Possibilities for the development of a decision support system for diagnosing heart failure

Linda Olsson

2007-08-14
LiTH-IMT/FMT20-EX--07/457--SE
Heart failure is a common disease which is difficult to diagnose. To aid physicians in diagnosing heart failure, a decision support system has been proposed. Parameters useful to the system are suggested. Some of these, such as age and gender, should be provided by the physician, and some should be derived from electro- and phonocardiographic signals.

Various methods of signal processing, such as wavelet theory and principal components analysis, are described. Heart failure should be diagnosed based on the parameters, and so various forms of decision support systems, such as neural networks and support vector machines, are described. The methods of signal processing and classification are discussed and suggestions on how to develop the system are made.
I would like to thank Peter Hult for presenting the idea to me, Linda Rattfält for taking the time to answer my questions, Marcus Ressner for telling me about the ways of the research community, Jerker Karlsson for answering my medical questions and Martin Eneling for reading and commenting on the report. Above all, I want to express my gratitude to my family, for their never-ending support.
# TABLE OF CONTENTS

Abbreviations ______________________________________________________________ 1  
1 Introduction ______________________________________________________________ 3  
   1.1 Background ___________________________________________________________ 3  
   1.2 Aim _________________________________________________________________ 3  
   1.3 Materials and methods___________________________________________________ 3  
   1.4 Limitations____________________________________________________________ 3  
   1.5 Outline of the thesis_____________________________________________________ 4  
2 Anatomy and physiology of the heart __________________________________________ 5  
   2.1 Circulation of blood_____________________________________________________ 5  
   2.2 The conduction system __________________________________________________ 6  
   2.3 Electrocardiography ____________________________________________________ 7  
   2.4 Phonocardiography _____________________________________________________ 8  
3 Time frequency analysis____________________________________________________ 11  
   3.1 The wavelet transform__________________________________________________ 11  
   3.2 Multiresolution representation ___________________________________________ 12  
4 Characteristics of a heart failure patient________________________________________ 15  
   4.1 Heart failure__________________________________________________________ 15  
   4.2 Diagnosing heart failure ________________________________________________ 15  
   4.3 Common signs and symptoms____________________________________________ 16  
   4.4 Signs and symptoms present in systolic heart failure _________________________ 17  
   4.5 Signs and symptoms present in diastolic heart failure ________________________ 18  
   4.6 Summary of parameters_________________________________________________ 18  
5 Extracting visible features __________________________________________________ 21  
   5.1 Pre-processing of the ECG signal _________________________________________ 21  
   5.2 Left ventricular hypertrophy _____________________________________________ 21  
   5.3 Left bundle branch block________________________________________________ 22  
   5.4 Myocardial ischemia ___________________________________________________ 22  
   5.5 Arrhythmias__________________________________________________________ 23  
      5.5.1 Atrial fibrillation___________________________________________________ 23  
      5.5.2 Premature ventricular contractions_____________________________________ 23  
6 Extracting invisible features_________________________________________________ 25  
   6.1 Heart rate variability analysis ____________________________________________ 25  
      6.1.1 Spectral analysis____________________________________________________ 25
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1.2</td>
<td>Detrended fluctuation analysis</td>
<td>25</td>
</tr>
<tr>
<td>6.1.3</td>
<td>Multiresolution wavelet analysis</td>
<td>27</td>
</tr>
<tr>
<td>6.2</td>
<td>High frequency, low amplitude parameters</td>
<td>27</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Ventricular late potentials</td>
<td>28</td>
</tr>
<tr>
<td>6.2.2</td>
<td>P wave duration</td>
<td>28</td>
</tr>
<tr>
<td>6.3</td>
<td>T wave alternans</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>Extracting features with wavelets</td>
<td>31</td>
</tr>
<tr>
<td>7.1</td>
<td>Using wavelet coefficients</td>
<td>31</td>
</tr>
<tr>
<td>7.2</td>
<td>Using a tailored wavelet</td>
<td>32</td>
</tr>
<tr>
<td>8</td>
<td>Principal components analysis</td>
<td>33</td>
</tr>
<tr>
<td>9</td>
<td>Neural networks</td>
<td>35</td>
</tr>
<tr>
<td>9.1</td>
<td>Multilayer perceptrons</td>
<td>36</td>
</tr>
<tr>
<td>9.2</td>
<td>Radial-basis function networks</td>
<td>38</td>
</tr>
<tr>
<td>10</td>
<td>Support vector machines</td>
<td>41</td>
</tr>
<tr>
<td>11</td>
<td>Committee machines</td>
<td>47</td>
</tr>
<tr>
<td>11.1</td>
<td>Ensemble averaging</td>
<td>47</td>
</tr>
<tr>
<td>11.2</td>
<td>Bagging</td>
<td>47</td>
</tr>
<tr>
<td>11.3</td>
<td>Boosting</td>
<td>47</td>
</tr>
<tr>
<td>11.4</td>
<td>Mixtures of experts</td>
<td>47</td>
</tr>
<tr>
<td>12</td>
<td>Self-organising maps</td>
<td>49</td>
</tr>
<tr>
<td>13</td>
<td>Clinical decision support systems</td>
<td>51</td>
</tr>
<tr>
<td>13.1</td>
<td>Model selection</td>
<td>51</td>
</tr>
<tr>
<td>13.2</td>
<td>Some examples of decision support systems</td>
<td>51</td>
</tr>
<tr>
<td>14</td>
<td>Discussion</td>
<td>53</td>
</tr>
<tr>
<td>14.1</td>
<td>General demands</td>
<td>53</td>
</tr>
<tr>
<td>14.2</td>
<td>Feature selection</td>
<td>53</td>
</tr>
<tr>
<td>14.3</td>
<td>Using the features</td>
<td>54</td>
</tr>
<tr>
<td>14.4</td>
<td>System selection</td>
<td>54</td>
</tr>
<tr>
<td>14.5</td>
<td>Summary</td>
<td>55</td>
</tr>
<tr>
<td>14.6</td>
<td>Usefulness</td>
<td>55</td>
</tr>
<tr>
<td>14.7</td>
<td>Future work</td>
<td>55</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>57</td>
</tr>
<tr>
<td>Figures</td>
<td></td>
<td>61</td>
</tr>
</tbody>
</table>
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BNP</td>
<td>Brain natriuretic peptide</td>
</tr>
<tr>
<td>CWT</td>
<td>Continuous wavelet transform</td>
</tr>
<tr>
<td>DFA</td>
<td>Detrended fluctuation analysis</td>
</tr>
<tr>
<td>DSS</td>
<td>Decision support system</td>
</tr>
<tr>
<td>DWT</td>
<td>Discrete wavelet transform</td>
</tr>
<tr>
<td>ECG</td>
<td>Electrocardiogram</td>
</tr>
<tr>
<td>HF</td>
<td>High frequency</td>
</tr>
<tr>
<td>HRV</td>
<td>Heart rate variability</td>
</tr>
<tr>
<td>LBBB</td>
<td>Left bundle branch block</td>
</tr>
<tr>
<td>LF</td>
<td>Low frequency</td>
</tr>
<tr>
<td>LVEF</td>
<td>Left ventricular ejection fraction</td>
</tr>
<tr>
<td>LVH</td>
<td>Left ventricular hypertrophy</td>
</tr>
<tr>
<td>ME</td>
<td>Mixture of experts</td>
</tr>
<tr>
<td>MLP</td>
<td>Multilayer perceptron</td>
</tr>
<tr>
<td>MMA</td>
<td>Modified moving average</td>
</tr>
<tr>
<td>MWA</td>
<td>Multiresolution wavelet analysis</td>
</tr>
<tr>
<td>NN</td>
<td>Neural network</td>
</tr>
<tr>
<td>PCG</td>
<td>Phonocardiogram</td>
</tr>
<tr>
<td>PVC</td>
<td>Premature ventricular contraction</td>
</tr>
<tr>
<td>RBF</td>
<td>Radial-basis function</td>
</tr>
<tr>
<td>S₁</td>
<td>First heart sound</td>
</tr>
<tr>
<td>S₂</td>
<td>Second heart sound</td>
</tr>
<tr>
<td>S₃</td>
<td>Third heart sound</td>
</tr>
<tr>
<td>S₄</td>
<td>Fourth heart sound</td>
</tr>
<tr>
<td>SOM</td>
<td>Self-organising map</td>
</tr>
<tr>
<td>STFT</td>
<td>Short time Fourier transform</td>
</tr>
<tr>
<td>SVM</td>
<td>Support vector machine</td>
</tr>
<tr>
<td>TWA</td>
<td>T wave alternans</td>
</tr>
<tr>
<td>VCG</td>
<td>Vectorcardiogram</td>
</tr>
<tr>
<td>VLF</td>
<td>Very low frequency</td>
</tr>
<tr>
<td>VLP</td>
<td>Ventricular late potential</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

1.1 BACKGROUND

Roughly 2% of the population suffer from heart failure, which is a disease associated with high mortality. Since heart failure mainly affects the aging population, we can expect the number of sufferers to rise rapidly. According to Skänér [2004], heart failure consumes 2% of the national health care budget, and yet researchers believe that 30-50% of sufferers are not correctly diagnosed [Fonseca 2006].

Heart failure is not a well-recognised disease, but a rather complex mixture of symptoms. Today, heart failure is usually not discovered until the symptoms are severe, and by then the disease has progressed far. Earlier diagnosis of heart failure would clearly benefit the sufferers, but it would also make health care more cost-effective. It is difficult for a general practitioner to diagnose heart failure at an early stage, and so a decision support system for diagnosing heart failure has been proposed.

1.2 AIM

The aim of the thesis is to examine and evaluate methods for signal processing and classification, which may be useful when constructing a decision support system. More explicitly, these questions are to be answered:

- Which parameters are characteristic for a heart failure patient and might be of use for diagnosing heart failure?
- What kind of signal processing is necessary for diagnosing heart failure using these parameters?
- How should the information in the parameters be weighted to diagnose heart failure?
- Will the proposed methods improve the number of correctly diagnosed patients?

1.3 MATERIALS AND METHODS

All information on heart failure and methods for signal processing and classification described in the thesis are gathered from scientific papers or specialist literature. The credibility of the information sources has been evaluated by critical examinations of described studies and phenomena.

1.4 LIMITATIONS

Heart diseases are a vast and complex subject, where many diseases seem to be connected to each other in more or less intricate ways. A limitation of the thesis is that diseases other than heart failure are discussed only as possible causes of heart failure. Signs, symptoms or diseases that suggest that a patient does not suffer from heart failure, are not considered. The reason for this is that it would require medical training, which the author lacks.

Perhaps the most important part of signal processing of the electrocardiogram is detection of the QRS complex. This is mentioned briefly, but no techniques are described since this is a subject that could easily constitute a full thesis.

There are plenty of methods for extracting features from signals and then classifying them. The methods presented in the thesis are well-known ones that have frequently been used in recent literature. Less well-known methods have been left out, as not enough examples of
their use in processing or classifying the electro- or phonocardiographic signal have been found.

1.5 OUTLINE OF THE THESIS

Chapter 1 gives a brief introduction to the thesis.

Chapter 2 contains the basic anatomy of the heart as well as diagnostic methods.

Chapter 3 describes methods of signal analysis.

Chapter 4 covers heart failure and its symptoms. Parameters which may be used for diagnosing heart failure are identified.

Chapter 5 describes methods for extracting visible information from the electrocardiogram.

Chapter 6 describes methods for acquiring implicit information from the electrocardiogram.

Chapter 7 explains how wavelet theory can be used to extract information from the electro- and phonocardiogram.

Chapter 8 describes principal components analysis, which is used to obtain information from signals.

Chapter 9 introduces decision support systems and neural networks.

Chapter 10 describes another kind of decision support systems; support vector machines.

Chapter 11 covers combinations of decision support systems.

Chapter 12 presents a third kind of decision support systems; self-organising maps.

Chapter 13 contains examples of how decision support systems may be used.

Chapter 14 concludes the thesis by a discussion on the usefulness of the presented methods and the fulfilment of the aims.
2 ANATOMY AND PHYSIOLOGY OF THE HEART

2.1 CIRCULATION OF BLOOD

The heart can be divided into two parts: the right side and the left side, both of which have two chambers; an atrium and a ventricle. Each side receives blood from a circulatory system and pumps blood into another. The right atrium receives deoxygenated blood from the systemic circulation. When the tricuspid valve is open, the blood flows into the right ventricle. Then the tricuspid valve closes, the pulmonary valve opens and the blood is pumped into the pulmonary circulatory system, where it becomes oxygenated. The left atrium receives oxygenated blood, which when the mitral valve is open flows into the left ventricle. When the mitral valve has closed and the aortic valve opens, the blood is pumped into the systemic circulation. This is illustrated in Figure 1.

![Figure 1: The interior of the human heart](image)

The inflow of blood to the atria followed by the outflow from the ventricles is known as the cardiac cycle. The cardiac cycle is divided into a relaxation phase, diastole, and two contraction phases, atrial and ventricular systole. During diastole blood fills the relaxed chambers. At the end of diastole, the atria contract in order to push the remaining blood into the ventricles. After the atrial systole, the ventricles contract.

The blood volume in a ventricle just before ventricular systole begins is called the end-diastolic volume. A measurement often used to describe heart function is the left ventricular ejection fraction (LVEF). LVEF is the fraction of the left ventricular end-diastolic volume that is ejected during the systolic phase of an average heartbeat [Tortora & Derrickson 2006].
2.2 The Conduction System

The heart has its very own pacemaker; a network of specialised cardiac muscle cells which are called auto rhythmic fibres. This network is usually called the cardiac conduction system, and is illustrated in Figure 2. A normal heartbeat is initialised by the sinoatrial node, where an electrical impulse is generated. From there, it propagates via the atrioventricular node to the bundle of His and then through the bundle branches.

![Figure 2: The cardiac conduction system](image)

The impulses generated by the sinoatrial node are called action potentials. An action potential is a way for cells to communicate with each other by means of a certain change in membrane potential. This voltage change excites adjacent cells, and so the action potential propagates. In the heart, there are two types of action potentials; the “common” action potential that propagates through the conduction system, and the cardiac action potential that spreads through the contractile fibres in the atrial and ventricular walls. The cardiac action potential is illustrated in Figure 3. The depolarisation of a contractile fibre causes the fibre to contract, and the contraction lasts until the repolarisation begins [Tortora & Derrickson 2006].

![Figure 3: Schematic of a cardiac action potential](image)
2.3 ELECTROCARDIOGRAPHY

An electrocardiogram (ECG) is a recording of the action potential’s propagation through the heart, registered at the body surface. A schematic of a normal ECG signal is shown in Figure 4. The cardiac cycle starts when the action potential is generated by the sinoatrial node. The P wave represents the depolarisation of the atria, as the action potential propagates to the atrioventricular node. During the PQ interval, atrial systole occurs. The QRS complex represents the action potential’s propagation through the bundle branches and the subsequent depolarisation of the ventricles. Repolarisation of the atria occurs at the same time, but this is masked by the large QRS complex. During the ST interval, ventricular systole occurs. The T wave represents the repolarisation of the ventricles. During the T wave, diastole begins and then lasts until the end of next P wave [Tortora & Derrickson 2006].

![Figure 4: Schematic of an ECG signal](image)

The ECG is usually recorded by electrodes positioned on the arms, legs and chest. The electrode configuration shown in Figure 5, a 12-lead ECG, is the most common. From these ten electrodes, 12 leads are constructed. All leads are different; they all contain the waves mentioned above but the amplitudes and orientation of the waves differ. These signals can be processed and viewed separately, but they can also be linearly combined. In the latter case the signal will resemble the schematic in Figure 4.

![Figure 5: Electrode position for recording of a 12-lead ECG](image)

The standard limb leads I, II and III are constructed by measuring the potential difference between RA and LA, RA and LL, and LA and LL, respectively. The augmented limb leads aVR, aVL and aVF are trickier: they are constructed by measuring the potential difference
between RA, LA or LL and a combination of the other two electrodes. To construct aVR, the potential difference between RA and a combination of LA and LL is measured. For aVL LA and a combination of RA and LL are used, and for aVF LL and a combination of RA and LA are used. The precordial leads V₁-V₆ are constructed by measuring the potential difference between electrode V₁-V₆ and the union of the standard limb leads. The ground electrode can be placed anywhere on the body, but convention has it placed on the right leg [Pahlm & Sörnmo (ed) 2006].

The leads can be grouped according to which part of the heart they view best. The inferior leads have a good view of the lower parts of the heart whereas the left lateral leads view the left side and the anterior leads view the front of the heart [Thaler 2007]. Hence, when looking for electrical activity of the left ventricle the left lateral leads are useful, but when the right ventricle is of interest the anterior leads are more valuable. The grouping is presented in Table 1.

<table>
<thead>
<tr>
<th>Leads</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>V₁, V₂, V₃, V₄</td>
<td>Anterior</td>
</tr>
<tr>
<td>I, aVL, V₅, V₆</td>
<td>Left lateral</td>
</tr>
<tr>
<td>II, III, aVF</td>
<td>Inferior</td>
</tr>
<tr>
<td>aVR</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 1: Grouping of ECG leads

2.4 Phonocardiography

A phonocardiogram (PCG) is a recording of heart sounds. A healthy adult has two distinct heart sounds, S₁ and S₂. The first, S₁, is caused by the closure of the mitral and tricuspid valves, and the second, S₂, is caused by the closure of the aortic and pulmonary valves. Healthy children (and some adults up to 40 years of age) have a third heart sound, S₃. This sound is caused by vibrations in the left ventricular wall, which in turn are caused by rapid filling of the ventricle when the mitral valve opens. Some elderly persons have a fourth heart sound, S₄, but it is subject to debate whether this ought to be considered healthy or not. S₄, like S₃, is caused by vibrations in the left ventricular wall. However, in this case the vibrations arise by the ventricular filling that is due to atrial systole, and by the fact that the ventricular wall is abnormally stiff [Fuster et al (ed) 2004, Tilkian & Conover 2001]. The sounds’ location in the cardiac cycle is shown in Figure 6, and details concerning the sounds’ durations and frequencies can be found in Table 2 [Ahlström 2006].
The four sounds are best heard in different sites. S₁ can easily be heard at the mitral area (M), and S₂ is best heard at the aortic area (A). To hear S₃ and S₄, the patient should lie on her left side and the area of auscultation should be the mitral area (M) [Tilkian & Conover 2001]. See Figure 7 for an illustration of the auscultation areas.
3 Time Frequency Analysis

One common way of analysing signals is by looking at their frequency content. The most well-known method for analysing a signal’s frequency spectrum is the Fourier transform, which maps a function in the time domain unto a function in the frequency domain:

\[ X(j\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} \, dt \]

In applications, the signal is sampled at even-spaced points \( k \) and the discrete Fourier transform is used:

\[ X[n] = \sum_{k=0}^{N-1} x[k] e^{-j \frac{2\pi}{N} kn} \]

This is very useful for stationary signals. However, when analysing non-stationary signals such as the ECG and the PCG the Fourier transform is not very practical. There is often a need of locating the frequency content at a certain time in the signal, and the Fourier transform only shows the frequency content of the signal as a whole. One solution to this problem is to use the short time Fourier transform, in which a sliding window is used for analysing only a short segment of the signal at a time:

\[ X(\omega, k) = \sum_{l} x[l] w[l] e^{-j\omega l} \]

The width \( \gamma \) of the window sets the time and frequency resolution. For good frequency resolution the window should be wide, but for good time resolution it should be narrow. A parallel to Heisenberg’s uncertainty relation is often drawn; that there cannot be good resolution in both time and frequency [Gustafsson et al 2006].

3.1 The Wavelet Transform

A solution to the resolution problem is the wavelet transform, which uses a variable window width. A comparison of the resolution of the STFT and the wavelet transform is shown in Figure 8. The continuous wavelet transform (CWT) resembles the Fourier transform:

\[ W_{\psi}(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^{*} \left( \frac{t - b}{a} \right) \, dt \]

where \( \psi(t) \) is the mother wavelet, \( a \) is the scale (or dilation) and \( b \) is the location (or translation) [Addison 2005]. The wavelets which, using different values of \( a \) and \( b \), are derived from the mother wavelet form the set of basis functions with which the signal is analysed:

\[ \psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left( \frac{t - b}{a} \right) \]
When implementing the CWT, the integral is approximated by a summation. The summation is made for several values of $a$ and $b$, resulting in large computational costs and redundant information [Addison 2005].

The discrete wavelet transform (DWT) has less computational costs and does not generate redundant information. It is computed as an integral over discrete values of $a$ and $b$, where $a$ is discretised logarithmically and $b$ is proportional to $a$:

$$
\psi_{m,n}(t) = \frac{1}{\sqrt{a_0^m}} \psi\left(\frac{t - nb_0a_0^m}{a_0^m}\right)
$$

Hence, the DWT can be expressed as

$$
W_x(m,n) = \int_{-\infty}^{\infty} x(t) \psi_{m,n}(t) \, dt
$$

Perhaps the most common way of discretising $a$ and $b$ are by using a dyadic grid; i.e. substituting $a_0$ with 2 and $b_0$ with 1. Then, the wavelet basis functions are

$$
\psi_{m,n}(t) = 2^{-\frac{n}{2}} \psi(2^{-m} t - n)
$$

These are orthonormal, which implies that they may be inverted and that they reproduce the signal without redundancy [Addison 2005]. From now on, the term wavelet transform will refer to the dyadic DWT if not specified otherwise.

### 3.2 Multi-resolution representation

The values of the wavelet transform, $W_x(m,n)$, are called wavelet (or sometimes detail) coefficients. According to Addison [2005], there are scaling functions similar to the basis functions:

$$
\phi_{m,n}(t) = 2^{-\frac{n}{2}} \phi(2^{-m} t - n)
$$

The scaling functions are used to generate approximation coefficients, which for a certain scale $m$ gives the discrete approximation of the signal at that scale:

$$
C_x(m,n) = \int_{-\infty}^{\infty} x(t) \phi_{m,n}(t) \, dt
$$

To get the continuous approximation at scale $m$, the approximation coefficients are multiplied by the scaling function and summed:
\[ x_m(t) = \sum_{n=-\infty}^{\infty} C_s(m, n) \phi_{m,n}(t) \]

This approximation approaches the original signal for small values of \( m \). Using the same method for the wavelet coefficients and the basis functions gives the signal detail:

\[ d_m(t) = \sum_{n=-\infty}^{\infty} W_s(m, n) \psi_{m,n}(t) \]

Together, the approximation at scale \( m \) and the signal detail up to scale \( m \) represent the original signal:

\[ x(t) = x_m(t) + \sum_{k=0}^{m} d_k(t) \]

This way of representing the signal is known as multiresolution representation [Addison 2005]. The highest scale that is computed is sometimes referred to as the decomposition level, and the approximation and detail coefficients are sometimes referred to as wavelet coefficients.
4 CHARACTERISTICS OF A HEART FAILURE PATIENT

4.1 HEART FAILURE

“Heart failure is a pathophysiological state in which an abnormality of cardiac function is responsible for the failure of the heart to pump blood at a rate commensurate with the requirements of the metabolizing tissues.” [Task Force 2001] This is a rather wide definition, but then heart failure is a rather complex disease which cannot be easily defined. Most often, heart failure is defined by its symptoms, which will be covered in a later section.

Heart failure can be categorised as acute or chronic as well as systolic or diastolic. The term acute heart failure is mainly used for acute pulmonary oedema, and the term chronic heart failure refers to the state described above. If the disease is caused by reduced contractile ability of the left ventricle, which reduces the minute volume, it is called systolic heart failure. In diastolic heart failure the ventricle has preserved contractile ability, but the minute volume is still reduced [Task Force 1995].

This thesis focuses on chronic heart failure. Both systolic and diastolic heart failure are discussed, with a main interest in diastolic heart failure. Systolic heart failure has been studied more frequently [Fonseca 2006], and the term heart failure often implicitly means systolic heart failure in literature as well as in primary care [Skånér 2004]. Yet, studies have shown that 30-50% of heart failure patients suffer from diastolic heart failure [Fonseca 2006]. According to Zile & Brutsaert [2002], most well-known symptoms are more common in systolic than diastolic heart failure. Hence, diastolic heart failure is more difficult to diagnose than systolic heart failure.

4.2 DIAGNOSING HEART FAILURE

In order to suspect heart failure, the patient should show symptoms of heart failure at rest or during exercise, and there should be unbiased proof of heart failure at rest [Task Force 1995]. Some typical symptoms are dyspnoea, fatigue and peripheral oedema. If these criteria are met, the patient should be further investigated. The investigation should incorporate a regular clinical examination, echocardiography, ECG, chest X-rays and laboratory tests. If the patient already receives treatment for heart failure, the effects of the treatment should be established. Once heart failure has been diagnosed, the disease should be classified according to severity [Task Force 2001].

The guidelines above refer mainly to systolic heart failure, as the methods of investigation often do not reveal signs of diastolic heart failure. In 1998, guidelines for diagnosing diastolic heart failure were published [European Study Group 1998]. According to these, three criteria should be met to suspect diastolic heart failure: the patient shows symptoms of heart failure, LVEF is normal and there is proof of diastolic dysfunction. Several criteria for what constitutes diastolic dysfunction are mentioned. To examine the diastolic function according to these criteria, some kind of imaging technique (like echocardiography or heart catheterization) must be used.

The problems with the guidelines are that they need approximately five years to become established, and that physicians often do not adhere to them [Bates et al 2003]. If they do, the methods recommended for investigation are not optimal. Echocardiography is expensive; X-raying should not be performed unless necessary because of the radiation; heart catheterization is invasive. Also, patients with diastolic heart failure but normal diastolic
function will often be misdiagnosed, as diastolic dysfunction is not the only cause of diastolic heart failure [Hogg et al 2004].

According to Skånér [2004], Swedish general practitioners tend to rely on the echocardiographic measurement of LVEF when diagnosing heart failure. If echocardiography has not been performed, chest X-rays are used. The patient’s symptoms are often not taken into consideration, and the ECG is rarely used for diagnostic purposes. This means that heart failure is diagnosed when LVEF is low, even if the patient doesn’t show other signs of heart failure. Consequently, a normal LVEF is used as proof that the patient does not suffer from heart failure.

Many British general practitioners claim that they don’t have the time to perform a full investigation. Some of them would like to use echocardiography whenever they suspect heart failure, but the waiting lists are too long. Instead, they rely on chest X-rays and the patient’s reaction to diuretics, which are rather inexact methods [Khunti et al 2002]. However, if they had ready access to echocardiography chances are they would use the method in the same way Swedish physicians do.

Another problem when diagnosing heart failure is that the normal aging process might give symptoms similar to heart failure. On the other hand, the aging process might contribute to the development of heart failure. Also, several heart diseases might give symptoms similar to heart failure, as well as coexist with or cause heart failure. It is also considered more difficult to diagnose heart failure in women and obese patients [Fonseca 2006, Skånér 2004]. Diastolic heart failure with normal diastolic function seems to be more common in women [Hogg et al 2004, Regitz-Zagrosek et al 2007], and as of now, there are no guidelines for dealing with that.

4.3 COMMON SIGNS AND SYMPTOMS

The most well-known symptoms of heart failure are fatigue, dyspnoea, peripheral oedema, tachycardia, sudden weight gain and peripheral coldness. All of these symptoms are somehow related to the reduced minute volume [Andersson 2000, Task Force 1995].

The reduced minute volume increases the activity of the sympathetic nervous system. This, coupled with decreased activity of the parasympathetic nervous system, reduces the heart rate variability [Andersson 2000, Fuster et al (ed) 2004, Hombach 2002]. This is very often found in heart failure patients. Another common clinical finding is raised jugular venous pressure [Andersson 2000, Task Force 1995].

Heart failure patients often have arrhythmias. When the arrhythmia causes shortened diastole it contributes to the heart failure since the coronary arteries do not receive enough blood. According to Varela-Roman et al [2005] only 59% of systolic and 46% of diastolic heart failure patients have sinus rhythm. Especially premature ventricular contractions (PVC’s) are more abundant in heart failure patients than in healthy subjects [Davey et al 1994]. Hombach [2002] claims that over 90% of heart failure patients have PVC’s.

Left ventricular hypertrophy (LVH), a condition in which left ventricular mass is increased, is common in heart failure patients. There are two kinds of hypertrophy; concentric and eccentric. In the concentric kind the myocardium thickens, whereas in the eccentric kind the myocardium gets thinner and the chambers dilate. They are closely related, and one kind of hypertrophy may evolve into the other [Andersson 2000]. However, eccentric hypertrophy is mainly associated with systolic heart failure and concentric hypertrophy with diastolic heart failure [Regitz-Sagrosek et al 2007, Varela-Roman et al 2005].
Factors that influence the development of left ventricular hypertrophy are hypertension, diabetes, obesity and aging [Fuster et al (ed) 2004, Havranek et al 2002, Regitz-Zagrosek et al 2007]. Since hypertrophy is closely connected to heart failure, the same factors are often found in heart failure patients. Hypertrophy may affect the other chambers as well as the left ventricle. Hypertrophy in one part of the myocardium may induce hypertrophy in a different part, and as a whole, myocardial hypertrophy is associated with heart failure.

Myocardial ischemia is very common in heart failure patients, especially in those with systolic heart failure [Andersson 2000, Fuster et al (ed) 2004, Persson 2003]. Approximately 40% of diastolic and 55% of systolic heart failure patients show signs of ischemia or have a history of myocardial infarction [Varela-Roman et al 2005, Hawkins et al 2007, Hogg et al 2004]. The number of heart failure patients with myocardial ischemia decreases with increasing age, possibly because of the high mortality associated with myocardial ischemia [Havranek et al 2002].

Something that has recently become established is the measurement of brain natriuretic peptide (BNP). Several studies prove that the blood content of BNP increases markedly in heart failure [Andersson 2000, Krüger et al 2004].

4.4 SIGNS AND SYMPTOMS PRESENT IN SYSTOLIC HEART FAILURE

The third heart sound is often present in patients with systolic heart failure [Andersson 2000, Fonseca 2006, Fuster et al (ed) 2004, Hawkins et al 2007, Hogg et al 2004, Joshi 1999, Varela-Roman et al 2005]. In heart failure, the left ventricular wall is often stiff because of fibrosis, ischemia and hypertrophy. Therefore it vibrates when blood enters rapidly. The number of heart failure patients with a third heart sound differs greatly between studies; numbers ranging from 10% to 57% have been reported. A reason for this diversity might be that the sound is often hard to hear. This theory is supported by the fact that the highest number, 57%, was reported by Hult [2002] who detected the sound automatically.

According to Špinarová [2003], the third heart sound is heard more often in patients who also have left bundle branch block (LBBB) (35% vs. 22%). LBBB is part of a vicious circle which leads to heart failure: fibrosis and apoptosis affect the conduction system, resulting in LBBB. The block desynchronises the electrical activation of the left ventricle, and so the left ventricle is forced to remodel and become hypertrophic. This, in turn, affects the conduction system and contributes to heart failure. Approximately 25% of systolic heart failure patients have LBBB [Hombach 2002, Krüger et al 2004, Špinarová 2003, Zannad et al 2007]. In one study, 28% of patients with LBBB developed heart failure whereas only a few percent of patients without LBBB did [Zannad et al 2007].

Some prognostic markers of arrhythmia are associated with systolic heart failure. Studies concerning these markers and diastolic heart failure do not seem to have been performed. Such markers are ventricular late potentials (VLP’s) [Galinier et al 1996, Hombach 2002], T wave alternans (TWA) [Sarzi-Braga et al 2004] and the P wave duration [Dixen et al 2003]. The VLP’s and the TWA can be used to predict ventricular arrhythmias, and the P wave duration may predict atrial fibrillation. In the study by Sarzi-Braga et al [2004], 52% of the heart failure patients had TWA. However, there was no control group for comparison. A prolonged P wave duration has been proven linked to an enlargement of the left atrium [Dixen et al 2004], which is common in heart failure.
4.5 Signs and Symptoms Present in Diastolic Heart Failure

In general, diastolic heart failure patients are elderly women. In a study by Hogg et al [2004], the average ages of diastolic and systolic heart failure patients are 78 and 74, respectively. 69% and 41% of these are women. In another study, the ages are 72 and 67, with 51% and 31% respectively being women [Varela-Roman et al 2005]. A study concerning only patients over 65 years of age shows that the percentage of women increases with age, as does the average LVEF [Havranek et al 2002]. This last fact sounds strange, but then the mortality is rather high in patients with a low LVEF.

The fourth heart sound might be heard in patients with diastolic heart failure, where it is often the result of left ventricular hypertrophy. The hypertrophic myocardium relaxes abnormally slow, which makes the diastolic filling a very slow process. As a consequence the filling due to atrial systole becomes faster, inducing a fourth heart sound [Tilkian & Conover 2001].

The fourth heart sound is never present together with atrial fibrillation, as in this condition there is no atrial systole [Fuster et al (ed) 2004]. Yet, atrial fibrillation is common in heart failure patients. It is subject to debate whether atrial fibrillation causes heart failure or the other way around. In any case, the filling of the left ventricle is affected negatively by the arrhythmia and blood might flow back into the atrium instead of entering the aorta [Naccarelli et al 2003]. Atrial fibrillation is present in 23-38% of diastolic heart failure patients, and the percentage rises with older age [Havranek et al 2002, Hogg et al 2004, Varela-Roman et al 2005].

Heart valve disorders, such as aortic and mitral stenosis, lead to hypertrophy of the atria and ventricles [Persson 2003]. Therefore, they also lead to heart failure. According to Varela-Roman et al [2005], 33% of diastolic and 13% of systolic heart failure patients have a heart valve disorder.

4.6 Summary of Parameters

This thesis concerns a clinical decision support system that will support a physician in diagnosing heart failure. The system should not be overly complicated to handle, and it should work in real time. In order to make an accurate diagnosis, several parameters should be used as input to the system. The list below is based on the information from the earlier sections. However, a few parameters have been excluded.

A BNP sample has to be sent to a laboratory, and so this measure cannot be used in a real time application. Of possible arrhythmias, only PVC’s and atrial fibrillation are considered since they have been mentioned specifically in the literature. Heart valve disorders have been excluded since there are plenty of them, and they require the PCG to be studied extensively. This is considered too time-consuming to fit in the scope of this thesis. However, in further studies heart valve disorders should be considered.

Also, several parameters that have not been mentioned are missing from the list. They are the parameters that should not be present in a heart failure patient, i.e. the ones that rather suggest another diagnosis. This, too, is too time-consuming for this thesis, but it ought to be considered in future studies. The parameters that will be analysed in this text are:

- Age
- Gender
- Body mass index
- Blood pressure
- Diabetes
- Dyspnoea
- Fatigue
- Peripheral oedema
- Peripheral coldness
- Raised jugular venous pressure
- Sudden weight gain
- Tachycardia
- Heart rate variability
- Premature ventricular contractions
- Atrial fibrillation
- Left ventricular hypertrophy
- Myocardial ischemia
- Left bundle branch block
- Ventricular late potentials
- T wave alternans
- P wave duration
- Third heart sound
- Fourth heart sound

The first eleven parameters should be fed to the system manually, the next ten should be extracted from the ECG signal and the last two should be extracted from the PCG signal. Some parameters should be binary (yes or no; is the parameter present or not), the others should be expressed as real values.
5 EXTRACTING VISIBLE FEATURES

Some of the parameters that characterise a heart failure patient can be extracted from the ECG and PCG signals. In this chapter, and the following three, the most common methods of feature extraction are reviewed. This chapter mainly concerns the computation of amplitudes and durations and resembles the visual inspection of the ECG made by physicians.

5.1 PRE-PRE-PROCESSING OF THE ECG SIGNAL

The most important procedure when processing the ECG signal is the beat detection. When a heartbeat has been detected and significant points (such as the end of the QRS complex or the start of the P wave) have been marked, a lot of information can be gathered from the signal. When ECG and PCG signals are measured simultaneously, the ECG beat detection is useful when evaluating the PCG signal as well. There are several types of beat detectors, most of which begin by detecting the QRS complex.

To detect significant points correctly, it is important that the noise level is low. There are several sources of noise such as line voltage, muscle activity and bad contact between the skin and the electrodes. Sometimes conventional filtering is enough, but when analysing very low amplitudes signal averaging should be used. Another option is using wavelets, since they have noise-reducing abilities. For more information about basic pre-processing, refer to Pahlm & Sörnmo (ed) [2006].

5.2 LEFT VENTRICULAR HYPERTROPHY

To diagnose LVH, several criteria for the amplitudes of R and/or S waves in certain leads have been suggested [Fuster et al (ed) 2004]. However, Bacharova & Kyselovic [2001] declare that these criteria should not be used, since they are based on the assumption that the hypertrophic myocardium has the same electrical properties as the healthy myocardium. It is easy to realise that with the increased mass and the reorganisation of the myocardial cells, the electrical properties probably change. Also, Bacharova & Kyselovic [2001] claim that physicians have ceased to use the criteria for diagnostic purposes since their sensitivities are too low. According to Fuster et al (ed) [2004], none of the criteria have a sensitivity greater than 78%.

Still, the criteria are subject to several studies. The effects of gender and obesity have been studied, and it seems that overweight patients generally have lower QRS amplitudes than patients with normal weight [Okin et al 1996]. In a comparison between the performances of four different criteria, the researchers found that no criteria showed better performance for both men and women [Alfakih et al 2004]. They found that for men the Cornell criterion should be used, whereas for women the Solokow-Lyon product was better.

The Cornell criterion states that for LVH to be diagnosed in males, the sum of the R wave amplitude in lead aVL and the S wave amplitude in lead V3 should exceed 2.8 mV:

\[ RaVL + SV_3 \geq 2.8 mV \]

For women, the same sum should exceed 2.0 mV. The Solokow-Lyon product is calculated as

\[ (SV_1 + RV_5 or RV_6) \times QRSduration \geq 2940 mVms \]

[Alfakih et al 2004].

According to Thaler [2007], the most useful criteria are
\[ (RV_5 or RV_6) + (SV_1 or SV_2) > 3.5 mV \]

and

\[ RaVL > 1.3 mV. \]

Also, a depressed ST segment and/or an inverted T-wave are said to be signs of LVH. In a comparison between several criteria and their voltage-duration products (i.e. products of the criteria and the QRS duration), the voltage-duration products were found to have greater sensitivity than the criteria alone [Okin et al 1995]. This supports the assumption that the electrical properties of the hypertrophic and the healthy myocardium differ.

QT dispersion (i.e. the difference between the longest and the shortest QT intervals) has been shown to correlate with LVH [Davey et al 1994]. A later study shows that the QT dispersion does not have better sensitivity than the Cornell criterion, but that using both QT dispersion and an amplitude criterion improves the diagnosis [Salles et al 2005]. However, it is quite hard to tell where the T wave ends and get a correct QT measurement, and so this method might not be useful in a clinical situation. According to Pahlm & Sörnmo (ed) [2006], QT dispersion measurements suffer from low reproducibility and are therefore unreliable.

### 5.3 LEFT BUNDLE BRANCH BLOCK

A bundle branch block causes the depolarisation of the myocardium to slow down. Hence the most common way to diagnose it is to measure the QRS duration (QRSD). The value which the duration should exceed differs between authors; \( QRSD \geq 0.11s \) and \( QRSD \geq 0.12s \) are both used [Fuster et al (ed) 2004, Thaler 2007]. This measurement is easy to make, but there is a problem: it holds for both left, right and complete (i.e. both types) bundle branch blocks, and only LBBB is associated with heart failure.

One way of separating the different blocks is by measuring the width of the R wave in certain leads. A wide R wave (exceeding 60 ms) in the left lateral leads is a sign of left but not right bundle branch block [Fuster et al (ed) 2004, Rautaharju et al 1998, Thaler 2007].

In a study by Desai et al [2006], approximately 16% of patients with \( QRSD \geq 0.11s \) but without any bundle branch block had LVH. This is consistent with the results of Okin et al [1995], which state that the ORSD is important in diagnosing LVH. Also, LBBB seems to be connected with increased left ventricular mass [Dhingra et al 2005], as is LVH.

### 5.4 MYOCARDIAL ISCHEMIA

There are different types and stages of myocardial ischemia. To simplify matters for the purposes of this thesis, myocardial ischemia is divided into two kinds: with or without previous myocardial infarction. The reason for this is that a previous infarction can show in the ECG in several ways, or it might not show at all [Fuster et al (ed) 2004]. However, an infarction is rather hard to miss and will be filed into the patient’s medical history. Therefore, an input parameter to the decision support system should be a simple question: has the patient had a myocardial infarction?

When myocardial ischemia is present although the patient has not had an infarction, the most useful signs are changes in the ST segment and/or T wave [Persson 2003]. Papaloukas et al [2000] have developed a set of rules to decide if a heartbeat is ischemic or not. These rules are similar to those of others [Persson 2003, Rautaharju et al 1998]. A beat is considered ischemic if one of the following apply:

- \( J_{80} \leq -0.08mV \) and angle \( \geq 65° \)
• \( J^{80} \geq 0.08mV \)
• T wave is inverted

The point where the QRS complex ends is sometimes referred to as the J point, and so J80 means the point situated 80 ms after the J point. The angle is calculated between a line perpendicular to the isoelectric line and the slope of the ST segment. See Papaloukas et al [2000] for details. In order to know whether the T wave is inverted or not, the normal direction must be known. Therefore, the first two rules are more useful.

5.5 ARRHYTHMIAS

Detecting and classifying arrhythmias by visual inspection is not too difficult. To correctly classify arrhythmias automatically seems trickier, since one arrhythmia may manifest in different, individual ways.

5.5.1 ATRIAL FIBRILLATION

The distinguishing thing about atrial fibrillation is its irregularly irregular ventricular rhythm, which usually varies between 120 and 180 beats per minute. Another characteristic is the absence of P waves [Thaler 2007]. Usually, atrial fibrillation is analysed in regard to fibrillation frequency. One common method is average beat subtraction [Pahlm & Sörnmo (ed) 2006]. In this method, several QRST complexes are averaged and the result is subtracted from the beat to be analysed. The remaining signal contains individual fibrillation features and may be subject to spectral analysis.

5.5.2 PREMATURE VENTRICULAR CONTRACTIONS

PVC’s are characterised by their wide and weird looking QRS complexes, and the absence of P waves. There is often a delay before the initiation of the following beat. PVC’s appear by themselves among normal beats, or they are part of a pattern. Two common patterns are bigeminy, where there is one normal beat between each PVC, and trigeminy, where there are two normal beats between each PVC [Thaler 2007].
6 EXTRACTING INVISIBLE FEATURES

Heart rate variability and high frequency, low amplitude parameters cannot be found by a visual inspection. Still, they can quite easily be found by the methods explained in this chapter. QRS detection and filtering, covered in chapter 5, are crucial for the performance of the algorithms. In particular, for heart rate variability analysis all ectopic beats must be removed from the signal.

6.1 HEART RATE VARIABILITY ANALYSIS

The autonomic nervous system consists of two parts, the sympathetic and the parasympathetic nervous systems. In healthy individuals at rest, the parasympathetic nervous system dominates. When the healthy individual becomes active, the sympathetic nervous system takes over. However, in a heart failure patient, the sympathetic nervous system is activated even at rest, and this affects the heart rate variability [Andersson 2000, Tortora & Derrickson 2006].

The variability is considered in terms of frequency of the variations. There are three regions of interest; the high frequency (HF) region of 0.15-0.4 Hz, the low frequency (LF) region of 0.04-0.15 Hz and the very low frequency (VLF) region of 0.003-0.04 Hz. The HF variations are caused by breathing and the LF variations by pressure variations in the peripheral blood vessels. In both cases, pressure variations affect the baroreflex, which influences the parasympathetic modulation of the heart rate. In persons with heart failure, the HF and LF components are markedly decreased. The VLF variations are under dispute, but they might be associated with the body’s heat regulation [Francis et al 2002, Pahlm & Sörnmo (ed) 2006].

6.1.1 SPECTRAL ANALYSIS

One established way of analysing the heart rate variability is by computing the signal’s spectral density. This can be done either by using the Fourier transform, or by using autoregressive modelling.

When using the Fourier transform, the spectrum will look very noisy. This is due to the fact that the spectrum’s standard deviation is of the same size as the spectrum itself. To overcome this, a smoothing method such as Welch’s or Blackman-Tukey’s must be applied [Gustafsson et al 2006]. For this application, the frequency resolution must be 0.003 Hz or smaller. Since the frequency resolution depends on the length of the signal as resolution = 1/length [Gustafsson et al 2006], the signal must be at least 5 minutes and 34 seconds long.

An autoregressive model is a reproduction of the signal’s spectral density. How well this is reproduced depends on the order of the model. It is very important that the best possible order is used. If the order is too low, the spectrum will not be sufficiently detailed. If on the other hand the order is too high, false peaks might appear. To decide the optimal order, Akaike’s information criterion may be used [Pahlm & Sörnmo (ed) 2006].

6.1.2 DETRENDED FLUCTUATION ANALYSIS

Detrended fluctuation analysis (DFA) is a method for evaluating heart rate variability. The purpose is to compute fractal scaling exponents, \( \alpha_1 \) and \( \alpha_2 \), which describe fractal-like measures of the signal. These exponents, particularly \( \alpha_1 \), have been used to separate heart failure patients from healthy persons with good results [Penzel et al 2003, Willson et al 2002].
After pre-processing, the process of DFA can be divided into four steps [Penzel et al 2003]:

i. Integrate the \( N \) sample long signal \( x(k) \): 
\[
X(i) = \sum_{k=1}^{i} (x(k) - <x>), \quad i = 1, \ldots, N.
\]

ii. Divide the integrated signal \( X(i) \) into \( N_n \) segments, each of length \( n \). \( N_n = \text{int}(N/n) \).
When \( N/n \) is not an integer, a short segment at the end of the signal will be missed. To compensate for this, divide the signal into \( N_n \) segments starting at the end. There should be a total of \( 2N_n \) segments.

iii. Compute the variance for each segment \( s \), \( s = 1, \ldots, N_n \), by
\[
F_n^2(s) \equiv \frac{1}{n} \sum_{i=1}^{n} ((X((s-1)n + i) - p_s(i))^2, \quad \text{where } p_s \text{ is a polynomial approximating the trend in the segment.}
\]

iv. Compute the fluctuation \( F(n) \) by taking the square root of the averaged segment variances: 
\[
F(n) \equiv \sqrt{\frac{1}{2N_n} \sum_{s=1}^{2N_n} F_n^2(s)}.
\]

Repeat ii.-iv. for different values of \( n \). The fluctuation \( F(n) \) may be plotted against \( n \) in a log-log graph, as Figure 9 illustrates, and the slope of the curve is the fractal scaling exponent. The relation between \( F(n) \) and \( n \) is \( F(n) \propto n^\alpha \), hence the name. Usually, two different exponents are calculated. Then, \( \alpha_1 \) covers \( 4 \leq n \leq 16 \) and \( \alpha_2 \) covers \( 16 \leq n \leq 64 \) [Francis et al 2002].

![Figure 9: The fractal scaling exponent](image)

Willson et al [2002] show how the fractal scaling exponents can be calculated using frequency-weighted spectral analysis, which is a mathematical equivalent of DFA. Based on this, Francis et al [2002] show that
\[
\alpha_1 = 2 \frac{LFW}{LFW + HF_W}
\]
and
\[
\alpha_2 = 2 \frac{VLF_w}{VLF_w + LF_w},
\]
where \( w \) is for weighted frequency. These expressions may be approximated by similar ones where the frequencies are not weighted:

\[
\alpha_1 = 2 \frac{LF}{LF + HF} = 2 \frac{1}{1 + HF/LF}
\]

and

\[
\alpha_2 = 2 \frac{VLF}{VLF + LF} = 2 \frac{1}{1 + LF/VLF}
\]

From these expressions, the physical meanings of the fractal scaling exponents may be interpreted.

Normal values for the exponents are \( \alpha_1 = 1.20 \pm 0.18 \) and \( \alpha_2 = 1.00 \pm 0.12 \). In heart failure patients, the values are \( \alpha_1 = 0.80 \pm 0.26 \) and \( \alpha_2 = 1.13 \pm 0.22 \) [Penzel et al 2003]. The decrease in \( \alpha_1 \) relates to a decrease in \( LF/HF \), and the increase in \( \alpha_2 \) relates to an increase in \( VLF/LF \). This corresponds with the presumption that the LF component is reduced in heart failure patients.

In an analysis comparing the fractal scaling exponents and the approximations with un-weighted frequencies, correlations were found to be 0.94 between \( \alpha_1 \) and its approximation and 0.86 between \( \alpha_2 \) and its approximation [Francis et al 2002].

### 6.1.3 MULTiresolution WAVELET ANALYSIS

A similar method is multi-resolution wavelet analysis (MWA), which computes a scaling exponent using wavelet coefficients [Thurner et al 1998]. For a simple comparison, the MWA method can be explained in four steps:

i. Transform the signal \( w(i) \) using the WT: 

\[
W_{m,n}(i) = 2^{-m/2} \sum_{i=0}^{M-1} w(i) \psi(2^{-m} i - n),
\]

where \( m \) is the scaling variable, \( n \) is the translation variable and \( \psi \) is the mother wavelet.

ii. Now there are \( N_m \) wavelet coefficients at scale \( m \). \( N_m = \text{int}(M/2^m) \).

iii. Compute the variance of the wavelet coefficients:

\[
\sigma^2(m) \equiv \frac{1}{N_m - 1} \sum_{i=0}^{N_m-1} (W_{m,n}(i) - <W_{m,n}(i)>)^2.
\]

iv. Compute the standard deviation of the wavelet coefficients:

\[
\sigma(m) \equiv \sqrt{\sigma^2(m)}.
\]

Repeat ii.-iv. for different scales \( m \). The standard deviation \( \sigma(m) \) may be plotted against \( m \) in a lin-log graph, and the slope of the curve is the scaling exponent \( \alpha \) [Thurner et al 1998].

In healthy persons, \( \alpha = 1.40 \pm 0.37 \) when \( 1 \leq m \leq 3 \) and \( \alpha = 1.22 \pm 0.11 \) when \( 3 < m \leq 10 \). In heart failure patients, \( \alpha = 0.26 \pm 0.6 \) when \( 1 \leq m \leq 3 \) and \( \alpha = 1.57 \pm 0.17 \) when \( 3 < m \leq 10 \) [Thurner et al 1998].

### 6.2 High frequency, low amplitude parameters

Parameters of high frequency (50-200 Hz) and low amplitude (µV) are often hidden in high frequency noise, since the amplitude of the noise exceeds that of the parameters. To extract the parameters, the noise level must be reduced. This is done by averaging heartbeats, since
the noise is considered uncorrelated and therefore averaging equals out the noise. The parameters of interest will not be affected since they are correlated. To ensure quality in the averaging, a beat is chosen for comparison and all subsequent beats are compared to it. Only beats with a cross correlation coefficient greater than 0.97 are used for averaging. To sufficiently reduce the noise level, 200-300 beats should be averaged [Pahlm & Sörnmo (ed) 2006].

Since only frequencies over 50 Hz are of interest, a high pass filter with a cut-off frequency of 40 Hz should be used. A low pass filter with cut-off frequency 250 Hz can be used as well. However, the most important aspect when analysing high frequencies is that the sampling frequency is sufficiently high. To avoid aliasing, the sampling frequency must be at least twice the frequency of the signal [Gustafsson et al 2006]. In this case, the sampling frequency should be at least 500 Hz.

By convention, vectorcardiograms (VCG’s) are used to extract high frequency, low amplitude parameters. However, there is nothing in the averaging method that requires the use of VCG’s, so standard ECG’s may be used just as well [Pahlm & Sörnmo (ed) 2006].

6.2.1 VENTRICULAR LATE POTENTIALS

VLP’s are found in the latter part of the QRS complex and in earlier part of the ST segment. Three standardised measurements are used to detect VLP’s [Pahlm & Sörnmo (ed) 2006]:

- The filtered QRS complex duration (FQRSD)
- The root mean square amplitude during the last 40 ms of the filtered QRS complex (RMS40)
- The duration of the latter part of the QRS complex in which the amplitude does not exceed 40 µV (LAS40)

There are three criteria for VLP’s: $FQRSD > 114ms$, $RMS40 < 20\mu V$ and $LAS40 > 38ms$. VLP’s are thought to be present if at least two of the criteria are fulfilled. In a patient with LBBB, the first criterion is fulfilled even if VLP’s are not present. Hence, this analysis is not reliable in patients with LBBB [Pahlm & Sörnmo (ed) 2006].

6.2.2 P WAVE DURATION

The P wave duration analysis is very straight-forward: the beginning and end of the P wave are located and the duration is computed. The reason why this requires signal averaging is simply that the P wave has low amplitude, and the start and end points are hard to pinpoint in the high frequent noise. A healthy person’s P wave typically has a duration of 120 ms [Pahlm & Sörnmo (ed) 2006]. According to Dixen et al [2003], a duration exceeding 149 ms is associated with increased mortality in heart failure patients.

6.3 T WAVE ALTERNANS

In TWA, the morphology and amplitude of the T wave change in every other beat. These changes are so small that they cannot be detected by visual inspection of the ECG. Traditionally TWA has been evaluated by spectral analysis, but that requires an exercise ECG or a paced rhythm. [Pahlm & Sörnmo (ed) 2006]. This method is not suitable for the application considered in this thesis.

Another method for evaluating TWA is the modified moving average method (MMA). To use the MMA algorithm, all ectopic beats must be removed and noise should be reduced. The algorithm then classifies the beats: the first beat is assigned to group A, the second to group
B, the third to group A and so on. Within the groups, the beats are ordered according to increasing amplitude. When the classifying and ordering is finished, the maximum absolute difference between the groups is calculated. This value constitutes the TWA [Nearing & Verrier 2002].
7 Extracting Features with Wavelets

The wavelet transform has found innumerable applications in the analysis of the ECG and the PCG [Addison 2005]. This chapter explains two ways of using wavelets to extract features that can be used for classification of the signals.

7.1 Using Wavelet Coefficients

One common method of extracting features from the ECG is by using the signal’s wavelet coefficients. Then, the selection of mother wavelet and decomposition level is highly important. In several studies [Ceylan & Özbay 2007, Güler & Übeyli 2005a, Güler & Übeyli 2005b, Übeyli 2007, Ölmez & Dokur 2003b] a Daubechies wavelet of order 2 is used due to its smoothing qualities, which makes it appropriate for analysing the ECG. However, when developing new algorithms several wavelets should be tested and the one that is best suited for the application should be selected. The level of decomposition should be selected so that important features in the signal are represented by the wavelet coefficients. When using a Daubechies wavelet of order 2, decomposition level 4 appears to be appropriate [Güler & Übeyli 2005a, Güler & Übeyli 2005b, Übeyli 2007, Ölmez & Dokur 2003b].

All wavelet coefficients up to the selected decomposition level (i.e. all detail coefficients up to scale $m$ and the approximation coefficients at scale $m$ when $m$ is the selected decomposition level) may be used as features. However, one purpose of extracting features from the signal is to reduce the computational cost of classifying the signal, and using all coefficients up to a certain scale doesn’t really fill this purpose. Also, the classification accuracy increases when the number of features are reduced [Übeyli 2007]. In the study by Übeyli [2007], the classification accuracy when using 265 wavelet coefficients as features is compared to the accuracy when 20 statistic features (derived from the coefficients) are used. When using 265 features the classification accuracy is 95.56%, and when using only 20 features it is 98.61%.

The statistical features derived by Übeyli [2007] are

- Maximum of the wavelet coefficients at every sublevel
- Mean of the wavelet coefficients at every sublevel
- Minimum of the wavelet coefficients at every sublevel
- Standard deviation of the wavelet coefficients at every sublevel

The sublevels used are 1-4 for detail coefficients and level 4 for approximation coefficients. Other statistical features, such as average power, are also used with good results [Güler & Übeyli 2005a, Güler & Übeyli 2005b].

Ölmez & Dokur [2003b] use dynamic programming and divergence analysis to extract the eight best features from the wavelet coefficients. In their study, the best features can be found among the approximation coefficients. For an explanation of divergence analysis, refer to Cohen [1986].

Wavelet coefficients may be used for extracting features from the PCG as well. When analysing the signal with respect to $S_1$ and $S_2$, Ölmez & Dokur [2003a] use a Daubechies wavelet of order 2. The signal is represented by detail coefficients at decomposition level 2, and every beat is divided into 32 windows of equal length. The signal power is computed for each window, and these 32 values are the features. Divergence analysis is used to select 16 of the 32 features, to reduce computational costs. The authors found that dividing the signal into
more than 32 windows increased the classification accuracy, but also increased computational costs.

### 7.2 Using a Tailored Wavelet

The third heart sound, S₃, has been analysed using a tailored mother wavelet; the impulse response of a sixth order Bessel filter which resembles S₃ [Hult 2002]. Since S₃ mainly has a frequency range of 15-60 Hz [Ahlström 2006], scalings that represent 17, 35 and 60 Hz are used for analysis. A scaling that represents 160 Hz is also analysed, mostly for comparison.

The signal is wavelet transformed using the scalings mentioned above. S₂ is localised by measuring the amplitude of the 60 Hz scaling in a time frame following the R peak (PCG and ECG are measured simultaneously, for easier and more accurate location of the heart sounds). When S₂ is located, S₃ is sought for in a time frame following S₂. In the search for S₃, the signal energies are analysed for all four scalings. The energies at scalings 17, 35 and 60 Hz are compared to the energies at 160 Hz, and if they are large enough an S₃ is located.*

This method could be used to locate S₄ as well, but in that case a different mother wavelet should be created and different scalings should be applied.

* Personal conversation with Niklas Gustafsson. Work not yet published.
One way of reducing the dimension and/or extract features from a data set is by principal components analysis (PCA), which is also known as the Karhunen-Loève transformation. The principle of PCA is to project a data set onto a linear space of lower dimension which is called the principal subspace, so that the variance of the projected data is maximised [Bishop 2006]. This is illustrated in Figure 10.

![Figure 10: The principle of PCA](image)

To find the principal components, the m-dimensional vector $\overline{X}$ is projected onto the principal subspace. The possible directions of the principal subspace are defined by the m-dimensional unit vectors $q_j$. The $m$ possible projections $A_m$ are then defined by

$$A_j = X^T q_j = q_j^T X, j = 1, ..., m.$$  

$\overline{X}$ is assumed to have zero mean, whereby $A_j$ has zero mean. The variance, $\sigma_j^2$, of $A_j$ is then

$$\sigma_j^2 = E(A_j^2) = E(q_j^T \overline{X}^T q_j) = q_j^T E(\overline{X} \overline{X}^T) q_j = q_j^T \overline{R} q_j = \psi(q_j),$$

where $\overline{R}$ is the m-by-m covariance matrix of $\overline{X}$. Now, the vectors $q_j$ for which $\psi(q_j)$ has stationary points are sought. These vectors are found through the eigenvalue problem

$$\overline{R} q_j = \lambda_j q_j,$$

where $\lambda_j$ are eigenvalues and $q_j$ are eigenvectors of $\overline{R}$ [Haykin 1999].

Since $q_j$ is a unit vector, $q_j^T q_j = 1$ and the variance is equal to the eigenvalues:

$$\sigma_j^2 = q_j^T \overline{R} q_j = q_j^T \lambda_j q_j = \lambda_j.$$

The eigenvectors are arranged corresponding to the eigenvalues, which are ordered in decreasing order. The eigenvector associated with the largest eigenvalue is called the first
principal component, the eigenvector associated with the next largest eigenvalue is the second principal component and so on [Bishop 2006]. The eigenvalues are assumed to decrease rapidly [Haykin 1999], so that only the first one or the first few are large enough to choose for feature extraction. The rest of the eigenvalues should be very small, approaching zero. A matrix containing all principal components to be used for feature extraction is called a feature vector.

Papaloukas et al [2002] used PCA to reduce the number of input features from 100 to four, thereby retaining 95% of the variance of the training data set. The results from the study, for which the aim was to find ischemic beats, were good with a sensitivity of 90%. Papadimitriou et al [2001] also used PCA in a study on ischemic beat classification, where a sensitivity of 83% was reached. In this study, the first five principal components, containing nearly 98% of the signal energy, were used instead of the 100 samples in the signal. It should be noted that for the classification of ischemic beats, the reported sensitivities are good when compared to other studies [Begg et al (ed) 2006]. In a study where PCA was compared to no feature reduction, wavelet theory, and a clustering algorithm, PCA was the fastest method. With respect to test error, PCA was as good as the other methods [Ceylan & Özbay 2007].
9 NEURAL NETWORKS

There are several ways of constructing decision support systems (DSS’s), which all have a common foundation and are often connected in many ways. DSS’s are based on Bayesian probability theory, in which knowledge of the problem is included in probability calculations. Bishop [2006] explains that Bayesian probability computations “provide a quantification of uncertainty”.

One category of DSS’s that has found much use in medical applications is neural networks (NN). An NN is constructed to resemble the brain, which with neurons and synapses works like a nonlinear, parallel computer. An NN is able to learn, and to store knowledge in synaptic weights. It is also capable of generalisation, i.e. classifying input data that has not been encountered before. Due to the vast number of nodes (neurons) and the connections between them, an NN is not easily damaged. Several nodes must be damaged for the performance of the NN to be reduced [Haykin 1999].

An NN consists of three or more layers of nodes: the first layer is called the input layer, then there is at least one hidden layer and at the end there is an output layer. The number of hidden layers depends on which method is used, and how complex the computations should be. See Figure 11 for a schematic illustration. Input data is fed to the input nodes; there is one node for every feature. At the connections between the input layer and the first hidden layer, the data is processed. The number of nodes in the hidden layers governs the complexity of the NN’s computations and can either be set in advance or acquired during the training process. At the connections between hidden layers, and between the last hidden layer and the output layer, the data is processed again. When the data has reached the output layer, a classification has been made. Generally, the output layer consists of only one node, but a different number of output nodes may be used as well [Bishop 2006].

NN’s are trained to make correct classifications. There are two ways of training a network; by supervised or unsupervised training. Chapter 9 through 11 focus on supervised training, while chapter 12 deals with unsupervised training. Supervised training simply means presenting a sufficient amount of input data together with the corresponding target data (i.e. the intended result) to the network. This way, the NN will learn to generalise and be able to classify data that does not have a known target. The amount of training data is of great importance. It is intuitive that too little training makes a bad classifier, but too much training causes something...
just as bad; overfitting [Bishop 2006]. This is illustrated in Figure 12. In unsupervised training, data is presented to the NN which, for instance, groups the data in clusters.

![Figure 12: Overfitting a model](Image)

There are several training methods, of which two of the most common are covered in the following sections: the multilayer perceptron (MLP) and the radial-basis function (RBF) network.

### 9.1 MULTILAYER PERCEPTRONS

As the name suggests, an MLP may have any number of hidden layers. In the MLP, every hidden layer has an activation function, and the output layer has an output activation function. All functions are nonlinear. Usually, sigmoid (i.e. S shaped) functions are used [Haykin 1999]. The two that are used most often are the logistic sigmoid function:

\[
\sigma(a) = \frac{1}{1 + e^{-a}}
\]

and the hyperbolic tangent function:

\[
\sigma(a) = \tanh(a).
\]

The activation functions give activations (outputs) of the nodes:

\[
z_j = \sigma(a_j)
\]

where \(a_j, j=1,\ldots,M\), is a sum of inputs \(z_i, i=1,\ldots,D\), somewhere in the network (see Figure 13):

\[
a_j = \sum_i w_{ji} z_i.
\]

![Figure 13: Nodes and weights](Image)

Every connection between two nodes has a weight parameter, \(w\), attached to it. As the training proceeds, the weights are updated. Thereby, the network learns to generalise. The function of an MLP with one hidden layer may be expressed as

\[
y_k(x,\bar{w}) = \sigma \left( \sum_{j=1}^{M} w_{kj} \sigma \left( \sum_{i=1}^{D} w_{ji} x_i \right) \right), k = 1,\ldots,K,
\]
where $\bar{x}$ is the vector of input data, $\bar{w}$ is the weight vector, $\sigma_o$ is the output activation function, $\sigma_a$ is the activation function of the hidden layer, $K$ is the number of output nodes, $M$ is the number of hidden nodes, $D$ is the number of input nodes and $y_k$ is the value of output node $k$ [Bishop 2006].

The weight vector is updated by an algorithm known as backpropagation [Haykin 1999]. Information that is used for adjusting the weight vector is passed backwards through the network, hence the name. Since the training is based on getting the correct output $z_{nj}$ (i.e. as close to the target $t_{nj}$ as possible) for input $z_{ni}$, where $n$ is the $n$th set of training data ($n=1,...,N$), the aim is to minimise the error function,

$$E(\bar{w}) = \sum_{n=1}^{N} E_n(\bar{w}),$$

where

$$E_n = \frac{1}{2} \sum_{j=1}^{M} (e_{nj})^2 = \frac{1}{2} \sum_{j=1}^{M} (t_{nj} - z_{nj})^2 = \frac{1}{2} \sum_{j=1}^{M} (t_{nj} - \sigma_o (a_{nj}))^2.$$

The information desired for adjusting $\bar{w}$ is the partial derivative of the error function for each input $n$:

$$\frac{\partial E_n}{\partial w_{nji}} = \frac{\partial E_n}{\partial e_{nj}} \frac{\partial e_{nj}}{\partial z_{nj}} \frac{\partial z_{nj}}{\partial a_{nj}} \frac{\partial a_{nj}}{\partial w_{nji}},$$

Computing the derivatives yields

$$\frac{\partial E_n}{\partial e_{nj}} = e_{nj}, \quad \frac{\partial e_{nj}}{\partial z_{nj}} = -1, \quad \frac{\partial z_{nj}}{\partial a_{nj}} = \sigma'(a_{nj}), \quad \frac{\partial a_{nj}}{\partial w_{nji}} = z_{ni}$$

and so the partial derivative of the error function can be expressed as

$$\frac{\partial E_n}{\partial w_{nji}} = -e_{nj} \sigma'(a_{nj}) z_{ni} = -\delta_{nj} z_{ni},$$

where $\delta_{nj}$ is called the local gradient [Haykin 1999]. This is used to adjust the weights according to

$$\Delta w_{nji} = -\eta \frac{\partial E_n}{\partial w_{nji}} = \eta \delta_{nj} z_{ni},$$

where $\eta$ is the learning rate. A large learning rate speeds up the training process, but the network may become unstable. Hence, there is a trade-off between speed and stability [Haykin 1999]. Using the expression of the error function, the local gradient can be written as

$$\delta_{nj} = -\frac{\partial E_n}{\partial z_{nj}} \sigma'(a_{nj}) = -\sum_k e_{nk} \frac{\partial e_{nk}}{\partial a_{nk}} \frac{\partial a_{nk}}{\partial z_{nj}} \sigma'(a_{nj}) = -\sum_k e_{nk} (-\sigma'(a_{nk})) w_{nkj} \sigma'(a_{nj})$$

which with the definition of the local gradient gives

$$\delta_{nj} = \sigma'(a_{nj}) \sum_k \delta_{nk} w_{nkj}.$$ 

This is known as the backpropagation formula [Haykin 1999].
At the output node, the local gradient is computed using $e_{nk} = t_{nk} - z_{nk}$ and the weights are adjusted accordingly. Then the local gradient at the previous node is computed using the backpropagation formula, and the corresponding weights are adjusted. This way, the adjustment propagates backwards through the network.

### 9.2 Radial-Basis Function Networks

The purpose of the RBF network is to fit a surface in multidimensional space to the training data. This surface is then used to interpolate the test data, and thereby classifying it. The RBF network has only one hidden layer, where a nonlinear transformation, from input space to multidimensional space, is made. At the output layer, there is a linear transformation from multidimensional space to output space, which makes the classification [Haykin 1999].

The basis for the RBF network is Cover’s theorem on separability of patterns, which states that a classification problem is more likely to be linearly separable in high-dimensional space than in low-dimensional space. A problem being linearly separable means that different classes may be completely separated by a hyperplane (i.e. a high-dimensional generalisation of a line) [Haykin 1999].

The fitting of the training data should satisfy

$$F(x_i) = d_i, \ i = 1, \ldots, N,$$

where $x_i$ are the training data points, $d_i$ are the target points and $F$ is the interpolating surface. $F$ is a sum of arbitrary RBF’s $\varphi$:

$$F(x) = \sum_{i=1}^{N} w_i \varphi(\|x - x_i\|),$$

where $x_i$ are the centres of the RBF’s and $w_i$ are the synaptic weights [Haykin 1999]. If $\Phi$ denotes an $N$-by-$N$ matrix with elements $\varphi_{ji}$, such that

$$\varphi_{ji} = \varphi(\|x_j - x_i\|),$$

$\overrightarrow{w}$ denotes the weight vector and $\overrightarrow{d}$ denotes the target vector, the data fitting can be expressed as

$$\Phi \overrightarrow{w} = \overrightarrow{d}.$$

If $\Phi$ has an inverse, the weight vector is

$$\overrightarrow{w} = \Phi^{-1} \overrightarrow{d}.$$

The classification of the test data is then made by the function $F$.

The centres of the RBF’s can be chosen in different ways, of which three are presented here [Haykin 1999]. The simplest way is to choose the centres randomly from the training data points. This method is suitable when the training data is known to be representative of the problem.

Another way of placing the centres is by self-organisation. In this approach, a clustering algorithm organises the training data in clusters, and the centre of each cluster becomes an RBF centre.

A third method, which is the most general one, is based on the error $e_j$: 
\[ e_j = d_j - \sum_{i=1}^{N} w_i \phi \left( \| x_j - \bar{t}_i \| \right). \]

Here, the \( \bar{t}_i \) that minimise \( e_j \) are chosen as centres. For more information on how to choose RBF centres, refer to Haykin [1999].
10 SUPPORT VECTOR MACHINES

Support vector machines (SVM’s) have the structure of a neural network with one hidden layer. Like RBF networks, they solve classification problems by constructing a hyperplane. The advantage of this method is that no prior knowledge of the problem structure is necessary; the classification is efficiently made anyway [Haykin 1999].

The most commonly used example is a linearly separable training data set, \( \{x_i, d_i\}_{i=1}^{N} \), where \( x_i \) is the training data, \( d_i \) is the target output and there are two classes, \( d_i = 1 \) and \( d_i = -1 \). The surface separating the classes is given by

\[
0 = -\mathbf{w}^T \mathbf{x} - b,
\]

where \( \mathbf{x} \) is a data vector, \( \mathbf{w} \) is a weight vector and \( b \) is a bias. The aim of an SVM is to maximise the distance between the surface and the closest data point, which is known as the margin of separation. The surface is then referred to as the optimal hyperplane [Haykin 1999]. This is illustrated in Figure 14.

\[\text{Figure 14: Possible hyperplanes}\]

It can be shown that the margin of separation is

\[
2 = \frac{2}{\|\mathbf{w}\|},
\]

where \( \mathbf{w} \) is the weight vector associated with the optimal hyperplane. Hence, the margin of separation is maximised when the weight vector is minimised. The weight vector, and the corresponding bias \( b \), satisfy the constraints

\[
\mathbf{w}^T x_i - b \geq 1, d_i = 1
\]

\[
\mathbf{w}^T x_i - b \leq -1, d_i = -1.
\]

The training data points for which these constraints hold with equality are called support vectors, since they are the points closest to the optimal hyperplane [Haykin 1999]. See Figure 15 for an illustration.
The problem can so far be summarised as an optimisation problem: to find the optimal weight vector and bias such that the following criteria are satisfied:

$$\min_w \Phi(w) = \min_w \frac{1}{2} w^T w$$

$$d_i(w^T x_i - b) \geq 1, i = 1, \ldots, N$$

The fact that the constraints are linear and $\Phi$ is a convex function is a great advantage of this method. In a convex optimisation problem, such as this, a local solution is always a global solution.

The primal problem stated above has a dual problem, and they share the same optimal solution. In the dual problem, the solution can be found using Karush-Kuhn-Tucker conditions and Lagrangian multipliers $\alpha$, which are introduced in the Lagrangian function:

$$J(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{N} \alpha_i \{d_i(w^T x_i - b) - 1\} = \frac{1}{2} w^T w - \sum_{i=1}^{N} \alpha_i d_i w^T x_i + b \sum_{i=1}^{N} \alpha_i d_i + \sum_{i=1}^{N} \alpha_i$$

The aim is now to maximise the Lagrangian function with respect to $\alpha$ as it is minimised with respect to $w$ and $b$ [Haykin 1999]. The latter is obtained by setting the partial derivatives to zero:

$$\frac{\partial J}{\partial w} = 0, \quad \frac{\partial J}{\partial b} = 0$$

and thereby gaining

$$\bar{w} = \sum_{i=1}^{N} \alpha_i d_i x_i$$

and

$$\sum_{i=1}^{N} \alpha_i d_i = 0.$$
\[
\max_{\alpha} Q(\alpha) = \max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j x_i^T x_j
\]

when

\[
\sum_{i=1}^{N} \alpha_i d_i = 0
\]

and

\[
\alpha_i \geq 0, i = 1, \ldots, N,
\]

where \( Q(\alpha) = J(\bar{w}, b, \alpha) \) and the aim is to find the optimal set of Lagrange multipliers \( \{\alpha_i\}_{i=1}^{N} \) [Haykin 1999].

When the optimal Lagrangian multipliers have been found, the optimum weight vector and bias may be found using

\[
\bar{w}_o = \sum_{i=1}^{N} \alpha_i d_i \bar{x}_i
\]

and

\[
b_o = 1 + \bar{w}_o^T \bar{x}_i
\]

where \( \bar{x}_i \) is a support vector on the “positive” side.

If the training data points are not linearly separable, there will be classification errors. There are two ways in which a classification error might appear; on the right side of the hyperplane or on the wrong side of it. In the former case, it cannot really be considered an error, since the data point will be classified correctly. In the latter case the data point will be misclassified.

The aim is now to minimise the probability of misclassification in nonseparable data sets [Haykin 1999].

To minimise classification errors, a set of non-negative slack variables \( \{\varepsilon_i\}_{i=1}^{N} \) are introduced to the constraints of the primal problem:

\[
d_i(w^T \bar{x}_i + b) \geq 1 - \varepsilon_i, i = 1, \ldots, N
\]

If the data point is inside the margin of separation but on the right side of the hyperplane, \( 0 \leq \varepsilon_i \leq 1 \). If, on the other hand, the data point is on the wrong side, \( \varepsilon_i > 1 \). The slack variables are used in the primal problem:

\[
\min_w \Phi(w, \varepsilon) = \min_w \frac{1}{2} w^T w + C \sum_{i=1}^{N} \varepsilon_i
\]

where \( C \) is a regularisation parameter [Haykin 1999].

The solution is obtained by solving the dual problem, in which only one change has been made. The last constraint is now

\[
0 \leq \alpha_i \leq C, i = 1, \ldots, N.
\]

After solving the dual problem, the weight vector is given by
\[ \overline{w}_o = \sum_{i=1}^{N_s} \alpha_i d_i x_i, \]

where \( N_s \) is the number of support vectors.

However, this construction of a classifying hyperplane is only part of an SVM. First, the input data must be transformed into high-dimensional feature space, just as is the case for RBF networks. By transforming the data set, chances are greater that it becomes linearly separable as long as the transform is nonlinear and the feature space has sufficiently high dimension [Haykin 1999].

With an \( m \)-dimensional vector of transformations \( \overline{\varphi} \), the feature space hyperplane is

\[ \overline{w}^T \overline{\varphi}(x) = 0, \]

where the bias is represented by \( w_0 \) \((\overline{\varphi}_0 = 1)\) for simplicity.

This gives the feature space weight vector

\[ \overline{w} = \sum_{i=1}^{N} \alpha_i d_i \overline{\varphi}(x_i) \]

which can be used to express the feature space hyperplane as

\[ \sum_{i=1}^{N} \alpha_i d_i \overline{\varphi}^T (x_i) \overline{\varphi}(x) = 0. \]

In order to use this theory without actually making computations in the high-dimensional feature space, the inner product \( \overline{\varphi}^T (x_i) \overline{\varphi}(x) \) is replaced by an inner product kernel \( K \), defined by

\[ K(x, \overline{x}) = K(\overline{x}_i, \overline{x}_j) = \overline{\varphi}^T (x_i) \overline{\varphi}(x_j). \]

The optimal hyperplane is then expressed as

\[ \sum_{i=1}^{N} \alpha_i d_i K(x, \overline{x}_i) = 0. \]

To find out whether a kernel is an inner product kernel and may be used in SVM applications or not, refer to Haykin [1999] or Bishop [2006].

Now, the problem may be formulated in its final form:

\[ \max_{\alpha} Q(\alpha) = \max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j K(x_i, x_j) \]

when

\[ \sum_{i=1}^{N} \alpha_i d_i = 0 \]

and

\[ 0 \leq \alpha_i \leq C, i = 1, \ldots, N, \]

which gives the weight vector of the optimal hyperplane as
\[
\overline{w_o} = \sum_{i=1}^{N} \alpha_i d_i \varphi(x_i).
\]

According to Haykin [1999], SVM’s have several advantages over MLP’s and RBF networks. Actually, MLP’s with one hidden layer and RBF networks are based on SVM theory. The similarity between the networks can be seen in the structure, and in the nonlinear transformations made in the hidden layer. When MLP’s and RBF networks are trained using SVM theory, the inner product kernels used in the transformations are \( \tanh(\beta_0 x^T x_i + \beta_1) \) and \( e^{-\frac{1}{2\sigma}||x-x_i||^2} \), respectively. The number of support vectors determines the number of hidden nodes, and the values of the support vectors determine the values of the weights or the locations of the centres.

The minimisation of the misclassification probability of an SVM is considered a better method than the backpropagation algorithm of an MLP, which is based on the classification error [Haykin 1999]. Also, the SVM has the advantage of always finding the global optimum whereas the backpropagation algorithm may converge towards a local minimum which is not the global minimum. There are disadvantages as well; the greatest being the amount of time needed for training. The number of support vectors cannot be set by the programmer, and no prior knowledge of the problem can be considered when programming an SVM. This means great computational costs. On the other hand, the generalising manner in which an SVM works is considered an advantage [Haykin 1999].
11 COMMITTEE MACHINES

A complex problem is often solved best by splitting it into smaller problems, solving them separately and combining the solutions. When several NN’s or SVM’s work together this way, they are known as a committee machine. The NN’s are usually called experts, since they are supposed to be experts at solving their part of the problem. Committee machines have proved to improve performance when compared to results from using a single network. There are virtually no limits for how many different networks may be combined, or of what type. The objective is to have each part perform as well as possible, and different networks have different advantages for different problems [Haykin 1999]. In this chapter, a few techniques for constructing a committee machine are presented briefly.

11.1 ENSEMBLE AVERAGING

In this method, the input is simply processed by several experts simultaneously. The results are then linearly combined, producing a solution. Ensemble averaging reduces the risk for overfitting as well as training time since there are fewer weights to adjust. Also, in ensemble averaging the variance of the result is lower than it is in the experts alone [Haykin 1999].

11.2 BAGGING

Bagging is short for bootstrap aggregating, in which one expert is trained using several data sets. Out of a training data set containing \( N \) data points, \( L \) points are chosen with replacement. This new set is called a bootstrap sample. Several bootstrap samples are taken, and the expert is trained using all of them. The results are then combined. This method also reduces variance and the risk of overfitting. However, reduction of variance only occurs when the model errors are uncorrelated. As they often are correlated, neither ensemble averaging nor bagging is as good in practice as they are in theory [Bishop 2006].

11.3 BOOSTING

In the boosting method, the system’s experts are trained sequentially. Adjustments to the weights are made according to the results of the previous experts; if an input is misclassified by one expert, it will have greater weight at the next expert. When all experts have been trained, the result is combined. According to Bishop [2006], this method gives good results even if the experts are not very good.

11.4 MIXTURES OF EXPERTS

In the mixture of experts (ME) method, a gating network is used together with several parallel experts. The objective of the gating network is to determine which expert is most useful in different regions of the problem. The experts’ results are then combined into an output [Haykin 1999]. The architecture of this method is illustrated in Figure 16.
There are as many nodes $u$ in the gating network as there are experts, and output $k$ of the gating network is

$$g_k = \frac{e^{u_k}}{\sum_{j=1}^{N} e^{u_j}}, k = 1, \ldots, N,$$

where $0 \leq g_k \leq 1, \forall k$, and $\sum_{k=1}^{N} g_k = 1$.

where $\underline{a_{k}}$ is a synaptic weight vector of the gating network and $\underline{x}$ is the input vector. The output from expert $k$ is

$$y_k = \underline{w_{k}}^T \underline{x},$$

where $\underline{w_{k}}$ is a synaptic weight vector and $\underline{x}$ is the input vector [Haykin 1999]. The output from the ME model is

$$y = \sum_{k=1}^{N} g_k y_k.$$

A hierarchical ME has the structure of Figure 16, except that the experts are replaced by ME’s. This model is the most flexible committee machine considered in this chapter.
12 SELF-ORGANISING MAPS

The previous three chapters have dealt with models trained by supervised learning. Self-organising maps (SOM’s) belong to a different class of decision support systems; those trained by unsupervised learning. One way of learning without a supervisor is by competition, which is how SOM’s learn. Another name for SOM’s is Kohonen networks [Haykin 1999].

A general SOM has a layer of input nodes and a one- or two-dimensional layer of output nodes, arranged in a lattice pattern as Figure 17 shows. The lattice pattern is often described as a topographic map over input data features, since the learning process adjusts the locations and weights of the output nodes. All synaptic connections have weights associated with them, and each output node has a discriminant function (i.e. a function assigning the input to a specific class). The classification is based on the winner-takes-all principle, which means that for each input data; only one output node is activated [Haykin 1999].

![Figure 17: Self-organising map with three input nodes](image)

The training process can be divided into three consecutive phases; competition, cooperation and synaptic adaptation. In the competitive phase, each output node’s discriminant function is computed with respect to current input data. The node which has the largest value wins. In the cooperative phase, this node’s neighbours are located. Finally, in the adaptive phase, the neighbours’ discriminant functions’ values increase by weight adjustment. This is done so that the probability of a similar input data being classified the same way increases [Haykin 1999].

Initially, the weights are set to low values. Then an index $i(\hat{x})$ is calculated for each node:

$$i(\hat{x}) = \arg \min_j |x - w_j|, \quad f = 1,...,l$$

where $l$ is the number of nodes, $\hat{x}$ is the input vector and $w_j$ is the weight vector. The aim of this is to maximise $w^T \hat{x}$. Hereafter, $i(\hat{x})$ denotes the winning neuron.

$$h_{ij}(\hat{x}) = e^{-\frac{d_{ij}^2}{2\sigma^2}}$$

is used to find the neighbours. Here, $d_{ij} = |r_j - r_i|$, where $r_j$ and $r_i$ are node locations, and

$$\sigma(n) = \sigma_0 e^{-\frac{n}{\tau_1}}$$

where $n$ is the number of iterations, $\tau_1$ is a time constant and $\sigma_0$ is the initial value of $\sigma$ [Haykin 1999]. Initially, $\sigma$ is large and a lot of neighbours are found. When $\sigma$ decreases, so does the number of neighbours.
During the synaptic adaptation, the weights are adjusted at learning rate \( \eta \), according to

\[
\bar{w}_j(n+1) = \bar{w}_j(n) + \eta(n) h_{j,\bar{x}}(n)(\bar{x} - \bar{w}_j(n)).
\]

Here, \( \eta(n) = \eta_0 e^{-\frac{n}{\tau_2}} \), where \( \tau_2 \) is a time constant and \( \eta_0 \) is the initial learning rate. The adaptive process causes \( \bar{w}_j \) to approach \( \bar{x} \), which in turn causes adjacent neurons to have similar weight vectors. This is the reason a SOM is often described as a topographic map [Haykin 1999].

The three phases described above are iterated in two phases; a self-organising phase and a convergence phase. In the self-organising phase the topographic map is drawn, and in the convergence phase the map is perfected. The first phase may take 1000 iterations of the three-phase algorithm, and the second phase takes approximately \( 500 \cdot n \) iterations [Haykin 1999]. Clearly, the SOM is a time-consuming method.
13 CLINICAL DECISION SUPPORT SYSTEMS

The methods for constructing DSS’s described in the previous chapters are just a few out of many. Even so, knowing the problem, how should the model be selected?

13.1 MODEL SELECTION

According to Mangiameli et al [2004], there is no available theory on how to select a model for a certain problem. In a study consisting of five different clinical problems, 23 models were used for classification. Different models performed differently for different problems, and the conclusion was that the best way to select a model is to try several models and choose the one that performs best [Mangiameli et al 2004]. This method of model selection has been applied in several other studies. According to Begg et al (ed) [2006], researchers often use models too complex for the classification problem, since they are expected to perform better than simpler models. Rather, the complexity of the model should be adapted to the complexity of the problem.

To find out which model performs better, the models need to be evaluated by tests performed on the same data set. The data set used for training and model selection is generally divided into three parts; one for training, one for validation and one for testing. This works well for large data sets, but often the data set is of very limited size. To achieve reliable results with a small data set, cross validation is often used. In cross validation, a small set is chosen for testing. The remaining data is divided into \( N \) sections, and the model is trained with \( N-1 \) sections and validated with the remaining section. This is repeated \( N \) times, so that each section is used for validation once. Then, the model is tested on the test data [Bishop 2006].

To avoid overfitting, the model must not be trained for too long. The easiest and most widely used way of regulating the number of training iterations is by looking at the validation set error. Training should be stopped at the iteration where the validation set error has a minimum [Bishop 2006].

13.2 SOME EXAMPLES OF DECISION SUPPORT SYSTEMS

In biomedical engineering, the MLP is a “classic” way of solving classification problems. However, several problems are associated with the MLP: it’s time-consuming, the algorithm might get stuck in a local minimum which is not the global minimum, initial values of parameters are often set randomly and the number of hidden nodes is determined initially [Mangiameli et al 2004, Olmez & Dokur 2003b, Ozbay et al 2006]. In recent studies, it seems that the MLP is used mostly for comparison to other methods.

The MLP is also used in combination with other methods, as in studies on arrhythmia classification by Engin [2004] and Ozbay et al [2006]. Here, a self-organising method called fuzzy c-means clustering is used as a preclassifier. The output from the preclassifier is classified by an MLP. Another classifier consisting of five MLP’s for the classification of four arrhythmias was created by Guler & Ubeysi [2005a]. First, four parallel MLP’s, each designed to be an expert for one of the arrhythmias, classify the input data. Their output is then used as input to a fifth MLP, which makes the final classification. The authors call their method “a more sophisticated version of cross validation” and claim that it “has been shown experimentally to effectively improve generalization ability”. Their model performs well; the arrhythmias are classified with 97% accuracy.
Sometimes, modified MLP’s are used. When classifying ten different arrhythmias, Ölmez & Dokur [2003b] achieved 98.6% accuracy with an MLP modification called an intersecting prisms network. Since SOM’s are said to be time-consuming [Papadimitriou et al 2001], modifications are made in order to reduce redundant information and speed them up. To improve classification performance when data are close to the classification boundaries, Papadimitriou et al [2001] used an SOM in combination with a supervised learning network. The SOM made the easier classifications itself, and for trickier classifications an SVM with a polynomial kernel was of assistance. The SVM was considered too complex to use for the entire problem, but as a small part of the model it improved performance by 10%.

SVM’s are considered superior to other methods by several authors [Haykin 1999, Papadimitriou et al 2001, Übeyli 2007]. However, the construction of a quadratic optimisation problem which is the basis of the method is complicated and time-consuming, and efforts have been made to create simpler and faster models. Çomak et al [2007] used a modified SVM where the optimisation problem was linear and not quadratic. When compared to an MLP, the modified SVM was shown to be 13 times faster. However, both sensitivity and specificity was higher for the MLP than for the modified SVM. In an other study, a regular SVM was compared to an MLP. Here, the classification accuracy of the SVM was 98.6% while for the MLP it was merely 91.4% [Übeyli 2007]. When implementing an SVM, finding a proper kernel is of great importance [Übeyli 2007].

Güler & Übeyli [2005b] created a modified ME, in order to reduce training time and obtain better classification performance. In their model, $N$ experts are divided into $K$ groups. There are $K$ input features, each of which is fed to a separate group. Also, all $K$ features are fed to the gating network, which gives $N$ outputs, one for each expert. The products of the experts’ output and the gating network’s output are then linearly combined, creating the overall output. When compared to regular ME’s and MLP’s, both classification accuracy and training time were shown to improve greatly.
14 DISCUSSION

In this chapter the different parts needed for a DSS for diagnosing heart failure are brought together and commented upon.

14.1 GENERAL DEMANDS

The system proposed must be usable in a clinical situation. This means that it should not depend on the environment, for example the PCG signal should not need to be registered in a silent room and the ECG electrodes should not need to be placed in the exact right positions for the system to give a reliable result. The system should give results in real-time, and it should not need long signals to be registered. Perhaps the most important feature of the system is that it must be very robust against noise of all kinds.

However, the system does not have to be 100% accurate since it should only aid the physician, not do her job for her. It has been suggested that the system’s response to input data should be the likelihood that the examined patient has heart failure, expressed as a percentage. This means demanding a lot from the system and leaving very little information for the physician. Instead, the response should be “yes”, “no” or “maybe”, and all information acquired from the signals should be presented. This way, the physician would have support from the system while being able to form a personal opinion based on information from the ECG and PCG recordings.

There would have to be an extensive database for developing and training the system. All parameters identified as important in diagnosing heart failure would need to be registered in the database; for healthy persons as well as for patients with heart failure of varying degree. All patients registered in the database would have to be examined very thoroughly, for the system to be trained properly.

14.2 FEATURE SELECTION

Several authors agree that the most important aspect when setting up a DSS is to use proper input features [Begg et al (ed) 2006, Mangiameli et al 2004]. A few different approaches to selecting the features have been described, which one should be applied?

A successful extraction of features by the computation of amplitudes and durations is dependant on a noise-free signal. Since it is highly unlikely to register a noise-free signal in a clinical situation, good noise-reducing algorithms must be applied. Even so, computing for instance an exact QRS duration requires the knowledge of the exact start and end points of the QRS complex. These are usually hard to find, and so the duration is unlikely to be computed exactly.

The wavelet transform has very nice noise-reducing abilities [Addison 2005]. It is also rather easy to implement and it is widely used, not only in medical applications. The most important part of the implementation of the wavelet transform is the choice of mother wavelet and decomposition level [Ceylan & Özbay 2007]. Different wavelets and decomposition levels should be tried and the most efficient combination should be selected. Since wavelet theory has been used in many areas, guidance to which wavelets to try might be given in literature concerning similar applications to the one of interest. However, a choice often has to be made as to which wavelet coefficients to use. Statistical information, which is easy to compute and has been shown to give good results, or PCA might be used to further reduce the number of features.
PCA, like wavelet theory, is widely used in applications of very different kinds. It doesn’t seem like a very straight-forward method, but a lot of information is retained in very few features. It may be applied to the signal directly, or to the wavelet coefficients.

Once the method of feature selection has been set, the number of features used for input to the system should be determined. Too many features will result in a slower, and perhaps more complex, system, and often it reduces the system’s performance [Übeyli 2007]. However, the number of features must be high enough to keep the significant characteristics of the signal. There is no theory as to the optimal number of input features to a system, and so trial and error should be used.

HRV is perhaps the most significant parameter, and so it is important that it is computed properly. DFA is often used to separate heart failure patients from healthy persons, and so it seems to be a particularly reliable method. MWA, on the other hand, is easy to implement if wavelet theory is used to extract other features as well. Spectral analysis seems to be an inexact and time-consuming method, compared to the other two.

Wavelet theory should be applicable when identifying parameters of low amplitude and high frequency, since they are often lost in noise. The method of using a tailored wavelet to identify third and fourth heart sounds is interesting and should be tried, evaluated and compared to the method of using wavelet coefficients and a classifying method.

14.3 USING THE FEATURES

Since the proposed system is really rather big, the principle of “divide and conquer” should be applied. A DSS should be constructed with the 22 parameters listed in chapter 4.6 as input features. Some are binary (yes/no) and some are continuous (have a set value).

An alternative approach would be to use some of the parameters from the list (the ones fed to the system by the physician, such as age and gender) together with features selected from the ECG and PCG signals. Then, however, the system would have to be very complex and perform difficult classifications. From the literature, it is clear that for example ischemia is a lot harder to find than arrhythmias. This approach would have the system use the same features to classify all parameters at the same time, and that would probably give poor results.

Instead, the parameters extracted from the ECG and PCG signals should all be extracted one at a time and by different systems. Both the features used as input and the type of system used should be adapted to the task at hand, since finding ischemia might require a different system than finding PVC’s. Then, the outputs from these systems can be used as inputs together with the inputs provided by the physician.

14.4 SYSTEM SELECTION

As was mentioned in the section 14.3, different systems should be used for different tasks. Several models have been described, but which is better for which task? There is no straight-forward answer to that question. The literature may be consulted for implementations similar to the one at hand, and the complexity of the classification problem may govern the complexity of the model chosen. For example, a complex model like the SVM might be a bad choice for a simple task.

Most implementations found in the literature are modifications of “standard” models like those described in chapters 9-12. There is nothing fix and final about the models; if the classification benefits from modifications, than modifications should be made. However,
according to Begg et al (ed) [2006] scientists often choose models that are overly complex. Modifications should not be made that do not improve the system’s performance.

A suitable classification model should be found by trial and error just as the best method of feature selection and the optimal number of features should. Different numbers of hidden nodes and different parameter values should also be tried, in order to find the best combination.

Except from choosing proper values for the system, the data used for training is of great importance. There should be enough training data to make the system’s generalising abilities as good as possible, and the distribution within the data set should be representative of the “real” data set (i.e. the set of patients for which the system will be used). The training data set should contain diverse kinds of patients with different pathologies. It is the training that will determine how the 23 parameters are weighted.

14.5 SUMMARY

To conclude the thesis, the parameters, their diagnostic values and methods of analysis are briefly summarised. 23 parameters have been identified; these are listed in section 4.6. There are three categories of parameters, one of which is symptoms and signs to be fed to the system manually. The other two categories are conditions to be extracted from the ECG and PCG signals. The first category can be divided into two sub-categories; on with actual symptoms of heart failure and one with important information about the patient, such as age and gender.

Symptoms of heart failure are important to include, since today they constitute the basis of a heart failure diagnosis. Parameters such as age and gender are important since there are differences in signs of heart failure that depend on them. HRV, which is extracted from the ECG, is of great importance since a reduced HRV is an independent marker of heart failure. The other parameters extracted from the ECG are of less importance, but for a correct diagnosis they should still be incorporated in the system. The third and fourth heart sounds are very important, particularly the third one (unless the patient is young) since it is a marker of heart failure.

Every parameter that is extracted from the signals should be classified separately. This should be done by extracting features from the signal using wavelet theory and/or PCA and then feeding the features to a neural network or other type of DSS. The results should then be fed to another DSS, together with the manually provided parameters. This DSS will classify the problem as heart failure or not.

14.6 USEFULNESS

Imagine that a huge, complicated system has been created. Does it matter? Well, DSS’s have been found to improve the efficiency of care through reduced costs, increased knowledge and the adherence of guidelines [Bates et al 2003, Dreiseitl & Binder 2005]. The more uncertain a physician is on the diagnosis, the more she relies on a DSS. In particular, young and inexperienced physicians tend to appreciate DSS’s [Dreiseitl & Binder 2005]. In local care centres, where physicians (often young) come across a multitude of pathologies on a daily basis, it seems that a DSS for diagnosing heart failure could do a lot of good. A correct diagnosis could be made earlier, thus improving the patient’s situation and reducing costs.

14.7 FUTURE WORK

Since the thesis merely contains a first study on the possibilities for a DSS for diagnosing heart failure, there is plenty of further work to be done. Parameters which suggest that a
patient does not suffer from heart failure should be identified and added to the list. A detailed
database containing all parameters for a large number of healthy persons and heart failure
patients with different pathologies and of different age and gender should be created.

If wavelet theory is to be used, mother wavelets that are suitable for analysing the ECG and
PCG signals should be identified, as should appropriate decomposition levels. If PCA is to be
used, the optimal number of principal components to use when analysing the signals should
be recognised. Appropriate classifying methods should be selected for the different problems,
and proper numbers of nodes and values of parameters should be identified.


FIGURES

Figure 2: Texas Heart Institute, http://texasheart.org/HIC/Topics/Proced/pacemake.cfm (2007-06-01).
Figure 3: own illustration.
Figure 6: Marcus et al 2005.
Figure 7: Monroe Community College, http://www.monroecc.edu/depts/pstc/backup/parashs1.htm (2007-06-04).
Figure 8: own illustration.
Figure 9: Willson et al 2002.
Figure 10: own illustration.
Figure 13: own illustration.
Figure 16: own illustration.
Figure 17: Johannes Lampel, http://johannes.lampel.net/bll137.html (2007-07-23).
**Svensk titel**
Möjligheter för utveckling av ett beslutsstödssystem för diagnostisering av hjärtsvikt

**Engelsk titel**
Possibilities for the development of a decision support system for diagnosing heart failure

**Författare**
Linda Olsson

**Uppdragsgivare:**
IMT, Linköpings Universitet

**Rapporttyp:**
Examensarbete

**Rapportspråk:**
Engelska

**Sammanfattning** (högst 150 ord).
Heart failure is a common disease which is difficult to diagnose. To aid physicians in diagnosing heart failure, a decision support system has been proposed. Parameters useful to the system are suggested. Some of these, such as age and gender, should be provided by the physician, and some should be derived from electro- and phonocardiographic signals.

Various methods of signal processing, such as wavelet theory and principal components analysis, are described. Heart failure should be diagnosed based on the parameters, and so various forms of decision support systems, such as neural networks and support vector machines, are described. The methods of signal processing and classification are discussed and suggestions on how to develop the system are made.

**Nyckelord** (högst 8)
Heart failure, signal processing, decision support systems.

**Bibliotekets anteckningar:**