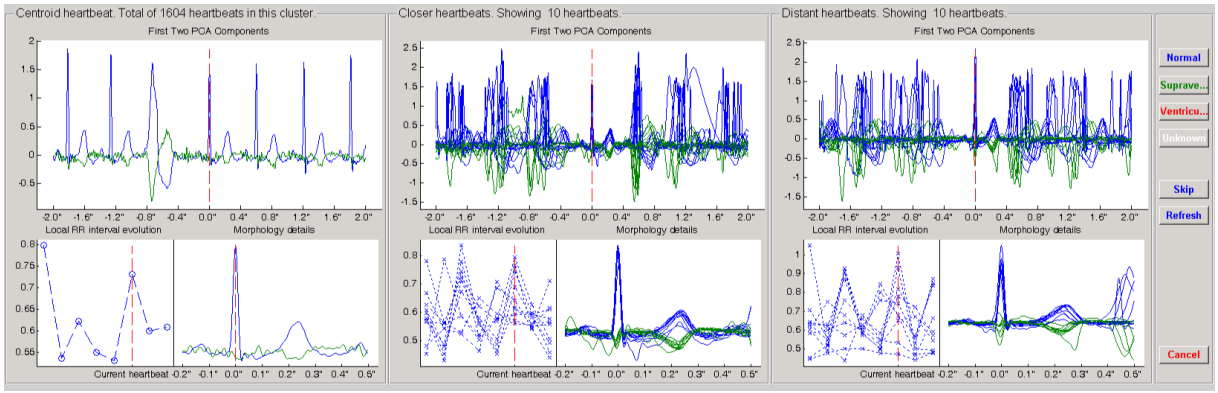


You must enter the recording file name and its format, the other controls you can leave with its default values by the moment. You can select the *208.dat* recording, included in the *example recordings* folder, and set the *MIT* format. Then press *Run!,* the script will start working. You can follow the evolution in the progress bar, and after a while, it ends and displays the classification results

```
Configuration
-------------
+ Recording: ... \example recordings\208.dat (MIT)
+ Mode: auto (12 clusters, 1 iterations, 75% cluster-presence)

  True             | Estimated Labels
  Labels           | Normal  Suprav  Ventri  Unknow| Totals
 ------------------|-------------------------------|-------
  Normal           | 1567       6      13       0  | 1586
  Supraventricular|    2       0       0       0  |    2
  Ventricular      |  255       8    1102       0  | 1365
  Unknown          |    2       0       0       0  |    2
 ------------------|-------------------------------|-------
  Totals           | 1826      14    1115       0  | 2955

Balanced Results for
--------------------
| Normal    || Supravent || Ventricul ||           TOTALS            |
|  Se   +P  ||  Se   +P  ||  Se   +P  ||   Acc   |   Se    |   +P    |
|  99%  45% ||   0%   0% ||  81%  99% ||   60%   |   60%   |   48%   |

Unbalanced Results for
----------------------
| Normal    || Supravent || Ventricul ||           TOTALS            |
```

```
|   Se    +P   ||   Se    +P   ||   Se    +P   ||    Acc    |    Se     |    +P     |
|   99%   86%  ||    0%    0%  ||   81%   99%  ||    90%    |    60%    |    62%    |
```

This is possible because this recording include the expert annotations, or *ground truth*, for each heartbeat. The manual annotations in MIT format are typically included in *.atr* files (in this case *208.atr*). Now you can check other operation modes, as the *slightly-assisted*. Click on *Run!* and then, eventually, the algorithm may ask you for help. In case of needing help, a window like this will appear:



In this window the algorithm is asking you to label the centroid of the cluster, that is showed in the left panel. In the top of each panel some information is showed, as the amount of heartbeats in the current cluster. In the middle panel, you have some examples of heartbeats close to the centroid in a likelihood sense. The same is repeated in the right panel, but with examples far from the centroid. This manner you can have an idea of the dispersion of heartbeats within a cluster. Large differences across the panels indicate large cluster dispersion. If you decide to label the cluster, you can use one of the 4 buttons on your right. The unknown class is reserved for the cases where you can not make a confident decision. At the same time, in the command window, a suggestion appears:

```
Configuration
-------------
+ Recording: .\example recordings\208.dat (MIT)
+ Mode: assisted (3 clusters, 1 iterations, 75% cluster-presence)
Suggestion: Normal
```

This means that the centroid heartbeat in the *.atr* file is labeled as *Normal*. You will see this suggestion for each cluster analyzed, if there are annotations previously available. You are informed about the percentage of heartbeats already labeled with a progress bar, in the bottom of the control panel window.

In case you believe that a cluster includes several classes of heartbeats, you can decide to *skip* the classification, and try to re-cluster those heartbeats in the next iteration. You are free to perform as many iterations as you decide, by skipping clusters. The refresh button resamples heartbeats close and far from the centroid, and then redraws the middle and right panels. This feature is useful for large clusters.

There are two possible ways of using A2HBC, in a single desktop PC or in a high performance cluster of computers.

A.3.1 The power of the command-line

You can control all the features described up to the moment (and more) from the command-line of Matlab. This is particularly useful for integrating *a2hbc* in your scripts. You can find several examples in the script *examples.m*, this is probably the best way of getting familiar with it. Here I reproduce some worked examples included in this script. First let's start executing

```
a2hbc( ...
    'recording_name', [ '.' filesep 'example recordings' filesep '208.dat'], ...
    'recording_format', 'MIT', ...
    'op_mode', 'auto');
```

As you can see, the parameter interface of *a2hbc* is by *name-value* parameters. In the previous example, the name of the parameters are self-explanatory, the only comment is for the third, which is the operating mode. In Table A.1 you can check the complete list of parameters. As a result, you will get similar results to the obtained in the first example using the GUI.

```
Configuration
-------------
+ Recording: .\example recordings\208.dat (MIT)
+ Mode: auto (12 clusters, 1 iterations, 75% cluster-presence)

  True             | Estimated Labels
  Labels           | Normal  Suprav  Ventri  Unknow| Totals
 ------------------|---------------------------|-------
  Normal           | 1575       3       8       0  | 1586
  Supraventricular|    2       0       0       0  |    2
  Ventricular      |  250       7    1108       0  | 1365
  Unknown          |    1       0       1       0  |    2
 ------------------|---------------------------|-------
  Totals           | 1828      10    1117       0  | 2955

Balanced Results for
---------------------
| Normal     || Supravent || Ventricul ||            TOTALS             |
|   Se    +P  ||   Se   +P  ||   Se   +P  ||    Acc    |    Se     |   +P     |
|   99%   46% ||    0%   0% ||   81%  99% ||    60%    |    60%    |   48%    |

Unbalanced Results for
----------------------
| Normal     || Supravent || Ventricul ||            TOTALS             |
|   Se    +P  ||   Se   +P  ||   Se   +P  ||    Acc    |    Se     |   +P     |
```

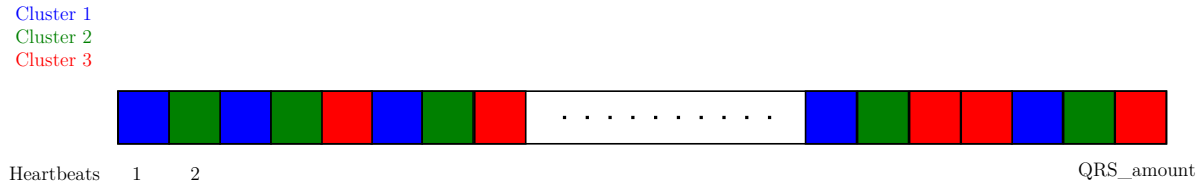
```
|  99%  86% ||   0%   0% ||  81%  99% ||   91%   |   60%   |   62%   |
```

In the case that you would like to integrate a2hbc to your software, or you have a proprietary ECG format not allowed by a2hbc, the best choice is that you pass the ECG samples directly. For doing this, you will have follow the requirements in Table A.1 indicated with a [2], and

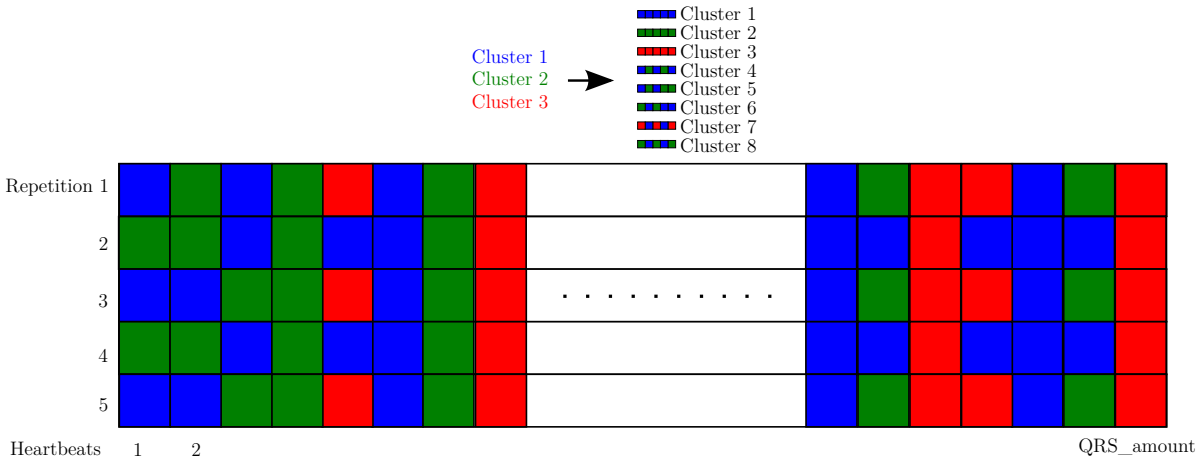
- *ECG* is the ECG signal matrix of *nsamp*× *nsig*, in ADC samples.
- *ECG_header* is a *struct* with the ECG properties, with fields:
  - *freq* is the sampling rate of the ECG
  - *nsamp* is the number of samples
  - *nsig* is the amount of leads.
  - *gain* is a vector of *nsig*×1 with the gain of each lead ($^{\text{ADC samples}}/_{\mu V}$).
  - *adczero* is a vector of *nsig*×1 with the offset of each lead in ADC samples.
- *QRS_annotations* is a *struct* with the location of the QRS complexes, with fields:
  - *time* is a vector of *QRS_amount*×1, with the sample value where the QRS complexes are.
  - *ann_type* [optional] is a *char* vector of QRS_amount×1, with each heartbeat label. This field is for evaluating the performance of a classifier, as a result *a2hbc* generates the confusion matrix seen in the examples above.

The parameters 'cant_pids' and 'this_pid' are explained in the next section, since were designed for partitioning and multiprocessing of recordings. The parameter *SimulateExpert* was designed to simulate the expert input by using the expert annotations provided in the annotations files, or via the *QRS_annotations* parameter. *ClusterPresence* is a threshold for evaluating the qualified majority in operating modes *auto* or *slightly-assisted*. The lower this threshold the more confident the algorithm in labeling all heartbeats in a cluster as the centroid. *Repetitions* was designed to evaluate multiple times the algorithm performance in a particular recording. As a result, the computed confusion matrix is a 3-D cube with the amount of repetitions as the third coordinate. With this kind of confusion matrix it is possible to estimate the dispersion of the results presented in [Llamedo Soria, 2012, Llamedo and Martínez, 2012b]. Finally, the *ClusteringRepetitions* parameter requires a special explanation, since it was not described in the bibliography. In few words, it is a trick for increasing the clustering resolution for complex or long-term recordings. The higher this parameter, the higher the amount of cluster found and, at the end, the assistance required by the algorithm.

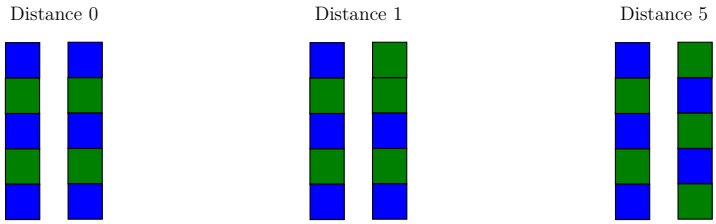
From the user point of view, you should be satisfied with this clue, however if you are interested in the trick, I will explain it with a toy example. Consider the following clustering problem, where we are interested in finding 3 clusters.



Now if we repeat the same process two times, our clustering algorithm does not guarantee to find the same partition. However our intuition tells us that in case where the classes are well separated, the partition is likely to remain very similar through the repetitions. In the opposite case, the partition can change. Then after N repetitions, the *QRS_amount* heartbeats were assigned N cluster labels.



So the number of clusters increases according to N. In order to keep this number in the order of tens, it was used a merging criterion for *similar clusters*. The similarity is measure in terms of labeling differences across the repetitions. For example:



Then we can group together some clusters based on this labeling distance. The *a2hbc* group together clusters within a distance of $0.2 \cdot N$.

A.3.2 The power of a high performance computing cluster

Maybe one of the most useful features of *a2hbc* is that was developed for being used in a high performance computing cluster. The parameters *cant_pids* and *this_pid* controls the partitioning of the work for each recording.

Table A.1: List of parameters of *a2hbc*

Name	Value default	Value validation	Description
'recording_name'	–	*char*	The file name of the ECG recording[1]
'recording_format'	–	{MIT, 'ISHNE', 'AHA', 'HES', 'MAT'}	The format of the ECG recording[1]
'ECG'	–	*numeric*	The ECG samples[2]
'ECG_header'	–	*struct*	Struct with ECG features[2,3]
'QRS_annotations'	–	*numeric* && $x_i \geq 1$, $\forall i$	Sample occurrence of QRS complexes[2]
'op_mode'	'auto'	{'auto', 'slightly-assisted', 'assisted'} || $1 \leq x \leq 3$	
'cant_pids'	1	$x \geq 1$	How many processes in total to compute this recording[3]
'this_pid'	1	$x \geq 1$ && $x \leq$'cant_pids'	Which of the processes is this.[3]
'CacheData'	true	*logical*	Save intermediate results to speed-up re-processing.
'InteractiveMode'	false	*logical*	Show the control panel after processing the recording.
'SimulateExpert'	false	*logical*	Use expert annotations to simulate expert interaction[3]
'tmp_path'	–	*char*	Path to store intermediate results.
'NumOfClusters'	12	$x > 1$	Number of cluster to search.
'ClusteringRepetitions'	1	$1 \leq x \leq 10$	Repetitions of the clustering process.[3]
'ClusterPresence'	75	$0 \leq x \leq 100$	Threshold for the qualified majority.[3]
'Repetitions'	1	$x \geq 1$	Repetitions to evaluate this recordings.[3]

[1] and [2] indicates groups of parameters that can not be mixed. You can specify file name and format, or pass the ECG samples, QRS annotations, etc.
[3] See a complete explanation in the text, or in the references [Llamedo Soria, 2012, Llamedo and M

```
% Computer 1

lab1 = a2hbc( ...
    'recording_name', [ '.' filesep 'example recordings' filesep '208.dat'], ...
    'recording_format', 'MIT', ...
    'this_pid', 1, ...
    'cant_pid', 2, ...
    'op_mode', 'auto');

% Computer 2

lab2 = a2hbc( ...
    'recording_name', [ '.' filesep 'example recordings' filesep '208.dat'], ...
    'recording_format', 'MIT', ...
    'this_pid', 2, ...
    'cant_pid', 2, ...
    'op_mode', 'auto');


% Somewhere: results collection and processing
lab = [lab1; lab2];
...
```

It is recommended to adapt this features to the batch manager available in your computing facilities. In the case of our University, the batch manager used is Condor [Condor, 2010]. You can ask me for the condor implementation for multiprocessing, but the details of this are outside of the scope of this tutorial.

A.4 Acknowledgments

This software was supported by the University of Zaragoza, Spain and the National Technical University of Buenos Aires, Argentina.

List of acronyms

A Accuracy

A-V Atrio-ventricular

AAMI Association for the Advancement of Medical Instrumentation

AAMI2 AAMI class labeling modification, see section 2.1.

AHA American Heart Association ECG Database

ANN Artificial neural networks

AP Action potential

APB Atrial premature beats

ARP Absolute refractory period

Biosigna Biosigna Database

bpm Beats per minute

CVD Cardiovascular diseases

CWT Continuous wavelet transform

DS1 Subset of MITBIH-AR for training purposes. Comprises the recordings 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223 and 230.

DS2 Subset of MITBIH-AR for testing purposes. Comprises the recordings 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223 and 230.

DWT Discrete wavelet transform

ECG Electrocardiogram

EMC Expectation-maximization clustering algorithm

ESTTDB European ST-T Database

F Fusion of normal and ventricular beats

FP Fiducial point. QRS complex location

FS Feature selection

INCART St. Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database

ISCS Intraventricular specific conduction systems

LDC Linear discriminant classifier

LDC-C Linear discriminant classifier compensated for class imbalance

LTSTDB The Long-Term ST Database

MAHB Manually annotated heartbeats

MCD Minimum covariance determinant estimator

MIT-BIH-SUP MIT-BIH Supraventricular Arrhythmia Database

MITBIH-AR MIT-BIH Arrhythmia Database

MITBIH-ST The MIT-BIH ST Change Database

MLP multilayer perceptron

MoG Mixture of Gaussians

N Normal beats

QDC Quadratic discriminant classifier

QRS Complex of waves Q, R and S on the ECG. Related to the ventricular depolarization.

RR Interval between two consecutive heartbeats.

RRP Relative refractory period

S Supraventricular beats

SCD Sudden cardiac death

SCS Specific conduction system

SFFS Sequential floating feature selection algorithm

SNR Signal to noise ratio

TRP Total refractory period

V Ventricular beats

V' Ventricular and fusion beats

VCG Vectocardiogram

VPB Ventricular ectopic beats

WT Wavelet Transform

P^+ Positive predictive value

S Sensitivity

List of Figures

List of Tables

Bibliography

AAMI-EC57. Testing and reporting performance results of cardiac rhythm and ST-segment measurement algorithms. American National Standard, ANSI/AAMI/ISO EC57, 1998–2008.

Burak Acar and Hayrettin Köymen. SVD-based on-line exercise ECG signal orthogonalization. 46:311 – 321, 1999.

American Heart Association. American heart association ECG database. URL `https://www.ecri.org`.

Claus Bahlmann. Directional features in online handwriting recognition. *Pattern Recognition*, 39(1):115 – 125, 2006. ISSN 0031-3203. doi: DOI:10.1016/j.patcog.2005.05.012.

Antoni Bayés de Luna. *Electrocardiografía clínica de las arritmias*. Publicaciones Permanyer, 2010.

P. de Chazal, M O'Dwyer, and R. B. Reilly. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 51:1196–1206, 2004.

P. de Chazal and R. B. Reilly. A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 53: 2535–2543, 2006.

I. Christov, German Gómez-Herrero, Vessela Krasteva, I Jekova, A Gotchev, and Karen Egiazarian. Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification. *Elsevier Medical Engineering & Physics*, 28:876–887, 2006.

V Chudácek, G Georgoulas, L Lhotská, C Stylios, M Petrík, and M Cepek. Examining cross-database global training to evaluate five different methods for ventricular beat classification. *Physiological Measurement*, 30(7):661, 2009. URL `http://stacks.iop.org/0967-3334/30/i=7/a=010`.

Albert Cohen and Jelena Kovačević. Wavelets: The mathematical background. *Proceedings of the IEEE*, 84(4):514–522, April 1996.

Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological measurements*, 20:37–46, 1960. doi: 10.1177/001316446002000104.

Condor. Condor high throughput computing system, 2010. URL `http://www.cs.wisc.edu/condor/`.

Dirección de Estadísticas e Información en Salud, Feb. 2012. URL `http://www.deis.gov.ar/`.

R.P.W. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. deRidder, D.M.J. Tax, and S. Verzakov. PR-tools, a matlab toolbox for pattern recognition, 2008. URL `http://www.prtools.org`.

W Einthoven. Weiteres über das elektrokardiogram. *Pflüger Arch. ges. Physiol.*, 122: 517–48, 1908.

Peter Filzmoser, Ricardo Maronna, and Mark Werner. Outlier identification in high dimensions. *Computational Statistics & Data Analysis*, 52:1694 – 1711, 2008. doi: 10.1016/j.csda.2007.05.018.

R Fischer, MF Sinner, R Petrovic, E Tarita, S Kääb, and T K Zywietz. Testing the quality of 12 lead holter analysis algorithms. In *Computers in Cardiology*, volume 35, pages 453–456, 2008.

Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B Moody, Chung-Kang Peng, and H. Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

SD Greenwald. *Improved detection and classification of arrhythmias in noise-corrupted electrocardiograms using contextual information.* PhD thesis, Harvard-MIT Division of Health Sciences and Technology, 1990.

Neil R Grubb and Steve Furniss. Radiofrequency ablation for atrial fibrillation. *BMJ*, 322(7289):777–780, 3 2001. doi: 10.1136/bmj.322.7289.777.

Arthur Guyton and John Hall. *Textbook of Medical Physiology.* Elsevier Saunders, 11th edition, 2006. ISBN 0-8089-2317-X.

F. van der Heijden, R.P.W. Duin, D. de Ridder, and D.M.J. Tax. *Classification, Parameter Estimation and State Estimation.* John Wiley & Sons, 2005. doi: 10.1002/0470090154.

G. E. Hinton, S. Osindero, and Y. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–1554, 2006.

Y. H. Hu, S. Palreddy, and WJ Tompkins. A patient-adaptable ECG beat classifier using mixture of experts approach. *IEEE Transactions on Biomedical Engineering*, 44: 891–899, 1997.

OT Inan, L Giovangrandi, and GTA Kovacs. Robust neural-netwok-based classification of premature ventricular contractions using wavelet transform and timing interval features. *IEEE Transactions on Biomedical Engineering*, 53:2507–2515, 2006.

Turker Ince, Serkan Kiranyaz, and Moncef Gabbouj. A generic and robust system for automated patient-specific classification of ECG signals. *IEEE Transactions on Biomedical Engineering*, 56:1415–1426, 2009.

Instituto Nacional de Estadística. Inebase, Feb 2012. URL `http:/www.ine.es/inebase/`.

Ovidiu Ivanciuc. Applications of support vector machines in chemistry. *Rev. Comput. Chem.*, 23:291–400, 2007.

F. Jager, A. Taddei, G. Moody, M. Emdin, G. Antoli, R. Dorn, A. Smrdel, C. Marchesi, and R. Mark. Long-term ST database: A reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia. *Medical and Biological Engineering and Computing*, 41:172–182, 2003. ISSN 0140-0118. URL `http://dx.doi.org/10.1007/BF02344885`. 10.1007/BF02344885.

Wei Jiang and Seong Kong. Block-based neural networks for personalized ECG signal classification. *IEEE Transactions on Biomedical Engineering*, 18:1750–1761, 2007.

Thorsten Joachims. Transductive inference for text classification using support vector machines. pages 200–209. Morgan Kaufmann, 1999.

Serkan Kiranyaz, Turker Ince, Jenni Pulkkinen, and Moncef Gabbouj. Personalized long-term ECG classification: A systematic approach. *Expert Systems with Applications*, 38 (4):3220 – 3226, 2011. ISSN 0957-4174. doi: DOI:10.1016/j.eswa.2010.09.010. URL `http://www.sciencedirect.com/science/article/pii/S0957417410009140`.

Vessela Krasteva and Irena Jekova. Qrs template matching for recognition of ventricular ectopic beats. *Annals of Biomedical Engineering*, 35:2065–2076, 2007. ISSN 0090-6964. URL `http://dx.doi.org/10.1007/s10439-007-9368-9`. 10.1007/s10439-007-9368-9.

M. Lagerholm, C Peterson, G Braccini, L Edenbrandt, and L Sörnmo. Clustering ECG complexes using hermite functions and self-organizing maps. *IEEE Transactions on Biomedical Engineering*, 47:838–848, 2000.

P Laguna, R Jané, and P Caminal. Automatic detection of wave boundaries in multilead ECG signals: Validation with the cse database. *Computers and Biomedical Research*, 27(1):24–60, 1994. URL `http://www.physionet.org/physiotools/ecgpuwave/`.

G.de Lannoy. Feature relevance assessment in auntomatic inter-patient heartbeat classification. In *BIOSIGNALS 2010 - International Conference on Bio-inspired Systems and Signal Processing*, 2010.

Cuiwei Li, Chongxun Zheng, and Changfeng Tai. Detection of ECG characteristic points using wavelet transforms. 42(1):21–28, January 1995.

Mariano Llamedo Soria. *Automatic Processing and Classification of Electrocardiogram for the Detection of Risk Indexes*. PhD thesis, Universidad de Zaragoza, June 2012.

Mariano Llamedo. Argentine-aragonese heartbeat classifier (a2hbc), 2012. URL `http://code.google.com/p/a2hbc/`.

M. Llamedo and J.P. Martínez. Heartbeat classification using feature selection driven by database generalization criteria. *IEEE Transactions on Biomedical Engineering*, 58(3): 616–625, march 2011a. ISSN 0018-9294. doi: 10.1109/TBME.2010.2068048.

M. Llamedo and J. P. Martínez. Cross-database evaluation of a multilead heartbeat classifier. *IEEE Transactions on Information Technology in Biomedicine*, In press:–, 2012a.

M. Llamedo and J. P. Martínez. An automatic patient-adapted ECG heartbeat classifier allowing expert assistance. *IEEE Transactions on Biomedical Engineering*, Currently under review, with major revision:–, 2012b.

M. Llamedo and J. P. Martínez. An ECG classification model based on multilead wavelet transform features. In *Computers in Cardiology 2007*, volume 34, pages 105–108. IEEE Computer Society Press, 2007.

S. Mallat. *A Wavelet tour of signal processing*. Academic Press, second edition, 1999.

Stephane Mallat. Multifrequency channel decompositions of images and wavelet models. 37:2091–2110, December 1989.

S. Mallat and S. Zhong. Characterization of signals from multiscale edge. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 14(7):710–732, 1992.

J Malmivuo and R Plonsey. *Bioelectromagnetism*. Oxford University Press, 1995. ISBN 0-19-505823-2.

T. Mar, S. Zaunseder, J. P. Martínez, M. Llamedo, and R. Poll. Optimization of ECG classification by means of feature selection. *Biomedical Engineering, IEEE Transactions on*, 58(8):2168 –2177, aug. 2011. ISSN 0018-9294. doi: 10.1109/TBME.2011.2113395.

R. Mark, G. Moody, and SD. Greenwald. Mit-bih supraventricular arrhythmia database. http://www.physionet.org/physiobank/database/svdb/, 1990.

J P Martínez, R Almeida, S Olmos, AP Rocha, and P Laguna. A wavelet-based ECG delineator: Evaluation on standard databases. *IEEE Transactions on Biomedical Engineering*, 51:570–581, 2004.

R Mason and L Likar. A new system of multiple leads exercise electrocardiography. *Am. Heart J.*, 71:(2):196–205, 1966.

GB Moody and RG Mark. The impact of the MIT-BIH arrhythmia database. *IEEE Eng in Med and Biol*, 20(3):45–50, May-June 2001.

Andrea Natale and Oussama Wazni. *Handbook of Cardiac Electrophysiology.* Informa Healthcare, 2007. ISBN-10: 1 84184 620 1.

K.S. Park, B.H. Cho, D.H. Lee, S.H. Song, and J.S. Lee. Hierarchical support vector machine. In *Computers in Cardiology 2008*, volume 35, pages 229–232. IEEE Computer Society Press, 2008.

P Pudil, J Novovicova, and J Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11):1119–1125, 1994.

P.J. Rousseeuw. Least median of squares regression. *Journal of the American Statistical Association*, 79:871–881, 1984.

P.J. Rousseeuw and K. Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41:212–223, 1999.

D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. pages 318–362, 1986.

J. S. Sahambi, S.N. Tandon, and R. K. P. Bhatt. Wavelet based ST-segment analysis. 36 (9):568–572, September 1998.

Zeeshan Syed, John Guttag, and Collin Stultz. Clustering and symbolic analysis of cardiovascular signals: discovery and visualization of medically relevant patterns in long-term data using limited prior knowledge. *EURASIP J. Appl. Signal Process.*, 2007:97–97, January 2007. ISSN 1110-8657. doi: http://dx.doi.org/10.1155/2007/67938. URL `http://dx.doi.org/10.1155/2007/67938`.

L. Sörnmo and P. Laguna. *Bioelectrical Signal Processing in Cardiac and Neurological Applications.* Elsevier, 2005. ISBN 0-12-437552-9.

A. Taddei, G. Distante, M. Emdin, P. Pisani, G. B. Moody, C. Zeelenberg, and C. Marchesi. The european ST-T database: standard for evaluating systems for the analysis of ST-t changes in ambulatory electrocardiography. *European Heart Journal*, 13 (9):1164–1172, 1992. URL `http://eurheartj.oxfordjournals.org/content/13/9/1164.abstract`.

G. J. Taylor. *150 Practice ECGs: Interpretation and Review*. Blackwell Science, 2002. ISBN 0-632-04623-6.

S. Verboven and M. Hubert. Libra: a matlab library for robust analysis. *Chemometrics and Intelligent Laboratory Systems*, 75:127–136, 2005. URL `http://wis.kuleuven.be/stat/robust/LIBRA.html`.

AD Waller. A demonstration on man of electromotive changes accompanying the heart's beat. *J. Physiol. (Lond.)*, 8:229–34, 1887.

R. Watrous and G. Towell. A patient-adaptive neural network ECG patient monitoring algorithm. In *Computers in Cardiology 1995*, pages 229 –232, sep 1995. doi: 10.1109/CIC.1995.482614.

FN Wilson, AG Macleod, and PS Barker. Potential variations produced by the heart beat at the apices of einthoven's triangle. *Am. Heart J.*, 7:207–11, 1931.

World Health Organization. Cardiovascular diseases, 2012. URL `http://www.who.int/cardiovascular_diseases/en/`.