

# Acquisition and Quantitative Analysis of the Arterial Pulse

A thesis submitted in partial fulfillment of  
the requirements for the degree of

Doctor of Philosophy

by

**Aniruddha J. Joshi**  
(Roll No. 03405802)

Under the guidance of  
**Prof. Sharat Chandran**



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING  
INDIAN INSTITUTE OF TECHNOLOGY BOMBAY

2011



# THESIS APPROVAL

The thesis entitled

## **Acquisition and Quantitative Analysis of the Arterial Pulse**

by

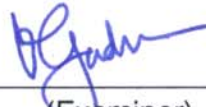
**Aniruddha J. Joshi**

(Roll No. 03405802)

is approved for the degree of  
Doctor of Philosophy



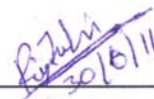
(Examiner)



(Examiner)

Sharat  
30.6.2011

(Guide)



(Chairman)

Date: June 30, 2011

Place: IIT Powai, Mumbai



## Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

*Joshi*

\_\_\_\_\_  
(Signature)

*Aniruddha J. Joshi*

\_\_\_\_\_  
(Name of the student)

*03405802*

\_\_\_\_\_  
(Roll no.)

Date: July 12, 2011



# Abstract

Pulse-based diagnosis is a prominently explored methodology in traditional medicine. Our work is inspired by *Ayurveda*. The *Ayurvedic* practitioner diagnoses subjects by feeling their pulse on wrist, and it is the belief that this information is useful in detecting imbalance or disorders in the entire body. This notion has not been fully accepted; one reason could be the absence of quantitative and objective characterization.

We develop a portable, and inexpensive system *Nadi Tarangini* to acquire noise-free pulse signals. As the study of pulse signals is in the nascent stage, we develop seven broad approaches (individually and in symbiosis) to study two characteristic properties of the pulse signal, namely nonlinear dynamics and period-to-period variability. Each approach helps revealing different ‘property’ of the pulse signal.

Many important biomedical signals originating from the self-regulating, and rhythmic heart action contain a pseudo-periodic nature, where the nature of shrinking and stretching is an indication of the information content. Though adjacent pulse beats look similar in shape, they show different patterns, and these evolutionary (period-to-period) changes in the behavior are medically significant. We show that different parts of the pulse signals have different scaling properties, and thus it is natural to expect the need for many exponents to characterize them.

Using the pseudo-periodic nature, we further propose a novel approach to compute the useful time domain parameters of a pulse signal, based on the pitch synchronous wavelet transform, and dynamic time warping. The approach is particularly useful in ill-behaved signals when the required parameters have either missing or confusing values. Further, we propose a novel Hybrid Support Vector Machine approach to reduce the false negatives in the classification. The improvements in the results show the efficacy of both the novel approaches.

In summary, the properties, and the progress of pulse analysis confirms the importance of its consideration for diagnosis. Years of development in the analysis of ECG-based methodologies has resulted in advanced automated tools. We establish the efficacy of the pulse-based methodology in this thesis with a positive note that it will receive appropriate recognition in future.



---

# Contents

---

<b>List of Tables</b>	<b>ix</b>
<b>List of Figures</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Problem Definition . . . . .	2
1.2.1 Methodology . . . . .	2
1.3 Related Work . . . . .	4
1.4 Our Contributions . . . . .	5
1.4.1 Acquisition of the Pulse Signals . . . . .	5
1.4.2 Analysis of the Pulse Signals . . . . .	8
1.4.3 Novel Hybrid Support Vector Machine . . . . .	13
1.4.4 Classification . . . . .	14
1.5 Organization of Thesis . . . . .	15
<b>2 Pulse Signals in Medicine</b>	<b>19</b>
2.1 A Cardiac Cycle . . . . .	20
2.1.1 Modeling the Pumping Action of the Heart . . . . .	21
2.1.2 Acquisition of the Heart Action: An Electrocardiogram . . . . .	23

2.1.3	Cardiac Output . . . . .	24
2.1.4	Blood Flow . . . . .	25
2.2	Formation of the Pulse Signal . . . . .	26
2.2.1	Elastic Property of the Arterial Wall . . . . .	27
2.2.2	Hemodynamics of Pulsating Stream . . . . .	28
2.2.3	Forward-Backward Wave . . . . .	29
2.2.4	Law of Conservation of Mass . . . . .	30
2.2.5	Hierarchy of Factors . . . . .	31
2.3	Traditional Pulse-based Methodology . . . . .	32
2.3.1	<i>Panch Mahabhuta</i> . . . . .	33
2.3.2	<i>Tridosha Theory</i> . . . . .	34
2.3.3	Importance of the Pulse-based Diagnosis in <i>Ayurveda</i> . . . . .	35
<b>3</b>	<b>Acquisition</b>	<b>41</b>
3.1	Components in a Typical Data Acquisition System . . . . .	42
3.1.1	Design Aspects . . . . .	46
3.2	Our Previous Systems of Pulse Acquisition . . . . .	47
3.3	<i>Nadi Tarangini</i> . . . . .	48
3.3.1	Pressure Sensor . . . . .	49
3.3.2	Signal Conditioning Unit . . . . .	51
3.3.3	Digitizer . . . . .	52
3.3.4	Storage . . . . .	52
3.3.5	Schematic of <i>Nadi Tarangini</i> . . . . .	53
3.4	Measurement Methodology . . . . .	54
3.4.1	Deciding VPK Locations . . . . .	54
3.4.2	Placement of the Sensors . . . . .	54
3.4.3	Deciding the Contact Pressure . . . . .	55
3.5	Pulse Database . . . . .	57
<b>4</b>	<b>The Nature of Pulse Signals</b>	<b>63</b>
4.1	Prior Work in Pulse Acquisition . . . . .	64
4.2	Pulse Signals Obtained Using <i>Nadi Tarangini</i> . . . . .	72
4.2.1	Differential Mode of Acquisition . . . . .	72

4.2.2	<i>Tridosha</i> Pulse . . . . .	73
4.3	Comparison of Pulse with Electrocardiogram . . . . .	76
4.4	Important Properties of the Pulse Signals . . . . .	78
4.5	Prior Work in Pulse Analysis . . . . .	82
4.5.1	Circulation Parameters . . . . .	82
4.5.2	Extraction of Important Features . . . . .	83
4.5.3	Problem Specific Observations . . . . .	84
4.5.4	Pulse Analysis in India . . . . .	86
4.6	Challenges in Pulse-based Analysis . . . . .	87
<b>5</b>	<b>Capturing Pseudo-periodic Behavior</b>	<b>91</b>
5.1	Pseudo-periodicity . . . . .	93
5.1.1	Estimation of Pseudo-periods . . . . .	94
5.2	Extracting Essence of the Pulse signal . . . . .	96
5.2.1	Pitch Synchronous Representation . . . . .	97
5.2.2	Computation of the PSWT . . . . .	99
5.3	Time Domain Parameters . . . . .	103
5.3.1	Challenge in Computation . . . . .	104
5.3.2	Related Work . . . . .	106
5.3.3	Prior Zero-crossing Approach for TDP Computation . . . . .	107
5.3.4	Our Maxima-Minima Approach . . . . .	108
5.3.5	Our Approach for Ill-behaved Pulse Cycle . . . . .	110
5.4	Novel PSWT-DTW Approach . . . . .	111
5.4.1	Proposed Method Overview . . . . .	113
5.4.2	Dynamic Time Warping . . . . .	113
5.4.3	Observations . . . . .	116
5.5	Pulse Rate Variability Measures . . . . .	124
5.5.1	Time-domain Measures . . . . .	125
5.5.2	Frequency-domain Measures . . . . .	126
5.5.3	Nonlinear Poincaré-based Measures . . . . .	128
<b>6</b>	<b>Capturing Nonlinear Dynamics</b>	<b>135</b>
6.1	Fractal Nature . . . . .	136

6.1.1	Self Similarity of the API Signal . . . . .	136
6.1.2	Hierarchical Organization of Singularities . . . . .	139
6.1.3	Local Hölder Exponents . . . . .	140
6.1.4	Multifractal Spectrum . . . . .	145
6.2	Chaotic Nature . . . . .	149
6.2.1	Recurrence plot . . . . .	149
6.2.2	Visualization . . . . .	150
6.2.3	Observations . . . . .	151
6.2.4	Recurrence Quantification Analysis . . . . .	152
6.3	Topological Structures . . . . .	154
6.3.1	Nonlinear Dynamics in Pulse . . . . .	156
6.3.2	Determining Period of a Pulse cycle . . . . .	160
6.3.3	Topological Invariants . . . . .	161
6.3.4	Observations . . . . .	168
6.3.5	Visualization . . . . .	170
<b>7</b>	<b>Classification</b>	<b>175</b>
7.1	Support Vector Machine . . . . .	175
7.1.1	Multiclass Support Vector Machine . . . . .	178
7.2	Novel Hybrid Support Vector Machine . . . . .	180
7.2.1	Prior Work . . . . .	180
7.2.2	Our Approach . . . . .	181
7.3	Classification Results . . . . .	184
7.3.1	Efficacy of Our Hybrid-SVM . . . . .	186
7.3.2	Efficacy of Our PSWT-DTW approach . . . . .	186
<b>8</b>	<b>Conclusions</b>	<b>193</b>
8.1	Applications . . . . .	193
8.2	Limitations and Future Work . . . . .	195
8.3	Pictorial Summary of Our Work . . . . .	201
<b>A</b>	<b>ECG-based Arrhythmia Classification</b>	<b>207</b>
A.1	Introduction . . . . .	207

A.2	Local Hölder Exponents . . . . .	208
A.2.1	Efficiency Considerations . . . . .	209
A.2.2	Stable Computation . . . . .	210
A.3	Results and Discussion . . . . .	211
A.3.1	Binary Classification . . . . .	212
A.3.2	Multiclass Classification . . . . .	212
<b>B</b>	<b>Hybrid-SVM for Imbalanced Data</b>	<b>215</b>
B.1	Introduction . . . . .	215
B.2	Classification Method . . . . .	219
B.2.1	Hybrid-SVM . . . . .	220
B.3	Experimental Setup . . . . .	222
B.4	Validity of Features . . . . .	224
B.4.1	Combination of Wavelet Scales . . . . .	226
B.4.2	ECG Lead Selection . . . . .	226
B.5	Validity of Hybrid-SVM . . . . .	227
B.5.1	Comparison with Conventional Multiclass SVM . . . . .	227
B.5.2	Performance of Hybrid-SVM for Enhanced Features . . . . .	229
B.5.3	Performance of Hybrid-SVM for Other Features . . . . .	231
B.5.4	Compatibility of Hybrid-SVM with Prior Solutions . . . . .	234
<b>C</b>	<b>Example Pulse Waveforms</b>	<b>237</b>

---

# List of Tables

---

5.1	Comparison of the success or failure of two approaches on individual pulse cycles. When the zero-crossings are missed, pseudo-location need to be identified. Our maxima-minima approach reduces manual corrections. . . . .	112
5.2	Comparison of statistics in the computation of TDP. We observe that the PSWT cycle captures the trend values. . . . .	120
5.3	Time-domain measures of API. . . . .	126
5.4	Frequency-domain measures. . . . .	127
5.5	Nonlinear Poincaré-based measures. . . . .	130
6.1	Ratios of areas under the multifractal spectra for multiple age groups, and varying disorders. . . . .	148
6.2	RQA descriptors with their meaning and the equations for computation.	155
6.3	Comparison of binary (healthy-disorder) classification accuracies. . . .	170
7.1	Confusion matrix for the multiclass classification using feature set FS1. Each entry corresponds to the conventional case ( <i>opM</i> ), and the Hybrid-SVM (R=0) case (in bold). . . . .	187

7.2	Confusion matrix for the multiclass classification using feature set FS2. Each entry corresponds to the conventional case ( <i>opM</i> ), and the Hybrid-SVM (R=0) case (in bold). . . . .	188
7.3	Comparison of the classification performances of our novel approaches with the conventional approach. . . . .	189
A.1	Confusion matrix (average case). . . . .	213
B.1	Confusion matrix for the conventional multiclass classification using local Hölder exponents ( <i>opM</i> case in Figure B.3). Six abnormal and one normal class were considered. . . . .	225
B.2	Confusion matrix for the binary classification using local Hölder exponents ( <i>opB</i> case in Figure B.3). . . . .	226
B.3	Confusion matrix for six abnormal classes, using local Hölder exponents ( <i>opN</i> case in Figure B.3). Only abnormal rhythms were considered. . . .	227
B.4	Confusion matrix for the multiclass classification using local Hölder exponents with our Hybrid-SVM approach considering threshold 0.1. See Table B.1 to contrast with the earlier approach. . . . .	228
B.5	Confusion matrix for multiclass classification using local Hölder exponents with our Hybrid-SVM approach considering threshold 0.5. See Table B.1 and Table B.4 to contrast. . . . .	228
B.6	Confusion matrix for multiclass classification using local Hölder exponents with our Hybrid-SVM approach considering threshold 0.9. See Table B.1 and Table B.4 to contrast. . . . .	229
B.7	The comparison of areas under ROC plots for the multiclass classification using local Hölder exponents. . . . .	230
B.8	Confusion matrix for the multiclass classification using local Hölder exponents for scales '1 & 2'. Each entry corresponds to the conventional case, and the Hybrid-SVM (R=0.1) case (in brackets). . . . .	231
B.9	MCC values for the considered seven classes in different scenarios. 'S' indicates the number of scales used (1 or 2) for the computation of local Hölder exponents 'L' indicates the number of ECG leads used (1 or 2) in the feature extraction. The Hybrid-SVM used the threshold value of 0.1.	232

B.10	Confusion matrix for the multiclass classification using Fourier coefficients [140]. Each entry corresponds to the conventional case, and the Hybrid-SVM (R=0) case (in brackets). . . . .	232
B.11	Confusion matrix for the multiclass classification using wavelet coefficients [140]. Each entry corresponds to the conventional case, and the Hybrid-SVM (R=0) case (in brackets). . . . .	233
B.12	Confusion matrix for the multiclass classification using AR coefficients [36]. Each entry corresponds to the conventional case, and the Hybrid-SVM (R=0) case (in brackets). . . . .	233
B.13	MCC values for the considered seven classes in different scenarios. . . . .	234
B.14	Confusion matrix using the principle of under-sampling data and then applying the Hybrid-SVM approach. Each entry corresponds to the conventional case, and the Hybrid-SVM (R=0) case (in brackets). . . . .	235
B.15	Confusion matrix using the principle of over-sampling data, and then applying Hybrid-SVM approach. Each entry corresponds to the conventional case, and the Hybrid-SVM (R=0) case (in brackets). . . . .	235



---

# List of Figures

---

1.1	Summarized in this figure is a typical data analysis procedure which involves problem definition, data acquisition and understanding, and finally data analysis. The other important blocks in the procedure of database, visualization, and storage are not displayed here. . . . .	3
1.2	Summary of work in this thesis. Top Row: Based on understanding the domain, we design an acquisition system, and provide a corpus of data. Middle Row: The pulse signal is complex and can be studied using a variety of known mathematical techniques. Examples of these include multifractals, and topological invariants. Bottom Row: Based on these techniques, features can be obtained as output for a plethora of purposes including classification, and can also be suitably visualized. . . . .	6
1.3	Pulse-based acquisition: Pictorial comparison of the traditional methodology with our system. . . . .	7
1.4	Sample VPK pulse signal from a subject's left hand. We can observe the difference between the amplitudes and patterns of the three <i>dosha</i> . . . .	8
1.5	Typical behavior of a pulse signal. A continuous, rhythmic P-T-V-D pattern is observed. Even though the pulse cycles look similar, they vary in amplitude and pseudo-period. . . . .	8

1.6	A global snapshot of the analysis portion of the thesis. Our work involves development of a variety of approaches to study different properties and patterns of the pulse signals. For example, multifractal analysis can eventually help understand variations with disorder. Details of this figure appear in the text. . . . .	9
1.7	Time domain parameters of a cycle in the pseudo-periodic pulse signal. $w$ refers to the width of the P-subwave at 80% of its height). Significant locations termed <i>TDP locations</i> are marked on the X-axis. . . . .	12
1.8	The steps in our novel PSWT-DTW approach to compute time-axis related parameters on individual pulse cycles. The dotted, black path displays the conventional approach. . . . .	13
1.9	The method of our Hybrid-SVM approach. The outputs of SVM blocks ( $opM$ , $opB$ , and $opN$ ) are suitably compared to increase the overall efficacy. . . . .	14
1.10	Organization chart for the thesis. . . . .	17
2.1	A cross-section of the heart showing its important portions [250]. . . . .	21
2.2	The pressure-volume loop [31] explaining the four phases in a cardiac cycle. The first two phases are known as diastole and the last two phases are known as systole. . . . .	22
2.3	Recording of one cardiac cycle using an ECG lead [118]. The left subfigure denotes different parts in a human heart. Different color codings are used in the right subfigure to indicate the location and strength of the cardiac action in a particular heart portion. Finally, all these activities sum up to provide complete ECG cycle at the right bottom subfigure. . . . .	23
2.4	Sample electrocardiogram composed of 5 cardiac cycles. This is signal number '100' from the arrhythmia database of NIH [63]). . . . .	24
2.5	The effect of pressure from the piston $P$ on the elastic arterial walls. This resembles the actions in the human circulatory system. . . . .	27
2.6	The effect of sudden ejection of a fluid (blood) into an elastic, distensible tube (e.g. an artery). The arrows indicate the immediate effect. This resembles the actions in the human circulatory system. . . . .	28

2.7	Important points on a pulse cycle form various segments indicating the stages of systole and diastole. . . . .	29
2.8	Relationship between blood velocity and cross-sectional area of the blood vessels. Because of the law of conservation of mass, flow at any given point in a closed system will be equal to flow at any other point. Thus, velocity is inversely proportional to the cross-sectional area. . . . .	31
2.9	The pulse signal shows very complex behavior as it depends upon many interrelated factors. The final outcome depends upon the values of all the factors [162]. . . . .	32
2.10	Sanskrit shloka defining <i>Ayurveda</i> as: The science that expounds on – the supportive of life ( <i>hitayu</i> ), the non-supportive for life ( <i>ahitayu</i> ), the happy life ( <i>sukhayu</i> ), the unhappy life ( <i>dukhayu</i> ), the measurement of life ( <i>manam</i> ), and the quality of life. . . . .	32
2.11	Sanskrit shloka explaining <i>Panch Mahabhuta</i> : Earth, water, fire, air and sky (ether or space). When they are combined in different combinations, make all the conscious or non-conscious entities of world. . . . .	33
2.12	Sanskrit shloka explaining <i>Tridosha</i> : <i>Vata</i> , <i>Pitta</i> and <i>Kapha</i> are the root for the origin of the body. When in balanced state they keep the body in its essential or true state. But when imbalanced, they can even cause death. . . . .	34
2.13	The positioning of <i>Ayurvedic</i> practitioner’s three fingers to sense the pulse of a subject. Respectively index, middle and ring finger are used to sense <i>vata</i> , <i>pitta</i> and <i>kapha dosha</i> . . . . .	36
2.14	Sanskrit shloka explaining a <i>disease</i> : Disease is distribution or imbalance in <i>dhatu</i> s (and <i>doshas</i> ) and health is balance between them. Health is defined as <i>sukh</i> (happiness) and disease as <i>dukha</i> (sorrow). The aim of <i>Ayurveda</i> is to balance the <i>doshas</i> . . . . .	37
3.1	Essential components of a typical data acquisition methodology. . . . .	43
3.2	The steps involved in analog to digital conversion. . . . .	45
3.3	Our previous system to acquire the arterial pulse. It comprised of a pressure sensor with diaphragm, a transmitter, a digitizer, and an interface to computer at the parallel port. . . . .	48

3.4	The pulse signal on the top is acquired using our previous system comprising a '1 PSI' sensor and digitizer of 100 Hz sampling rate. The pulse signal on the bottom is acquired using our current system <i>Nadi Tarangini</i> (refer section 3.3). The signal on the bottom is replete with details and variations. . . . .	49
3.5	This is our setup to mimic the traditional pulse-based methodology. The three tiny diaphragms are placed exactly at the VPK locations, where the pressure values are captured, denoised, digitized, and finally stored on the computer. . . . .	50
3.6	Figure 2.13 shows the positioning of the <i>Ayurvedic</i> practitioner's fingers on subject's wrist. We built <i>Nadi Tarangini</i> to mimic this methodology. The three sensors are exactly of the size of fingertips. The assembly is properly shielded for the best performance. The display on computer shows the captured pulse signals. . . . .	50
3.7	The schematic diagram of <i>Nadi Tarangini</i> displaying important specifications of the components: Three sensors, three transmitters (or signal conditioning units), a digitizer, and storage. . . . .	53
3.8	Procedure to decide VPK locations. (a) The vertical axis is decided by manually sensing the pressure values at the three locations. (b) The first wrist line is considered as the horizontal axis. Their intersection point is the <i>vata</i> location. (c) VPK locations of acquisition. . . . .	55
3.9	Changes observed in the pulse signals as the applied contact pressure increases from Left to Right. . . . .	57
4.1	A pulse from [240]. This method was described for beat-to-beat estimation of changes in stroke volume. . . . .	64
4.2	A pulse from [178]. Difference curves between the peripheral pulses and the most proximal one in the thoracic were studied. . . . .	65
4.3	A pulse from [247]. Aortic flow pulsations were computed from arterial pressure by simulating a nonlinear, time-varying three-element model of aortic input impedance. . . . .	65
4.4	A pulse from [127]. Age-related changes were evaluated in pulsatile arterial function. . . . .	66

4.5	A pulse from [266]. The variations in pulse with increasing contact pressure were studied. . . . .	66
4.6	A pulse from [111]. A pulse contour cardiac output algorithm was developed based on frequency analysis studies of the arterial system. . .	67
4.7	A pulse from [6]. The age-related changes were demonstrated through pulse shape characteristics. . . . .	67
4.8	A pulse from [229]. The acute effect of cigar smoking was investigated on aortic elastic properties and wave reflection. . . . .	68
4.9	A pulse from [41]. The indices from systolic and diastolic pulse contour analysis were compared to correlate with traditional risk factors in cardiovascular disease. . . . .	68
4.10	A pulse from [258]. A classification approach using discrete wavelet model was proposed for pulse waveforms. . . . .	69
4.11	A pulse from [218]. Detection of pulse cycle's onset and end was carried successfully on the data. . . . .	69
4.12	A pulse from [237]. A method was demonstrated to distinguish rhythmic and arrhythmic pulse patterns using the Lempel-Ziv decomposition. . . . .	70
4.13	A pulse from [233]. A quantitative system for pulse diagnosis was constructed based on Bayesian networks to build the mapping relationships between pulse waves and pulse types. . . . .	70
4.14	A pulse from [259]. The comparison was made for pulse contour variability for before and after exercise conditions. . . . .	71
4.15	A pulse from [253]. The cardiovascular risk was investigated using the radial pulse. . . . .	71
4.16	The actual pulse signal (bottom) is obtained by integrating the pulse signal (top) acquired using the differential mode. The dotted lines indicate the corresponding important locations on both the signals. . . . .	73
4.17	An example of a pulse signal obtained from a subject's left hand. The three colors represent VPK locations of acquisition. . . . .	74
4.18	Typical behavior of a pulse signal. This is a <i>vata</i> pulse from a subject's left hand. It continuously and rhythmically follows P-T-V-D pattern. . . . .	74

4.19	<i>Vata</i> pulse signals of different disorders acquired using <i>Nadi Tarangini</i> . Different patterns show possibility of application of the rigorous data analysis algorithms for the identification. . . . .	75
4.20	Two examples of simultaneously obtained ECG and pulse signals. The top row, middle row and the last row respectively display the ECG, the pulse in differential mode, and the pulse after integration. . . . .	77
4.21	Fourier spectrum of a pulse signal (healthy behavior). The fundamental frequency provides the pulse rate. More subtle second order effects are the secondary peaks within each pulse cycle reflected in the subsequent harmonics. . . . .	78
4.22	Fourier coefficients for two pulse signals showing different behaviors. The variations in the two spectrums show possibility of classification. . .	79
4.23	Fourier spectrums of two different pulse signals showing different coefficient patterns. The secondary peaks in the pulse signals are usually reflected in the harmonics, and thus can be further used for the analysis.	80
5.1	Typical behavior of a pulse signal. A continuous, rhythmic P-T-V-D pattern is observed. Even though the pulse cycles look similar, they vary in amplitude and pseudo-period. . . . .	93
5.2	The top block provides the input pulse signal. Subsequent two blocks are thresholded values corresponding to positive & negative numbers of respectively real and imaginary coefficients. We redraw the input again in the last block but also overlay the detected peaks using our method. .	96
5.3	The top row provides the input pulse signal with five pulse cycles. The bottom figure shows the conventional PS representation using peak-to-peak segments. . . . .	98
5.4	The top row provides an input pulse signal with 203 pulse cycles. After peaks detection, the bottom subfigure shows our PS representation. Unlike Figure 5.3, this method captures the nature of the systole-diastole process. . . . .	100

5.5	Comparison of the PS representations. We show top views of the PS matrix, and colors represent the amplitudes (green has value zero). Notice the predominant green in the zero-padded region. (Figure best viewed in color.) . . . . .	101
5.6	The PSWT coefficients of the pulse signal from Figure 5.4. The A4, D4, D3, D2, D1 are the approximation and detail wavelet coefficients at various scales. . . . .	101
5.7	The PSWT cycle from the second row of the A4 matrix (from Fig. 5.6). We observe that the signal is smooth, and representative of the entire signal. . . . .	102
5.8	The signal of Figure 5.4 is split into four parts and the resulting four trend (PSWT) cycles are superimposed. The four trend cycles appear very similar. . . . .	103
5.9	An ill-behaved pulse signal is split into 3 parts. The second row displays three sets of three randomly selected pulse cycles. Finally, their trend (PSWT) cycles are superimposed. Though the individual cycles are ill-behaved, their trend cycles appear similar. . . . .	104
5.10	Time domain parameters of a cycle in the pseudo-periodic pulse signal. $w$ refers to the width of the P-subwave at 80% of its height). Significant locations termed <i>TDP locations</i> are marked on the X-axis. . . . .	105
5.11	The computation of the TDP locations becomes non-trivial for ill-behaved pulse cycles. Manual efforts may be needed. . . . .	106
5.12	Figure from [255] indicating missing zero-crossings. A search is made between the first and second zero crossing looking for values close to zero to obtain pseudo $h_2$ and $h_3$ parameters. . . . .	107
5.13	Maxima and minima information of a pulse cycle $O$ and its first and second derivatives are pruned to reveal TDP locations. . . . .	108
5.14	Our methodology for identifying TDP locations. Assessment is made in tandem on $O$ , $F$ and $S$ as described in the text. . . . .	109
5.15	Pseudo-locations may need to be identified for the ill-behaved pulse cycles depicted here. . . . .	111

5.16	The novel PSWT-DTW approach. Our approach can be used for analyzing any pseudo-periodic signals for computing domain-specific parameters; the only change will involve changing the method used in the lightly shaded brown block on the right. (Figure best viewed in color.)	114
5.17	The PSWT cycle (P1) is on x-axis, while a sample pulse cycle (P2) is on y-axis. The warping path indicates the maximum coincidence between the two. . . . .	117
5.18	Cycle on top is one of the cycles of pulse signal, while PSWT cycle is at bottom. Using DTW, we achieve nonlinear elastic matching between the two cycles. . . . .	117
5.19	Method overview. We wish to extract the TDP of individual cycles in a pseudo-periodic signal. The dotted, black path displays the conventional approach which employs a generic method. On the other hand, we use the PSWT (bold blue arrow) to capture the trend values (dashed, red arrow). Now this trend information is further used along with the dynamic time warping (green arrows), to obtain the TDP of all cycles. . . . .	118
5.20	TDP of a pulse signal with 197 pulse cycles. The black points indicate the TDP computed from individual pulse cycles. The TDP of the PSWT cycle is indicated by red circles. (y-axis values are normalized.) . . . . .	119
5.21	Comparison between the two approaches of computing TDP. Black points correspond to the second column (mean values) of Table 5.2 and red points to the third column (PSWT-based approach). The two colors follow each other, which makes PSWT cycle acceptable for mapping in the next step. (Figure best viewed in color.) . . . . .	121
5.22	Comparison showing the effectiveness of our novel PSWT-DTW approach (Approach 2). The green star, blue filled circles, and red circles (respectively) indicate the ground truth, results from Approach 1 (black path in Figure 5.19) and Approach 2 (red path in Figure 5.19) respectively. Approach 2 detects the TDP locations accurately even for ill-behaved pulse signals. . . . .	122



5.23	Typical behavior of a pulse signal. The API signal is the sequence of its peak-to-peak distances (696–678–700–718–724 milliseconds and so on). . . . .	125
5.24	Power Spectrums of pulse signals acquired from different subjects, computed using the nonparametric FFT spectrum approach. . . . .	128
5.25	Power Spectrums of pulse signals acquired from different subjects, computed using the parametric AR spectrum approach. . . . .	128
5.26	A sample API signal and its Poincaré plot. $SD1$ and $SD2$ are the dispersions of points perpendicular and along the axis of line-of-identity.	129
5.27	The elliptical area on the Poincaré plot shifts as age increases. . . . .	131
5.28	Poincaré plot of an arrhythmic pulse shows two distinct areas instead of a typical elliptical behavior. . . . .	131
6.1	(a) An example of an API signal. (b) $y(k)$ (in green), and local trends for $n = 20$ (in red) and $n = 40$ (in black). (c) Log-log plot of $F(n)$ versus $n$ shows a linear trend indicating the presence of self-similar nature. . . . .	137
6.2	Linear DFA plots for absolute and signed values of an API signal. . . . .	138
6.3	A sample pulse signal, its CWT, and corresponding WTMM tree. The branching structure of the WTMM tree is useful in extracting the hierarchical organization of the singularities. . . . .	140
6.4	(a) Sample pulse signal of diabetes behavior, (b) Continuous wavelet transform (CWT), (c) Wavelet transform modulus maxima (WTMM) tree, (d) Local Hölder exponents (LHE) at scale one, and (e) Probability density of LHE at scale 1. Note that the local Hölder exponents at red circles capture, e.g. , the singularity strengths at the peaks in the pulse signal. . . . .	143
6.5	(a) Sample pulse signal of skin disorder behavior, (b) Continuous wavelet transform (CWT), (c) Wavelet transform modulus maxima (WTMM) tree, (d) Local Hölder exponents (LHE) at scale one, and (e) Probability density of LHE at scale 1. Note that the local Hölder exponents at red circles capture, e.g. , the singularity strengths at the peaks in the pulse signal. . . . .	144

6.6	Comparison of probability densities of the local Hölder exponents of various pulse signals. The blue color is for the skin disorder behavior and the red color is for the diabetes behavior. Note the differences in the patterns of the densities for different wavelet scales. . . . .	146
6.7	Variations in multifractal spectra for three age-groups with normal behavior. . . . .	147
6.8	Variations in multifractal spectra for subjects in group G3 with normal, diabetes and skin abnormalities behaviors. . . . .	148
6.9	Recurrence plots of two pulse signals with different behaviors. We used embedding dimension 5, time delay 8, and radius 0.3. The structures of black points show the recurrence in the pulse signal, and is subsequently captured using RQA descriptors. . . . .	152
6.10	Sample pulse with typical behavior. It follows a continuous P-T-V-D rhythm. . . . .	156
6.11	Attractors of three different pulse signals using the DPS embedding. The viewpoint of the 3D rendering is such that the holes in the attractors are clearly visible. Though all the signals have P-T-V-D nature, the attractors show distinctive patterns. We want to capture these patterns quantitatively using topological invariants. . . . .	157
6.12	The projection of the DPS embedding onto the Y-Z plane clearly shows the flow directions. The plane bounded by X axis with Y = 0 serves as a Poincaré section. . . . .	158
6.13	A sectional cut on the attractor is shown through three views – XYZ, XY and XZ. Red dots indicate orbits going into the plane, while blue dots indicate orbits coming out of the plane. . . . .	159
6.14	Embedding of a pulse using (a) time-delay embedding with delay 6. (b) $x(t) \rightarrow (\int x, x, dx/dt)$ . Refer Figure 6.13 for comparison. . . . .	160
6.15	Pulse beats with different periods follow different trajectories in the attractor. Two pulse beats with period of 6 and 7 are highlighted. . . . .	161
6.16	Intersections between pulse beats of period 3 and 4. Systematic study of them reveals underlying patterns and information. . . . .	162
6.17	Two possibilities of intersection of the orbits. . . . .	164

6.18	LN matrix for periods 3,4,5,6. . . . .	165
6.19	Permutation matrix. . . . .	167
6.20	Crossing matrix. . . . .	167
6.21	Varying LN numbers for pulse signals with normal (black) and abnormal (red) behaviors indicate possibility for classification. . . . .	169
6.22	Contour diagrams indicating LN of pulse signals for respective periods. The colors in the contours indicate the amplitudes of LN, and the background color depends upon the range of the amplitudes. . . . .	171
6.23	Contour diagrams indicating RRR of pulse signals for respective orbits. The colors in the contours indicate the amplitudes of RRR. The symmetric RRR matrices form diagonal and parallel contours. . . . .	172
7.1	Possible classifying hyperplanes for a two-class problem. The bold line is the best hyperplane exactly at the center from both the classes. . . . .	176
7.2	Optimal hyperplane for two classes with labels +1 & -1. The two dashed lines are the margins and the points on them are the support vectors (SV). 176	176
7.3	Kernel action: Transformation of 2D space into 3D feature space . . . . .	177
7.4	One-against-all approach for multiclass classification. Here, class with label L4 is separated from all L1, L2 and L3. . . . .	178
7.5	One-against-one approach for multiclass classification. All the four classes L1, L2, L3 and L4 are separable (may not be the best hyperplane shown here for better visibility) from each other. . . . .	179
7.6	DDAG approach for classification of 4 classes, which are at the leaves. At each node of SVM binary classification, 2 classes (highlighted bottom numbers of each node) are selected of all the classes available at that node (top numbers of each node). . . . .	180
7.7	Handling imbalanced data in (a) conventional SVM. Prior methods (b) emphasize the creation of the model while building the SVM. In contrast, our method (c), while compatible with earlier methods, emphasizes the testing phase. . . . .	181

7.8	Intuition for Hybrid-SVM. (a) A representation of multiclass classification for four classes L1, L2, L3 and L4. L1 is huge as compared to the other three classes, and thus the classifying hyperplanes are biased. (b) When the three smaller classes are combined into one L5 class, the new hyperplane is not biased for L1 class. . . . .	182
7.9	Our Hybrid-SVM method. See text for details. . . . .	183
8.1	Our novel approach of splitting a pulse cycle into five subwaves (instead of conventional four subwaves) for capturing the variations at the v2 location using HMM segmentation. . . . .	196
8.2	The top sequence shows HMM model for normal behavior, modeled using 5-state HMM, while the bottom sequence shows that for a disorder with missing T-subwave, modeled using 4-state HMM. . . . .	196
8.3	The systematic methodology of our work to study different properties and patterns of the pulse signals. . . . .	202
8.4	Summary of our work. . . . .	203
A.1	A normal consistent <i>P-QRS-T</i> rhythm is exhibited on the left. Abnormal rhythm for a patient in every alternate beat appears on the right. . . . .	207
A.2	Sample ECG waveform, its continuous wavelet transform (CWT) and the corresponding local Hölder exponents at first scale. . . . .	209
A.3	probability density of local Hölder exponents. . . . .	213
B.1	Handling imbalanced data in (a) conventional SVM methodology. Prior methods (b) emphasize the creation of the model while building the SVM. In contrast, our method (c), while compatible with earlier methods, emphasizes the testing phase. . . . .	218
B.2	Sample ECG waveform, its continuous wavelet transform (CWT) with scales 1 to 20 in which the maxima lines are visible, and corresponding local Hölder exponents at scale 1. . . . .	219
B.3	Our Hybrid-SVM method. See text for details. . . . .	221
B.4	Highlighted points on the probability density function of the local Hölder exponents are used as features. . . . .	223

B.5 Receiver operating curves for the considered seven classes (N, P, AFIB, NOD, VF, BI and B) in our database. The abscissa in the subfigures are different for better clarity. The curve for the Hybrid-SVM case (with threshold 0.1) is above all others, and hence significantly better. . . . . 230

---

# Chapter 1

## Introduction

---

Analysis and use of the electrocardiogram (ECG) in personal medicine is a well established technique in the world. In contrast, pulse-based diagnosis, a prominently explored methodology in alternative medicine such as *Ayurveda* [104, 225] and Traditional Chinese Medicine (TCM) [50], has not yet received sufficient scientific recognition due to the lack of quantitative basis.

### 1.1 Motivation

Our work is inspired by the healing treatment in *Ayurveda*. The *Ayurvedic* practitioner diagnoses a subject by feeling the pulse at three predefined locations on wrist. There are differences between pulse signals of any two persons, varying with rhythm and patterns. The meaning of these differences is what *Ayurveda* has refined over 5000 years of its existence. The main aim of the traditional pulse-based diagnosis has always been to obtain useful information about what goes on inside the ‘entire’ body, what should have caused the disorder in the body, what might be done to rectify the problem, and to know what are the chances of success.

Even though the pulse-based methodology is very effective in the detection of symptoms and disorders in the entire body, there is still no quantitative information

available. Also, the methodology is person driven and requires high levels of concentration, and skill of “feeling” the pulse. Further, the information available in the form of books, lessons from *Ayurvedic* teachers, and our own experiences are qualitative.

In modern medicine, the examination of the pulse signal has only been used for computing the heart rate and blood pressure measurements. Further development of the quantitative analysis of pulse signals has many challenges such as heterogeneity of the data, volume and complexity, importance of the *Ayurvedic* practitioner’s interpretation, and most importantly, mathematical characterization of the pulse signals. As effective models of the automatic classification using pulse signals have not been reported, practical applications of the pulse-based methodology are severely limited.

## **1.2 Problem Definition**

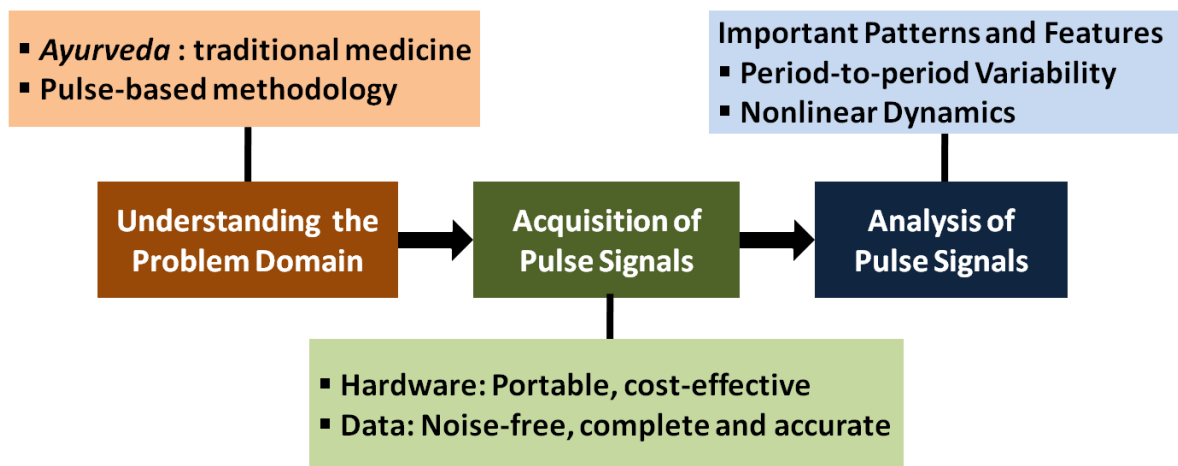
The focus of this thesis is to introduce and establish the importance and efficacy of the pulse signals through a scientific basis.

We wish to develop

- A portable and inexpensive acquisition system for recording noise-free pulse signals
- A variety of approaches to systematically study the variations in the pulse signals considering different dimensions such as age, gender, and disorders.

### **1.2.1 Methodology**

We wish to extract meaningful patterns from the pulse signals that are of interest, and hence useful to the user. For this purpose, we use visualization and analysis methods from the fields of pattern recognition, dynamical systems, and chaotic attractors. As the analysis of pulse signals is in the nascent stage, we follow the established procedure of the data analysis, as summarized in Figure 1.1.



**Figure 1.1:** Summarized in this figure is a typical data analysis procedure which involves problem definition, data acquisition and understanding, and finally data analysis. The other important blocks in the procedure of database, visualization, and storage are not displayed here.

In passing we note that, though the origin of the arterial pulse is in the rhythmic heart action, it is believed that the pulse also provides [54, 236] information about other organs. In other words, it is a complex system to understand.

1. **Acquisition of the pulse signals.** One of the most important requisites of any data analysis procedure is to acquire the noise-free, and complete data. A well understood acquisition (e.g. an electrocardiogram, X-ray machine) helps to know exactly what is wrong in a particular section of the body (which has caused the disorder). Similarly, it is critical that the information and structure of the pulse signal be faithfully preserved. The design should also consider important aspects such as the isolation of the subject, range of operation, sensitivity, signal-to-noise ratio, stability, and reproducibility.

Our goal is to develop a portable and inexpensive acquisition system to record noise-free pulse signals, considering all the above aspects. Also, we wish to mimic the measurement methodology of *Ayurveda* – ‘feeling’ the pressure values on wrist. At this stage, the pulse signals will be available in a quantitative form for visualization, and subsequent rigorous analysis.

2. **Analysis of the pulse signals.** Pulse signals vary in rhythm, amplitude, frequency, slopes and so on. A pulse signal with healthy behavior shows a typical



nature with regular rhythm, stable amplitude, and balanced shape. A disorder will cause alterations in the physiological processes and thus in the acquired pulse signals, leading to different patterns in one or several respects.

Our goal is to develop a variety of pattern recognition and data analysis methodologies, to extract useful features from the pulse signals, and to study variations with disorders (imbalance) in the body. Furthermore, the developed methodologies should be generic and conveniently usable in other disciplines.

### 1.3 Related Work

The beginnings of the noninvasive arterial pulse signal recording can be traced to 1555, where Strus [201] had proposed that the arterial pulse possesses a waveform. Although the instrumentation that he used was simple, he suggested that the changes in the arterial pulse shape and strength may be related to disorder conditions. However not much attention has been given so far, in this useful domain. Also, most of this research has been done based in Traditional Chinese Medicine (TCM).

In *acquisition*, photoelectric sensors [6, 10, 19, 71, 108, 114, 131, 170, 179] have been widely used to obtain signals from fingertips; this differs from the method adopted in *Nadi Vigyana* [104] in *Ayurveda*. Mechanical (pressure) sensors [2, 10, 35, 54, 105, 191, 263] are also popular due to their size, and their response to low frequency movements. The remaining blocks in the acquisition systems have usually been the same.

For *analysis*, different pattern recognition and data analysis algorithms have been applied to extract useful features (or hidden patterns) in the pulse signals. Examples include Lempel-Ziv complexity-based parameters, [262], Lyapunov exponents [265], approximate entropy (ApEn) [235], and fuzzy theory [107]. All such approaches study the correlation between the computed features and some 'property' of the pulse signal. We now turn towards work explicitly based on *Ayurveda*. Examples include computing [183] the pulse rate, detecting [168] peaks and other significant feature points, studying [208] the position specific nature, and studying [160] blood flow in various subjects.

However, the accurate and clean quantitative acquisition of the pulse signal based on *Nadi Vigyana*, followed by its extensive analysis to completely “understand” the pulse is still in the initial stages.

## 1.4 Our Contributions

The specific accomplishments are summarized in Figure 1.2. The work flow is spanned over all important steps of a typical data analysis procedure of Figure 1.1, including understanding, acquisition, preparation, and analysis of the data, followed by visualization of the results.

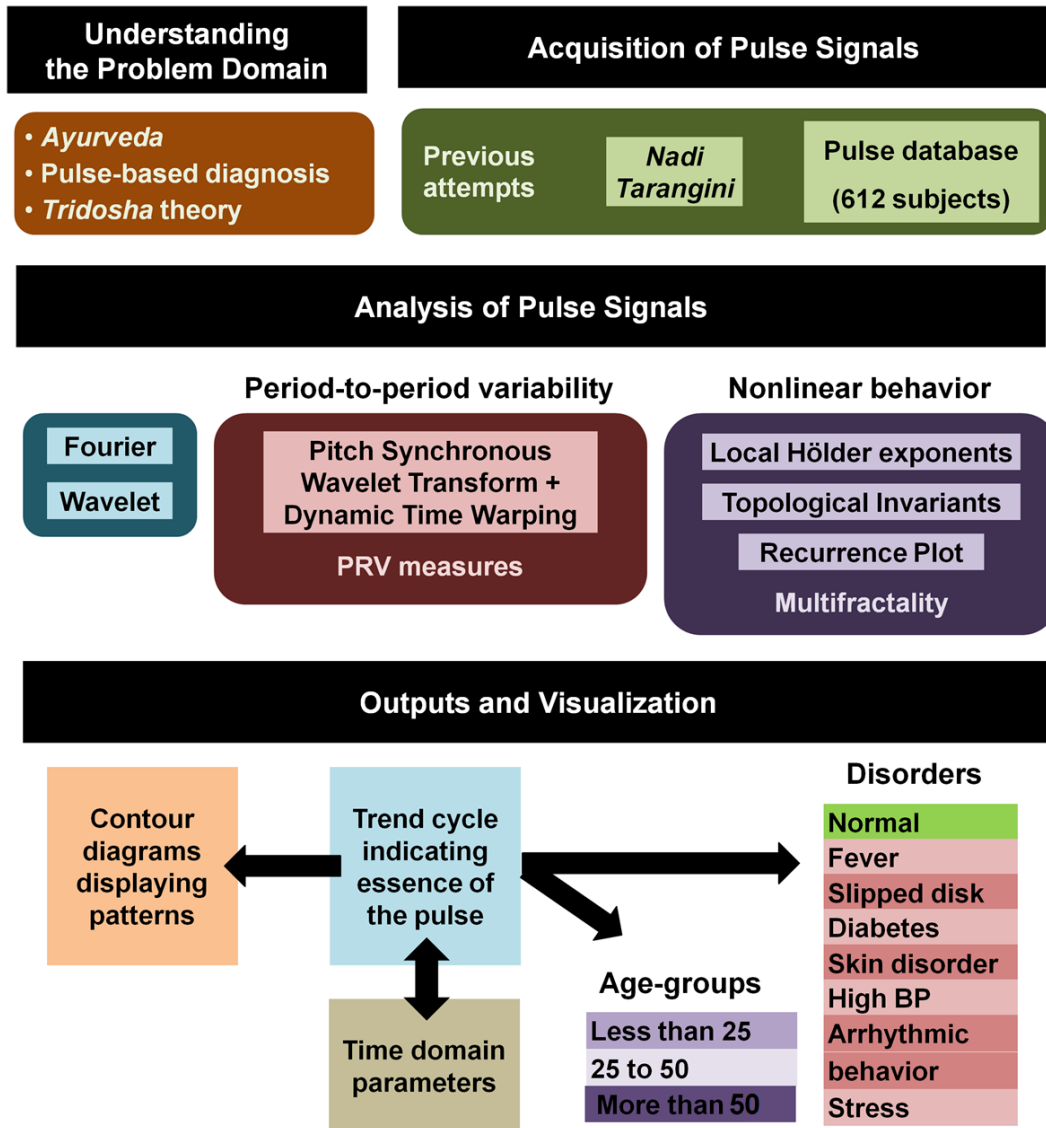
Fig. 1.2 is presented in the specific context of *Nadi Vigyana*. The techniques and methods are quite generic and can be potentially considered for other signals viewed as mathematical time series data. For example, the idea of the combinations of pitch synchronous wavelet transform and dynamic time warping (PSWT-DTW), in Section 5.4, is applicable to any pseudo-periodic signal. We now briefly provide more details of the various components of this thesis.

### 1.4.1 Acquisition of the Pulse Signals

The index, middle and ring fingers are used by the *Ayurvedic* practitioners to sense the arterial pulse, respectively at *vata* (V), *pitta* (P) and *kapha* (K) locations, as shown in Figure 1.3(a). To mimic this methodology, we have developed ‘*Nadi Tarangini*’, a data acquisition system using pressure sensors as shown in Figure 1.3(b).

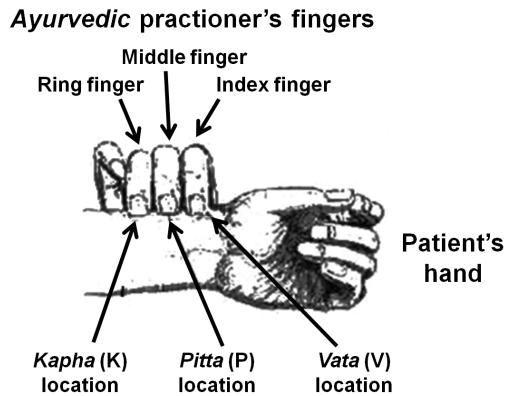
The three pressure sensors are placed at the VPK locations instead of the fingertips. They pick-up the pressure values at the respective locations and generate proportional electrical signals. The digitizer converts the analog values into discrete numbers. Finally the signals are stored on the computer in the time series format.

Very high attention was given to the significant blocks of the data acquisition process, comprising of sensors, the signal conditioning unit, digitizer, and storage. The size

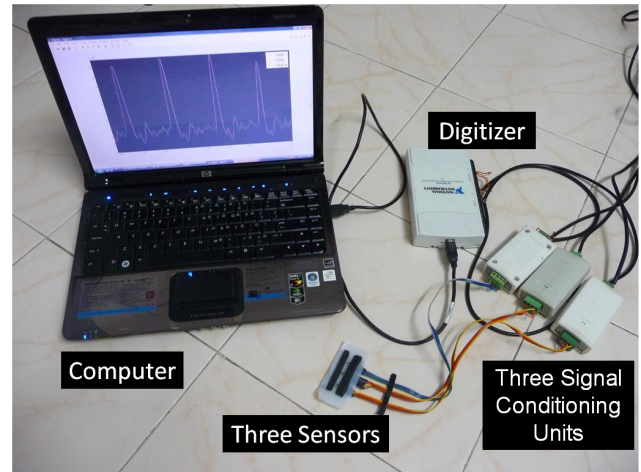


**Figure 1.2:** Summary of work in this thesis. Top Row: Based on understanding the domain, we design an acquisition system, and provide a corpus of data. Middle Row: The pulse signal is complex and can be studied using a variety of known mathematical techniques. Examples of these include multifractals, and topological invariants. Bottom Row: Based on these techniques, features can be obtained as output for a plethora of purposes including classification, and can also be suitably visualized.

(9×7 mm), and the pressure range (0–4 inch  $H_2O$ ) of each of our strain gauge sensors is optimized, to replace the *Ayurvedic* practitioner’s fingertip, to sense very minute pressure details (which otherwise requires high concentration and experience). In the signal conditioning unit, filtering and amplification is employed for the removal of the obvious noise, and providing strength to the signal respectively. Then the analog



(a) Positioning of the practitioner's fingers for sensing the arterial pulse



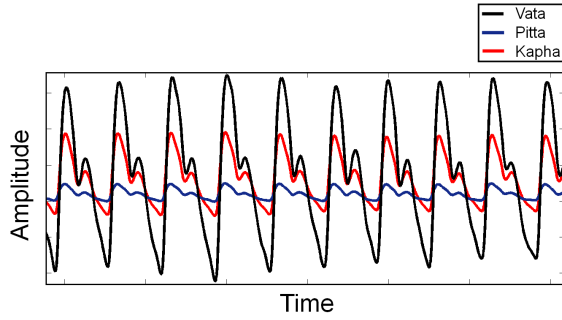
(b) *Nadi Tarangini* comprising of three pressure sensors, three signal conditioning units, a digitizer and computer

**Figure 1.3:** *Pulse-based acquisition: Pictorial comparison of the traditional methodology with our system.*

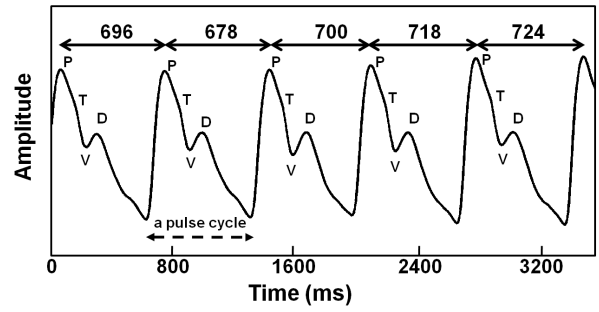
signal is converted into discrete values, using a digitizer of 16-bit resolution and 500 Hz sampling rate. In addition, to increase the accuracy of the data, we use the *differential mode* of acquisition with two inputs as the consecutive pressure values at each *dosha* location.

The three colors indicate three separate time series simultaneously obtained from VPK locations (refer Figure 2.13)

**Important properties of our data.** We have collected pulse signals of 612 subjects of varying age-groups (from 4 months to 75 years), gender, and disorders (such as high BP, slipped disk and diabetes). We have also stored the analysis and diagnosis details by our domain experts Dr. Ashok Bhat and Dr. Jategaonkar (these serve as the ground truth for our further analysis). The three colors in Figure 1.4 indicate three separate time series obtained simultaneously from VPK locations. Figure 1.5 shows a typical behavior of a pulse signal (*vata* location), comprised of a rhythmic pattern. As shown, this pattern is usually composed of the subwaves [272] a percussion subwave (P), a tidal subwave (T), a valley (V), and a dicrotic subwave (D).



**Figure 1.4:** *Sample VPK pulse signal from a subject’s left hand. We can observe the difference between the amplitudes and patterns of the three dosha.*



**Figure 1.5:** *Typical behavior of a pulse signal. A continuous, rhythmic P-T-V-D pattern is observed. Even though the pulse cycles look similar, they vary in amplitude and pseudo-period.*

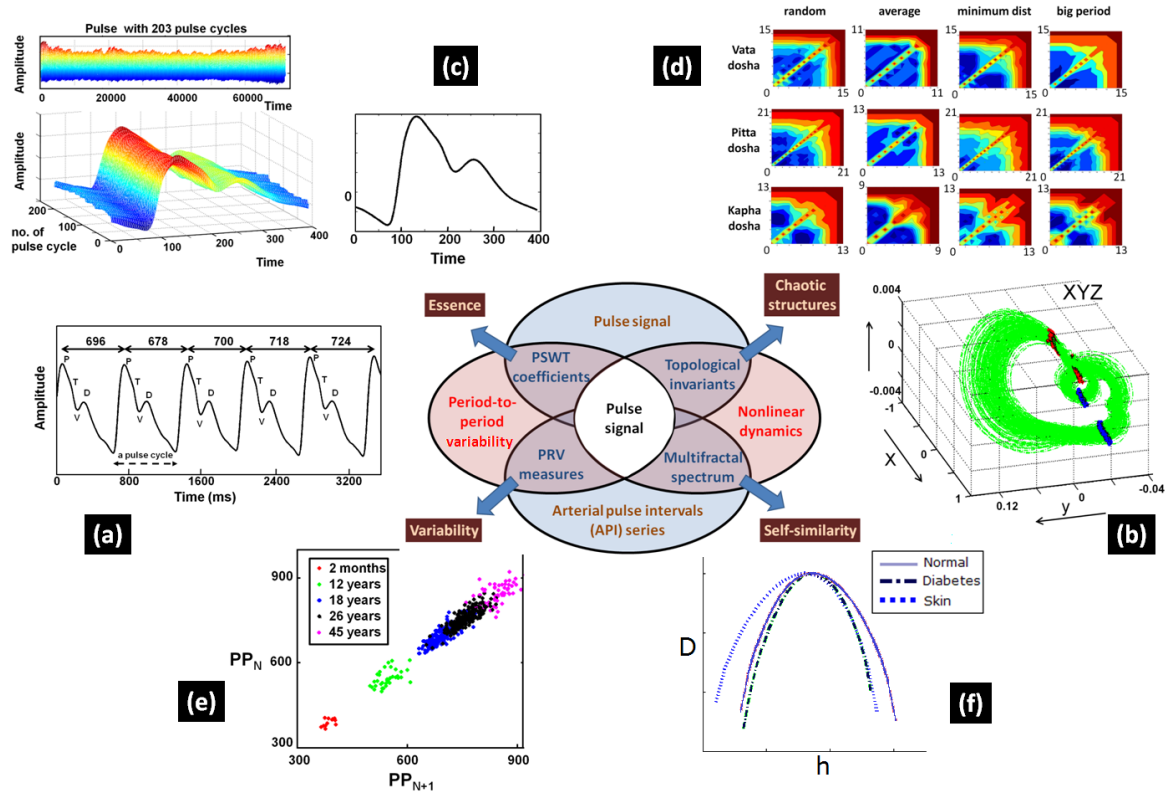
Our acquired signals are rich in harmonics. All the important properties including rhythm, rate, stability (amplitude and intensity), reproducibility and repleteness are captured in our pulse signals. There are distinctly observable patterns for different disorders which we consider for analysis in order to establish significant conclusions.

### 1.4.2 Analysis of the Pulse Signals

In this thesis, we have particularly examined two characteristic properties of the pulse signals, namely period-to-period variability and nonlinear dynamics. Figure 1.6 shows various aspects of our methodology.

In the figure, these two characteristic properties are shown in red text. Further, we have adopted two different ways (text in brown) of providing data for understanding: (i) the complete pulse signal and (ii) the sequence of arterial pulse intervals (API) (sequence of peak-to-peak distances 696–678–700–718–724 milliseconds and so on). We have analyzed the pulse signals in each of the combinations of the characteristic properties, and ways of input. For example, for studying the nonlinear dynamics in the pulse signal, we study its inherent topological structures.

We have developed a variety of approaches (text in blue), individually and in symbiosis, to make the pulse-based analysis useful and effective. Each method reveals



**Figure 1.6:** A global snapshot of the analysis portion of the thesis. Our work involves development of a variety of approaches to study different properties and patterns of the pulse signals. For example, multifractal analysis can eventually help understand variations with disorder. Details of this figure appear in the text.

a different ‘property’ (brown boxes) of the pulse, such as the pulse rate through Fourier spectrum, rhythm (Figure 1.6(c)) through pitch synchronous wavelet transform, inherent structures (Figure 1.6(d)) (through topological invariants), chaotic nature through recurrence plot based descriptors, period-to-period variability (Figure 1.6(e)) through various pulse rate variability (PRV) measures, and self-similar nature (Figure 1.6(f)) through multifractal spectrum. Each of these is now described in the subsequent chapters of this thesis. The outputs and progresses of the analysis, on the pulse signals, confirm its importance in the diagnosis.

### 1.4.2.1 Property: Nonlinear Dynamics

Intuitively the pulse signals display nonlinear dynamics due to the constant interaction between the sympathetic & parasympathetic systems [214], along with effect of other

physiological components. We quantitatively confirmed this with the Lyapunov exponents which are positive.

We showed that pulse signal exhibits a (chaotic) self-similar nature, and thus requires a large number of exponents to characterize their scaling properties. We showed the hierarchical organization of the singularities through the branching structure of the wavelet transform modulus maxima (WTMM) skeleton in the (scale,time) plane. We proposed a combination of wavelet scales in the computation of local Hölder exponents, resulting in better capture of the strength of singularities. In Appendix A, we show similar ideas applied for ECG signals. To extract the self-similar nature of the API signals, we computed the  $h$  versus  $D$  multifractal spectra. *Significantly, we showed that as the age of subjects increases, the spectrum drifts to higher values of  $h$ .* Furthermore, the spectra also vary for different disorders, and hence there is an opportunity for classification.

A century earlier Poincaré had observed that the key to understand a dynamical system, like an arterial pulse signal, lay in identifying its unstable periodic orbits. We established the pseudo-periodic P-T-V-D rhythm of the pulse signals, in terms of clusters of points on the Poincaré section. We further studied the intersections and crossings of orbits, of different pseudo-periods, through two topological invariants, namely linking numbers (LN) and relative rotation rates (RRR). *In terms of techniques* we proposed novel ‘minimum distance’ and ‘big orbit’ paradigms for selecting the representative pulse orbit of each pseudo-period. They help in improving the classification performance. We have also introduced ‘contour diagrams’ for the topological invariants to further help in their visualization, and to study their varying and intersecting patterns.

#### **1.4.2.2 Property: Period-to-Period Variability**

The pulse signal in Figure 1.5 seems to repeat the P-T-V-D pattern in each pulse cycle, but actually the signal contains some variations due to its pseudo-periodicity. These variations in the pulse cycles (and thus the peak-to-peak distances) are related to health [104], and a lot of information is conveyed in such pseudo-periodic patterns.

To obtain individual pulse cycles, we have developed the complex frequency B-spline wavelet ( $\psi$ ) based methodology, which exploits the relatively high amplitudes and steep slopes. Our methodology has achieved 100% detection accuracy. *In terms of techniques* this method could well be applied to other types of bio-medical signals.

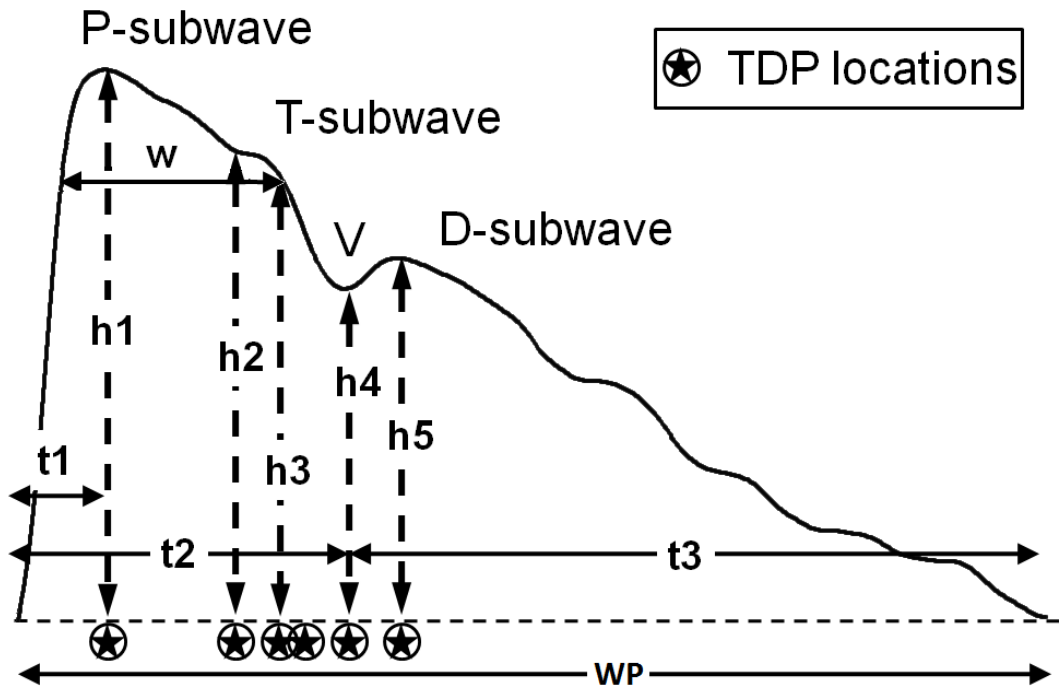
We have established the period-to-period variability to study the nonlinearly variable API signal by computing several PRV measures. Time-domain measures (such as mean and standard deviation) are computed directly from the raw API, and the differences between API. Frequency-domain measures (such as peak frequencies and normalized power) consider the thermoregulatory, vasomotor as well as respiratory activities in three frequency bands. Finally, the nonlinear Poincaré-based measures graphically capture the correlation between the consecutive API to summarize the entire (short or long) nonlinear signal. We showed that these PRV measures indicate the presence and absence of the arrhythmic patterns, and various disorders (such as diabetes and fever). We further showed that the measures have different ranges for age-groups of 'below 25 years', '25–50 years', and 'above 50 years'. *Our results showed the efficacy of the PRV analysis akin to the heart rate variability in clinical use.*

#### **1.4.2.3 Novel PSWT-DTW Approach**

A novel PSWT-DTW approach is established to compute various time-axis related parameters of a pseudo-periodic signal. We explain its effectiveness with an example of computing the useful time domain parameters (TDP) [257, 272] of a pulse signal shown in Figure 1.7. TDP are the amplitudes and widths at certain important locations, and are useful for diagnosis [153].

Providing more details of our novel approach, the pitch synchronous wavelet transform (PSWT) helps in representing such signals better in terms of a periodic trend plus period-to-period fluctuations. The dynamic time warping (DTW) enables a nonlinear mapping of one signal to another, even if they are out of phase in the time axis, by minimizing the distance between the two. The steps of our novel approach are shown in Figure 1.8. First, the PSWT cycle (essence) of the complete pulse signal is extracted using the PSWT. Then, trend TDP locations are computed on this smooth cycle, and



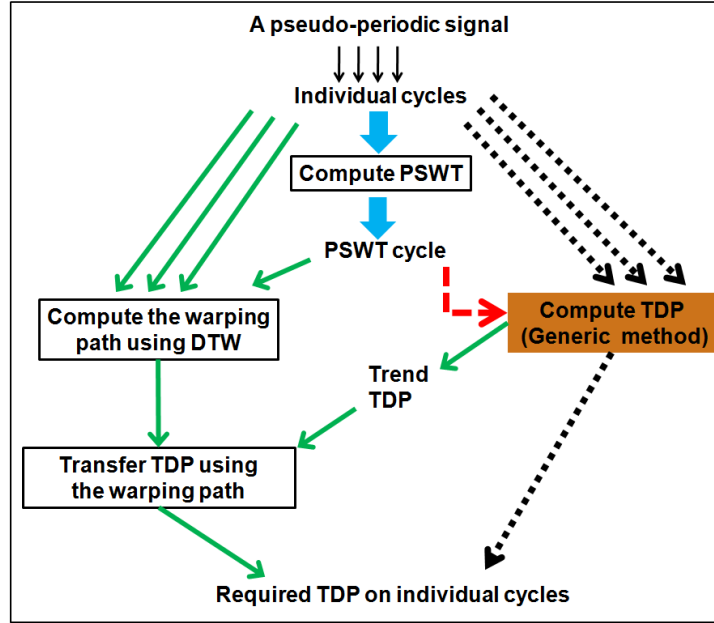


**Figure 1.7:** Time domain parameters of a cycle in the pseudo-periodic pulse signal.  $w$  refers to the width of the P-subwave at 80% of its height). Significant locations termed TDP locations are marked on the X-axis.

finally are transferred back on each of the pulse cycles using the nonlinear DTW map.

Specifically, the advantage of our novel approach is in reducing the tedious efforts of searching the parameters (especially in the ill-behaved cases) at two stages. In the computation of TDP, we have developed a maxima-minima based approach to avoid the searching of pseudo parameters arising due to the false negatives of missing zeros. Furthermore, our novel PSWT-DTW approach helps for two reasons. First, the resulting trend (PSWT) cycle is smooth in nature and second, the efforts (if any) of tuning the TDP are only for one PSWT cycle.

Our results show that the novel approach is feasible and effective. *Its applications can be extended to other signals with pseudo-periodic behavior, to study both average values as well as period-to-period variations.*



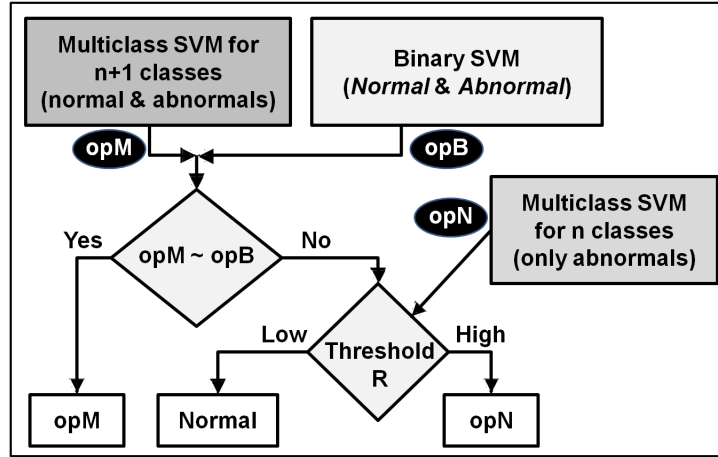
**Figure 1.8:** The steps in our novel PSWT-DTW approach to compute time-axis related parameters on individual pulse cycles. The dotted, black path displays the conventional approach.

### 1.4.3 Novel Hybrid Support Vector Machine

We observe that biomedical signals of normal behavior are more common, and thus are widely available as compared to signals with abnormalities. Thus, in the multiclass classification, the classifier usually gets trained with more bias towards the huge *Normal* class.

We have introduced the notion of Hybrid Support Vector Machine – a combination of binary and multiclass SVM – to increase the testing accuracy for such imbalanced datasets. When all the abnormal (say  $n$ ) classes are combined into one class (say *Abnormal* class), the *Normal–Abnormal* (binary) classification is unbiased when compared to the individual (multiclass) classification. Our novel approach is explained in Figure 1.9.

When a test case enters the system, we first compare  $opM$  and  $opB$  coming from the above mentioned two sub-systems. If the predicted outputs are consistent, then we use  $opM$  as the final predicted output. On the other hand, if they are not consistent, then we assume that the incompatibility is due to the bias in the trained SVMs. Therefore,



**Figure 1.9:** The method of our Hybrid-SVM approach. The outputs of SVM blocks ( $opM$ ,  $opB$ , and  $opN$ ) are suitably compared to increase the overall efficacy.

we devote more care to the  $n$  abnormal classes by inserting a threshold value  $R$  in the decision box. We present a detailed description of selecting  $R$  in Appendix B.

Our Hybrid-SVM approach particularly reduces the false negatives (situations in which abnormal signals are erroneously classified as normal).

#### 1.4.4 Classification

We performed the classification of the pulse signals of various subjects, with TDP as input vectors to two classifiers (conventional SVM and our Hybrid-SVM). To illustrate our approach, we used the TDP computed using our PSWT-DTW approach. Our dataset contains a total 4987 pulse cycles of 10 different behaviors including normal, diabetes, high BP, and skin disorder. We first showed that the feature sets perform well in classification. Further the Hybrid-SVM improves the performance, by reducing the false negatives.

Our investigations show the immense potential benefit of both the novel PSWT-DTW and Hybrid-SVM approaches, respectively for the pseudo-periodic signals, and the imbalanced datasets. We believe the two approaches can be suitably extended to other biomedical as well as non-biomedical applications.

In this thesis, we have combined both the traditional and modern approaches – by modeling the pulse-based approach into the developed hardware (*Nadi Tarangini*), and by developing algorithms to extract useful patterns. Thus, we have brought the subjective experience (of “feeling” the pulse) to the objective level (analysis of the quantitative data).

As a matter of fact, years of development in the analysis of the ECG-based methodologies has resulted in advanced automated tools. In the future, pulse signals can be equally effective for diagnosis. Further attention is required to achieve high respect for the automated pulse-based analysis.

## 1.5 Organization of Thesis

In Chapter 1, we introduced the pulse-based methodology in Indian traditional medicine, provided the problem definition, and the goals of this thesis, and summarized the contributions of our work.

In Chapter 2, we briefly explain a cardiac cycle, and different theories dictating the formation of the pulse signals. We provide some of the domain knowledge of Indian traditional medicine, pulse-based diagnosis (PBD), and the importance of PBD in medicine. The purpose of this chapter is to elaborate the basic concepts to a novice – to understand the complexity involved in PBD, the need of quantitative analysis, and also to understand the efficacy of our system explained in the following chapters.

In Chapter 3, we first describe the steps and salient features of a typical data acquisition assembly. Then we elaborate on our system *Nadi Tarangini*, to acquire noise-free pulse signals, enabling it to reflect the “feeling” information used in the *Ayurveda*. Finally, we provide the details of our pulse database.

In Chapter 4, we provide the important properties of our pulse signals such as rate, rhythm and frequency. We compare our pulse signals with those of prior approaches to show that our data contains high details. This chapter prepares the background of *possible distinguishing patterns* for the subsequent chapters of the data analysis. We list

out the important prior works in this domain.

In Chapter 5, we study the physiologically important period-to-period variability of the pulse signals. We present our novel PSWT-DTW approach to compute the time-axis related parameters, taking advantage of the pseudo-periodicity of the signals. We explain its efficacy with an example of computing useful time domain parameters of the pulse signal. Next, we explore the variability by computing various time-domain, frequency-domain and nonlinear Poincaré based measures. We further give evidence for the classification of disorders in subjects using these PRV measures.

In Chapter 6, we study the inherent nonlinear dynamics of the arterial pulse signals. We systematically study the chaotic properties, fractal nature, and structures of the pulse signals, respectively through the recurrence plot-based analysis, local Hölder exponents & multifractal spectrum-based analysis, and the topological analysis. We also analyze their variations for different subjects.

In Chapter 7, we show the efficacy of our novel PSWT-DTW and Hybrid-SVM approaches through the improvements in the classification.

In Chapter 8, we present the pictorial summary of our work. We describe applications and future directions of our work. We conclude the thesis on a positive note that we have successfully established the methodology of acquisition and quantitative analysis of the pulse signals.

In Appendix A, the approach of multiclass classification of arrhythmia signals is presented using the idea of the local Hölder exponents. In Appendix B, we present our novel Hybrid-SVM approach to reduce the false negatives in the classification of imbalanced arrhythmia datasets. We also show that the combination of wavelet scales for the computation of Hölder exponents increases the classification accuracy. Finally in Appendix C, we provide a few examples of the *Tridosha* (VPK) pulse signals from our database.

An organizational chart is given in Figure 1.10 and is repeated at the beginning of each chapter to highlight the contents.

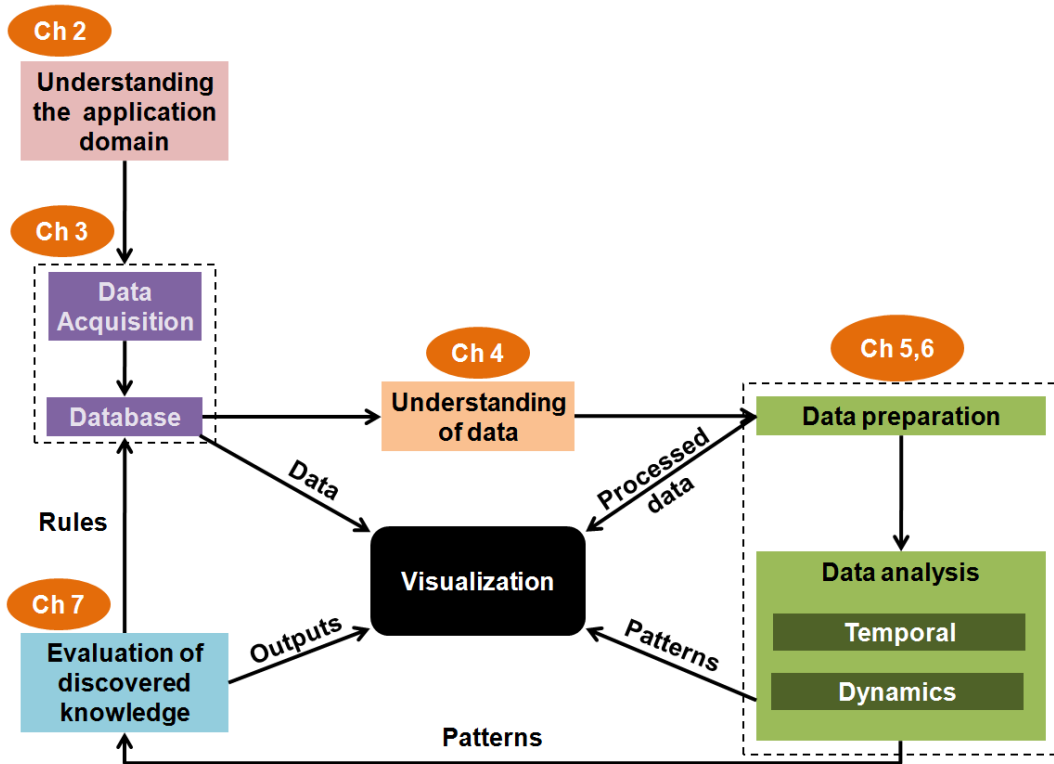
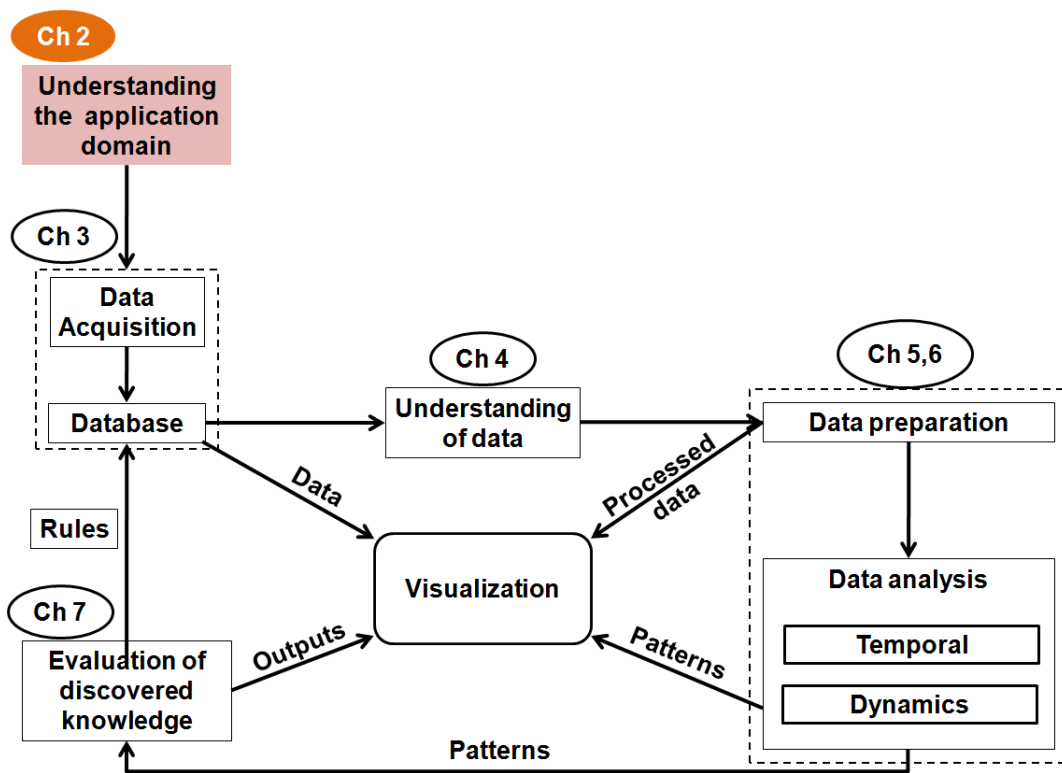


Figure 1.10: Organization chart for the thesis.



---

## Chapter 2

# Pulse Signals in Medicine

---

*The circulatory system of a human body (a) carries nutrients, gases, and wastes to and from cells, (b) helps fight disorders, and (c) helps [68] stabilize the body temperature and pH.*

*In this chapter, first we briefly explain a cardiac cycle, an electrocardiogram (ECG) that quantitatively captures the heart actions, blood flow, and different theories dictating the formation of the pulse. Then we provide the required domain knowledge about the Indian traditional medicine, pulse-based diagnosis (PBD), Tridosha theory, and the importance of PBD in Ayurveda.*

*The purpose of this chapter is to elaborate the basic concepts to a novice, to understand the complexity involved in PBD, the need of the quantification, and also to understand the efficacy of our system explained in the subsequent chapters.*



The circulatory system is composed of the heart and blood vessels, including arteries, veins, and capillaries. The human body has two circulatory systems [68, 149] as:

- The pulmonary circulation is a short loop from the heart to the lungs and back again.
- The systemic circulation, usually described as the circulatory system that sends blood from the heart to all other parts of the body and back again.

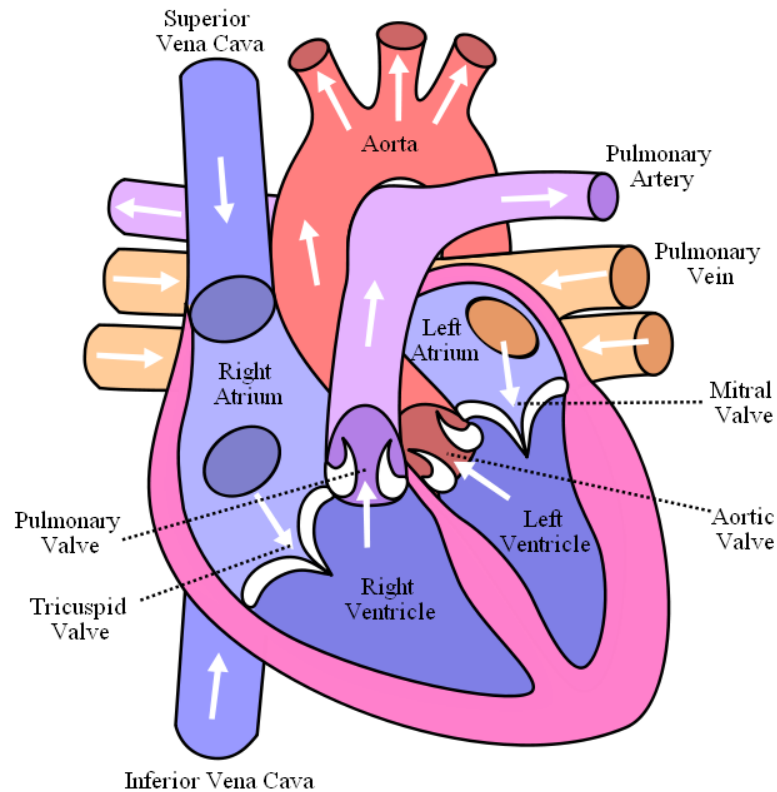
The function of the circulatory system is to carry blood throughout the body. To achieve this goal, the heart muscle contracts rhythmically and systematically.

## 2.1 A Cardiac Cycle

Cardiac cycle is the term referring to all events related to the flow of blood that occurs from the beginning of one heartbeat to the beginning of the next. Each of these beats takes approximately 0.8 seconds and the same sequence of events occurs in every beat. A cross-section of the heart is shown in Figure 2.1.

There are four chambers in the heart – two atria and two ventricles [68]. A cardiac cycle can be divided into three separate segments: a diastole (relaxation), an atrial systole (contraction of atria), and a ventricular systole (contraction of ventricles).

- **Diastole:** During this segment, blood flows into the heart. Blood comes into the atria through the pulmonary veins (left atrium) and the superior and inferior vena cava (right atrium). From the left and right atria the blood flows directly into both ventricles because both the mitral and tricuspid valves are open. The ventricles accept this blood and passively dilate (expand). By the end of the diastole, the ventricles are filled to about 80%. The cardiac muscle is relaxed throughout this segment. The aortic and pulmonary valves are closed during this segment to prevent the reverse flow of blood from the aorta and pulmonary artery.
- **Atrial Systole:** This is a relatively short segment. Atrial systole is the contraction



**Figure 2.1:** A cross-section of the heart showing its important portions [250].

of the heart muscle (myocardia) of the left and right atria. Normally, both atria contract at the same time. As the atria contract, the blood pressure in each atrium increases, forcing additional blood into the ventricles.

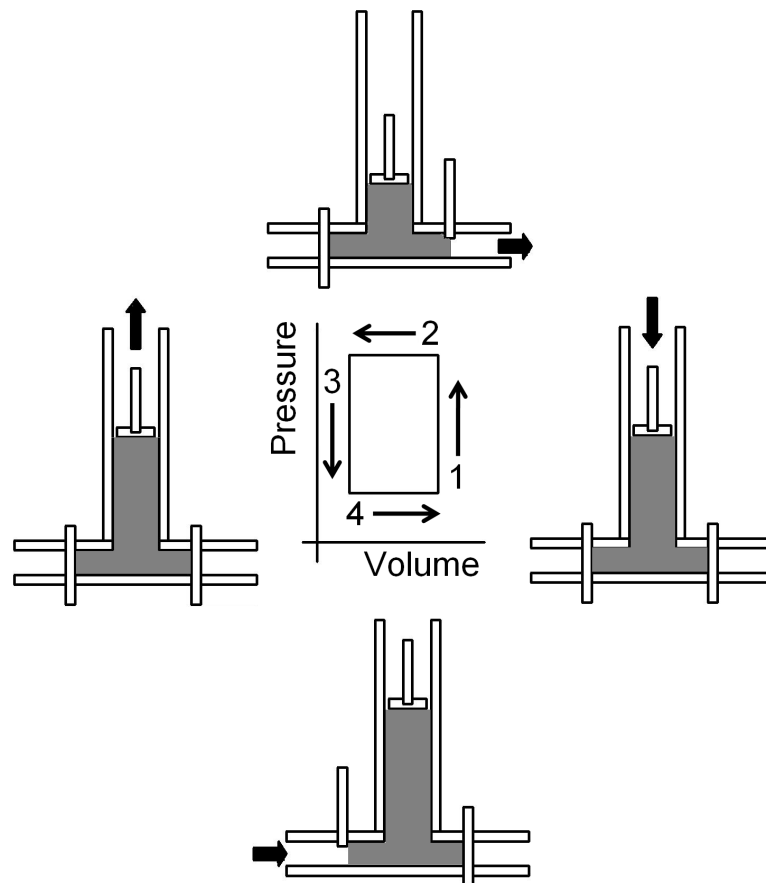
- **Ventricular Systole:** During this segment the ventricles contract. As they begin to contract, the mitral and tricuspid valves close to prevent any back flow into the atria. At the end of this segment the pulmonary and aortic valves close and diastole begins again.

### 2.1.1 Modeling the Pumping Action of the Heart

Figure 2.2 shows the modeling of one complete cardiac cycle.

The four phases are [68]:

1. **Isovolumic contraction phase:** In this phase, the mitral and aortic valves are closed. Pressure is being built up in this phase, until the left ventricular pressure



**Figure 2.2:** The pressure-volume loop [31] explaining the four phases in a cardiac cycle. The first two phases are known as *diastole* and the last two phases are known as *systole*.

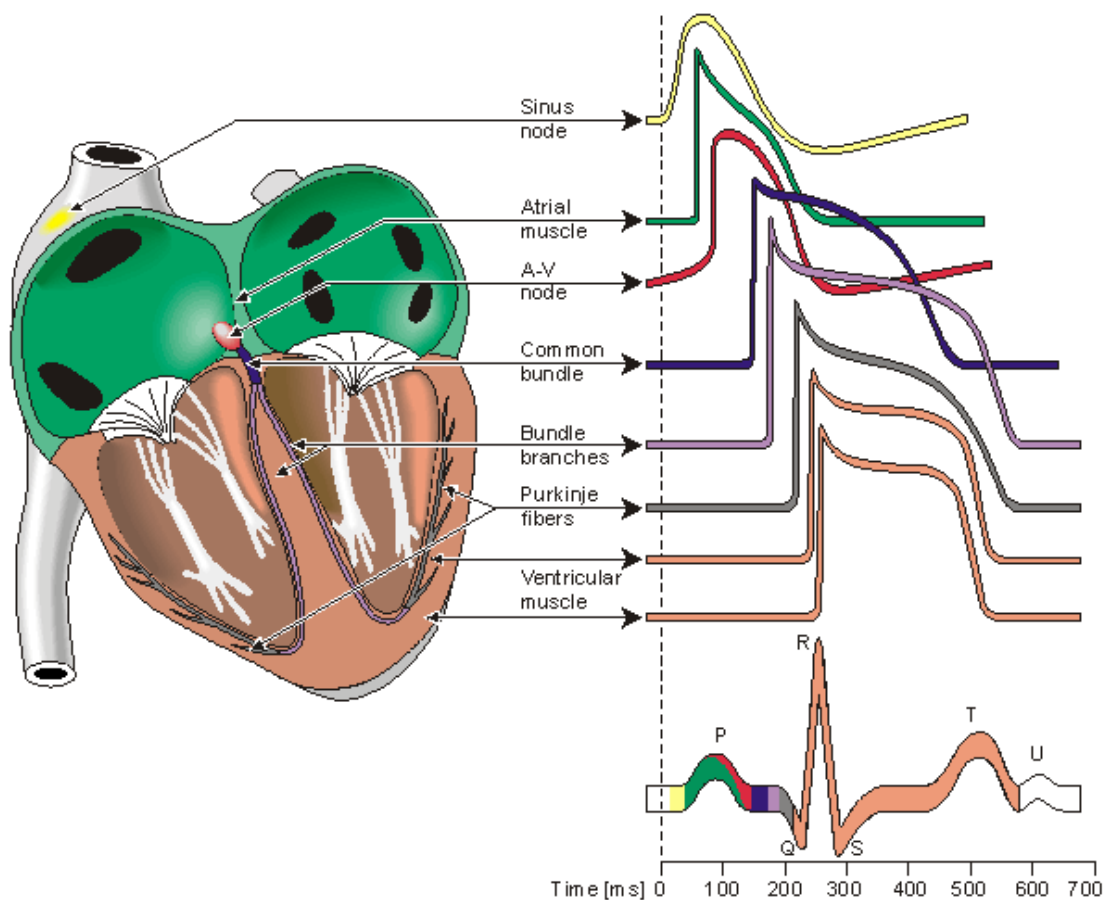
risers sufficiently above the aortic pressure to open the aortic valve.

2. **Ejection phase:** In this phase, the mitral valve is closed and the aortic valve is open. Blood flows out of the chamber into the aorta.
3. **Isovolumic relaxation phase:** In this phase, the mitral and aortic valves are closed. Pressure in the chamber decreases until it is so low that the mitral valve opens.
4. **Filling phase:** In this phase, the mitral valve is open and the aortic valve is closed. Blood flows into the chamber, and the cycle then repeats itself.

The first two phases are known as *diastole* and the last two phases are known as *systole*.

## 2.1.2 Acquisition of the Heart Action: An Electrocardiogram

The electrical waves in both the systole and diastole can be measured at selectively placed electrodes (electrical contacts) on the skin. Electrodes on different sides of the heart measure the activity of different parts of the heart muscle. An ECG displays the voltage values between pairs of these electrodes [68]. The ECG waveform indicates the overall rhythm of the heart, and weaknesses in different parts of the heart muscle. Figure 2.3 explains how a cardiac cycle is recorded in one particular ECG lead.



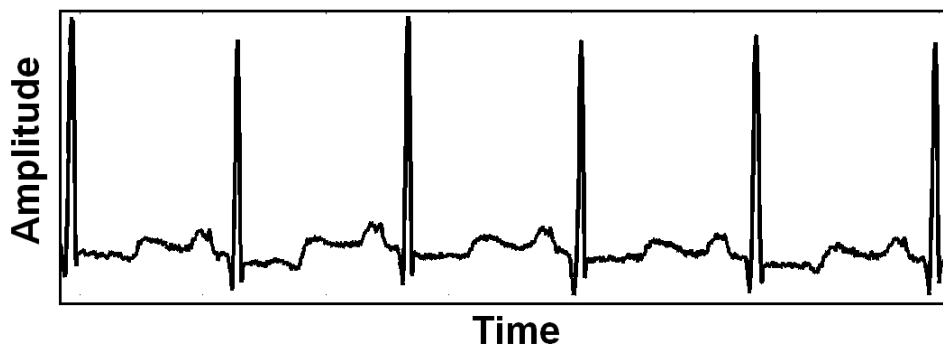
**Figure 2.3:** Recording of one cardiac cycle using an ECG lead [118]. The left subfigure denotes different parts in a human heart. Different color codings are used in the right subfigure to indicate the location and strength of the cardiac action in a particular heart portion. Finally, all these activities sum up to provide complete ECG cycle at the right bottom subfigure.

The *P-wave* is the result of the atrial depolarization. The *QRS-complex* is the average of the depolarization waves of the inner (endocardial) and outer (epicardial) cardiomyocytes. The *T-wave* represents the repolarization of the ventricles. *U-waves* are thought

to represent the repolarization of the papillary muscles or Purkinje fibers.

Therefore, one complete heart beat is consisted of an atrial depolarization → atrial contraction → P-wave, ventricular depolarization → ventricular contraction → QRS-complex → the resting phase (including the repolarization during the T-wave) between two heart beats.

Cardiac muscle has automaticity, which means that it is self-exciting. The heart's rhythmic contractions occur spontaneously. An important thing to note that the rate of contraction can be changed by nervous or hormonal influences, exercise and emotions and is responsible for the heart rate variability (HRV – explained in Chapter 5). Figure 2.4 provides an example of the ECG signal composed of several cardiac cycles one after the other. The frequency of such cardiac cycles is the heart rate.



**Figure 2.4:** *Sample electrocardiogram composed of 5 cardiac cycles. This is signal number '100' from the arrhythmia database of NIH [63]).*

Currently, ECG acquisition and its analysis is in the advanced stage. In this thesis, we introduce and establish the arterial pulse-based analysis.

### 2.1.3 Cardiac Output

Cardiac output (CO) is the amount of blood ejected from the ventricles and is a measure of the pumping ability of the heart [251]. It is determined by the cardiac rate and stroke volume.

$$\begin{array}{ccccc} \text{cardiac output} = & \text{cardiac rate} & \times & \text{stroke volume} & \\ (\text{ml}/\text{min}) & (\text{beats}/\text{min}) & & (\text{ml}/\text{beats}) & \end{array}$$

where,

- Cardiac rate is based on the natural rhythm of the SA node [120]. Sympathetic and parasympathetic stimulation continuously modify the cardiac rate by increasing or decreasing the rate of depolarization.
- Stroke volume is the amount of blood pumped per beat by each ventricle. It is regulated by three parameters
  - ◊ **End-diastolic volume (EDV):** An increase in the volume of blood in the ventricles prior to contraction, results in an increase in the ejection volume.
  - ◊ **Contractility:** A more forceful ejection results in a more complete emptying of the ventricles.
  - ◊ **Total peripheral resistance:** A higher resistance, or increase in arterial pressure, results in less blood entering the aorta.

Thus, stroke volume is directly proportional to EDV, contractility, and inversely proportional to total peripheral resistance.

#### 2.1.4 Blood Flow

The pumping action of the heart (refer to Figures 2.3 and 2.2) generates blood flow that passes through the arteries and veins. The blood flow (mainly) depends directly on the pressure difference between two ends of the blood vessel, and indirectly on the resistance to the blood movement.

*Blood Pressure (BP)* is the force produced on the walls of the blood vessels as a result of the blood flow. The pressure gradient (difference in pressure) is what keeps blood flowing, always from high to low pressure. The term blood pressure used by common man refers to the arterial pressure (the pressure in the larger arteries) measured via a sphygmomanometer [249].

For each heartbeat, blood pressure varies between systolic and diastolic pressures. Systolic pressure is the peak pressure in the arteries, which occurs near the beginning

of the cardiac cycle when the ventricles are contracting. Diastolic pressure is the minimum pressure in the arteries, which occurs near the end of the cardiac cycle when the ventricles are filled with blood. Actually, the pressure is generated when the flow is opposed by some resistance.

**Peripheral Resistance** is the total resistance to the flow of blood in the systemic circuit [68]. It is a measure of the amount of friction, blood encounters as it passes in the blood vessels. The resistance depends on:

- **Blood viscosity:** Internal resistance to the flow is related to the thickness of blood. As the blood viscosity increases, the peripheral resistance increases and thus the blood flow decreases.
- **Blood vessel length:** As the length of the travel of blood in a vessel decreases, its resistance decreases and thus the blood flow increases.
- **Blood vessel diameter:** Smaller diameter results in more friction because more fluid is in contact with the blood vessel wall, which increases the resistance and thus decreases the blood flow.

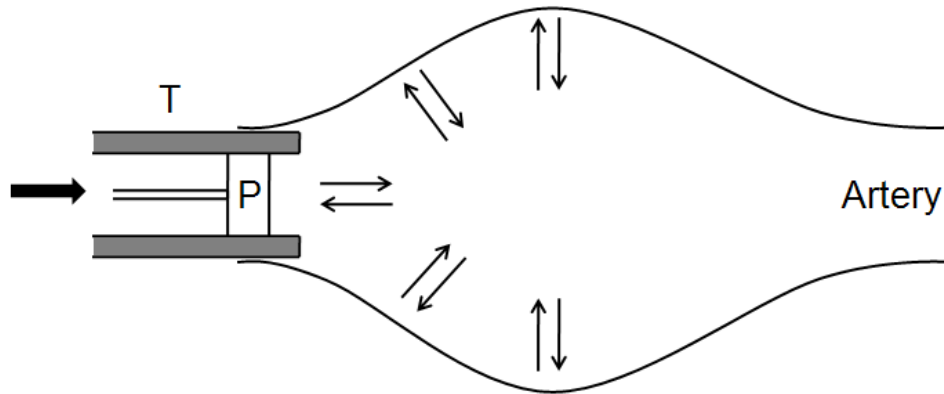
It is assumed in the Indian traditional medicine that all such factors (and many others including blockage) are generated due to the malfunction of different organs or the unbalanced *dosha*. They are ultimately sensed using the arterial pulse on wrist.

## 2.2 Formation of the Pulse Signal

As explained earlier, the maintenance of the blood flow and pressure gradients in the arterial circulation is managed from the energy transfer between the heart and arteries [133, 195]. There are many ways to explain the complex behavior of the pulse (i.e. throbbing of the arteries). These theories have different approaches with their own assumptions and explanations.

## 2.2.1 Elastic Property of the Arterial Wall

Consider an arrangement of Figure 2.5 [225], where a tube ( $T$ ) is fastened with the piston ( $P$ ) and further tied into an artery.



**Figure 2.5:** The effect of pressure from the piston  $P$  on the elastic arterial walls. This resembles the actions in the human circulatory system.

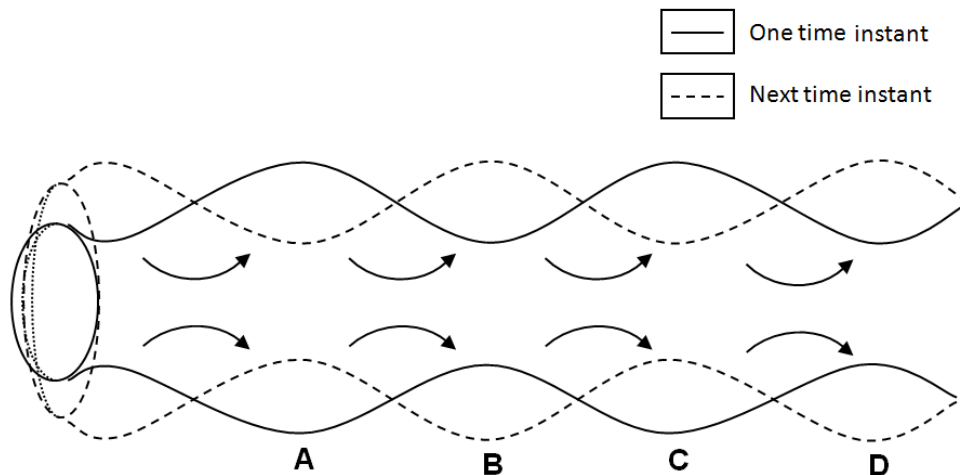
If the piston is pushed forward towards the right into the tube, the vessel distends locally. This occurs simply because the blood is *incompressible*. This immediate incompressible column of blood produces a local increase in pressure that has a little effect on the advancing piston, which in fact is a mass of incompressible blood. Instead, the next section of the artery is stretched so that a wave of pressure travels along the vessel wall without involving actual transmission of the blood.

This means that the pulse is actually a wave set-up in the walls of the vessels by the systole of the ventricles and it is not only due to the passage of blood along the arteries.

As shown in Figure 2.6, the sudden ejection of the large quantity of fluid into a distensible tube (*point A*) causes bulging of the tube at the point of ejection, and the pressure in the area of the tube rises.

Immediately after the vessel wall is distended at *point A*, the elevated pressure at this point forces a small amount of fluid along the vessel, as illustrated by the arrows. This causes the vessel wall to distend at *point B*, while at the same time the vessel begins shrinking at *point A* (dashed line). After *point B* is well distended (high pressure), the elevated pressure at this point causes fluid to flow to *point C*, distending the vessel at





**Figure 2.6:** *The effect of sudden ejection of a fluid (blood) into an elastic, distensible tube (e.g. an artery). The arrows indicate the immediate effect. This resembles the actions in the human circulatory system.*

point C and relaxing the vessel wall at point B.

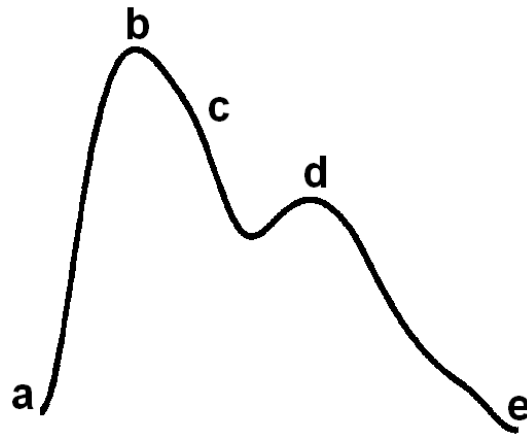
It means that *the elastic property of the arterial wall* takes part in the generation and transmission of the pressure wave of the pulse. Although the pressure wave or the pulse may reach the vessel of the foot in less than a second, it requires several heart beats for the blood which enters the ascending aorta to reach the foot vessel. This is because the speed of the blood depends on factors such as the *blood pressure gradient, viscosity, and cross-sectional area of the blood vessel.*

## 2.2.2 Hemodynamics of Pulsating Stream

Figure 2.7 [134] displays the important points on a pulse cycle.

There occurs a great forward movement of blood [194] together with steep increase in pressure (*a-b*). During the next segment (*b-c*), the pressure mounts with a slower gradient to a summit but the velocity of flow gradually decreases. During remainder of the systole (*c-d*), both the pressure and flow decrease. The time of closure of semilunar valves decreases the beginning of the diastole as shown at (*d-e*).

It is quite obvious that the conversion of the potential energy stored during the systole



**Figure 2.7:** Important points on a pulse cycle form various segments indicating the stages of systole and diastole.

into the kinetic energy of the flow during the diastole, increases the continued flow through the capillaries [152].

Actually, the percussion (P) subwave (refer to Figure 1.5) first rises due to rapidly transmitted shock of the left ventricular contraction and then falls towards the end of the systole. It also represents that the escape of blood into the blood vessels. The fall towards the end of systole is due to the reduced ejection phase of the ventricle delivering less blood than is passing out of the arterioles. The tidal (T) subwave is caused [53] due to the summation of the uncompleted P-wave and waves reflected from the periphery. The descending limb of the pulse is due to a more gradual fall in the pressure corresponding to the elastic recoil of the arterial walls.

### 2.2.3 Forward-Backward Wave

The heart generates forward-traveling wave energy [147] that propagates through the arteries to maintain tissue and organ perfusion for metabolic homeostasis [267].

At the beginning of each cardiac systole, an individual forward-traveling wavefront initiates the flow and thus increases pressure in the arteries. Although most of the wave energy in this initial compression wave travels distally into smaller arteries, some is reflected back toward the heart at sites of impedance mismatch. These include focal sites such as bifurcations [186] and the blockages in the arteries. *Therefore the forward-*

*backward combination will depend upon the conditions of the malfunctioning organs that affect the partial or complete blockage to the blood flow.*

These reflected waves can further be re-reflected [97]. At the end of systole, deceleration of left ventricular contraction results in reduction in both pressure and flow in the ascending aorta mediated by a forward-traveling expansion wave [158]. Interactions between the forward and backward-traveling waves result in complex patterns of the blood flow and pressure change at different points in the arterial circulation. These complex patterns are acquired at the wrist location, and the subsequent analysis of the forward and backward-traveling wave energy therefore provides important information [88, 175].

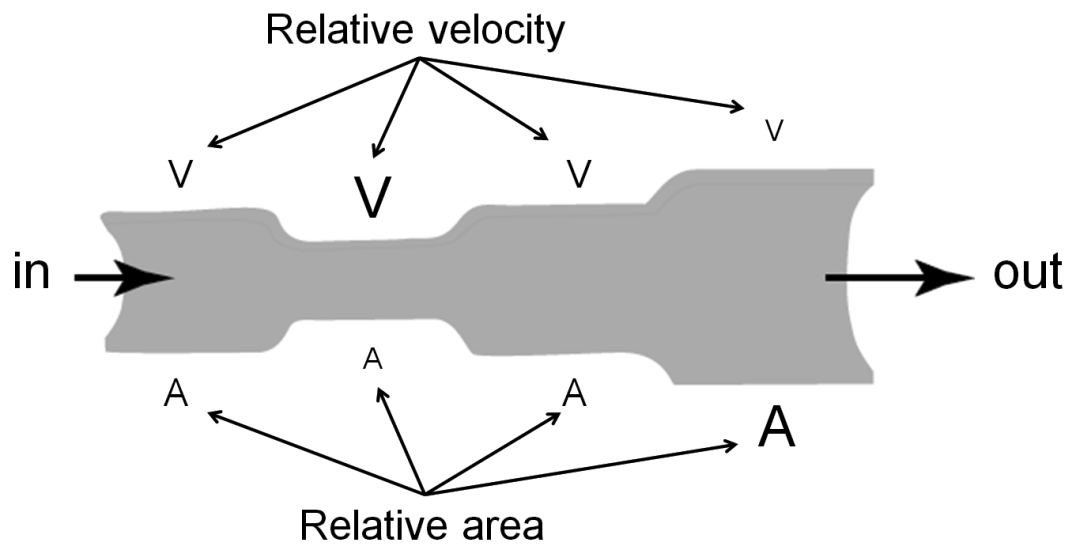
#### **2.2.4 Law of Conservation of Mass**

An important hemodynamic concept is the continuity equation, which is derived from the law of conservation of mass [172]. This equation is based on the principle that the flow at any given point in a closed hydraulic system will be equal to the flow at any other point. Thus, since the flow is constant, velocity is inversely proportional to the cross-sectional area (Figure 2.8).

$$Q = A_1 \times V_1 = A_2 \times V_2 = A_3 \times V_3 \quad (2.1)$$

where  $A$  is area and  $V$  is velocity at any given point within the system.

As velocity increases, its corresponding cross-sectional area decreases (e.g. at the site of an arterial stenosis). In the human body, the blood velocity decreases as it flows from the aorta into the capillary system but then increases again as it coalesces from venules to veins to the vena cava [149, 171]. An average velocity of blood at any given point within the circulation can be calculated knowing the volume of blood flowing past a given cross-sectional area of the blood vessel.

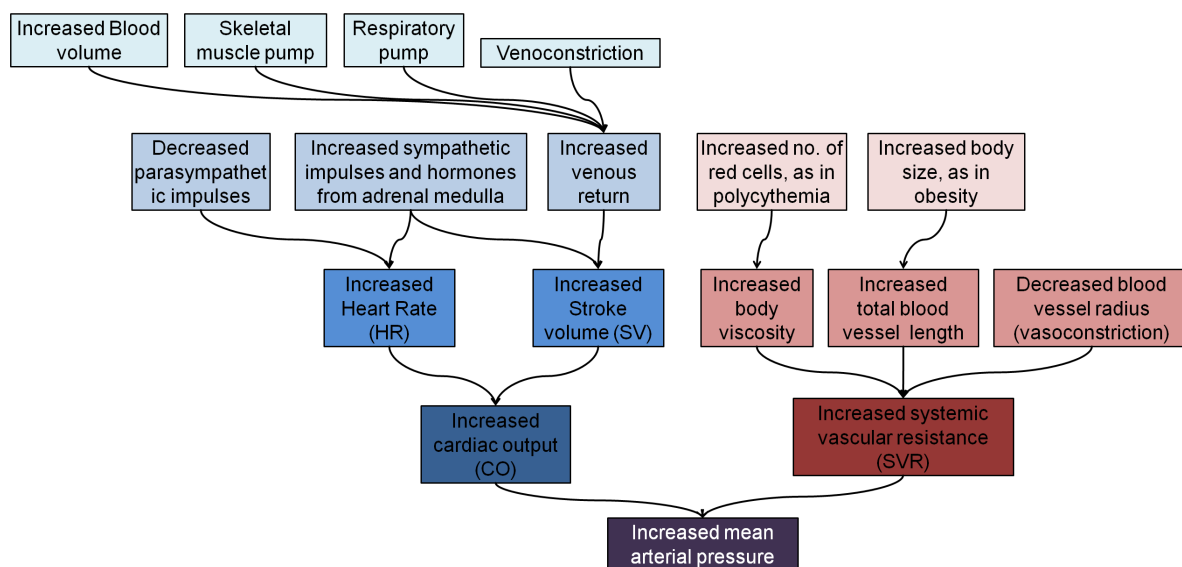


**Figure 2.8:** Relationship between blood velocity and cross-sectional area of the blood vessels. Because of the law of conservation of mass, flow at any given point in a closed system will be equal to flow at any other point. Thus, velocity is inversely proportional to the cross-sectional area.

### 2.2.5 Hierarchy of Factors

Blood flow to an organ is controlled by the constriction and dilation of vessel walls. The changes in the vessel diameter are regulated by the sympathetic interaction, and by local conditions within the blood vessel and organ. Other than wave energy, area, arterial thickness, there are many more factors (refer Figure 2.9) [162] responsible in making the non-dying pulse so complex, including respiratory effect and body size, which in a way affects the blood or pulse pressure.

There are different locations on the body where the above mentioned pressure variations can be felt, by pressing the surface of the skin against the underlying bone. For example, the *carotid pulse* is located on both sides on the neck, the *brachial pulse* is located on the inside of arm and the *pedal pulse* is located on the top of the foot. We are interested in the *arterial pulse*, felt on the wrist near the root of thumb. It is very convenient to “sense” the pulse at this location and thus is highly appreciated. Three locations on the wrist are considered in *Ayurveda* for sensing the pulse, and for subsequently analyzing unbalanced states of the body.



**Figure 2.9:** The pulse signal shows very complex behavior as it depends upon many interrelated factors. The final outcome depends upon the values of all the factors [162].

## 2.3 Traditional Pulse-based Methodology

The word *Ayurveda* is made up of two Sanskrit words *Ayur* and *Veda* meaning ‘life’ and ‘knowledge’ respectively. Taken together they mean the science of life [104, 225]. In a more limited sense, the term is used to imply the science of medicine. According to *Ayurveda* (Figure 2.10), health is not merely freedom from disease. It is essential that body, mind and soul are in an excellent state so that the individual can perform the functions and fulfill his/her role in life.

हिताहितं सुखं दुःखमायुस्तस्य हिताहितम्  
मानं च तच्च यत्रोक्तमायुर्वेदः स उच्यते  
च. सू. 1/41

**Figure 2.10:** Sanskrit shloka defining Ayurveda as: The science that expounds on – the supportive of life (hitayu), the non-supportive for life (ahitayu), the happy life (sukhayu), the unhappy life (dukhayu), the measurement of life (manam), and the quality of life.

There are originally four main books of spirituality, which included, among other topics, health, astrology, spiritual business, government, army, poetry and ethical living. These books are known as the *Vedas* – *Rig*, *Sama*, *Yajur* and *Atharva*. *Ayurveda* was used in conjunction with Vedic astrology. At later dates, *Ayurveda* was organized

into its own compact system of the health and considered as an auxiliary branch of the Vedas, called an *Upa-veda* (limb of the Veda), because it dealt with the healing aspects of spirituality, and not directly discussing spiritual development.

### 2.3.1 *Panch Mahabhuta*

All of the creations, including humans, are made up of the combination of five elements [104, 225]. These building blocks of life forms are known as *Panch Mahabhuta*. They are ether (*akash*), air (*vayu*), fire (*agni*), water (*jal*), and earth (*prithvi*). These elements are the subtlest aspects of the human life, finer than the molecular, atomic, or subatomic levels. This is the level that *Ayurveda* healing works on.

पृथ्वी अप् तेजो वाय्वाकाशानां  
समुदायात्सर्वद्रव्याभिनिर्वितिः।  
सु. शा. 41

**Figure 2.11:** *Sanskrit shloka explaining Panch Mahabhuta: Earth, water, fire, air and sky (ether or space). When they are combined in different combinations, make all the conscious or non-conscious entities of world.*

One of the descriptions of *Panch Mahabhuta* for a complex, multi-cellular organism, such as human being can be briefly stated as follows. Ether corresponds to spaces within the body (mouth, nostrils, abdomen); air denotes the movement (essentially muscular); fire controls the functioning of enzymes (intelligence, digestive system, metabolism); water is in all body fluids (as plasma, saliva, digestive juices); and finally earth manifests itself in the solid structure of the body (bones, teeth, flesh, hair).

Focusing on the cause of the grosser levels of life, the denser aspects will be taken care of since they are made of these five elements. Just as a strong foundation supports a strong building, the five elements (the foundation of all matter) when strong and balanced in a person, they will automatically balance the more material levels. Thus, *Ayurveda* does not need to look at isolated parts of the human anatomy, or at the vitamin, chemical, or nutritional level of health. It simply balances the elements, and this balances the more physical levels.

### 2.3.2 Tridosha Theory

*Dosha* is a Sanskrit word that translates as constitution into English, but actually has no direct equivalent. A *dosha* is actually a physical manifestation of more subtle forces and substances at work in the body. So a *dosha* is an 'output'. These outputs enable us to measure and follow the workings of subtle forces and substances. The *Panch Mahabhuta* manifest in human body as the three *doshas*, collectively known as *Tridosha* [104, 192, 225, 230].

वातपित्तश्लेष्माण एव देहसंभवहेतवस्तै रेवाऽव्यापन्नैः  
शरीरमिदं धार्यते । त एव व्यापन्नाः प्रलयहेतवः

**Figure 2.12:** Sanskrit shloka explaining Tridosha: Vata, Pitta and Kapha are the root for the origin of the body. When in balanced state they keep the body in its essential or true state. But when imbalanced, they can even cause death.

From the ether and air elements, the bodily air principle called *vata* (V) is manifested. The fire and water elements exist together as the fire principle called *pitta* (P). The earth and water elements exhibit as the water principle, *kapha* (K). The functioning and existence of the body entirely depends on VPK combined. A *dosha* is dominant in a human body individually *Vata*, *Pitta*, and *Kapha* or in combination as *Vata-Pitta*, *Vata-Kapha*, *Pitta-Kapha* or a combination of all three.

#### 2.3.2.1 Vata Dosha

The combination of air and ether gives us *Vata* (V) or the kinetic biological mode. *Vata* is responsible for all the movements. *Vata* is cold in nature and causes the diseases related to cold. Because of its minute nature, it cannot be visualized in any form in the body, but its presence can be proved by its actions (breathing, urination, defecation, menstruation). It is the sole power to control all the movements of nerves, tissues, cells and systems.

### 2.3.2.2 *Pitta Dosh*

The combination of fire and water gives us *pitta* (P) or the transformative biological mode. Heat is the active principle whereas liquid acts as a vehicle. *Pitta* represents energy or heat (calorie) in the body. Digestion entirely depends on *pitta* (*agni*). *Pitta* governs the digestion of physical, mental, as well as emotional elements.

### 2.3.2.3 *Kapha Dosh*

The combination of earth and water gives us *kapha* (K) or the structuring biological mode. *Kapha* molecules are heavy, stable, smooth, soft, viscous, shiny and moist, white in color, and predominantly have a sweet and slightly salty taste. *Kapha* constitutes the main body mass, and is responsible for the shape and form of the body. All the cells, tissues and organs of the body are predominantly composed of *kapha* elements. The biological combination of solid and liquids in varying proportions is responsible for the varying structure and composition of various tissues like body fluids (*rasa*), semen, blood, muscle, fat, bone marrow and brain. Because of its heavy and stable qualities, it is responsible for strength, stability, and firmness of body and mind. Because of its oily nature, it prevents friction between the two parts of the body.

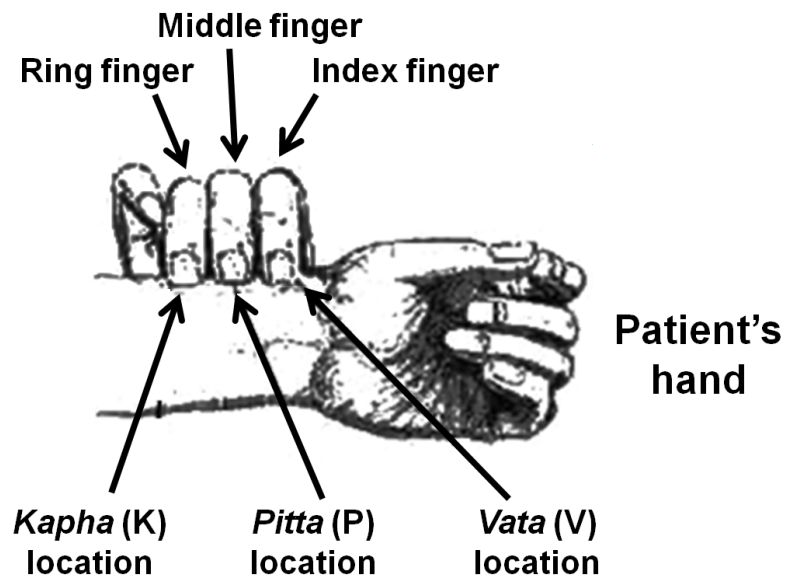
## 2.3.3 Importance of the Pulse-based Diagnosis in *Ayurveda*

The pulsation of blood carries nutrients to the cellular level (Section 2.1) throughout the body. There is a continuous flow of communication between cells, and thus this flow of communication is the intelligence. The pulse signal can reveal this cellular intelligence through a person's constitution. The amount of pressure will vary with the individual, depending upon the constitution and the volume of blood in the radial artery.

**Methodology.** As the Figure 2.13 shows, the pulse of a subject is felt by *Ayurvedic* practitioner with the first three fingers: the index, middle and ring fingers.



## Ayurvedic practitioner's fingers



**Figure 2.13:** The positioning of Ayurvedic practitioner's three fingers to sense the pulse of a subject. Respectively index, middle and ring finger are used to sense vata, pitta and kapha dosha.

The pulse is sensed on both the hands on wrist near the root of thumb. The positions of the index, middle and ring fingers denote the locations of the *vata*, *pitta* and *kapha* doshas respectively.

First, the dominance of the *dosha*(s) is sensed to know the *prakriti* (constitution) of the subject. Then depending upon his/her *prakriti* and the imbalance in the *Tridosha*, the *Ayurvedic* practitioner determines the diagnosis. The medicines (and sometimes meditation) are prescribed in such a way that, the *dosha* will come back to their (original) balanced state, as stated in Figure 2.14. Thus, the pulse-based methodology directly finds out the root cause of the disorder (location and extent), and thus is considered as effective for long-term removal of the cause.

In the subsequent chapters, we present our methodology (comprising both the hardware and software parts) to efficiently capture the pulse signals at three predefined locations. We will show that pulse signals of our system contain high details. Further, we analyze these acquired signals for studying distinguishable patterns for identifying possible disorders.

विकारो धातुवैषम्यं साम्यं प्रकृतिरुच्यते |  
सुख संज्ञकमारोग्यं विकारो दुःखमेव च ॥  
च. सू. 9/4

**Figure 2.14:** *Sanskrit shloka explaining a disease: Disease is distribution or imbalance in dhatus (and doshas) and health is balance between them. Health is defined as sukh (happiness) and disease as dukha (sorrow). The aim of Ayurveda is to balance the doshas.*

## Other Active Healing Systems

Our work is inspired by *Ayurveda*. Similar to *Ayurveda*, there are other traditional systems, also known as alternative medicine, complementary medicine, holistic medicine, and integrative medicine.

The national institutes of health (NIH) Office of alternative medicine adopted the definition of complementary and alternative medicine (CAM) as follows [151]. *A broad domain of healing resources that encompass all health systems, modalities and practices and their accompanying theories and beliefs, other than those intrinsic to the politically dominant health care system of a particular society or culture in a given historical period. CAM includes all such practices and ideas self-defined by their users as preventing or treating illness or promoting health and wellbeing. Boundaries within CAM and between the CAM domain and the domain of the dominant system are not always sharp or fixed.*

CAM includes *Ayurveda*, Traditional Chinese Medicine (TCM) based on yin-yang theory [39, 50, 69, 145], naturopathy, homeopathy based on principle of similars, Korean medicine, traditional Tibetan medicine, and samkhya with philosophical base that underlies yoga.

Even though all these alternative medical systems differ in their geographic locations, diverse cultures, languages, and philosophical approaches to the prevention and treatment of disease, they share a number of common elements. They offer patients a viable alternative to the western medicine.

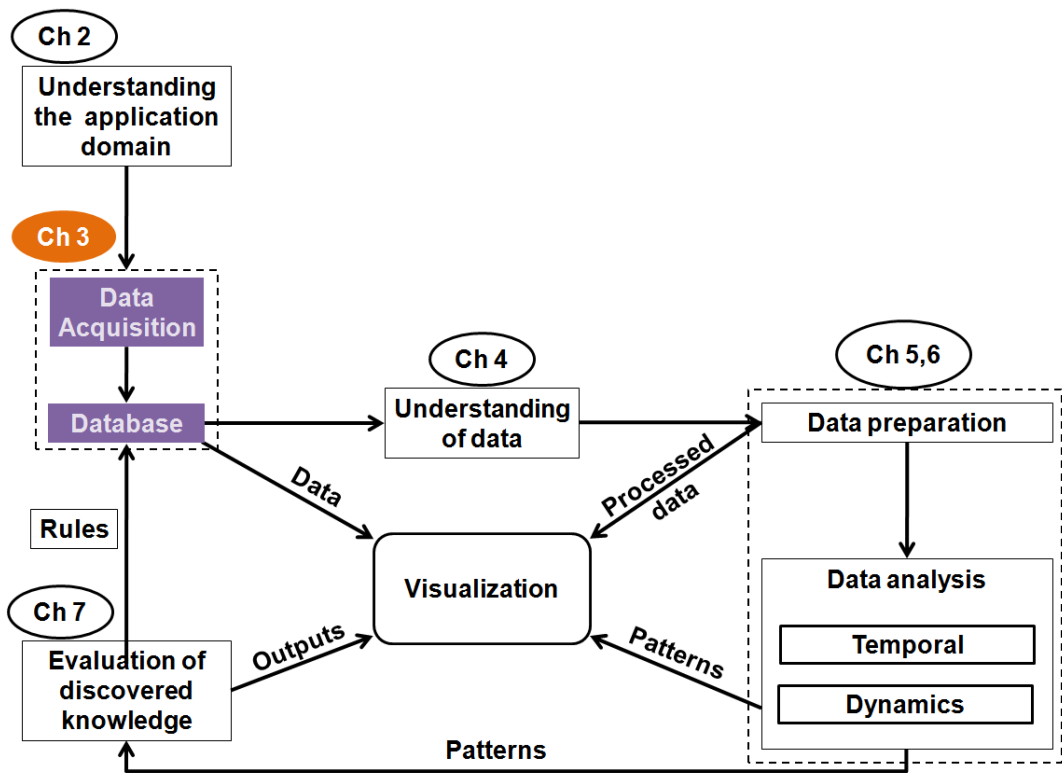
- These systems are based on the belief that one's body has the power to heal itself. Thus, they try to maximize the body's inherent healing ability.

- These systems emphasis more on the disease prevention versus disease treatment.
- These systems treat the “whole” patient (physical, mental and spiritual) versus a single pathology.
- These systems perform the treatment often individualized and dependent on the presenting symptoms.

## Conclusions

This chapter contributed to the first block of Figure 1.2. We dictated the domain knowledge about the traditional methodology. We explained the importance of the pulse-based methodology in *Ayurveda*. We also explained the formation of the pulse signals through five different theories (in medicine), and the factors affecting its behavior. We will study the complex behavior of the pulse signals quantitatively in the subsequent chapters.





---

## Chapter 3

# Acquisition

---

*In the previous chapter, we studied the importance of the pulse signals and how they are used in detecting the doshic imbalance in the body, for physical examination.*

*In this chapter, we first motivate the need for the quantitative study of the pulse signals. We describe the steps and salient features of a typical data acquisition assembly. We elaborate on each step for our previous attempts of acquisition, and then our current system Nadi Tarangini. Our methodology reflects the “feeling” information. Finally, we provide the details of our database.*

### **Need of Quantitative Analysis of the Pulse Signals**

There are differences between pulse signals of any two persons, varying with rhythm and patterns. The meaning of these differences is what *Ayurveda* has refined over 5000 years of its existence. The main aim of the traditional pulse-based diagnosis has always been to obtain useful information about what goes on inside the ‘entire’ body, what should have caused the disorder in the body, what might be done to rectify the problem, and to know what are the chances of success.

The *Ayurvedic* practitioner “feels” the pulse of a subject to understand the physical

conditions. However, the main challenge in this traditional pulse-based reading is that it depends upon the subjective observation and the quality of perception of the *Ayurvedic* practitioners. Therefore, many times there is an inconsistency in the diagnosis process of the two practitioners for the same subject. Furthermore, they treat the subject depending upon the methodologies learnt from their teachers, and their own experience. Thus, the whole process of traditional pulse-based diagnosis is very *subjective*.

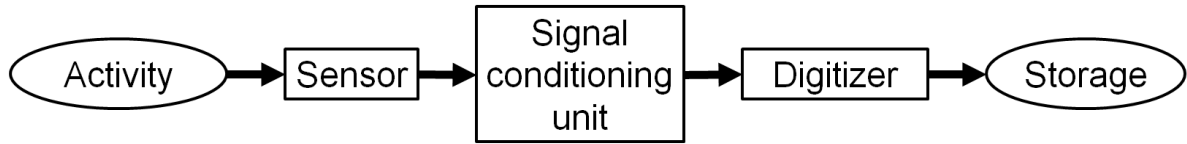
Currently, only qualitative information is available in the form of books, lessons from the *Ayurvedic* teachers and own experience. As explained, the methodology requires very high concentration and skill of “feeling”. Thus, the whole process is too difficult when practiced by a novice. Due to this difficulty, many people still think of it as a mystery.

In the modern medicine, the examination of the arterial pulse signals has only been used for computing the heart rate and blood pressure measurements. As the effective models of the automatic classification or detection using pulse signals have not been reported till now, this undoubtedly limits the practical applications of the pulse-based methodology.

The basic requirement of any data analysis procedure is to collect noise-free data with very high details. We have developed *Nadi Tarangini* to acquire pulse signals from the *Tridosha* (VPK) locations in the time series format. In subsequent sections, we explain our system.

### **3.1 Components in a Typical Data Acquisition System**

A data acquisition system is designed to measure and log desired parameters. The representation of acquired signals in the electronic form facilitates further processing and analysis of the data with the help of a computer. Figure 3.1 shows essential components of a typical data acquisition system.



**Figure 3.1:** *Essential components of a typical data acquisition methodology.*

We wish to record the non-dying pulsations of the arterial pulse on wrist. It is very important to capture the repeatability and consistency of the activity. Now, we elaborate on the properties of each of these components, and list the prior work in the advancement of the respective component.

- **Sensor:** Electrodes, sensors, and transducers are used to transform a physical activity into the recordable form. In this thesis, we are concerned only with the sensors which convert the pressure energy into the electrical energy. Also, we concentrate on the non-invasive type of sensors which minimize risk to the subject.

The beginnings of the noninvasive arterial pulse can be traced to the year 1555, when the physician Strus [201] had proposed that the arterial pulse possesses a waveform. Although instrumentation that he used was simple, he suggested that changes in the arterial pulse shape and strength may be related to disease conditions. Till now, there have been various acoustic, photoelectric, and mechanical sensors proposed for the acquisition of the pulse signals.

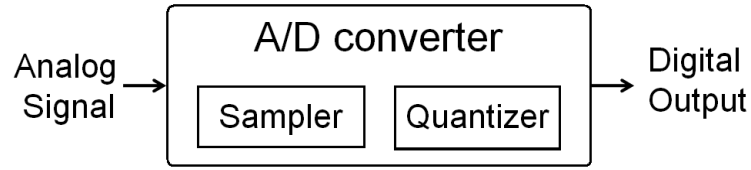
- ◇ **Acoustic Sensors:** Usually, a non-contact-type phonic sensor [18, 110, 232] is attached to the wrist with a coupling chamber, which is to insulate the interference of muscular activities and outside sounds. Then, the air in that close chamber is evoked by the sound of pulsatile blood flow signals. Such vibrations will be converted to the varying electrical signal as the track of pulsatile blood flow signals. However, these sensors are very rarely used due to their large size and low efficiency compared to the piezoelectric sensors.
- ◇ **Photoelectric Sensors:** It has been established that transmission of light in the living tissue is affected by blood. The optical power reaching to the



photo-detector is related to blood volume changes in the arterial tree [114]. Such phenomenon has been used to determine the volume of arterial blood [6, 10, 19, 71, 131, 170, 179]. It provides a feasible way to depict the track of pulsatile blood flow signals, which can be calibrated for subsequent pulse-based analysis. It is found that the skin and the vessel wall's scattering, reflection, absorption, and movement, all play a role in determining the signal [19, 108].

- ◇ **Applanation Tonometry:** It measures the force needed to flatten a superficial artery, at the site of the radial artery. When this force is exerted with a piezoelectric sensor, the recorded track is a representative for arterial pulsatile blood flow signal [49, 23, 72, 93, 132, 184, 243, 267]. *SphygmoCor system* are commercially available for noninvasively recording the arterial stiffness [55, 243].
- ◇ **Pressure Sensors:** Piezo-film based sensors [2, 10, 35, 54, 99, 105, 112, 115, 191, 193, 256, 259, 263] are used to capture the arterial pulse signals, because they are very sensitive to the minute mechanical movements, responsive to low frequency movements of interest [157], and weigh little. Also the size of such pressure sensors is adequate, and thus are most widely used.
- **Signal conditioning unit:** Signal conditioning is required to amplify and filter the signal, in order to remove the noise, and to satisfy the requirements of the subsequent hardware such as the dynamics of the digitizer. Another important requirement is to preserve the information contained in the original signal.
- **Digitizer:** Analog signal from the sensor (and signal conditioning unit) is converted into discrete values for storage. Though it is possible to deal with continuous time series signals, it is usually convenient to convert them into a digital form for easy implementation, even of complex algorithms. As shown in the Figure 3.2, the complete A/D conversion is implemented in two steps.

Firstly, uniform sampling process converts the continuous signal into a discrete-time series. Then a quantization procedure assigns the amplitude value of



**Figure 3.2:** *The steps involved in analog to digital conversion.*

each sample within a set of determined discrete values. In this procedure, to adequately capture the continuous time signal, the sampling instants  $t_k$  must be selected carefully so that information is not lost. The minimum sampling rate is twice the highest frequency content of the signal (based on the sampling theorem from communication theory) [189].

12-bit of resolution is most popularly used for acquisition of the pulse. Few systems such as [6, 218] have used 16-bit of resolution, which results in more data storage.

- **Storage:** Enormous data coming out of an acquisition system is finally stored on the computer for visualization and further processing.

Professional softwares are readily available for interface and easy manipulation of digitized signals, e.g. LabVIEW (National Instruments, TX, USA) and PVX system software [55]. The accuracy and size of the stored data depends upon the sampling rate. Sampling rates from 200 Hz [218] to 2500 Hz [6] are used for digitization, however the rate of 500 Hz is most commonly used. The softwares also help in easy selection of the data for further analysis.

Other steps or components of a data acquisition system are a display equipment for visualization, control devices such as power supply and stabilizers, and data cleaning and processing algorithms.

Though there have been many attempts to capture the useful pulse signals as mentioned above, the acquisition of pulse signals of high details, followed by its rigorous analysis is not reported (or is in the initial stages). In the development of our system, we have considered important aspects, for the efficient acquisition. We compare our signals with the previous works later in Section 4.1.

### 3.1.1 Design Aspects

As said earlier, we want to establish the *Ayurvedic* pulse through the patterns & visualization techniques. Throughout the data acquisition procedure, it is critical that the information and structure of the pulse be faithfully preserved. While designing our system, we have considered following aspects:

- **Range of operation:** The range of the pressure sensors was carefully selected to detect even the minute details of the pressure values. The *Ayurvedic* practitioners can sense very high details due to their experience and high concentration. Therefore, to mimic their methodology, we should collect all the details hidden in the pulse signal.
- **Sensitivity:** To note even the smallest measurable signal variations, we have used the 'differential mode' of acquisition (as explained in Section 4.2.1).
- **Signal-to-noise ratio (SNR):** The errors from component tolerance, movements of subject, drift due to changes in temperature, humidity, and reading errors have been successfully eliminated to achieve high SNR in pulse signals. Furthermore, due to optimized properties of sensor, signal conditioning unit, and elimination of most of the error causing components, the accuracy of our system is high.
- **Stability and reproducibility:** Our system is able to provide stable and reproducible outputs which are essential for accurate analysis.

All the above aspects are further elaborated in Section 3.3, where we discuss *Nadi Tarangini* in detail.

The pulse signals may be obtained at other points on the body. Examples include carotid pulse, brachial pulse, posterior tibial pulse, and dorsalis pedis pulse. But, the arterial pulse on wrist is the most convenient to read and is more readily available. Hence it is prominently been studied to reveal the characteristics of *doshic* imbalance, the nature of diseases and expected prognosis.

## 3.2 Our Previous Systems of Pulse Acquisition

### Previous attempt: Using a microphone

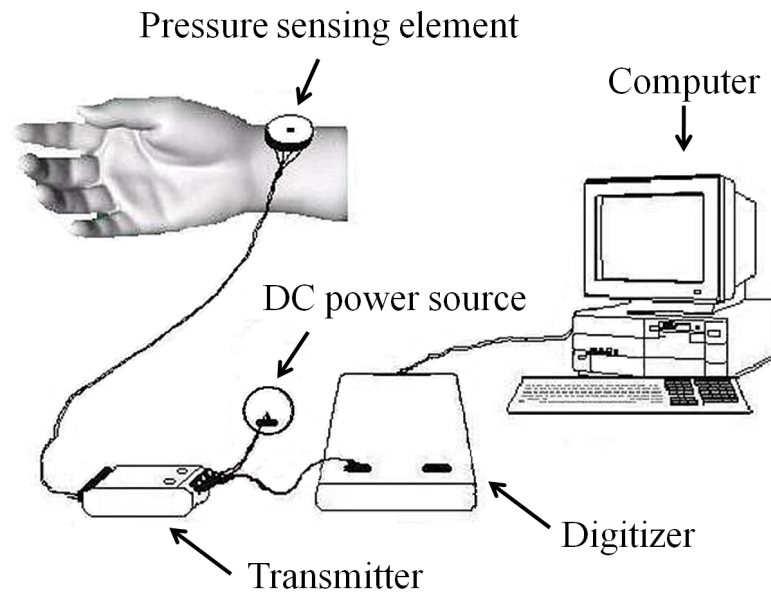
In our first attempt, we used the 'sound' quality of the pulse for the acquisition. We enclosed the *Tridosha* (VPK) locations on the wrist with a cuff for insulating the pulse from the outside sounds. A microphone was carefully attached inside of the enclosure, which could pick up the pulse sounds. The sound was converted into the electrical energy, denoised, digitized, and finally stored on the computer.

The disadvantage of these pulse signals were that they were not clean, and rich in contents. The signal was mainly consisted of a strong peak, and one or two secondary peaks; but the high details were missing (refer Chapter 4). Also, there were sometimes errors occurring due to sounds generated by the patient's movement. The scope of this methodology was limited to compute the pulse rate of the subject.

### Previous attempt: Using one pressure sensor

Our next system comprised of a pressure sensing element to record the pulsations (mechanical activity) as shown in Figure 3.3. We used a '1 PSI' pressure sensor from 'Sensym Products' [187]. It was coupled with a transmitter to amplify the electrical signal. This electrical signal was then digitized using a 8-bit digitizer, having an interface with the personal computer at the LPT1 port (parallel port). The pressure sensing element was supplied with the excitation voltage by using the DC power source.

The sensor was circular with a tiny diaphragm at the center. For capturing the three *doshas* on the wrist, the sensor (diaphragm) has to be placed at the exact locations one at a time. The bandwidth of the digitizer was 100 kHz, which was adequate to capture the rapid changes in the input signal from the transmitter. For interface, we used the software Picoscope [163], which supported the digitizer, to save the digitized output for a prefixed time on the computer.



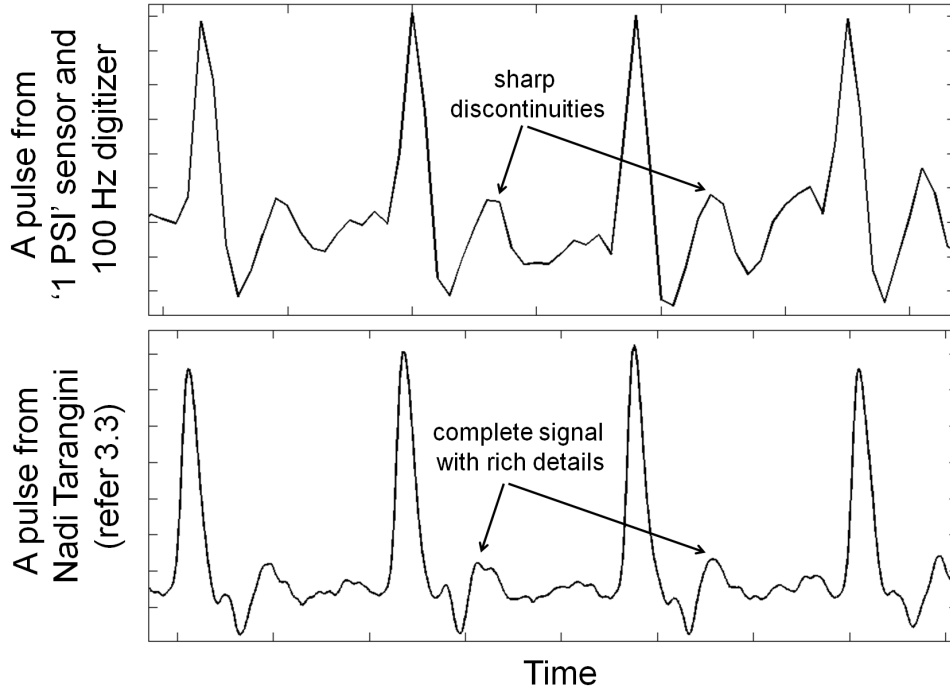
**Figure 3.3:** Our previous system to acquire the arterial pulse. It comprised of a pressure sensor with diaphragm, a transmitter, a digitizer, and an interface to computer at the parallel port.

After recording the pulse signal using this system, we immediately measured the pulse rate manually. We also computed the pulse rates from the acquired pulse using Fourier transform (refer Section 4.4). The pulse rates were consistent, which showed successful acquisition. However, it is clear from the Figure 3.4 that even this accuracy was not enough for rigorous analysis using modern data analysis tools.

The figure compares the pulse signals from this system with those of our current system *Nadi Tarangini* (Section 3.3). Note that the two pulse signals are of the same subject, but at different locations and at different times. We explain *Nadi Tarangini* in the next section, where the acquired signals are replete with all the rich details in the form of secondary peaks.

### 3.3 *Nadi Tarangini*

We now describe our portable and inexpensive system, *Nadi Tarangini*, which captures the arterial pulse in the format of time series. Figure 3.5 and Figure 3.6 display the line diagram and the actual photo of *Nadi Tarangini* respectively.

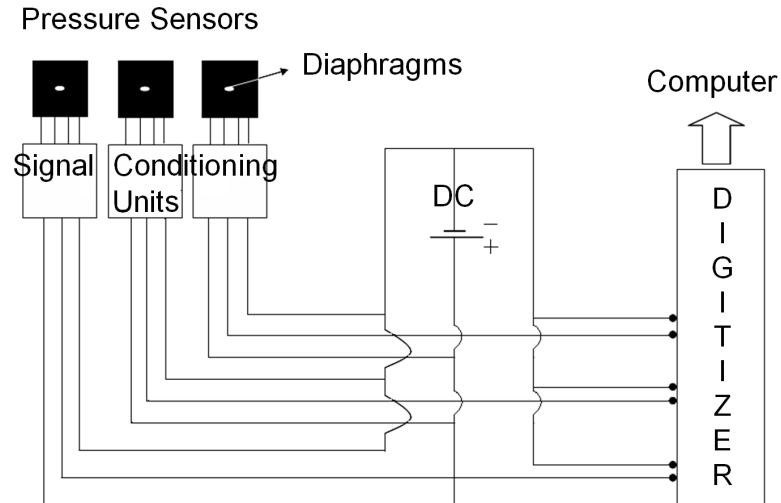


**Figure 3.4:** The pulse signal on the top is acquired using our previous system comprising a '1 PSI' sensor and digitizer of 100 Hz sampling rate. The pulse signal on the bottom is acquired using our current system Nadi Tarangini (refer section 3.3). The signal on the bottom is replete with details and variations.

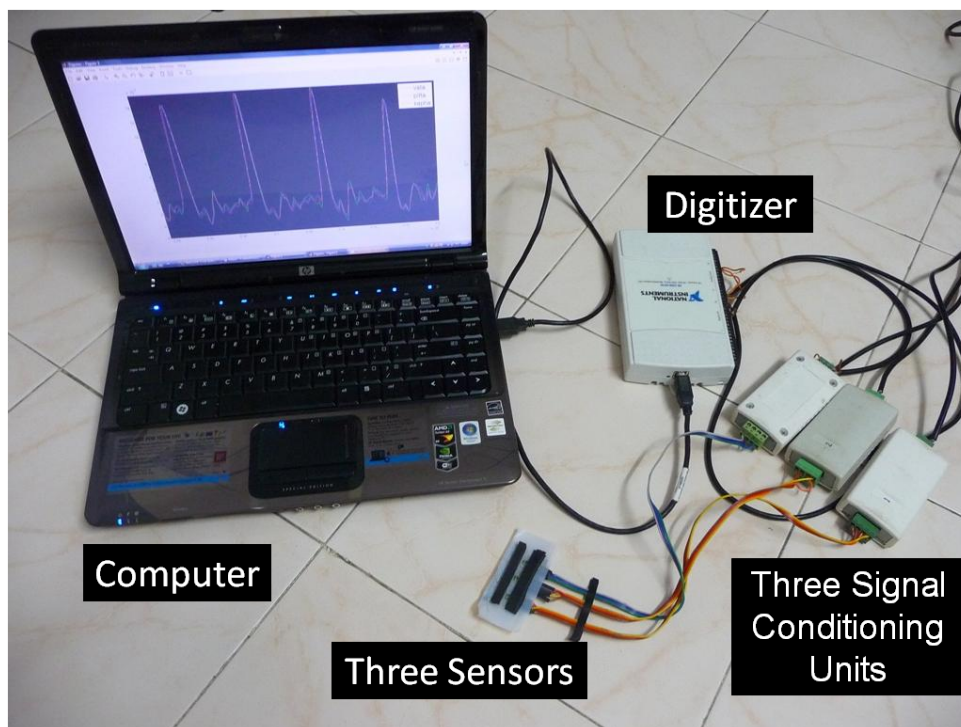
The important components of *Nadi Tarangini* are three pressure sensors, three signal conditioning units, a digitizer, an interface, and a storage. The three sensors are placed exactly at *Tridosha* locations on wrist. They pick-up the pressure values and provide equivalent electrical energy to the signal conditioning units. These energies are standardized in signal conditioning units, digitized, and finally stored on computer for the subsequent analysis.

### 3.3.1 Pressure Sensor

A set of three pressure sensors is mounted on the wrist to sense the pressure values at three locations (Section 2.3.2), namely *vata* (V), *pitta* (P) and *kapha* (K) [225]. We have adopted 'millivolt output medium pressure sensor' (Mouser Electronics, Inc. [141]) having '0–4 inch  $H_2O$ ' pressure range. The physical dimensions of the strain gauge element are approximately  $9mm \times 7mm$  (nearly the size of a fingertip). We have further



**Figure 3.5:** This is our setup to mimic the traditional pulse-based methodology. The three tiny diaphragms are placed exactly at the VPK locations, where the pressure values are captured, denoised, digitized, and finally stored on the computer.



**Figure 3.6:** Figure 2.13 shows the positioning of the Ayurvedic practitioner’s fingers on subject’s wrist. We built Nadi Tarangini to mimic this methodology. The three sensors are exactly of the size of fingertips. The assembly is properly shielded for the best performance. The display on computer shows the captured pulse signals.

used the ‘differential mode’ of acquisition for better resolution of the signals (Please refer to Section 4.2.1 for more details on the ‘differential mode’).

Each sensor consists of a flexible and tiny diaphragm at the center (as shown in Figure 3.5), which is a force-gathering element that gets deformed by the arterial pressure values. A Wheatstone bridge circuit methodology is used as follows. When the pressure (at a particular location) is exerted across the diaphragm, it will cause the diaphragm to deflect. Then, two of the fine-wire strain gauges get extended by a small amount, and the other two contract by the same amount. When this long narrow electrical conductor (fine gauge wire) is stretched within its elastic limit, it will increase in length and decrease in cross-sectional area.

$$R = \rho \frac{l}{A} \quad (3.1)$$

where  $\rho$  is the electrical resistivity of the wire material,  $l$  is its length, and  $A$  is its cross-sectional area. This stretching results in an increase in the resistance. The change in the length can only be very small for the foil to remain within its elastic limit, so the change in electric resistance will also be small. The relative sensitivity of this device is given by its gauge factor  $\gamma$ , which is defined as

$$\gamma = \frac{\delta R/R}{\delta l/l} \quad (3.2)$$

where  $\delta R$  is the change in resistance when the structure is stretched by an amount  $\delta l$ . When the strain is applied in the sensitive direction, the direction of the individual elements of the strain gauge, the length of the gauge will be slightly increased, and this will result in an increase in the electrical resistance seen between the terminals. *Therefore, the piezoelectric strain gauges are particularly useful as the variations in the pressure values are of short duration, and rapidly changing in value.*

### 3.3.2 Signal Conditioning Unit

The signal conditioning unit is a standard industrial amplifier with 4–20 mA output, which is used to linearize and denoise the acquired signal. While connecting to 16-bit digitizer, the 4–20 mA is converted to 2–10 V through a resistor of 500  $\Omega$ .



We have also provided zero and span adjustments. The zero adjustment is calibrated to adjust the zero, such that at zero pressure (atmospheric pressure), the output is 4 mA. The span adjusts the degree of amplification for output of the signal.

The data obtained in this way is usually corrupted because of implicit and explicit electronic and electrical noise. However, in our developed system, the noise level is negligible, after proper shielding. Furthermore, we have used a battery-powered equipment to avoid the common problem of interference at 50 Hz from mains wiring.

### **3.3.3 Digitizer**

The digitizer or analog to digital converter (ADC) works as the heart of a data acquisition system. In *Nadi Tarangini*, the (amplified and denoised) analog signal is digitized using a 16-bit multifunction data acquisition card NI USB-6210 (National Instruments, TX, USA), having an interface with the personal computer at the USB port. The data is captured at a sampling rate of 500 Hz (which is sufficiently higher than the Nyquist criteria) for a predetermined length of time.

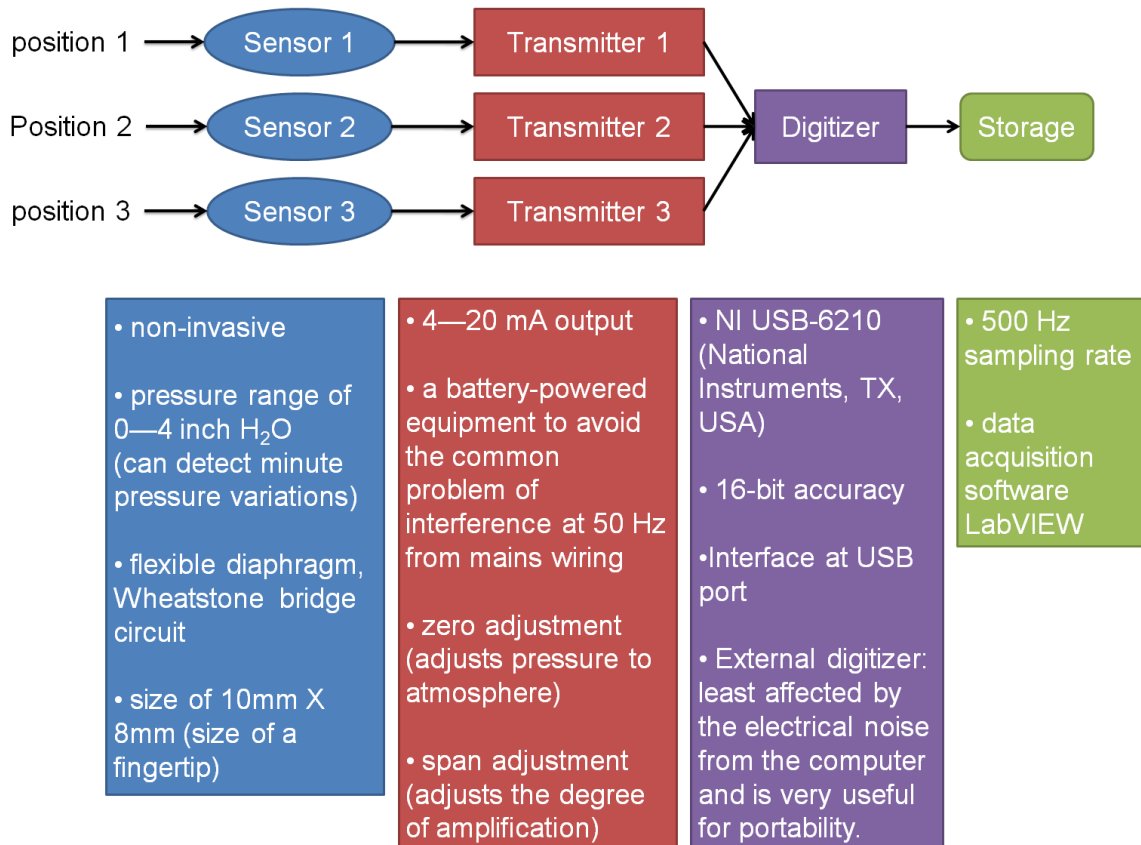
We have used an external ADC, as it is the least affected by the electrical noise from the computer. Also it is very useful for the portability, as it can be connected to any computer for storage.

### **3.3.4 Storage**

We have used the data acquisition software LabVIEW (National Instruments, TX, USA) to store the pulse signals. LabVIEW provides all the functions needed to build a complete system, namely control of the analog to digital converter, data collection and storage, user interface controls, graphical displays, and a good range of advanced signal processing algorithms. The minimum change in the signal, which can be measured, depends solely on the resolution of the digitizer.

### 3.3.5 Schematic of *Nadi Tarangini*

In *Nadi Tarangini*, three pulse signals are recorded, one each from the three sensors at predefined locations on wrist. They are stored in the time series format, which can be easily imported into a spreadsheet or a graph-drawing program for easy visualization. The summary of all the above components is displayed in Figure 3.7.



**Figure 3.7:** The schematic diagram of *Nadi Tarangini* displaying important specifications of the components: Three sensors, three transmitters (or signal conditioning units), a digitizer, and storage.

The selection and arrangement of all the components result in the best acquisition of the pulse signals. This is very important for the subsequent analysis. E.g. in Chapter 6.1.1, we show that the pulse signal contains a self-similar nature, which was possible due to the high details in our data.

## 3.4 Measurement Methodology

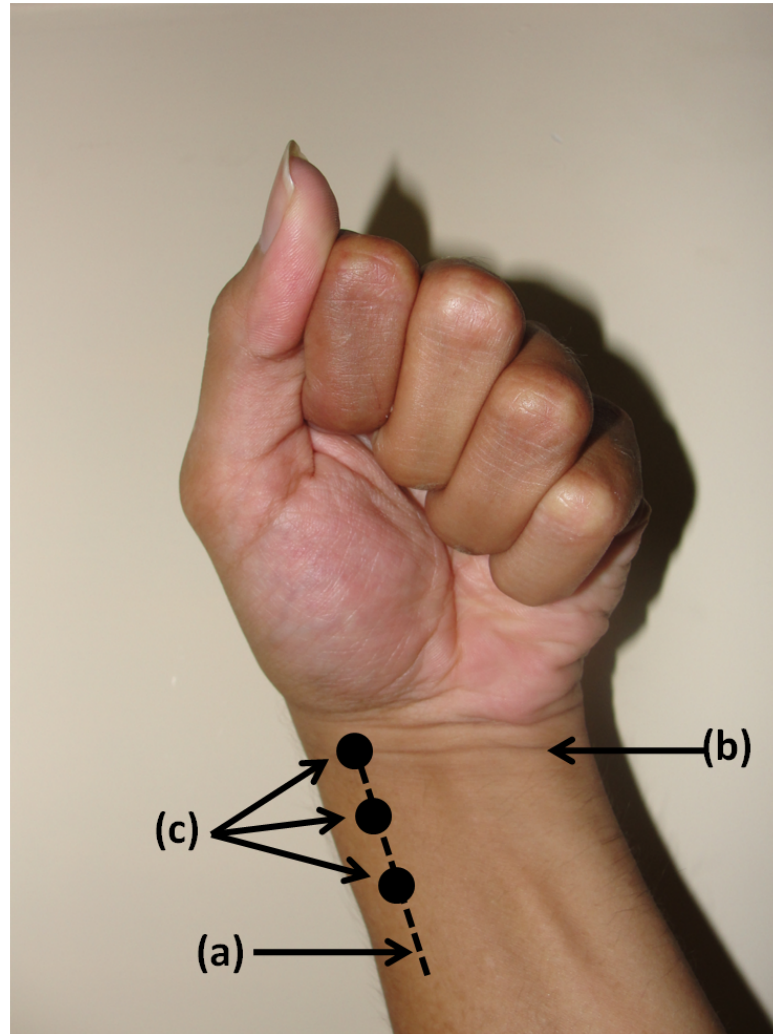
Our procedure of acquiring the pulse signals from a subject is as follows. The subject is asked not to undergo strenuous exercise, for at least two hours before taking the reading. Also, just before acquiring the pulse signal, the subject is asked to relax in a room, and clear his/her mind. This is our basic assumption that the pulse signal is not affected by the psychological thoughts of the subject. The pulse recordings are made whilst the subject is breathing regularly. The subject is asked to keep his/her hand (and also all the other body parts) steady and in the comfortable positions.

### 3.4.1 Deciding VPK Locations

The procedure to decide locations of the VPK *dosha* is shown in Figure 3.8. The three locations will be available when both the horizontal and vertical axes on the wrist are known. The vertical axis is manually decided by sensing the pressure values at the three locations using the fingertips. Then, we consider the first wrist line (as shown) as the horizontal axis. Their intersection point (*vata* location) is where the first sensor is placed. The other two sensors along the vertical axis pick up the pulse signals of the *pitta* and *kapha* locations.

### 3.4.2 Placement of the Sensors

Once the three locations (VPK) are properly decided, the sensors are carefully placed on the wrist, so that the three diaphragms come exactly above those locations. As the acquisition is real-time, the position of the sensors could be readjusted, if required, by looking at the computer screen. In such cases, we readjust the sensor positions in the vicinity to confirm the three locations. If two *dosha* are correctly captured, while the third signal does not show the pulse behavior – then either the location is incorrect or the pulse itself has negligible energy. Once, the VPK locations are correctly picked up by the sensors (diaphragms), we start the data collection. The pulse signals are usually recorded on each hand of the subject for at least 1 minute of duration.



**Figure 3.8:** Procedure to decide VPK locations. (a) The vertical axis is decided by manually sensing the pressure values at the three locations. (b) The first wrist line is considered as the horizontal axis. Their intersection point is the vata location. (c) VPK locations of acquisition.

### 3.4.3 Deciding the Contact Pressure

In *Ayurveda*, the pulse is sensed at the VPK locations by the *Ayurvedic* practitioners on the wrist at a particular contact pressure. By increasing the contact pressure, the pulse under that fingertip might get partially blocked [104]. At different contact pressures, different amplitudes, energies, and patterns are sensed, which are then correlated with the body conditions.

By adopting the pressure sensors, we have efficiently mimicked the *Ayurvedic* methodology of sensing the pulse on wrist. In our work, we keep the contact pressure constant

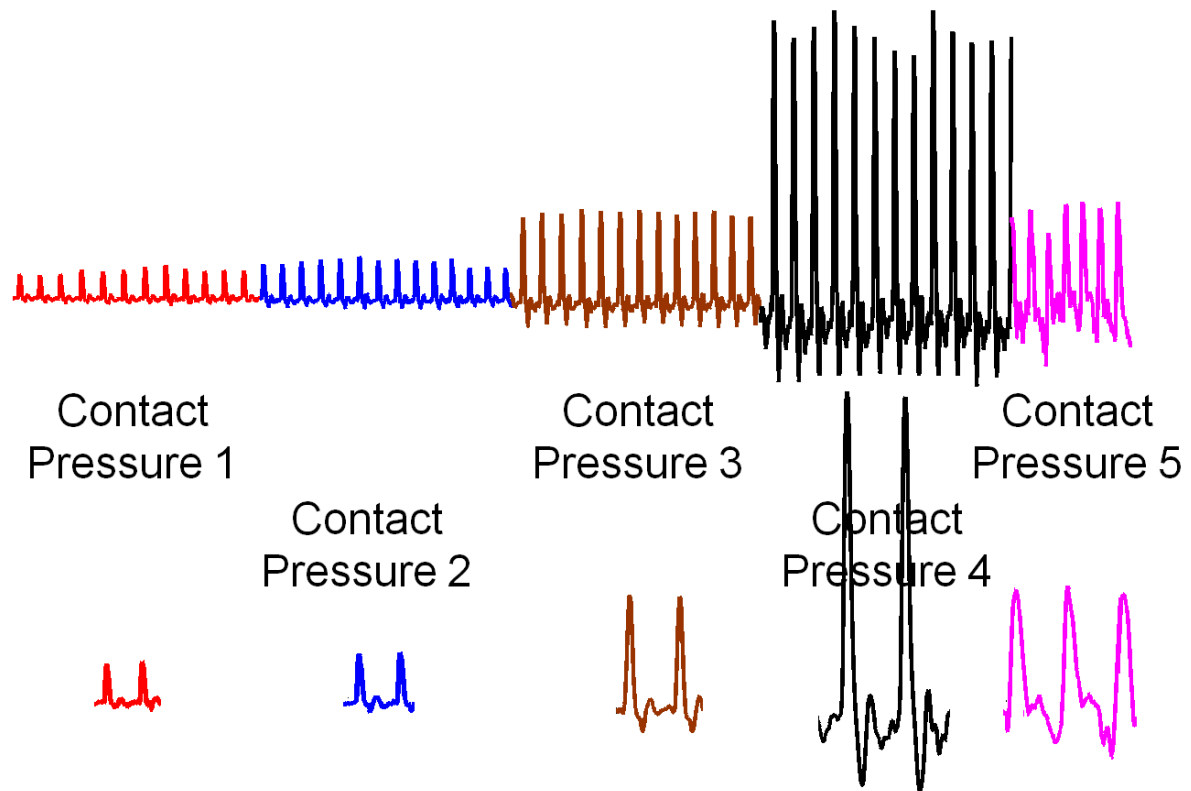
at one particular value. Of course, it is difficult to prove that the applied contact pressure remains constant throughout the acquisition. But with the visual impression of a long pulse signal (say 200 pulse cycles) where the overall pattern and amplitude of the pulse series remains almost the same, we claim that the applied contact pressure is constant (within manual approximations). The advanced study of varying contact pressure and observing the effect on pulse signals is beyond the scope of this thesis.

**Methodology:** As discussed before, we place the sensors exactly at the VPK locations. We start with the minimum contact pressure, where the acquired pulse signal is negligible. We increase the contact pressure on the sensors in such a way that, the pulse behavior is seen on the screen (real-time acquisition) with considerable amplitude. The contact pressure is not increased too much, which might block the *nadi* itself.

Our observation is that, this contact pressure depends primarily upon the skin of the subject, hair on the wrist, and the actual depth of the pulse. Therefore, the contact pressure is different for each subject, and is decided while acquiring the pulse signal.

Even though the methodology explained above contains manual judgement, we assume that the collection is adequate due to following observation. We have recorded the pulse signals of a subject at five different contact pressures. Figure 3.9 provides the pulse signals at each contact pressure.

As the contact pressure on the sensor increases, the amplitude and the overall shape of the pulse signal increases. After reaching a maximum (contact pressure 4 in this case), the amplitude decreases. After a particular threshold value, the pulse dies. Our observation are consistent with the literature [225]. Both the extreme cases of very little contact pressure (negligible pulse), and very high contact pressure (where the pulse is blocked, or dies) are not shown in the figure. At each contact pressure, the acquired pulse signal gives different insights about the body. The pattern remains almost the same between consecutive applied contact pressures, even though the amplitude increases. Thus as said earlier, we consider our data to be adequate for further analysis.



**Figure 3.9:** *Changes observed in the pulse signals as the applied contact pressure increases from Left to Right.*

### 3.5 Pulse Database

**Advantages of the established ECG database:** Due to very wide variability in ECG rhythms, and in details of the waveform morphology, it is necessary to collect huge collection of long-term ECGs for their analysis. But, their acquisition and characterization is a tedious and expensive process.

Until 1980, it was necessary for the researchers to collect their own data, for studying the abnormalities of cardiac rhythm. For removing this limitation, MIT-BIH arrhythmia database [63] was generated, and the recordings were made available to the research community. The most important advantage of this open database is the reproducible and objectively comparable evaluations of different algorithms on the same data. It helped in rapid and quantifiable improvements in the technology of automated arrhythmia analysis [139]. Now, there are many ECG databases including AHA Database for the evaluation of Ventricular Arrhythmia detectors [73]

and European ST-T Database [209].

Huge pulse database should be generated in a similar fashion to enable studying together for better understanding of the pulse signals. We now provide the details about the important entries in our database.

## Important Entries in Our database

There are five major categories of entries.

1. **Patient's information:** This category is to know the background of the subject.

This can be further categorized as

- Personal details such as name, age, gender, height, weight, birth date, marital status, and contact details
- Medical details such as blood group, usual body temperature, and body structure
- Other personal details such as family background, status, appetite, acidity, food, sweating, movements, and sleep
- Habits such as smoking, drinking and exercise
- Details noted at the time of acquisition such as medicine or food before acquisition, blood report, and urine report

2. **Analyzer's entries:** This category is to consider various important dimensions while correlating the pulse signals with the physical condition of the subject. This can be further categorized as

- Atmospheric details such as room temperature and humidity
- Patient related entries such as skin of the subject at the wrist, contact pressure applied on the sensors, and time taken to find the pulse

3. **Doctor's entries:** In our work, we have noted down the entries by Dr. Ashok Bhat and Dr. Jategaonkar. They are the domain experts in our work to provide information about the traditional methodology. These are the most important entries, and are considered as the ground truth for the final correlation with the pulse signals.

These can be further categorized as

- Observations such as condition of tongue, color of eyes, and voice
  - *Dosha* entries
  - Analysis and diagnosis entries
4. **Pulse data:** Finally, the three pulse signals (*Tridosha* locations) are stored in the time series format, with three amplitude columns against a single time column. All the three signals can be visualized on the screen even using simple programs such as Excel and Matlab.

For subsequent pulse-based analysis (in Chapters 5 and 6), we choose the input pulse in the following manner. (a) a complete pulse time series or some portion of it, (b) 3D attractor (e.g. time-delayed embedded format), (c) a sequence of peak-to-peak distances, and (d) single *dosha* or all the three combined from the database.

5. **Results:** After the application of different methodologies (discussed in subsequent chapters), the hidden information is extracted in terms of useful feature sets or classification numbers. They are stored in the database for visualization, comparison, and future use.

## Pulse Data

Currently our database contains data from 612 subjects. Appendix C displays a few examples of the signals from our database. Approximately half of the data is from the male subjects, while remaining is from the female subjects. We have divided the



subjects into three age groups as 'below 25 years', '25–50 years' and 'above 50 years'. Respectively 102, 308 and 202 subjects fall into these three age-groups. In our results, we show that the pulse patterns show variations with these three age-groups. The pulse signals of our database can also be categorized as those obtained from healthy subjects (176), and subjects having disorders (436) such as fever, diabetes, fits, cough, hypertension, acidity, skin and heart related disorders. In Chapters 5 and 6, we apply various approaches to capture their distinguishable patterns.

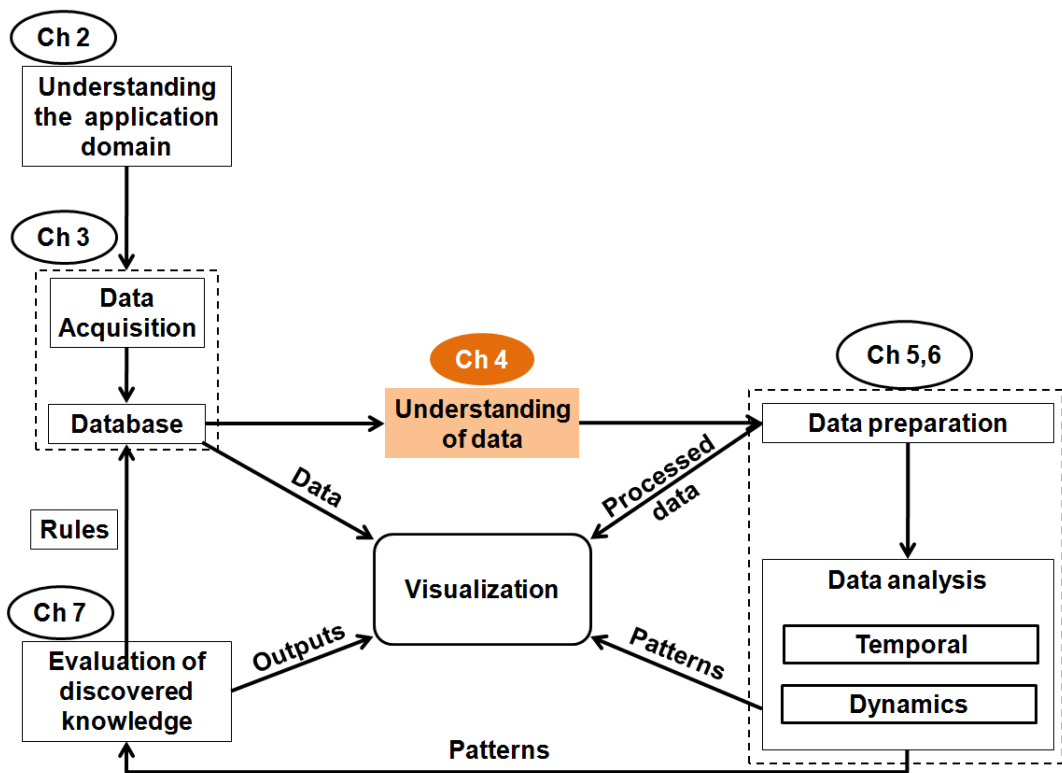
## Conclusions

This chapter contributed to the second block of Figure 1.2. Though the useful arterial pulse shows enormous scope in the diagnosis of traditional medicine, it lacks the quantitative base. Therefore, we have built *Nadi Tarangini* (NT) to acquire noise-free pulse signals. We have taken into consideration all the important aspects of the acquisition process for the best performance.

The design of NT reflects the “feeling” of fingertips. The system is comprised of three pressure sensors (pressure range of 0–4 inch  $H_2O$ ), three signal conditioning units with proper shielding, an external digitizer (sampling rate of 500 Hz and 16-bit resolution), and a storage. We also explained the important steps in the acquisition methodology, namely deciding the VPK locations, the placement of the sensors, and deciding the contact pressure. Along with the data in the time series format, we also store the patient’s information and doctor’s entries (ground truth) in our database.

In modern medicine, there is a range of health care tools which provide a diagnosis for a wide variety of disorders and injuries, by opening the “inner man” to the medical inspection, such as x-rays and electrocardiogram (ECG). In this thesis, we analyze the pulse signals of various behaviors to identify meaningful patterns that can be of interest and hence useful to the user.





---

## Chapter 4

# The Nature of Pulse Signals

---

*In this chapter, we explain the nature of pulse signals acquired using Nadi Tarangini. First, we display pulse signals of earlier approaches. We show that our data contains high details, one reason could be the 'differential mode' of acquisition. We compare a pulse signal with simultaneously recorded ECG. We elaborate on the important properties of the pulse signals including rate, shape, rhythm and frequency.*

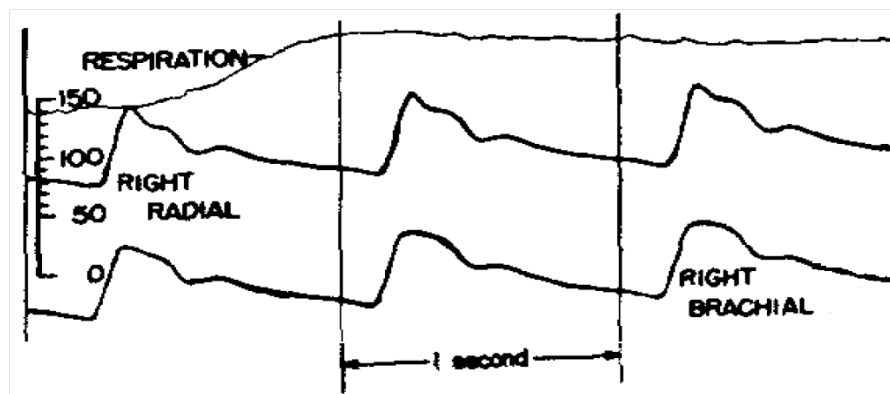
*This chapter prepares the background of possible distinguishing patterns for subsequent chapters involving data analysis. We briefly list the important prior work in this domain Finally we explain the challenges in the pulse-based analysis, before explaining our approaches in subsequent chapters.*

## 4.1 Prior Work in Pulse Acquisition

Figure 4.1 through Figure 4.15 show examples of the pulse signals obtained from previous publications. We have obtained these examples directly by copying the figures from the respective publications. These pulse signals are from different subjects of varying age-groups, using various techniques, and are of different resolutions.

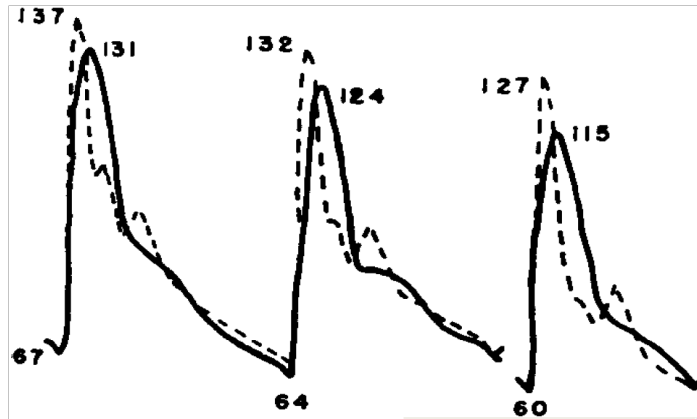
The examples are arranged in the chronological order of dates of the publication. We also mention the methodology of acquisition, and a one-line description of the way pulse signals were analyzed and used in the publication.

- (1953) Central arterial pressure pulse were recorded in [240] by means of small radiopaque plastic catheters advanced into the left subclavian artery through a thin-walled 18-gauge needle. The proximal end of the catheter was attached to a two-way stopcock so that it could be connected interchangeably to either a strain-gauge manometer or a capacitance manometer. The output of the manometers was recorded photographically by means of galvanometers of the d'Arsonval type, using a chart speed of 85 mm/sec.



**Figure 4.1:** A pulse from [240]. This method was described for beat-to-beat estimation of changes in stroke volume.

- (1956) Pulse were recorded in [178] by strain gauges of sensitivity such that a pressure change of 100 mm of mercury produced a deflection of about 7 cm on the record.



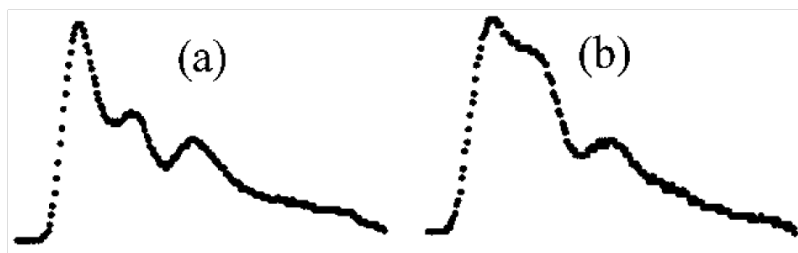
**Figure 4.2:** A pulse from [178]. Difference curves between the peripheral pulses and the most proximal one in the thoracic were studied.

(1993) Pressure in the radial artery was recorded in [247] using a standard Hewlett-Packard 78534A arterial pressure channel with a 15 Hz built in cut-off frequency.



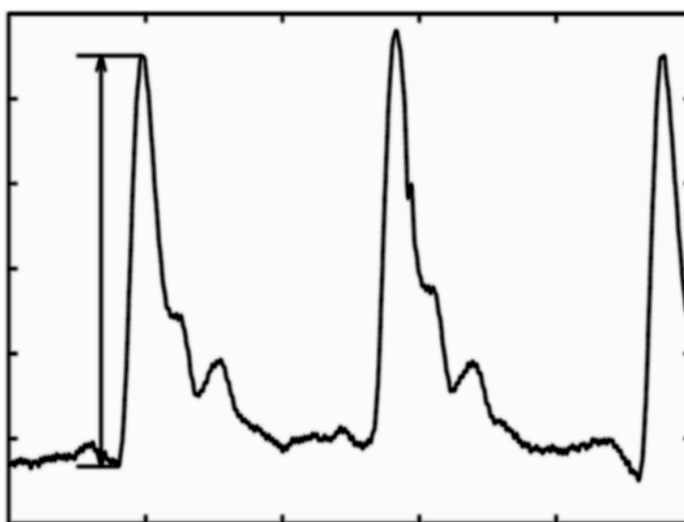
**Figure 4.3:** A pulse from [247]. Aortic flow pulsations were computed from arterial pressure by simulating a nonlinear, time-varying three-element model of aortic input impedance.

(1999) For the noninvasive studies, radial artery pressure pulse waves were recorded in [127] with an arterial tonometer sensor array (model N-500, Nellcor Inc). An invasive methodology was also incorporated using an 18-gauge catheter.



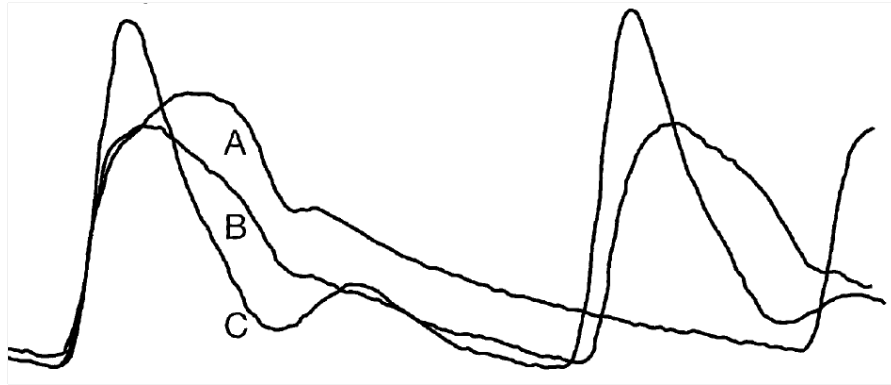
**Figure 4.4:** *A pulse from [127]. Age-related changes were evaluated in pulsatile arterial function.*

(2000) For each subject, the left radial artery pulse was recorded in [266] at each value of contact pressure with 700 Hz sampling rate. The contact pressure exerted on the pulse point was increased in increments of 4 kP, going from 4 kP to 136 kP. At each level of applied pressure, local maxima and local minima amplitudes were extracted, and the difference between them were calculated to give pulse amplitude.



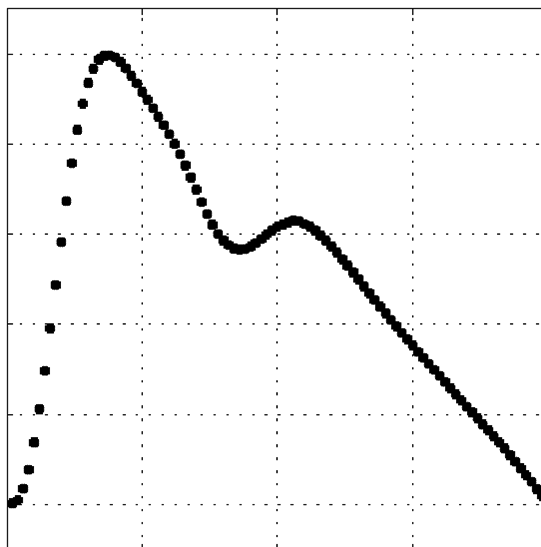
**Figure 4.5:** *A pulse from [266]. The variations in pulse with increasing contact pressure were studied.*

(2001) The arterial pressure was measured in [111] using a Baxter 150 cm pressure tube connection and Baxter Truware PX-600F sensors (Baxter Edwards Critical-Care, Irvine, CA, USA).



**Figure 4.6:** *A pulse from [111]. A pulse contour cardiac output algorithm was developed based on frequency analysis studies of the arterial system.*

(2003) Pulse wave was recorded in [6] using a system comprised of photoplethysmography (PPG) pulse amplifier channels with bandwidth of 0.15 to 20 Hz. All responses were recorded on a computer for subsequent off-line pulse wave analysis with sampling rate of 2500 Hz and 16-bit resolution.

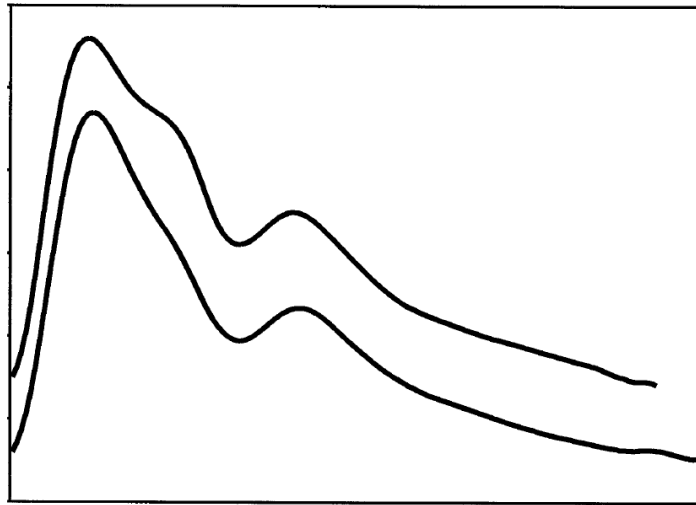


**Figure 4.7:** *A pulse from [6]. The age-related changes were demonstrated through pulse shape characteristics.*

(2004) Carotid femoral pulse wave velocity was calculated in [229] from measurements of pulse transit time and the distance traveled between two recording sites (pulse wave velocity = distance [in meters]/transit time [in seconds]) using a validated

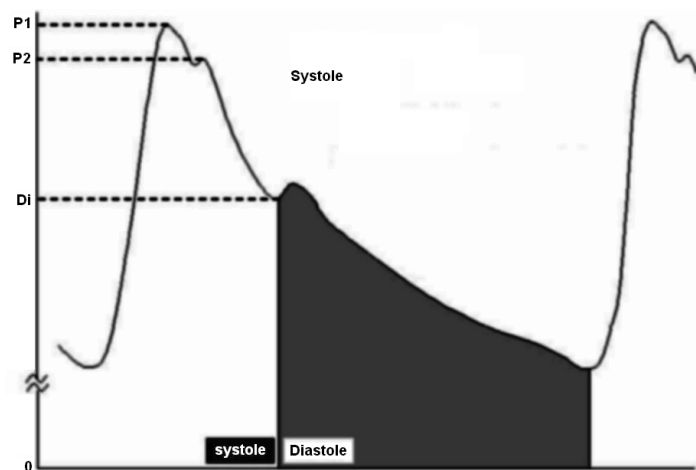


noninvasive device (Complior; Artech Medical, Pantin, France), which allows online pulse wave recording and automatic calculation of pulse wave velocity.



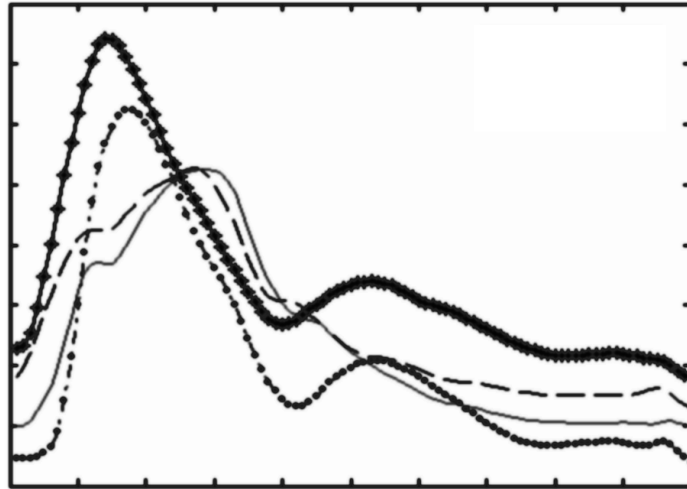
**Figure 4.8:** A pulse from [229]. The acute effect of cigar smoking was investigated on aortic elastic properties and wave reflection.

(2004) Radial arterial pulse waves were recorded in [41] noninvasively by an acoustic sensor using the CR-2000 device (HDI Inc.).



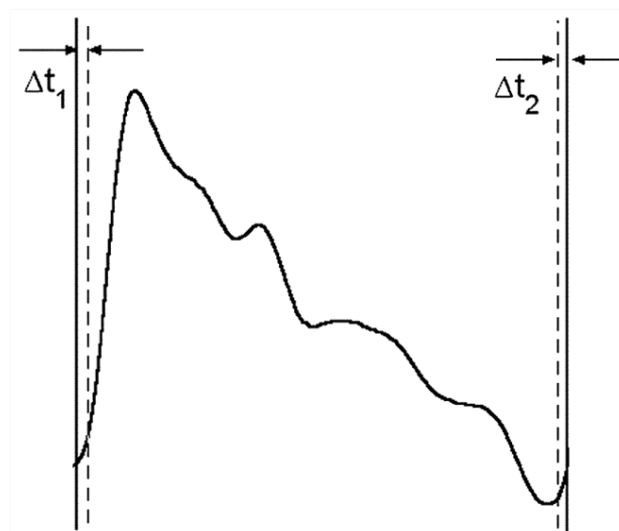
**Figure 4.9:** A pulse from [41]. The indices from systolic and diastolic pulse contour analysis were compared to correlate with traditional risk factors in cardiovascular disease.

(2005) The 4-order Daubechies wavelet was used in [258] for the analysis.



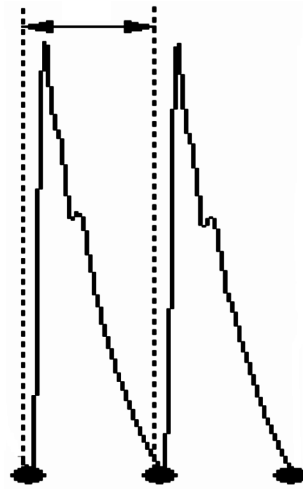
**Figure 4.10:** *A pulse from [258]. A classification approach using discrete wavelet model was proposed for pulse waveforms.*

(2005) Pulse was acquired in [218] using Impedance Plethysmography and Recording System. Two metallic electrodes (ECG standard type) were placed over the forearm artery line, and impedance was measured with a custom-made laboratory apparatus. Digital acquisition (sampling frequency of 200 Hz, at 16 bits) was carried out using a commercial system (BIOPAC System Inc, AcqKnowledge II for MP100WSW).



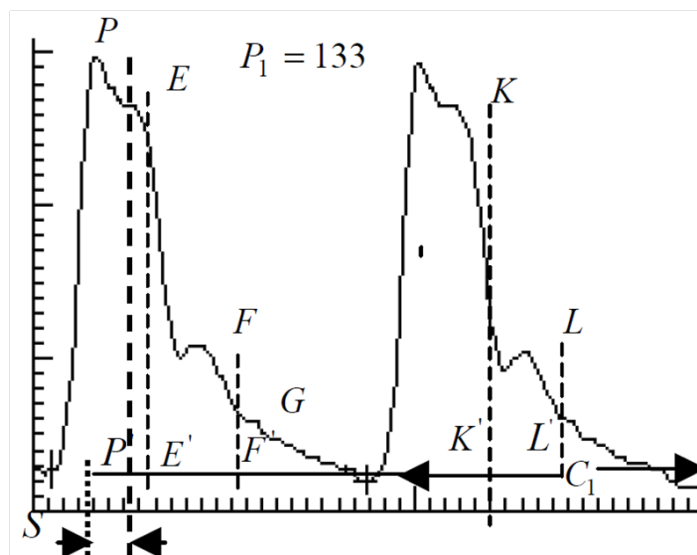
**Figure 4.11:** *A pulse from [218]. Detection of pulse cycle's onset and end was carried successfully on the data.*

(2005) Pulse signal in [237] was analyzed for detecting arrhythmic patterns.



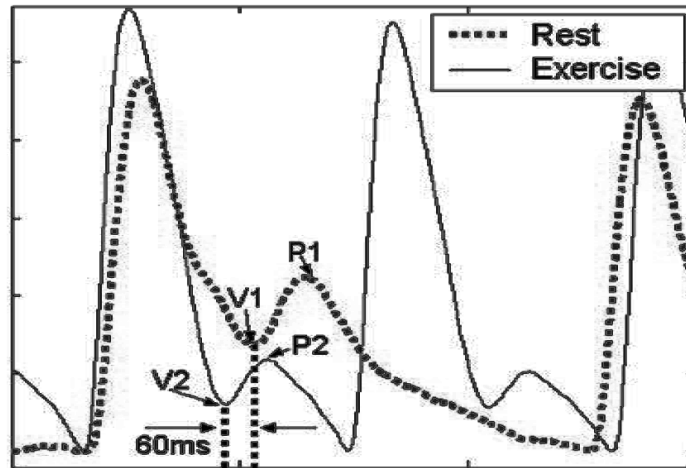
**Figure 4.12:** A pulse from [237]. A method was demonstrated to distinguish rhythmic and arrhythmic pulse patterns using the Lempel-Ziv decomposition.

(2005) The acquisition system in [233] comprised a pressure-adjusting pulse sensor. Pressure from 25 ~ 150 g could be regulated, and the pulse wave was recorded when the amplitude reaches the maximum.



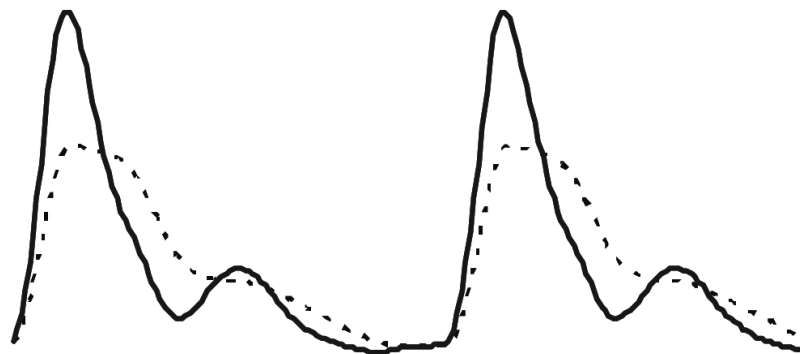
**Figure 4.13:** A pulse from [233]. A quantitative system for pulse diagnosis was constructed based on Bayesian networks to build the mapping relationships between pulse waves and pulse types.

(2006) Pulse was recorded in [259] using a typical data acquisition system comprising a pulse sensor, a strain cantilever beam sensor. The system was tied at wrist location using a band. The pulse was also recorded after performing progressive cycle ergometer for five minutes.



**Figure 4.14:** A pulse from [259]. The comparison was made for pulse contour variability for before and after exercise conditions.

(2006) Pressure waveforms were recorded in [253] from radial artery using applanation tonometry, with a probe applied to the wrist. Further, radial waveform was transformed into corresponding aortic waveform, in SphygmoCor system, through generalized mathematical transfer function.



**Figure 4.15:** A pulse from [253]. The cardiovascular risk was investigated using the radial pulse.

Very preliminary observations have been produced in the prior work (also explained in Section 4.5). One reason could be the noisy or less accurate data. The data analysis methods strongly depend not only on the clinician's experience but also on the quality of pulse signals.

## 4.2 Pulse Signals Obtained Using *Nadi Tarangini*

As explained in Chapter 3, we have used the optimized specifications of the components in *Nadi Tarangini*. In addition, we have used the *differential mode* of acquisition which results in the increase of details of the pulse signals.

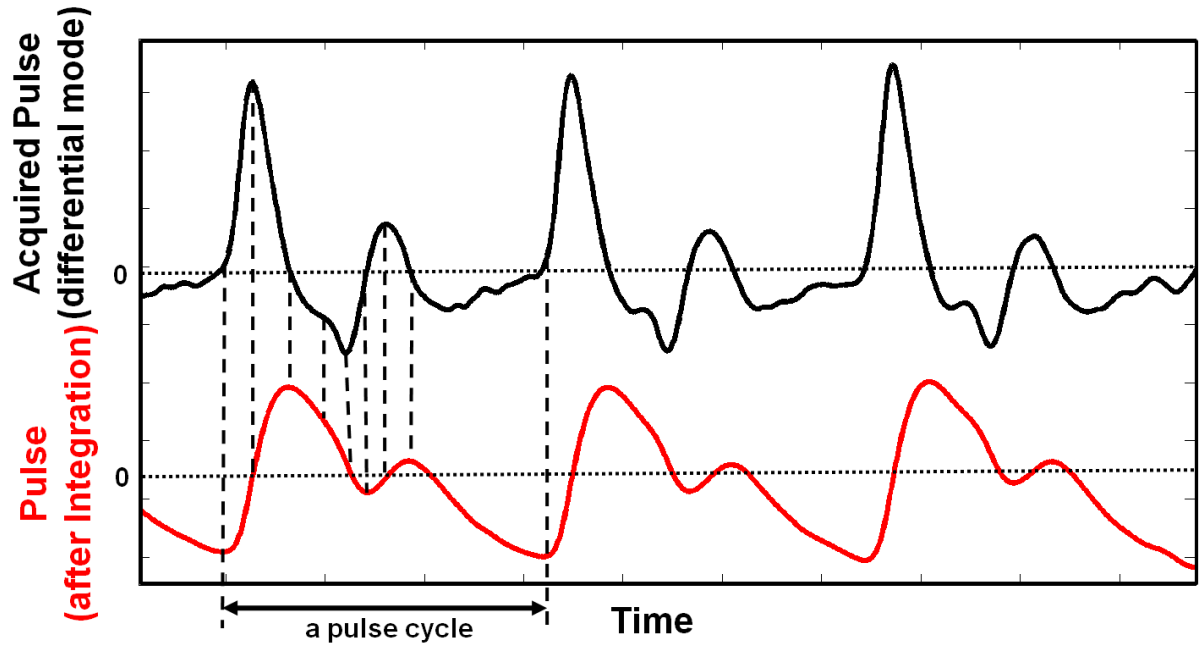
### 4.2.1 Differential Mode of Acquisition

There are two different modes of an acquisition methodology. A single-ended input requires only one input relative to the ground. On the other hand, a differential mode input requires two inputs that respond to the *difference* signal between them. In our system, we have used the differential mode with two inputs as the consecutive pressure values.

**Advantage.** The resolution of the digitizer is fixed (16-bit). If we consider just the single-ended input mode, then the whole 0–10 volts range (of the signal conditioning unit) will be spanned by the resolution of the digitizer. Instead, we use the difference between the consecutive numbers, which are smaller values and so now, the difference-range will get spanned by the digitizer. It helps in increased details of the data.

The actual pulse signal at each *dosha* can then be obtained by integrating the one acquired using *Nadi Tarangini*. Figure 4.16 shows an example of how the acquired pulse signal looks after integration.

The acquired pulse signal is rhythmic, and shows resemblance with an electrocardiogram (which is also a recording of the difference between values recorded by two electrodes). The figure shows corresponding important time domain properties



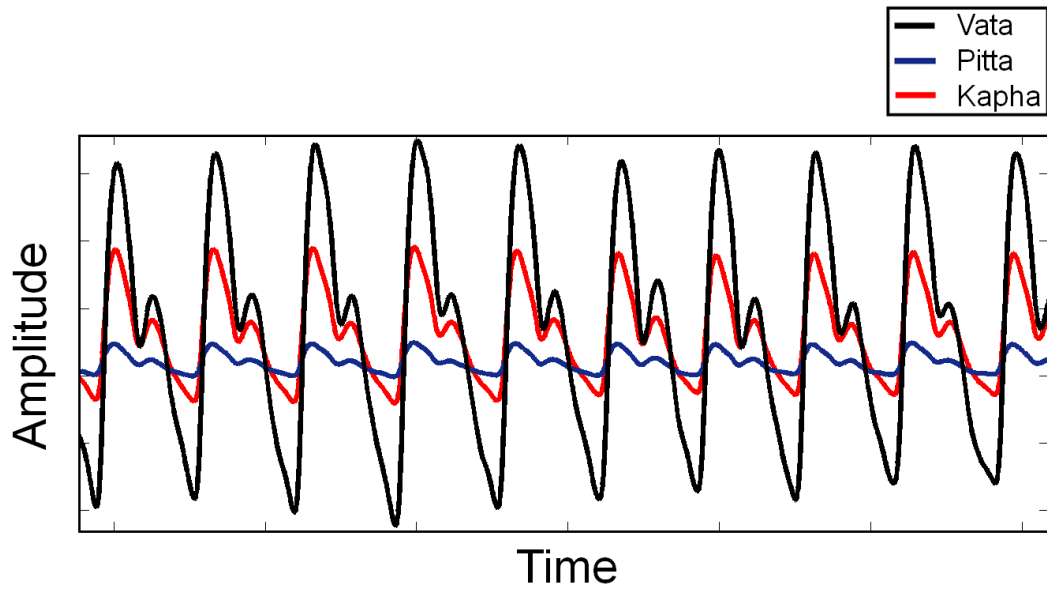
**Figure 4.16:** The actual pulse signal (bottom) is obtained by integrating the pulse signal (top) acquired using the differential mode. The dotted lines indicate the corresponding important locations on both the signals.

(dashed lines) on both the pulse signals. Here onwards, we have used both the models of pulse signals (acquired differential pulse and the integrated pulse) for the subsequent analysis. E.g. we use the acquired differential pulse for computing pulse rate variability measures (Section 5.5) and capturing multifractal nature in arterial pulse intervals signals (Section 6.1), while we use the integrated pulse for the PSWT-based analysis (Section 5.2) and topological analysis (Section 6.3).

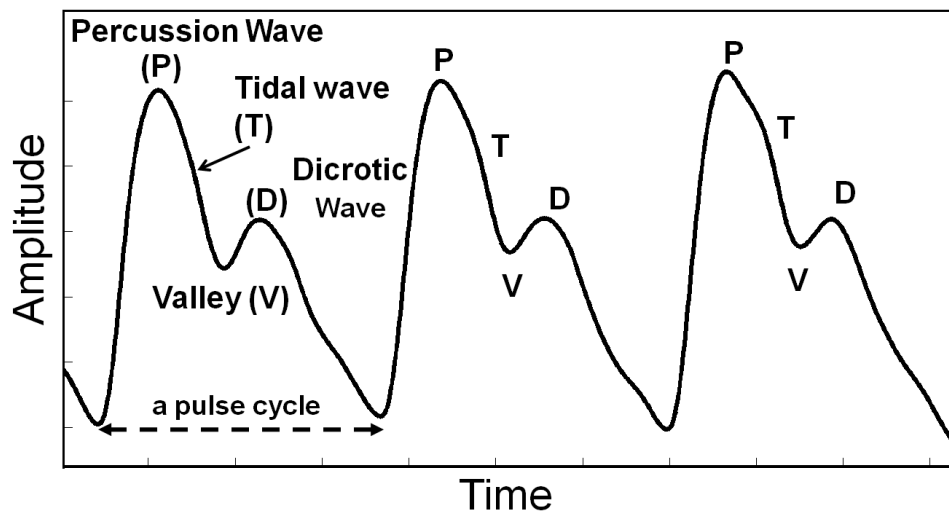
#### 4.2.2 *Tridosha* Pulse

Figure 4.17 shows a sample *Tridosha* pulse from our database. Appendix C displays many more examples obtained from different subjects. The three colors indicate three separate time series obtained simultaneously from VPK locations (refer Figure 2.13). In other words, at each time instant the output is a 3-dimensional vector.

**Pulse subwaves.** Figure 4.18 shows an example of a pulse signal acquired at the *vata* location.



**Figure 4.17:** An example of a pulse signal obtained from a subject's left hand. The three colors represent VPK locations of acquisition.



**Figure 4.18:** Typical behavior of a pulse signal. This is a vata pulse from a subject's left hand. It continuously and rhythmically follows P-T-V-D pattern.

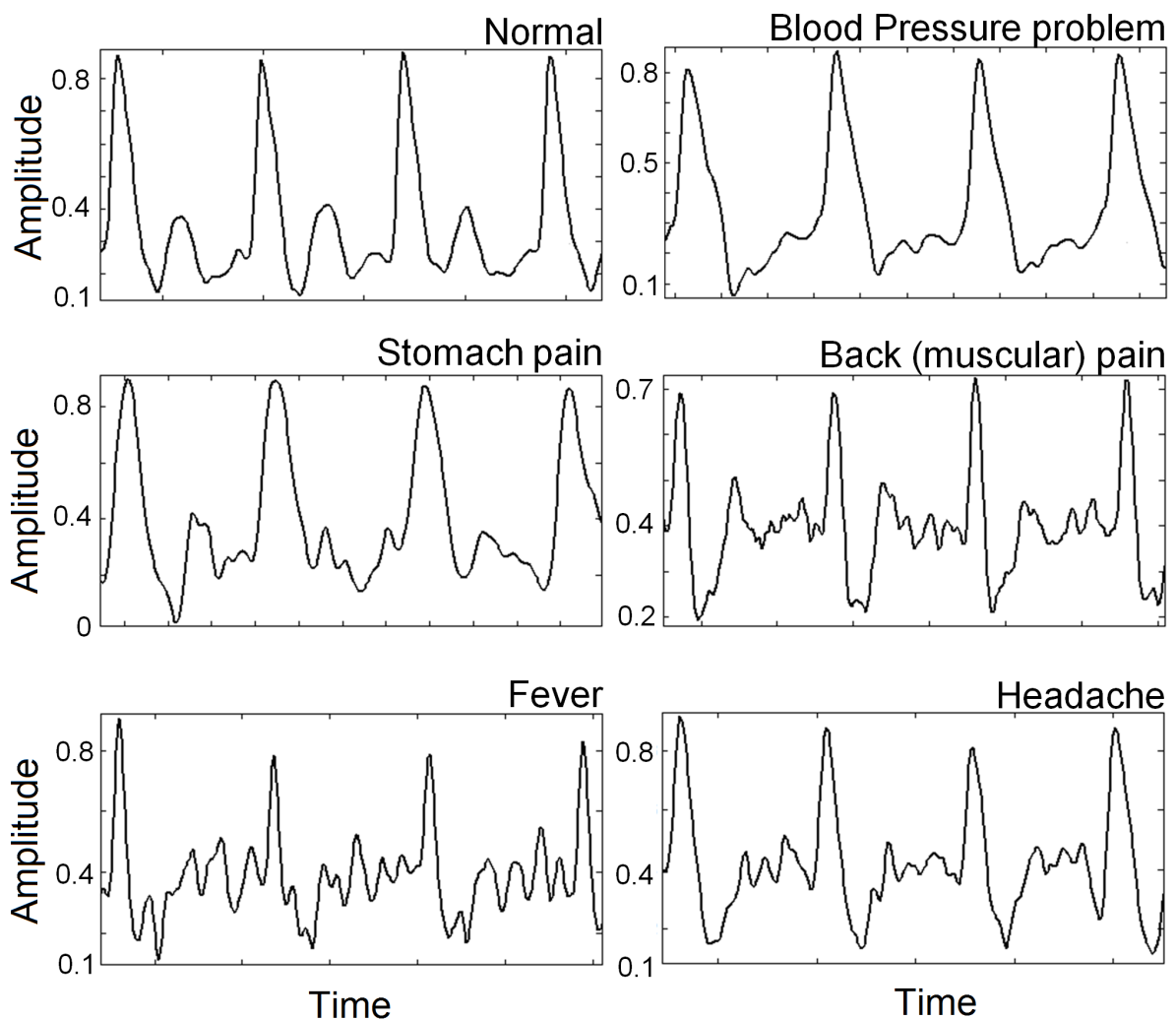
A pulse cycle is usually composed of the following subwaves [272]:

- A percussion subwave (P)
- A tidal subwave (T)
- A valley (V)

- A dicrotic subwave (D)

These subwaves have their origin and properties as explained in Chapter 2. The subwaves should be present in a typical pulse cycle with definite amplitude and duration to indicate proper functioning of the heart and other body organs.

We now provide 6 more pulse signals (only *vata dosha* location) from our database in Figure 4.19. They are from different subjects of different age groups. These examples exhibit different patterns and nature of different disorders.



**Figure 4.19:** Vata pulse signals of different disorders acquired using Nadi Tarangini. Different patterns show possibility of application of the rigorous data analysis algorithms for the identification.

It can be observed that there are distinctly observable patterns in each of the six pulse



signals. For these examples, our observations on the visible patterns are:

- A pulse cycle of healthy behavior typically has a strong main peak (at P-subwave) with high amplitude, and two distinct secondary peaks indicating the presence of T, V and D-subwaves. The behavior follows a continuous P-T-V-D rhythm (refer Figure 4.18).
- The pulse cycles showing fever behavior contain many secondary peaks of varying amplitudes and durations. These differences in such behaviors will be more clear when we discuss about the frequency spectrum in Section 5.5.2.
- The pulse cycle of high blood pressure behavior does not contain any secondary peaks (or contains with negligible amplitude). Thus, it shows non-typical behavior with missing T and/or V-subwaves.

**Motivation for rigorous data analysis:** We observe variations in various parameters such as the amplitudes of secondary peaks, rising & falling slopes, systolic & diastolic energies, and velocities. These variations (visible or hidden) in the patterns of the pulse signals (or pulse cycles) are sometimes hard to extract. Therefore, an advanced and rigorous data analysis is required for the efficient identification.

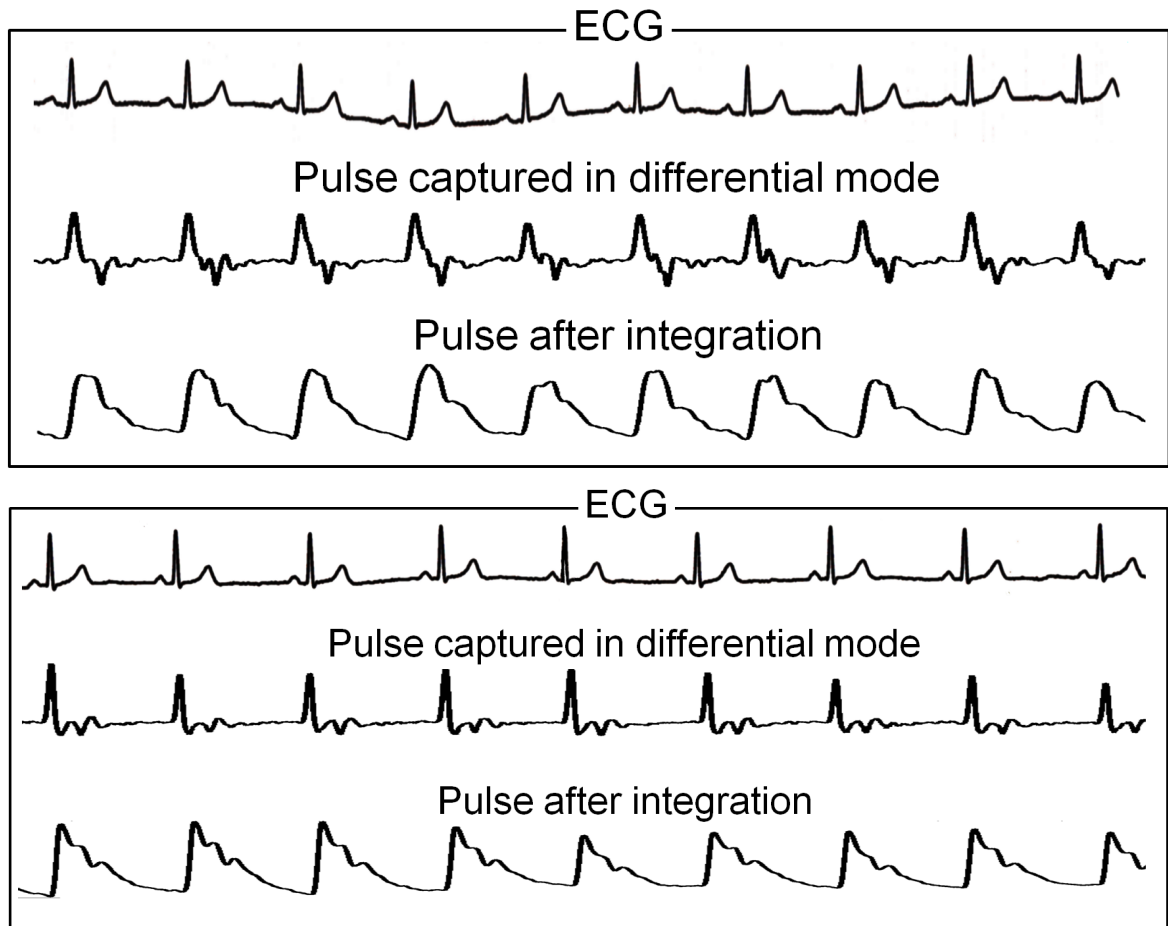
### 4.3 Comparison of Pulse with Electrocardiogram

As explained in Chapter 2, the pulse is generated due to the continuous and rhythmic pumping action of the heart. However, it is further transmitted by the blood flow through arteries, and goes a long way from the heart. It is believed in *Ayurveda* that because of many factors affecting the flow (such as Section 2.2), the pulse signal has more information, and thus is useful in the diagnosis.

An electrocardiogram (ECG) quantitatively captures the heart actions. Currently, this acquisition and its subsequent analysis is in the advanced stages, such as automatic detection of various arrhythmias shown in Appendix A.

Studies in [91, 170] have shown the comparison of the heart rate variability (HRV)

with peak-to-peak variations obtained using photoplethysmography. They indicate that, the two time series are different in many situations. The comparison of the ECG signals with the pulse signals is beyond the scope of this thesis. Here, we provide two examples of simultaneously obtained ECG and pulse signals in Figure 4.20.



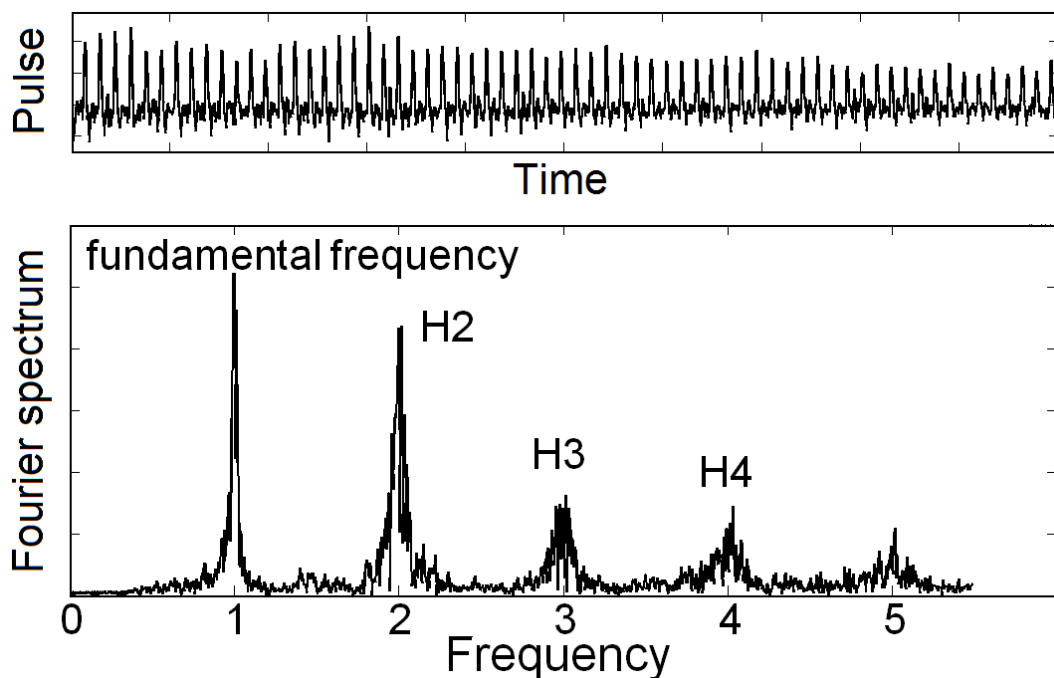
**Figure 4.20:** *Two examples of simultaneously obtained ECG and pulse signals. The top row, middle row and the last row respectively display the ECG, the pulse in differential mode, and the pulse after integration.*

Though the three signals show positive correlation (similar set of actions one after the other) with each other, their patterns are completely different. In this thesis, we study the arterial pulse signals to establish its importance and efficacy, through its accurate acquisition and analysis.

## 4.4 Important Properties of the Pulse Signals

In literature, the important properties of the pulse signals are depth, intensity, amplitude, frequency, rhythm, length, type or shape, temperature, quantity, texture, and width [104]. Many of them are interrelated, such as intensity and amplitude. We describe few properties related to the time series format, in which our data is stored. We also describe other properties essential for the signal processing, namely completeness and reproducibility.

- **Pulse rate:** A simple desirable feature in any physiological system is the rate (heart rate or pulse rate). Pulse rate is the average number of pulse cycles per minute. We compute it by finding the fundamental frequency in the Fourier spectrum as shown in Figure 4.21.

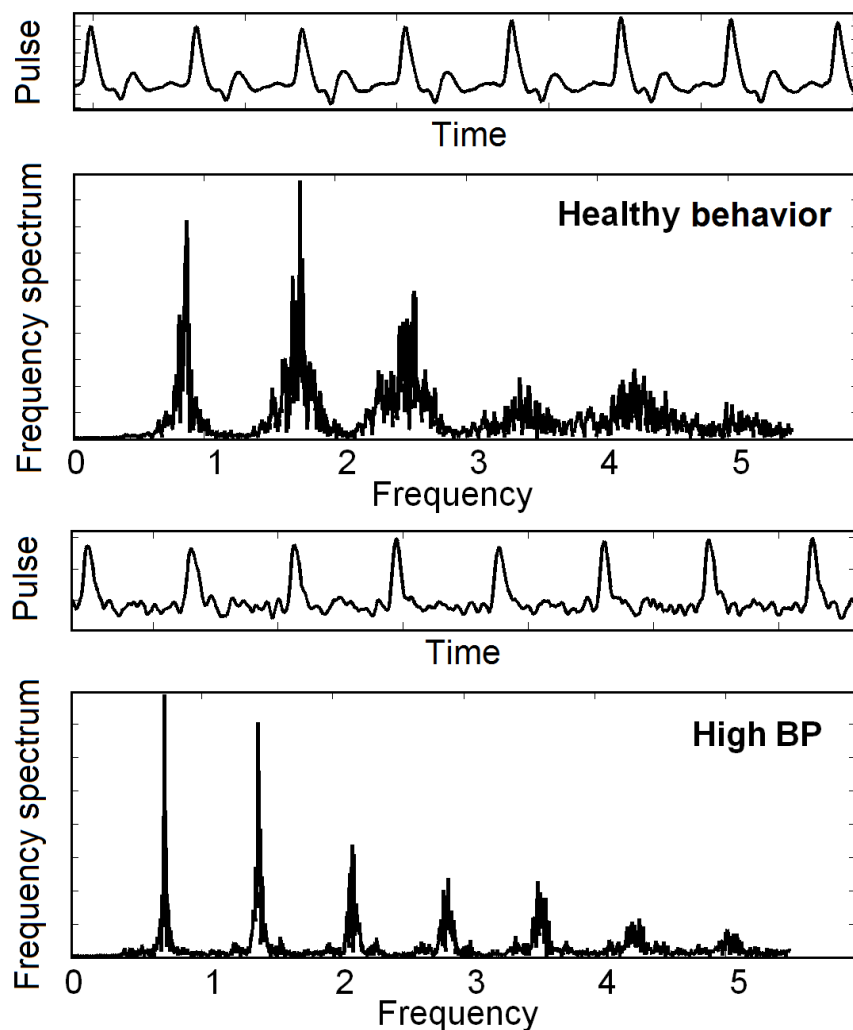


**Figure 4.21:** *Fourier spectrum of a pulse signal (healthy behavior). The fundamental frequency provides the pulse rate. More subtle second order effects are the secondary peaks within each pulse cycle reflected in the subsequent harmonics.*

In modern medicine, doctors sense the pulse on wrist of a subject to check his or her heart rate (by counting the pulse rate). We want to explore the pulse signal beyond just the computation of the heart rate, and analyze them to identify

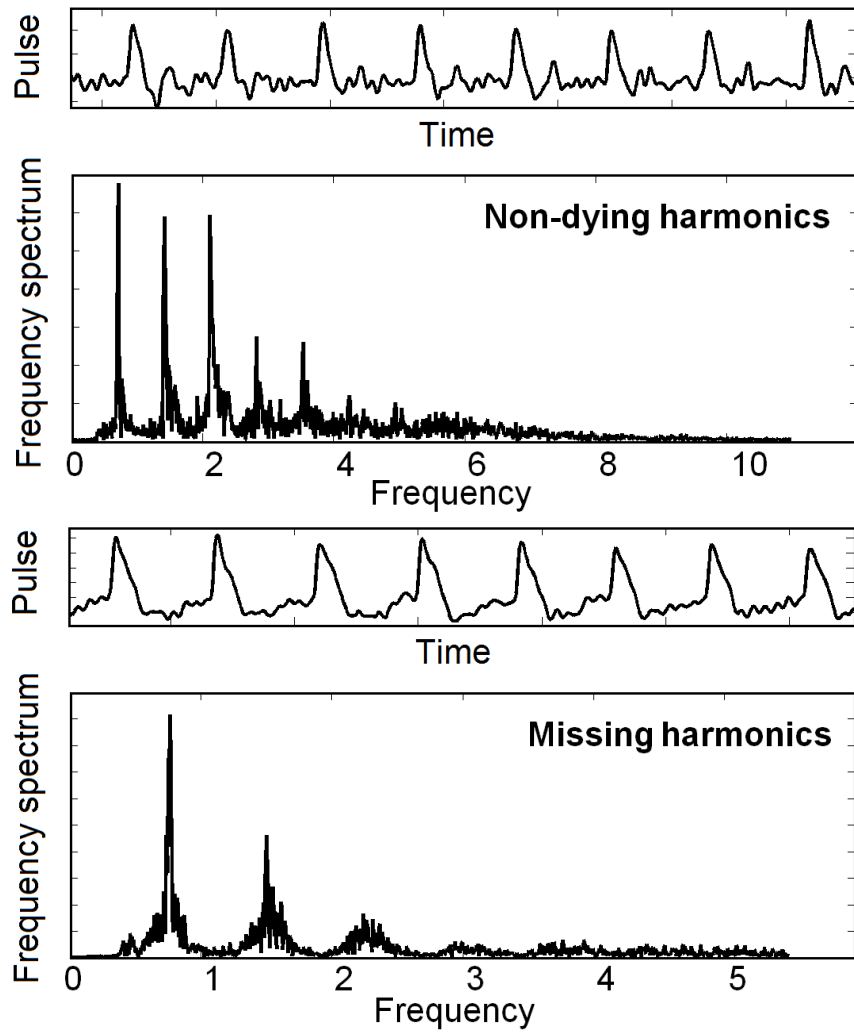
possible disorders and other important properties. For example, the subsequent harmonics and other frequencies in the spectrum can be further studied for correlating with the physiological conditions of the subject. Actually, we have used wavelets and nonlinear dynamics instead of Fourier spectrum for this purpose.

- **Frequency:** We have already used the fundamental frequency of the spectrum for computing the pulse rate. In Figure 4.22 and 4.23, we provide Fourier coefficients of pulse signals showing different behaviors.



**Figure 4.22:** *Fourier coefficients for two pulse signals showing different behaviors. The variations in the two spectrums show possibility of classification.*

More subtle second order effects are the secondary peaks within each pulse cycle. They are primarily reflected in the harmonics H2 through H5 (and sometimes



**Figure 4.23:** *Fourier spectrums of two different pulse signals showing different coefficient patterns. The secondary peaks in the pulse signals are usually reflected in the harmonics, and thus can be further used for the analysis.*

more). An effective cross-correlation coefficient was established in [29] with these harmonics as:

$$\text{k ratio} = \frac{\sqrt{(H2)^2 + (H3)^2 + (H4)^2 + (H5)^2}}{H1}$$

We observe that the Fourier spectra vary with respect to the regularity and nature of secondary peaks. These coefficients can be used as important features for the classification. The strength of the spectral harmonics accounts for the morphology of the pulse, and can be used in detection. We have used effective

wavelet coefficients instead of Fourier, which extract important information at various frequencies (scales) keeping the time (space) information.

- **Rhythm:** Due to continuous pumping action of the heart, the pulse signal has a considerably regular rhythm. Similar set of actions (say P-T-V-Din Figure 4.18) appear on the time axis, separated by a pseudo-period.
- **Shape:** The overall shape of a pulse cycle is described by all the above properties including amplitude, rate, rhythm, and systolic & diastolic energies. Actually, the *Ayurvedic* practitioners sense the overall shape of the pulse (cycles) with high concentration, and relate them graphically to a more concrete images, such as *Sarpa* (Cobra) for *vata* pulse, *Manduka* (Frog) for *pitta* pulse and *Hamsa* (Swimming Swan) for *kapha* pulse. Further minute variations then help in diagnosing the subject. We have quantitatively analyzed the pulse signals in our database to find important patterns which help to classify.
- **Completeness:** In order to check the completeness of the acquired time series, pulse signals were acquired with digitizers having accuracies of 8-bit, 12-bit and 16-bit. The remaining components of the system (subjects, sensors of '0–4 inch  $H_2O$ ' pressure range, and signal conditioning units) were kept the same. In the experiments, the details captured by the 12-bit digitizer were more as compared to the 8-bit digitizer (Observe the slope variations in Figure 3.4). However, there was no significant new information from 12-bit to 16-bit upgradation, so we say that all the details have been captured. The said sensor with 16-bit resolution provides very high details, which are able to capture the self-similarity at various scales (refer Chapter 6). No other work has still been able to capture such property due to lack of details. Thus, we say that *Nadi Tarangini* is replete with details and variations.
- **Reproducibility:** Pulse signals of a single healthy subject were recorded at five different times (10.20 AM, 1.05 PM, 2.45 PM, 4.40 PM, 7.45 PM) for five consecutive days with the same settings. It was observed that the number of data points per pulse cycle, and the approximate entropy (ApEn) [165] (ApEn is a "regularity statistic" that quantifies the unpredictability of fluctuations in

a time series) of the complete signals remained almost constant for the same timings during the five days. Our experiments showed that our system is fairly reproducible. However, exact reproducibility of the pulse is not possible, as the pulse is sensitive to the mental status, stresses, and thoughts (and which results in variability as explained in Section 5.5).

## 4.5 Prior Work in Pulse Analysis

It is clear that, a pulse signal with healthy behavior will show a typical P-T-V-D pattern with above characteristics. Disorders or defects in the biological system cause alterations in its normal physiological processes. This affects the performance, health, and general well-being of the system, and thus acquired pulse signals. Therefore, the pulse signals will follow different patterns in one or all respects from the corresponding healthy signals.

Survey on the pulse-based analysis [113, 260] states that from a pulse signal, various physiologically, pathologically and psychologically significant parameters can be extracted. The pulse signal, originated by the pumping actions of the heart, is transmitted by the blood flow throughout the body. Thus it is affected not only by the condition of the heart, but also by the condition of other organs. By analyzing the pressure fluctuation signal of the pulse, we can detect and predict many symptoms to extract useful information [54, 236].

### 4.5.1 Circulation Parameters

Heavy contribution of the overall pulse-based analysis has been done in the area of hemodynamics to study the circulation parameters. *Hemodynamics analysis* [152] provides information about the interaction between physical properties of the arterial circulation and the left ventricle. This is studied by the descriptive and quantitative analysis of the arterial pressure pulse waveform to compute the risk factors [51, 94, 110, 126, 128] or through the theoretical analysis of the wave propagation [125, 215, 216].

**Cardiac output (CO)** is used to evaluate cardiac function in critically ill patients, to aid in the diagnosis of cardiovascular disease, and to guide therapeutic decisions in complex clinical situations. Continuous determinations of the cardiac output are important for monitoring acutely ill patients, since their hemodynamics status may vary significantly during a short period of time [111, 120]. **Augmented pressure (AP)** is defined as the difference between the second and the first systolic peak. **Augmentation index (AIx)** is AP expressed as a percentage of the total pressure [41, 154, 221, 243]. AIx is sensitive to many factors including heart rate [252], carotid atherosclerosis [55], and blood pressure [132]. AIx depends on ventricular ejection [153], the timing of reflected waves (determined mainly by pulse wave velocity in the aorta [247] and large arteries and hence, on the stiffness of these arteries), and the amount of reflection (determined mainly by arterial tone at the major sites of reflection).

#### 4.5.2 Extraction of Important Features

To diagnose a subject through the quantitative study of pulse signals, important information (resembling some disorders) is extracted in the form of features. These features are trained for considered classes and are then used in future for identifying unclassified samples. Different sets of features are possible depending upon the *property of the pulse signals under consideration* for the analysis.

Cleaning of the pulse signals is an essential prerequisite to the analysis, as the pulse signals suffer from manual, mechanical, and 1/f type noise as well as the subject-derived distortions known as the baseline drift [197, 261, 263].

*Contour analysis* is one of the elementary steps in data analysis, which visually displays different contour characteristics of the pulse [178, 191]. Examples include studying variations with different age-groups [127], and effect of heavy exercise [259] and heart-problem [105].

*Period-to-period variability.* The nonlinear variations in characteristic parameters from a pulse cycle to the next are endowed with important physiological significance, and have been studied on the finger pulse [60, 74, 170, 233, 255]. It was further examined in



[91, 170, 238] whether these intervals can be alternative data to the R-R intervals (ECG) for the same analysis. It was observed that the respiratory pulse rate variability does not precisely reflect the respiratory heart rate variability in standing healthy subjects, and in patients with low HRV.

Different pattern recognition and data mining algorithms have been applied to extract the hidden patterns of the pulse signals, such as Fourier coefficients [37], wavelet coefficients [17, 109, 258, 271], Lyapunov exponent-based analysis [265], Lempel-Ziv complexity-based analysis [237, 262], approximate entropy (ApEn) [235], fuzzy theory [107, 233]. All the approaches correlate the computed features to some pattern or 'property' of the pulse signal.

### 4.5.3 Problem Specific Observations

It is yet not possible to prove or achieve global rules, as the pulse-based scope is very wide. Here, a summary of some intermediate results (preliminary observations) is provided to understand different types of possible outcomes from the pulse-based analysis. Also, most of this research has been done based upon Traditional Chinese Medicine (TCM). *Ayurveda*, Korean medicine, and other alternative medicines, though started, lack the advanced quantitative research. The spectrums of pulse signals have been analyzed for patients with liver problems [112], Atherosclerosis [55], cardiovascular disease [41], and vascular abnormalities [128].

- **Variations with age:** The older subjects have increasing arterial stiffness resulting in changes in the propagation of the blood flow to the periphery, and thereby influencing the peripheral pulse timing and shape characteristics [52, 226]. Thus, *Aging alters arterial pulsatile function and produces consistent changes in the pressure pulse contours* [19, 127].

The analysis of the pulse characteristics with age is important as AIX increases with age (and blood pressure), and is elevated in subjects with other risk factors for cardiovascular disease, including diabetes and hypercholesterolemia [132]. The age-related changes in the carotid pulse waveform were explicable by the

increased pulsatile load and were quantified by AIx, the ratio of the augmented systolic pressure to the pulse pressure in central arteries [51, 94, 143].

The association between age and changes in pulse shape characteristics was studied at various pick-up locations including the ears, fingers, and toes in [5]. It was observed that the oldest subject group of the four groups (subjects younger than 30 years, 30-39 years, 40-49 years, and 50 years of age or older) was the most visually distinguishable. The power in the harmonics of the pulse wave decreases with age [190].

- **Variations with gender:** There have been many gender specific studies on the pulse signals, which is another important dimension to be considered in the analysis [38, 41, 59, 127]. There are gender-related differences on the left ventricular structure and on symptomatic heart failure [72]. Such differences were observed also in the pulsatile vascular load as indicated by AIx and other features of the central arterial pressure waveform. In particular, females differ significantly from males in central arterial pressure waveforms as assessed by carotid artery tonometry. The combination of hemodynamic load, higher relative wall thickness, and lower subendocardial ratio in women may further help to explain their susceptibility to heart failure in states of cardiac stress.
- **Cigar smoking:** The effect of cigar smoking was evaluated with carotid-femoral pulse wave velocity and wave reflection with augmentation index of the aortic pressure waveform [198, 199]. It was observed that the cigar smoking produces a significant increase in the pulse wave velocity (by 0.80 m/sec), denoting an increase in aortic stiffness [182, 229].

Actually, all the above mentioned experiments were carried on the pulse signals obtained from various locations on the body (forehead, ear, neck, wrist, finger, and toe). At each location, the shape and the characteristics of the pulse signals are different. In our studies, we use the pulse signal obtained on wrist, as it is easily available, and thus is studied traditionally by the experts.

#### 4.5.4 Pulse Analysis in India

Important works in Indian traditional medicine include computing the pulse rate [183], detecting peaks and other important points [168] using the wavelet method, studying the position-specific nature [208] of the pulse wave signal, studying the *Siddha*-based disease diagnosis [116], studying of gastric activities [264], and studying effect of upper arm cuff pressure on the pulse morphology [207]. Accurate and clean quantitative acquisition of the pulse, followed by its extensive analysis to “understand” the pulse is still in the initial stages.

**A survey driven approach** (without actually acquiring the pulse signals) has been implemented in [90], to understand the correlation between *Tridosha* (refer Section 2.3.2) and the constitution (*prakriti*) of a subject. A questionnaire was designed using a comprehensive list of 28 features for identification of the constitution. Each question (including general physique, type of skin, style of walking, eating habits, nature of frequent dreams) had three options for the response. Five different types of *prakriti*'s were considered – VP (predominance of V and P is secondary), PV, PK, KP and VK. By applying algorithmic heuristic approach to the exhaustive list of qualitative features/factors commonly used by *Ayurvedic* practitioners, the constitution of the subject was determined.

Another interesting work at Bhabha Atomic Research Centre (BARC) exploits the diagnostic utility of the impedance plethysmography (IPG). It is a recording of the blood volume changes obtained by measurement of the electrical impedance (EI) of any part of the body [86]. Blood is a good conductor of electricity and so the amount of blood in any part of the body is inversely proportional to the EI offered by the body segment. Thus a recording of instantaneous EI and its first time-derivative yields important information about the circulation. The EI is measured by passing a fixed (1–10 mA rms) sinusoidal carrier current and recording the voltage developed along the current path. These voltage signals resemble the pulse if taken from the wrist location. The technique was used to study the blood flow in patients with (1) arterial occlusive diseases, (2) valvular heart disease and (3) congenital heart diseases, and to evaluate myocardial function in coronary artery diseases [159, 160].

## 4.6 Challenges in Pulse-based Analysis

In the previous section, we mentioned various methodologies to extract important properties of the pulse signals. Some of the attempts were successful, and few presented preliminary observations with the future scope. Actually, the enormous hidden information in the pulse signal and its complexity itself, though important, is the largest challenge in the analysis of the pulse signals. Further challenges in the analysis of the pulse signals are:

**Complexity and heterogeneity of the data:** This is an important limitation due to a wide diversity in opinions of the physicians, approaches to diagnose, noting their respective diagnosis parameters and so on. Same set of rules are sometimes not applicable. In the acquisition of the pulse signals, data of one of the *dosha* locations could be negligible (insignificant as compared to other *doshas*), and is considered as a set of missing values.

**Size:** From each subject, three pulse signals are recorded (each with 500 Hz sampling rate in our system). Therefore, the size of the data is very huge, especially when the data is acquired for a longer duration. Nowadays, the memory and processing power are not much of a concern, although analysis of the huge data could be time consuming.

**Special status of medicine:** Medicine has a special status in science, philosophy and daily life. The outcomes of medical care are life or death, and they apply to everybody.

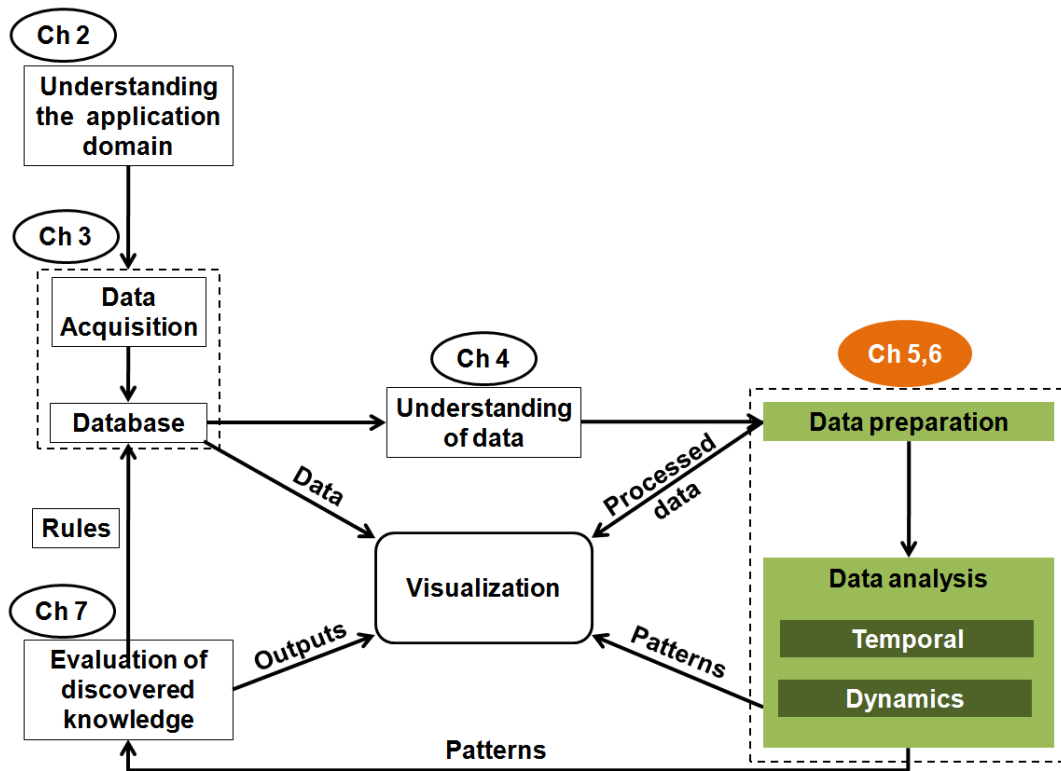
**Poor mathematical characterization of the pulse:** This is the most important challenge in our study. In many physical sciences, scientists collect data which they can put into formulas, equations, and models that reasonably reflect the relationships among the data. In contrast, the underlying data structures of the pulse signals are poorly characterized mathematically. The conceptual structures of the pulse consists of only word-descriptions (and sometimes poor images). Therefore, there is no comparable formal structure yet, into which an analyzer can organize the information, which might be modeled by classification, clustering, regression models, or sequence analysis.

## Conclusions

In this chapter, we explained the nature of pulse signals obtained using *Nadi Tarangini* through examples (more examples are provided in Appendix C). We showed that our data has high details with important properties including rate, frequency, rhythm, shape, reproducibility, and repletteness. Diseases or defects cause alterations in the pulse patterns from the corresponding typical behavior. We listed out important prior works to find effective ways of capturing such different behaviors.

The properties, challenges, and progresses of the pulse confirms the importance of its consideration for the diagnosis. However, *a variety of approaches (individually and clubbed together) are needed to make these computer-based decisions useful and effective*. Each method will reveal different 'property' of the pulse. In next two chapters, we present seven broad approaches to examine the pulse signals considering their characteristic properties, namely period-to-period variability and nonlinear dynamics.





---

## Chapter 5

# Capturing Pseudo-periodic Behavior

---

*In this chapter, we explain the physiologically important behavior of the period-to-period variability of the pulse signals. Due to the pseudo-periodicity, though adjacent pulse cycles look similar in shape, they show different patterns.*

*We study the pseudo-periodic nature of a pulse signal by constructing its pitch synchronous representation, to extract the overall trend. Based on this trend, we propose our novel PSWT-DTW approach to compute useful time domain parameters (TDP) of a pulse signal.*

*Finally, we formalize pulse rate variability (PRV) measures by computing various time domain, frequency domain and nonlinear Poincaré based measures. We provide evidence for the detection of disorders in patients using these measures, paving the way for future clinical use.*



The circulatory system is composed of the heart, and blood vessels including arteries, veins, and capillaries. To move blood throughout the body, the heart muscle contracts rhythmically (refer Chapter 2) and systematically. Therefore, most of the signals that arise from physical systems are oscillatory in nature. However, these signals, including pulse signal, are not periodic in a strict mathematical sense, but fluctuate with amplitude, frequency modulations.

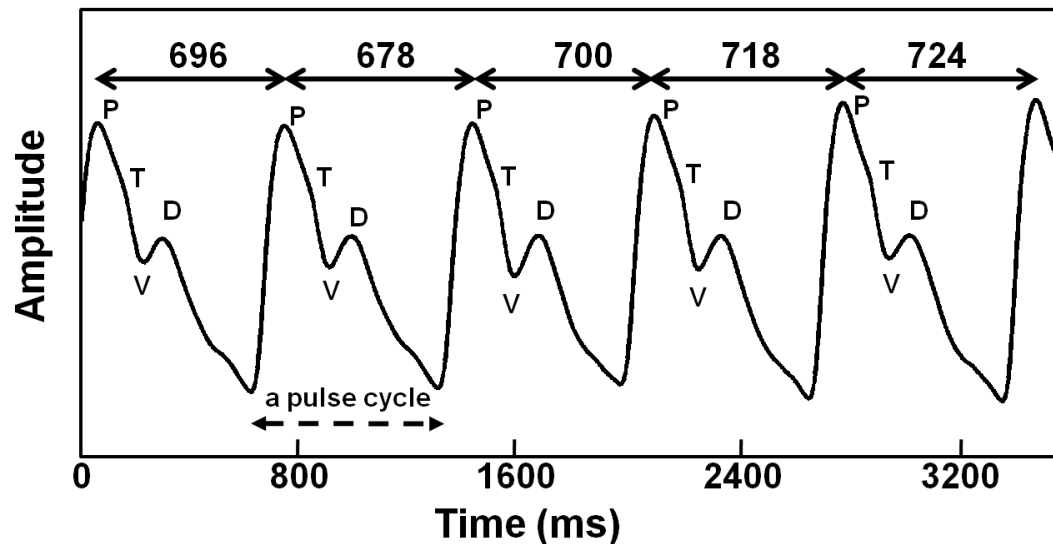
The peak-to-peak variations in the ECG signals (also known as R-R intervals) are termed as heart rate variability (HRV). The clinical relevance of the HRV was first appreciated [75] in the year 1965, in which it was noted that fetal distress was preceded by alterations in inter-beat intervals before any appreciable change occurred in the heart rate itself. The clinical importance of HRV became apparent in the late 1980s when it was confirmed that the HRV was a strong and independent predictor of mortality following an acute myocardial infarction [14, 100, 117].

Currently, the analysis of the HRV is in advanced stages to evaluate the equilibrium between the sympathetic system (which accelerates the heart rhythm), and parasympathetic system (which decelerates the heart rhythm) in cardiac pacemaker cells [3, 30, 78, 214]. It has been shown that the properties of the HRV play an important role in the better prognosis in patients with diabetes [66], autonomic dysfunction and heart failure [248].

Related, but inexpensive, physiological signals derived from the finger plethysmogram, and the arterial blood pressure have also been preliminarily studied for the variability [40, 91]. It was further examined whether these intervals can be alternative data [91, 170, 238] to the R-R intervals (ECG) for the same analysis. It was observed that the respiratory pulse rate variability does not precisely reflect the respiratory heart rate variability in standing healthy subjects, and in patients with low HRV. We establish the pulse rate variability (PRV) in this chapter to show its efficacy. In future, these alternative approaches will receive appropriate recognition in the diagnosis.

## 5.1 Pseudo-periodicity

As seen in Figure 5.1, a pulse signal usually (but not always) follows a pattern of P-T-V-D subwaves (refer Section 4.2.2). These evolutionary (period-to-period) changes in the behavior are medically significant and are explored in this chapter.



**Figure 5.1:** Typical behavior of a pulse signal. A continuous, rhythmic P-T-V-D pattern is observed. Even though the pulse cycles look similar, they vary in amplitude and pseudo-period.

The pulse signal in Figure 5.1 seems to repeat the P-T-V-D pattern in each pulse cycle, but actually the signal contains some variations which are sometimes not visible to the naked eye. In this example, the consecutive peak-to-peak durations are 696–678–700–718–724 milliseconds and so are different. The signal is not mathematically periodic; however, to classify it as non-periodic would do injustice to the systematic rhythm present.

There are many factors responsible for such pseudo-periodicity in the physiological signals [166]. The momentary heart rate, and the duration of the peak-to-peak intervals is a consequence of constant interaction between

- The intrinsic activity of the sinus node
- The influence of the autonomic nervous system, various substances circulating in the blood and present in the heart tissues

Breathing also appears to be an important factor modulating the heart rate. It causes heart rate acceleration during inspiration, and its deceleration during expiration. The changes in the blood pressure modulated by baroreflex are another example of a separate system regulating the heart rate. Further, the control of the heart rate is modulated by both sympathetic and parasympathetic branches of autonomic nervous system as well as many other autonomic reflexes. All these systems and reflexes are responsible for the changing of the durations from one cycle to another.

### 5.1.1 Estimation of Pseudo-periods

As the pulse cycles differ from each other, detection of the start and the end of each pulse cycle is very important for further analysis. If any one of the peaks is missed, then the peak-to-peak distance at that missing peak will be approximately twice the average. Conversely, the secondary peaks (e.g. D-subwave in Figure 5.1) should be avoided.

For physiological signals, the popular methods of peak detection [13, 45, 48, 65, 146, 218] are based on threshold cutoff, wavelet coefficients, maximum rising phase of pulse signal, and autocorrelation function. Usually with the autocorrelation function, pitch period of the signal is computed and then pseudo-periods are searched near the locations of the pitch period.

In our experiments, we compute the locations (start and end) directly. We make use of the relatively high amplitudes and the steep slopes of the pulse signal with a wavelet-based setting.

Wavelet transform (WT) helps in viewing a function at various levels of approximation or resolution [200]. Using this concept, a complicated function can be divided into several simpler ones and studied separately. It provides a set of wavelet coefficients ( $W$ ), which indicate how close the signal ( $f(x)$ ) is, to a particular mother wavelet ( $\psi$ ), which are basically oscillatory functions satisfying certain properties. WT is defined as:

$$W_{s,x_0}(f) = \int_{-\infty}^{+\infty} \frac{1}{s} \psi\left(\frac{x-x_0}{s}\right) f(x) dx \quad (5.1)$$

In other words, the wavelet coefficients essentially quantify the strength of contribution of the wavelets at the corresponding locations ( $x_0$ ) and scales ( $s$ ). In our experiments, we have used the complex frequency B-spline [87] wavelet as the mother wavelet  $\psi$ . This is defined as:

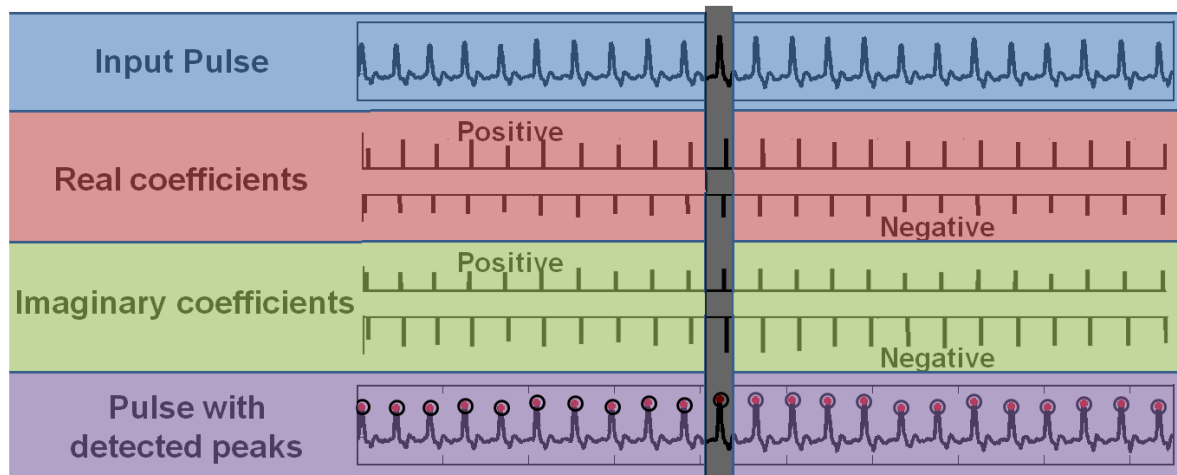
$$\psi(t) = \sqrt{f_b} \left( \text{sinc} \left( \frac{f_b t}{m} \right) \right)^m e^{2\sqrt{-1}\pi f_c t}$$

Here,  $f_b$  is a bandwidth parameter,  $f_c$  is the wavelet central frequency and  $m$  is an integer order parameter. The complex frequency B-spline wavelet has a good temporal localization properties. The values we used are (respectively) 1, 1.5, and 1. We computed the coefficients at scale ( $s$ ) 1.

The advantage of using this mother wavelet over Haar, Morlet or Daubechies wavelets is the observation that, whenever there is a strong singularity in the pulse signal, we get a negative and a positive spike in both the real as well as imaginary parts with high amplitudes. This increases the robustness of our method.

Specifically, we obtain four different coefficient series – real positive, real negative, imaginary positive, and imaginary negative. Next, we compute the thresholded coefficients of these four series, as shown in Figure 5.2. We observe that whenever there is a peak in the pulse signal with high amplitude, we obtain spikes in more than two of the coefficient series. Note that, these spikes have a lag due to the action of the mother wavelet. Finally, we select the common spikes (i.e. peaks) among these four combinations.

The extra efforts of selection from the four coefficient series increases the computational cost. Nevertheless, this also increases the precision in detecting peaks correctly. As the further analysis depends upon the accurate detection of the peaks, we use this method for better performance.



**Figure 5.2:** The top block provides the input pulse signal. Subsequent two blocks are thresholded values corresponding to positive & negative numbers of respectively real and imaginary coefficients. We redraw the input again in the last block but also overlay the detected peaks using our method.

Our methodology has achieved 100% detection accuracy on our database. We have used the peak detection as a preprocessing step in various methodologies including computing the essence of the pulse signal (Section 5.2), computing time domain parameters (Section 5.3), studying the pulse rate variability (Section 5.5), and quantifying self-similar nature (Section 6.1.4).

## 5.2 Extracting Essence of the Pulse signal

The pseudo-periodic pulse signals have a lot of redundancy between periods. The redundancy is due to the same set of points or patterns occurring after each pseudo-period. Therefore, pulse signals can be better represented in terms of a periodic trend plus period-to-period fluctuations.

Conventional wavelet model does not take advantage of the pseudo-periodicity of the signals. Therefore we consider pitch synchronous wavelet transform (PSWT) [48], where the wavelet transform is taken over samples that are *spaced period(s) apart*. The PSWT has been initially used to describe speech and musical sounds, taking advantage of the pitch of the signal [48, 124], and then also to analyze the ECG signals [45, 46, 101],

and in coding scheme [27].

## 5.2.1 Pitch Synchronous Representation

Once all the peaks of a pulse signal are accurately computed, its pitch synchronous (PS) representation is obtained. The PS representation consists of synchronizing the pulse segments which are alike. It has a special meaning only when the signal contains pseudo-periodic segments, otherwise it reduces to the whole signal itself. In our experiments, we have synchronized pulse segments of one period length. In the rest of this subsection, we contrast two different methods in computing the PS representation.

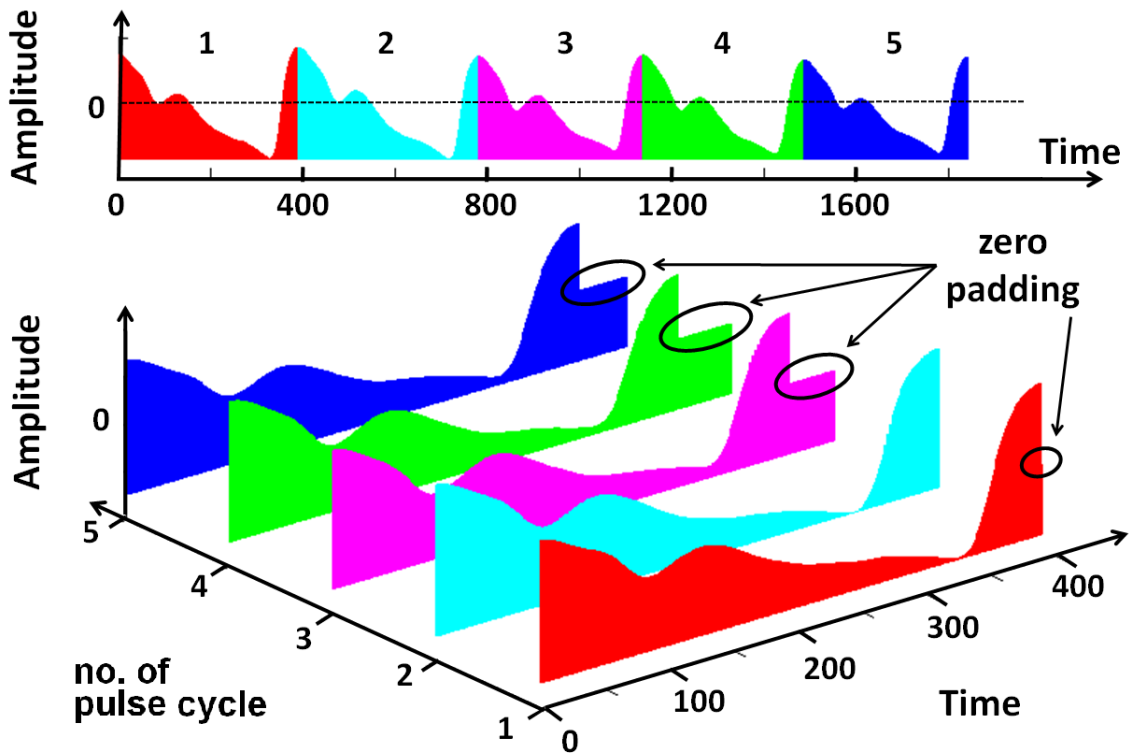
### 5.2.1.1 Conventional PS representation

The conventional way of obtaining the PS representation is to arrange the peak-to-peak segments in a two dimensional matrix termed as the PS matrix. Visually we see segments behind other segments as shown in Figure 5.3.

However, as observed earlier in Figure 5.1, these segments do not have the same time duration. Thus, it is necessary to remove these variations, prior to further processing, by equalizing the segment lengths. Discretized, this results in a PS matrix; the number of rows in the matrix is the number of pulse cycles, and the number of columns is the number of data points in the equalized segments. The difficulty in this conventional method for arterial signals stems from the equalization. Pre or post zero-padding is popularly used to equalize all the segments to the maximum period length [46].

Specifically, the post zero-padding method for a pulse signal  $X$  with  $N + 1$  pulse cycles is as follows. The pseudo-periods  $P(r)$  for  $r = 0, 1, 2, \dots, N$  vary, resulting in different lengths for the row vectors in the PS matrix. Let  $P_{\max}$  be the maximum value of  $P(r)$ . The PS matrix is obtained by inserting  $P_{\max} - P(r)$  zeroes at the end of each row vector. Mathematically the new row vectors are defined as the sequences:

$$\hat{v}_q(r) = X_q(r)v_q(r) \quad (5.2)$$



**Figure 5.3:** The top row provides the input pulse signal with five pulse cycles. The bottom figure shows the conventional PS representation using peak-to-peak segments.

where,

$$\begin{aligned}
 X_q(r) &= 1, \quad q = 0, 1, \dots, P(r) - 1 \\
 &= 0, \quad \text{otherwise}
 \end{aligned}$$

### 5.2.1.2 Our PS Representation

There are two drawbacks with the conventional peak-to-peak PS representation. First, the important P-subwaves get split into two halves at the two extremes, as shown in Figure 5.3; clearly the computed wavelet transform in Equation 5.3 results in vastly different approximation coefficients. Second, if the pulse signal has high variability amongst different cycles, then the resulting PSWT will get distorted at the end with the zero-padded region.

To overcome these drawbacks, we place the peak of each P-subwave in the middle and consider sections of the pulse cycle on two sides of the peak as shown in the bottom of Figure 5.4. If  $P_{\max}$  is the maximum peak-to-peak distance considering all cycles, then we use  $\frac{1}{3}P_{\max}$  of the cycle on one side of each peak, and the remaining  $\frac{2}{3}P_{\max}$  of the cycle on the other side of each peak for constructing the PS matrix. While other proportions of splitting can be chosen, our specific proportion has been chosen with the following in mind

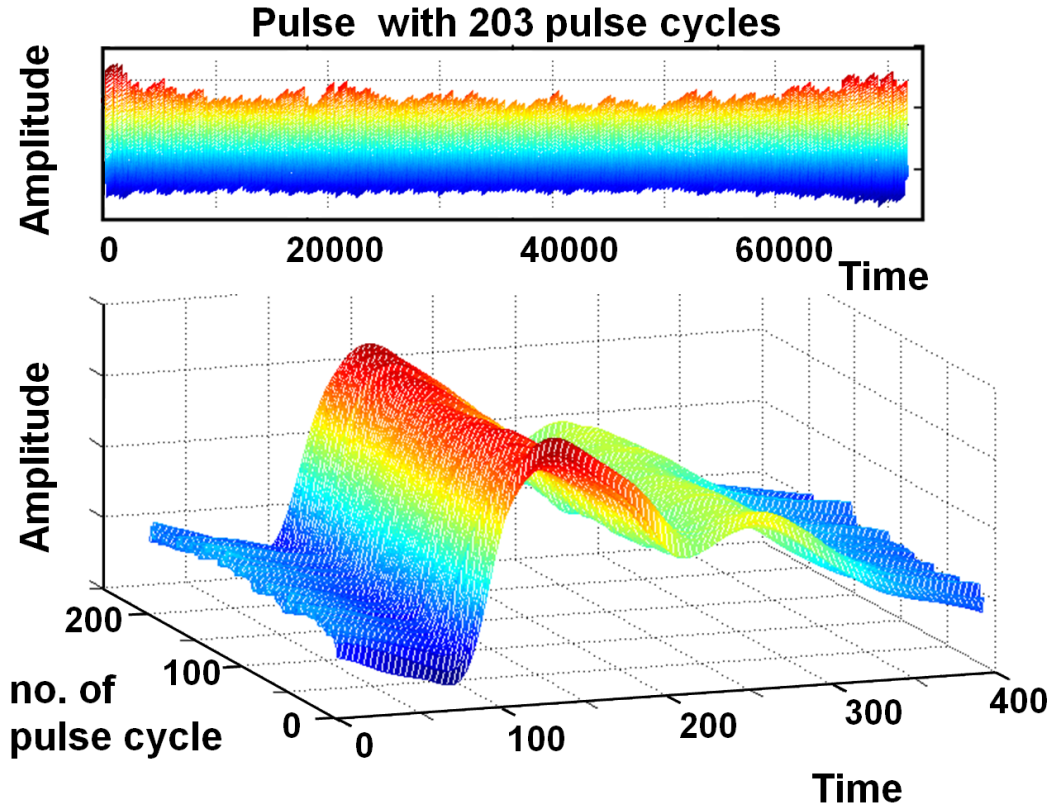
- All the P-T-V-D-subwaves are included in the resulting PS segments
- The splitting ensures the fidelity of the smoothness of the important P-subwave
- The splitting does not occur at the positive or negative peaks of the pulse cycles, where the change in slopes is the maximum. Because if there is change in slopes at locations  $k$  of the PS representation, then the wavelet transform in Equation 5.3 will result in incorrect approximation coefficients.

The negative effect due to uneven number of zero-padding is shown in Figure 5.5(b). As will become clear in the next subsection, the DWT is computed next. Due to the haphazard zig-zag nature of zero padding, null values will be combined with valid values resulting in a poor trend (PSWT) signal. The advantage of our method is in *period normalization*. Instead of using zero-padding of the conventional approach [26], we make use of the whole pulse signal.

## 5.2.2 Computation of the PSWT

After structuring the pulse signal in the PS matrix, the DWT is applied to each of the column vectors of the PS matrix





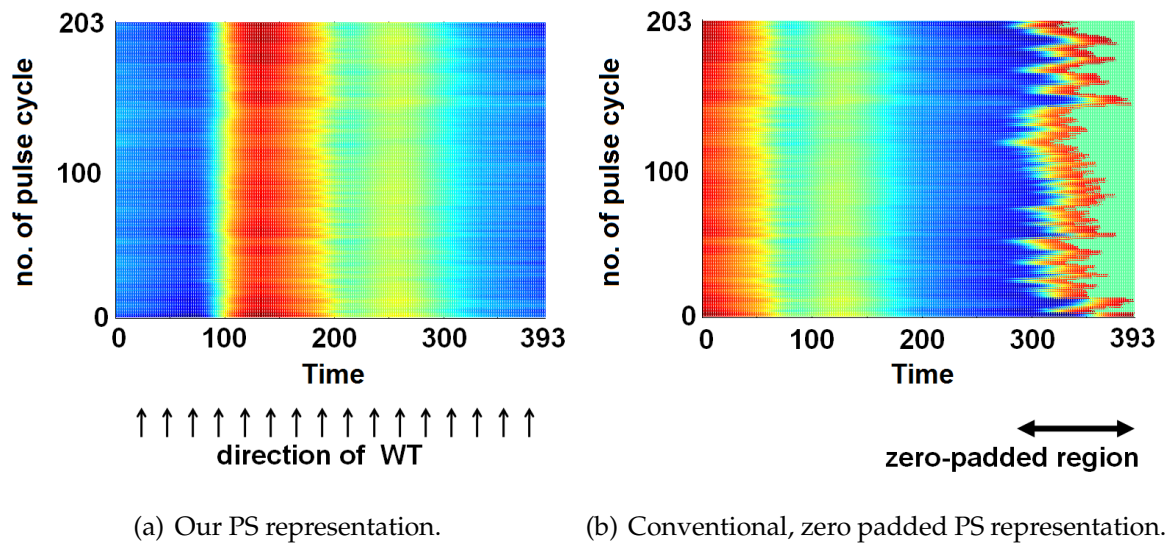
**Figure 5.4:** The top row provides an input pulse signal with 203 pulse cycles. After peaks detection, the bottom subfigure shows our PS representation. Unlike Figure 5.3, this method captures the nature of the systole-diastole process.

$$\hat{V}_{j,k,q} = \sum_r \hat{v}_q(r) \psi_{j,k}(r) \quad (5.3)$$

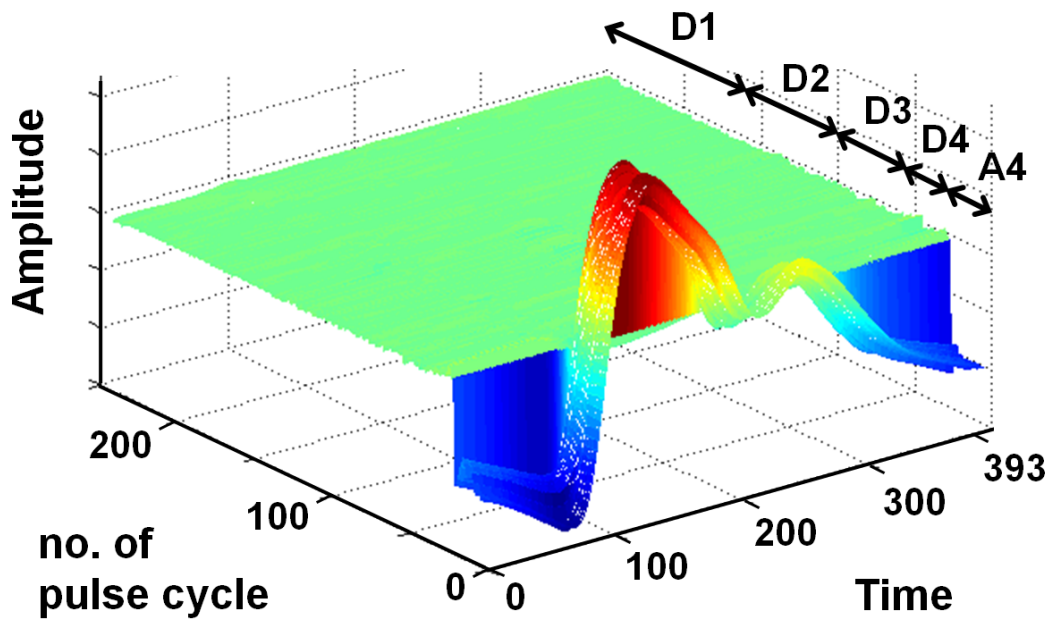
Signal reconstruction is possible using

$$\hat{v}_q(r) = \sum_{j,k} \hat{V}_{j,k,q} \psi_{j,k}(r) \quad (5.4)$$

Figure 5.6 shows the PSWT coefficients resulting from Equation 5.3. (In our experiments, we have used Matlab function 'wavedec' with 'db6' as the mother wavelet. As our database contains pulse signals with no more than 250 cycles, the DWT was computed up to 4 scales). In the figure, the detail coefficients (D1–D4) capture the peak-to-peak variations at the four levels. The detail surfaces are flat and contain very small amount of energy.

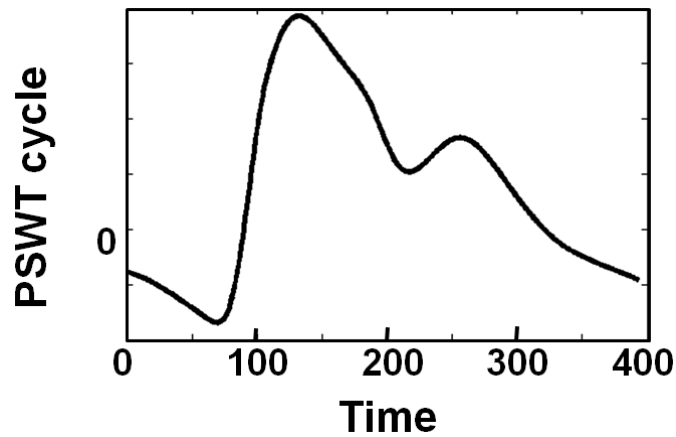


**Figure 5.5:** Comparison of the PS representations. We show top views of the PS matrix, and colors represent the amplitudes (green has value zero). Notice the predominant green in the zero-padded region. (Figure best viewed in color.)



**Figure 5.6:** The PSWT coefficients of the pulse signal from Figure 5.4. The A4, D4, D3, D2, D1 are the approximation and detail wavelet coefficients at various scales.

On the other hand, due to the pseudo-periodicity of the pulse signal, most of the energy appears in the residue or the approximate coefficients (A4). A4 contains the underlying pulse shape (essence) of the whole pulse signal. This is evident in Figure 5.7, where only second A4 coefficient vector is displayed.



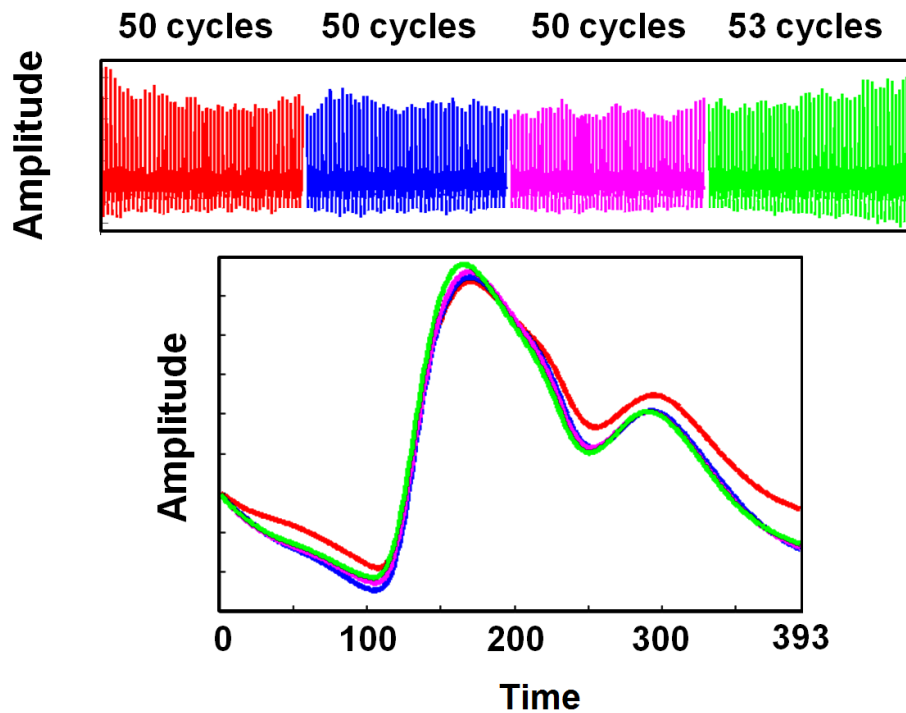
**Figure 5.7:** *The PSWT cycle from the second row of the  $A_4$  matrix (from Fig. 5.6). We observe that the signal is smooth, and representative of the entire signal.*

### The PSWT Intuition

How well does the trend (PSWT) cycle relate to the pseudo-periodic signal? For validating the computation of the PSWT, we split the pulse signal of Figure 5.4 into 4 parts of 50 cycles each. Then using the above steps, we computed the trend (PSWT) cycles of each part. Figure 5.8 displays the superimposition of the four trend cycles.

Though the four parts contain different pulse cycles of different amplitudes and variabilities, the four trend (PSWT) cycles are almost the same. It shows that each pulse cycle contains the basic structure of the whole pulse signal and is reflected in the trend (PSWT) cycle.

What happens to the trend (PSWT) cycle if the individual cycles of the pseudo-periodic signal do not indicate a predominant P-T-V-D pattern? We see, separately, three cycles of such a signal in Figure 5.9, where the individual pulse cycles show highly unpredictable and vibrating patterns. However, their trend (PSWT) cycles are, once again, smooth, and indicative of the overall behavior. Unlike the individual cycles in the middle row, it is considerably easier to compute the time domain parameters of the smoothed cycles as we see in the next section.



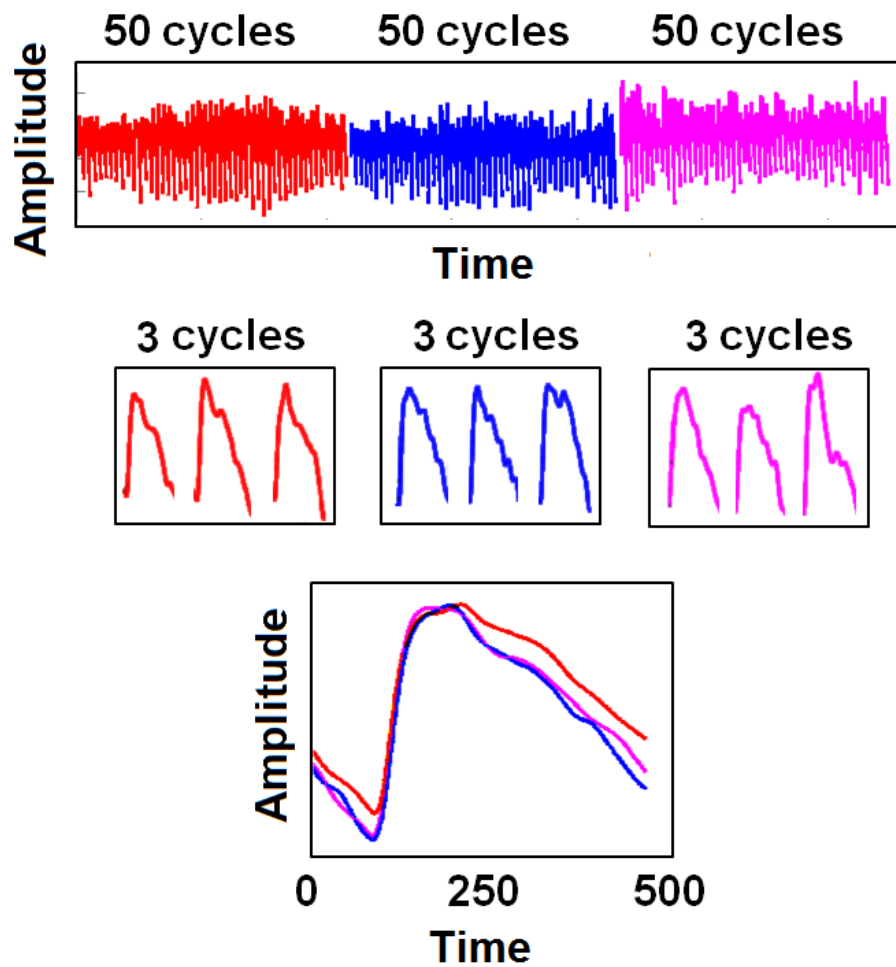
**Figure 5.8:** *The signal of Figure 5.4 is split into four parts and the resulting four trend (PSWT) cycles are superimposed. The four trend cycles appear very similar.*

### 5.3 Time Domain Parameters

The time domain parameters (TDP) [257, 272] of a pulse signal are the amplitudes and widths at certain important locations. They are marked in Figure 5.10.

- **Y-axis related parameters:** These parameters are amplitude parameters denoted by  $h_1$  (amplitude at maxima of P),  $h_2$  and  $h_3$  (amplitudes defined by T),  $h_4$  (amplitude at minima of V), and  $h_5$  (amplitude at maxima of D).
- **X-axis related parameters:** These parameters are width parameters denoted by  $t_2$  (systole width),  $t_3$  (diastole width),  $t_1$  (defined by the maxima of the pseudo-periodic cycle), and  $w$  (width of the pulse cycle at 80% height at the P-subwave (see Figure 5.10)). In addition, the pseudo-period,  $WP$  (peak-to-peak distance shown in Figure 5.1) is another parameter.

Important locations, on the X-axis, used to compute these parameters are called **TDP locations** in this thesis. They are also marked in the figure.

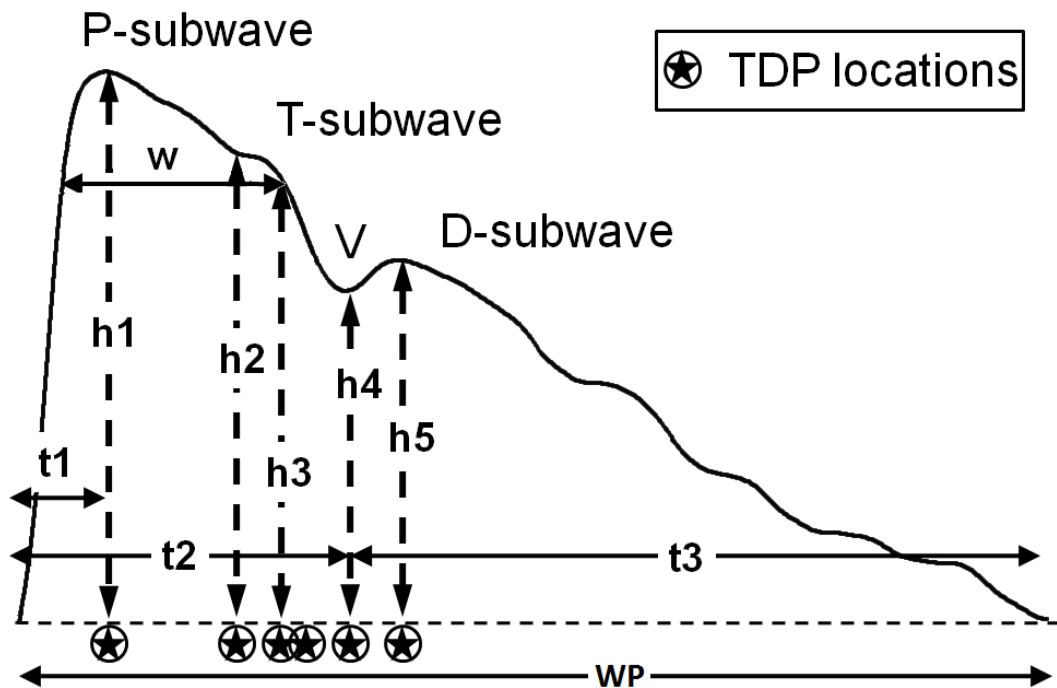


**Figure 5.9:** *An ill-behaved pulse signal is split into 3 parts. The second row displays three sets of three randomly selected pulse cycles. Finally, their trend (PSWT) cycles are superimposed. Though the individual cycles are ill-behaved, their trend cycles appear similar.*

The TDP on a pulse signal have been testified to be important for diagnosis [153, 272]. For example, the parameter  $h1$  relates to the peak amplitude of the P-subwave, reflecting the ejection function of the left ventricle and the resilience of the main artery. The parameter  $t1$  relates to the left ventricular ejection time [272].

### 5.3.1 Challenge in Computation

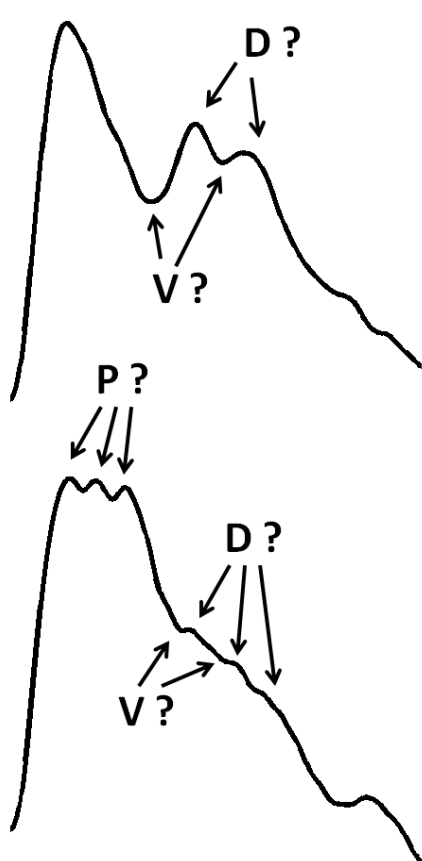
Note that the peak-to-peak distances (696–678–700–718–724 milliseconds) in Figure 5.1 vary due to the pseudo-periodicity of the pulse signal. Therefore, TDP also vary across different pulse cycles in the signal.



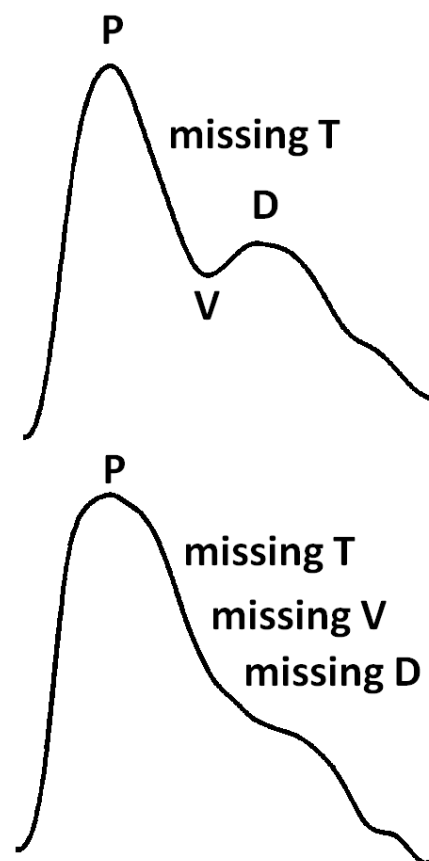
**Figure 5.10:** Time domain parameters of a cycle in the pseudo-periodic pulse signal.  $w$  refers to the width of the P-subwave at 80% of its height). Significant locations termed TDP locations are marked on the X-axis.

To extract the TDP of a pulse signal automatically, one idea is to start with the detection of the TDP locations. However, often the pulse signals are vibrating in nature. Worse, in a few cases, one or two subwaves may not be predominantly manifested. Such an ill-behavior is commonly observed in the pulse signals acquired from elderly subjects, or subjects with a disorder. Figure 5.11 shows nominal cases.

We emphasize that for generic pseudo-periodic signals, there could be different sets of axes related parameters for a signal depending upon the property under study. For example, two different sets of parameters [9] for an electrocardiogram (ECG) could be (i) Locations and amplitudes of P-wave, QRS-complex and T-wave (for purposes of segmentation) and (ii) Widths of P-wave, QRS-complex (for studying blood circulation).



(a) The pulse cycle is of the vibrating nature, and therefore multiple candidates are available for the TDP locations.



(b) The pulse cycle is atypical; TDP locations have to be qualitatively inferred.

**Figure 5.11:** *The computation of the TDP locations becomes non-trivial for ill-behaved pulse cycles. Manual efforts may be needed.*

### 5.3.2 Related Work

The TDP were obtained by a manual labeling process because of these difficulties. This becomes a huge obstacle in realizing automated pulse analysis. Recently, frameworks have been constructed [239, 255, 272] to compute the TDP.

However, along with the pseudo-periodicity (non-stationary characteristics and inconsistency in the cycles), the pulse signals also show the baseline wander [263] across the cycles due to the respiration and other physiological activities.

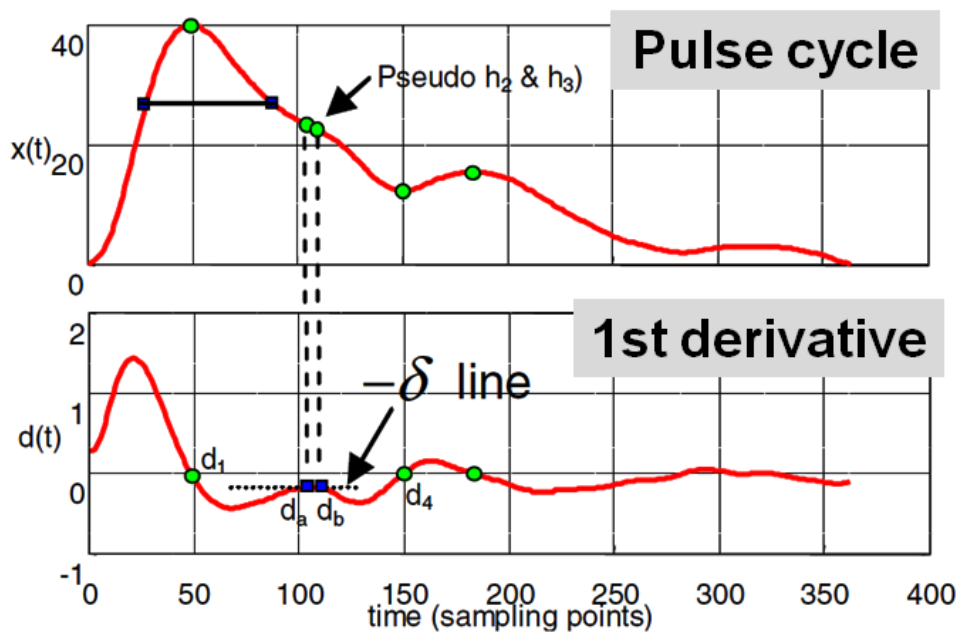
Therefore, though the methods in [239, 255, 272] can be claimed to be satisfactory on

a specific data set, it is difficult to use the same method over a long pulse signal. A primary concern is that, if a pulse signal (and thus its pulse cycles) shows ill-behavior exemplified in Section 5.3.1, extra efforts in tuning the algorithm is involved.

### 5.3.3 Prior Zero-crossing Approach for TDP Computation

One of the methods to compute the TDP in [255] is easy to understand and implement. In this method, the TDP locations (refer Figure 5.10) are obtained using the first derivative of the pulse cycle, and looking for the zero-crossings.

Consider, however, the example in Figure 5.12 where the pulse cycle and its first derivative are shown. It can be seen that due to the nature of the T-subwave, the zero crossing is not obtained. The algorithm therefore tries to search, or guess, for values by performing an automated moving  $\delta$  line crossing point search to obtain what may be called pseudo-locations of  $h_2$  and  $h_3$  values.



**Figure 5.12:** Figure from [255] indicating missing zero-crossings. A search is made between the first and second zero crossing looking for values close to zero to obtain pseudo  $h_2$  and  $h_3$  parameters.

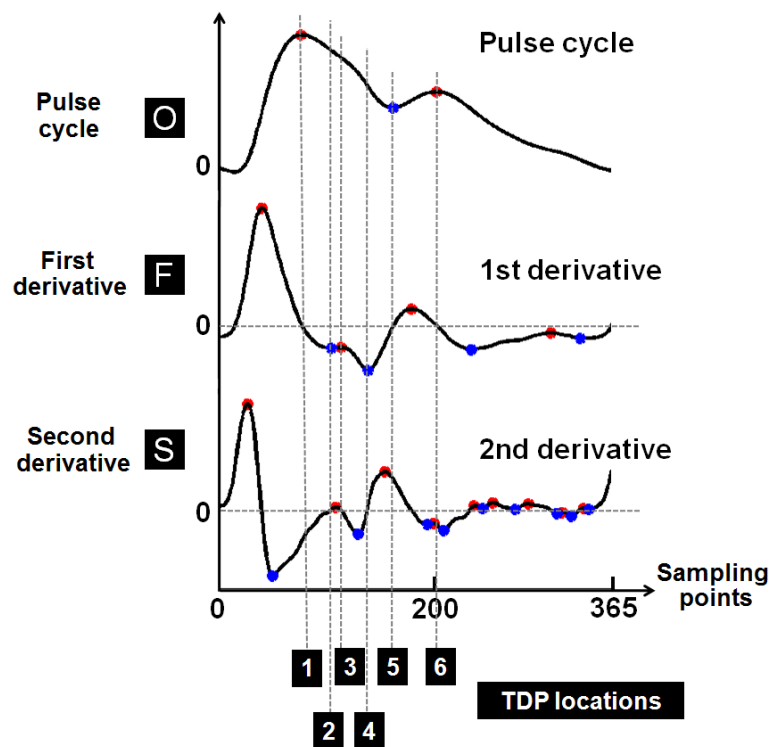
This situation of missing zero-crossings shown in Figure 5.12 was commonly observed (refer Table 5.1) for a variety of pulse signals whether ill-behaved, or well-behaved. We



develop an alternative method that, while not perfect, non-parametrically identifies TDP locations in a larger number of cases as seen in Table 5.1.

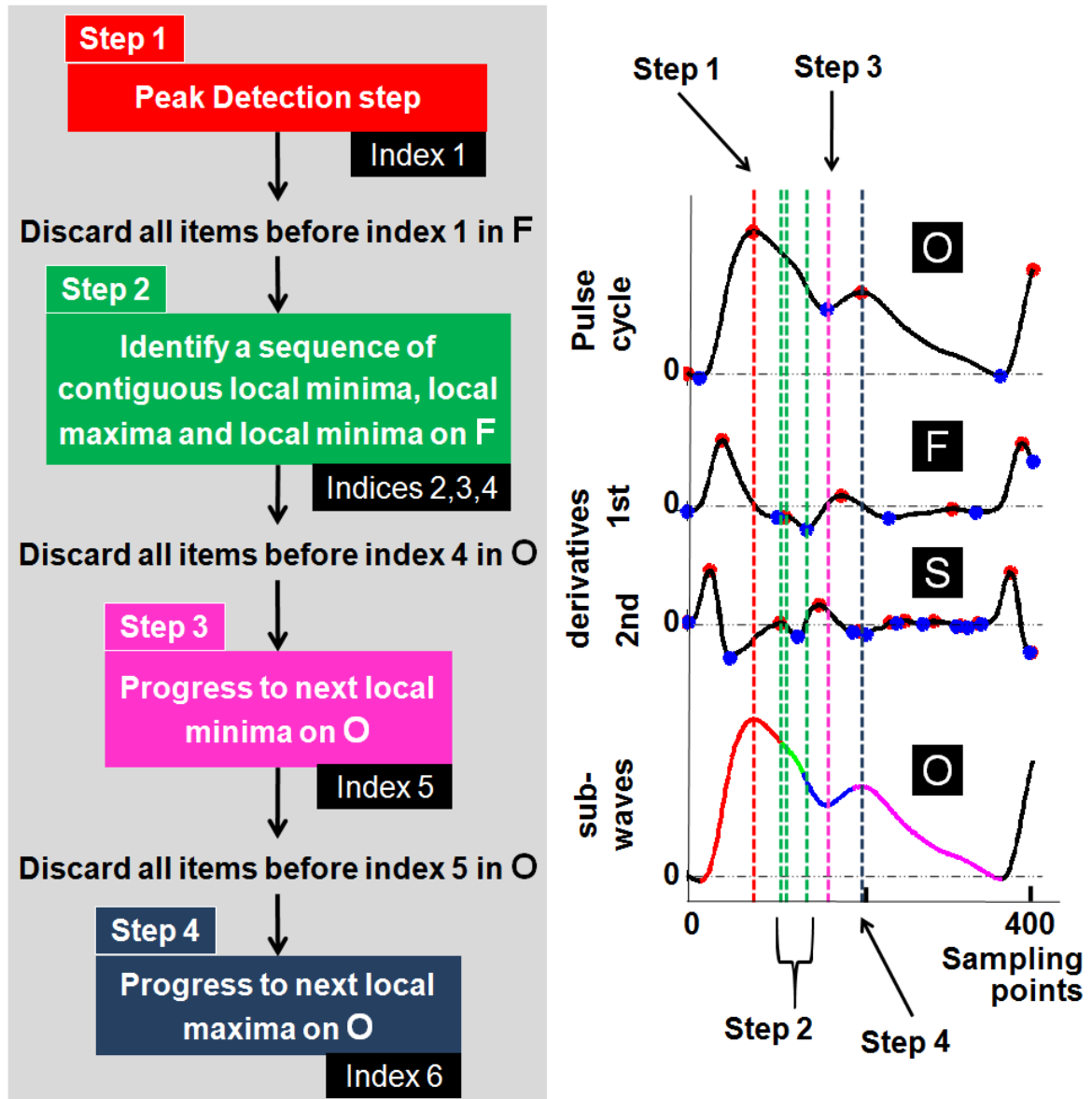
### 5.3.4 Our Maxima-Minima Approach

In our approach we consider three sequences corresponding to the individual pulse cycle  $O$ , and its first derivative  $F$  and second derivative  $S$ . Next, we identify the maxima and minima locations in all three sequences. Following Figure 5.13, a subset (locations 1 through 6) of these maxima-minima locations correspond to the desired TDP locations. These locations are computationally identified using the flowchart shown in Figure 5.14 where it is helpful to think of the sequences stored as arrays. We briefly describe these steps.



**Figure 5.13:** Maxima and minima information of a pulse cycle  $O$  and its first and second derivatives are pruned to reveal TDP locations.

- **Step 1:** Computation of the TDP location corresponding to the P-subwave, denoted by index 1 in Figure 5.13: This is readily achieved in the peak detection step (refer Section 5.1.1), and  $h_1$  is computed at this location.



**Figure 5.14:** Our methodology for identifying TDP locations. Assessment is made in tandem on  $O$ ,  $F$  and  $S$  as described in the text.

- **Step 2:** Computation of the TDP locations corresponding to the T-subwave, denoted by indices 2,3,4 in Figure 5.13: We discard items in the 1st derivative array  $F$  if they appear before the index 1. We look for a contiguous combination of a local minima, a local maxima and a local minima. If successful,  $h_2$  and  $h_3$  are computed at the locations denoted by indices 2 and 3 respectively.
- **Step 3:** Computation of the TDP location corresponding to the V-subwave, denoted by index 5 in Figure 5.13: Using the information computed in the

previous step, we discard items on  $O$  if they appear before index 4. By progressively looking for the next local minima, we succeed in identifying  $h_4$ .

- **Step 4:** Computation of the TDP location corresponding to the D-subwave, denoted by index 6 in Figure 5.13: We discard items in  $O$  if they appear before index 5. By progressively looking for the next local maxima, we identify  $h_5$ .

We have already indicated the potential failure in Step 2. Steps 3 and 4 can also fail if the input pulse cycle is ill-behaved. We consider these situations below.

### 5.3.5 Our Approach for Ill-behaved Pulse Cycle

Ill-behaved pulse cycles may produce insufficient, or too many maxima-minimas thereby confounding any attempt to define the TDP. Multiple (Figure 5.15(a)) local maxima and minima locations (vibrating nature of the pulse) in the range confuses the decision for the respective subwave. A similar situation for missing local maxima or minima locations is shown in Figure 5.15(b). A blind application of the steps mentioned earlier may result in out of order classification of the subwaves. For example a V-subwave may be identified as occurring prior to the T-subwave (this happens because of the negligible strength of the T-subwave).

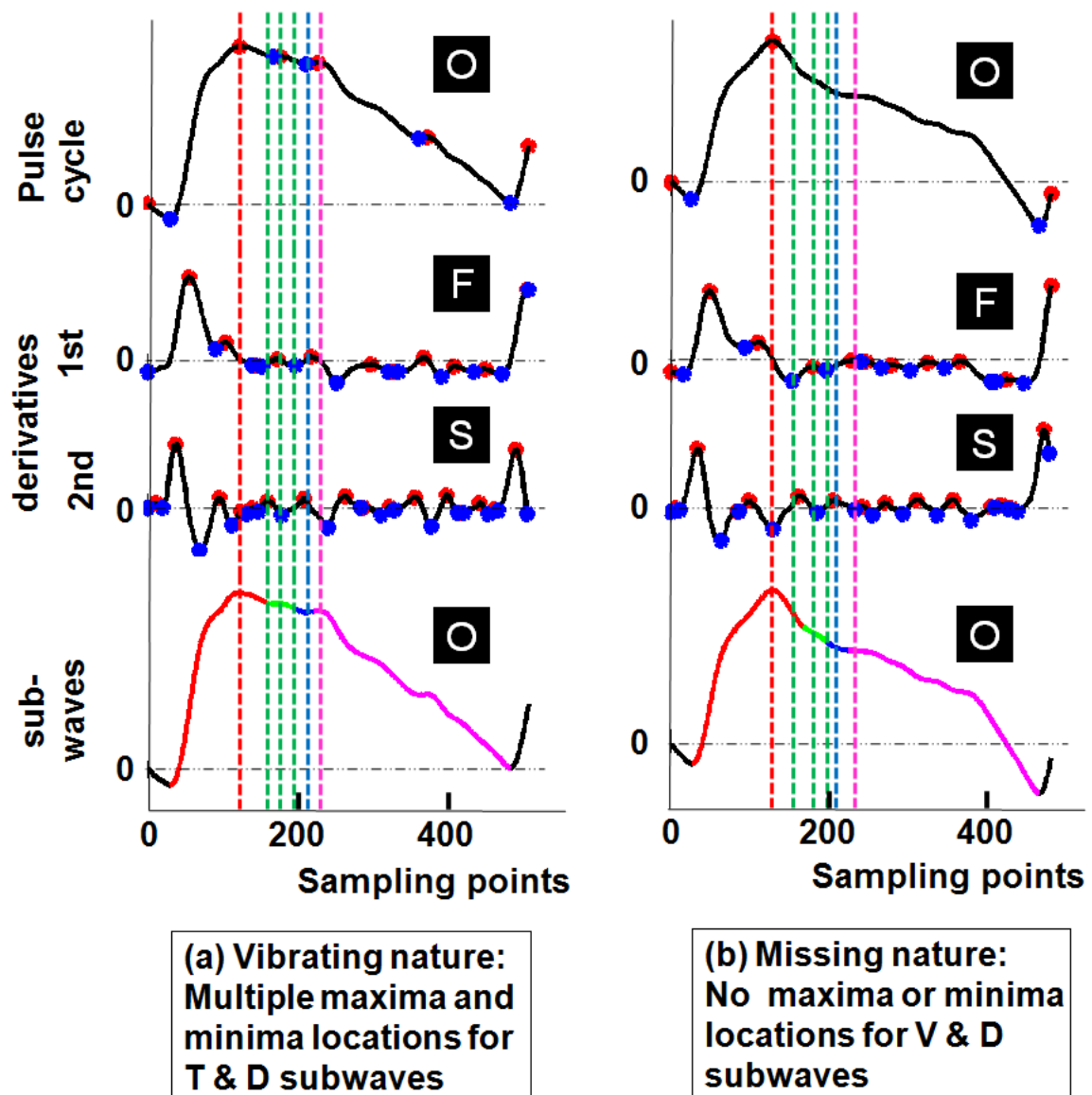
To reduce incorrect values, we make use of the observation that the pulse signal follows all the actions in a cardiac cycle (explained in Section 2.1). We postulate the following:

- **Postulate 1:** Subwaves P-T-V-D come in the sequential fashion.
- **Postulate 2:** Each subwave has a particular duration in the complete pulse cycle. The P-subwave cannot exceed 40% of the pseudo-period (WP). The T-subwave must be within the range of 25% to 45% of WP. Similar ranges of V-subwave and D-subwave are 35% to 50% and 40% to 55% respectively.

## 5.4 Novel PSWT-DTW Approach

### Need of the PSWT-DTW Approach

Table 5.1 shows the comparison of the above two approaches for the pulse cycles of both healthy as well as disordered labels. For each pulse cycle (total 33767) in our database, we note the number of times the zero-crossings were detected, or missed and then probed using the moving  $\delta$  line (refer Section 5.3.3). We also indicate the



**Figure 5.15:** *Pseudo-locations may need to be identified for the ill-behaved pulse cycles depicted here.*

number of times the local maxima-minima locations were detected, or missed and then probed (refer `subsub:ill-behaved-pulse-cycle`). As stated, the two approaches put an extra effort in computing the pseudo-locations in the ‘missed’ situations. Note also from the table that such ill-behavior present in pulse cycles (refer Figure 5.11) manifests in both subjects labeled as healthy, and subjects labeled as disordered.

		Healthy	Disordered	Total
zero-crossing [255]	maxima-minima	(10942)	(22825)	(33767)
missed and probed	missed and probed	7881	17387	25268
missed and probed	detected	3057	5386	8443
detected	missed and probed	4	52	56
detected	detected	0	0	0

**Table 5.1:** *Comparison of the success or failure of two approaches on individual pulse cycles. When the zero-crossings are missed, pseudo-location need to be identified. Our maxima-minima approach reduces manual corrections.*

We observe that the zero-crossing are ‘missed and probed’ in almost all the pulse cycles (especially at the T-subwave). Our maxima-minima based approach was successful to reduce the efforts of computing pseudo-locations for 8443 cycles of them. However, there is still need of a robust method to remove the efforts in computing pseudo-parameters of total 25268 cycles.

The explanation for the 52 cases for which the zero-crossing approach succeeds, but our approach fails is as follows. In Figure 5.13, we show that our approach detects the T-subwave, even though there is no zero-crossing. In the 52 cases, consider very same situation along with a bumpy behavior of the T-subwave. Now, in this case, the zero-crossing method will detect the T-subwave without the need of delta-parameter, while our approach will get confused between the two cases (one as shown in Figure 5.13, and the one due to bumpy behavior).

Our novel PSWT-DTW approach takes care of all ill-behaved situations, by first computing the essence of the whole signal (Section 5.2), then computing the trend TDP locations on this essence, and finally transferring them on each of the pulse cycles using

DTW.

### **5.4.1 Proposed Method Overview**

Our novel PSWT-DTW approach takes care of such situations. The key idea is that even though each pulse cycle could be different from others, the signal has an overall trend structure (refer Figures 5.8 and 5.9). This trend is captured in the PSWT computation and can be exploited. The trend is usually smooth; therefore its TDP can be obtained with no false negatives. Further, once the trend is obtained, the trend may be “applied” on a variety of pulse cycles – well-behaved, or ill-behaved – from the same input, to get individual TDP. Figure 5.16 shows the schematic of our proposed system.

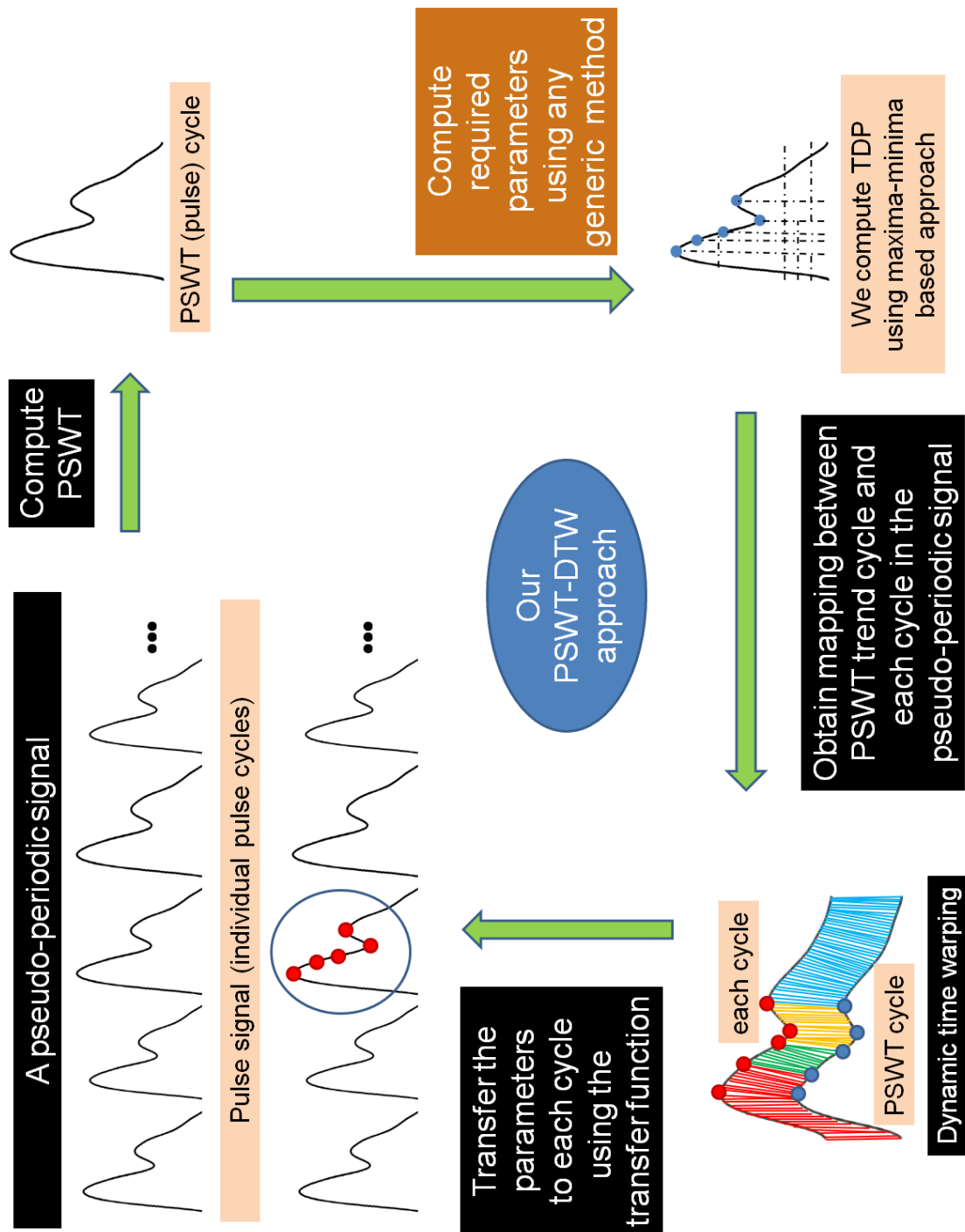
### **5.4.2 Dynamic Time Warping**

For mapping the TDP on PSWT cycle to each of the (subwaves of) pulse cycles in Figure 5.16, a greedy search algorithm can be applied. However in reality, due to nonstationarity and variability of the pulse signal, we need a method that allows an elastic shifting of the time axis. Therefore we have used the dynamic time warping (DTW) approach [96, 103, 176], as it allows a nonlinear mapping of one signal to another, even if they are out of phase in the time axis, by minimizing the distance between the two. It is a very robust distance measure for time series, allowing similar shapes to match.

Because of this flexibility, DTW was introduced by the speech recognition community [80] and has spread to other disciplines of application including biomedical data [223], signature analysis [142], and bioinformatics [1].

#### **5.4.2.1 Computation of Warping Path**

The fluctuation in time is modeled approximately by a nonlinear warping function with some carefully specified properties. Timing difference between two pulse



**Figure 5.16:** *The novel PSWT-DTW approach. Our approach can be used for analyzing any pseudo-periodic signals for computing domain-specific parameters; the only change will involve changing the method used in the lightly shaded brown block on the right. (Figure best viewed in color.)*

subwaves is minimized by warping the time axis of one frame such that the maximum coincidence is attained with the other. Then the time-normalization is calculated as the individual minimized residual distance between them. This minimization process is carried out efficiently by the DTW technique as follows [95, 180].

Consider two multivariate time series (e.g. pulse cycles)  $Q$  and  $C$ , of length  $n$  and  $m$  respectively,

$$Q = q_1, q_2, \dots, q_i, \dots, q_n \quad (5.5)$$

$$C = c_1, c_2, \dots, c_j, \dots, c_m \quad (5.6)$$

where,  $q_i$  and  $c_j$  are both vectors such that,  $q_i, c_j \in \mathfrak{R}^p$  with  $p \geq 1$ . To align these two sequences using DTW, a  $n$ -by- $m$  matrix is constructed where the  $(i^{th}, j^{th})$  element of the matrix corresponds to the squared distance

$$d(i, j) = \sum_{r=1}^{r=p} (q_{i,r} - c_{j,r})^2 \quad (5.7)$$

To find the best match between these two sequences, one finds a path through the matrix that minimizes the total cumulative distance between them. Such a path is called a *warping path* ( $W$ ). It is a contiguous set of matrix elements that characterizes a mapping between  $Q$  and  $C$ . The  $k^{th}$  element of  $W$  is defined as  $W(k) = (i, j)_k$ . The time-normalized distance [103] between the two time series is defined over the path as,

$$DTW(Q, C) = \min_W \left( \sqrt{\frac{\sum_{k=1}^{k=K} d(W(k)) \cdot \phi(k)}{\sum_{k=1}^{k=K} \phi(k)}} \right) \quad (5.8)$$

where,  $\phi(k)$  is the non-negative weighting coefficient and  $K$  is the length of the warping path  $W$ , which satisfies the condition,

$$\max(m, n) \leq K \leq m + n - 1 \quad (5.9)$$



The normalization is done to compensate for  $K$ , which can be different for different cases; i.e. number of steps will be very few if too many diagonal steps are followed and hence the objective function in Equation 5.8 will prefer such paths. Using path normalization, the path selection procedure is made independent of number of steps taken. The symmetric normalization [103] has been used for this purpose.

$$\phi(k) = (i(k) \cdot i(k-1)) + (j(k) \cdot j(k-1)) \quad (5.10)$$

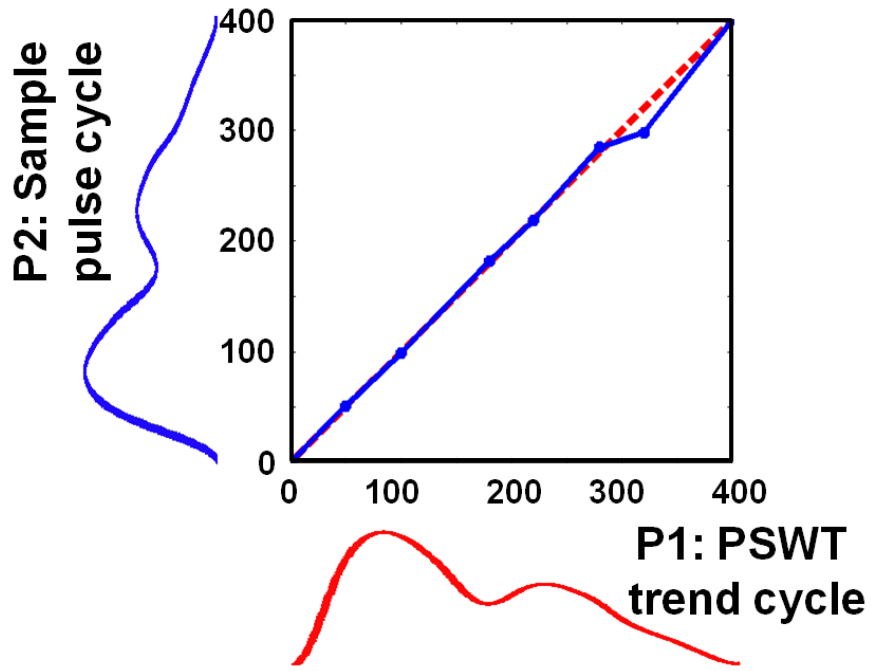
Note that for a horizontal or a vertical step, the value of  $\phi(k)$  is 1, but for the diagonal step, it is 2. The path, however, is constrained to the following conditions [103]:

- **Boundary conditions:** The path must start at  $W(1) = (1, 1)$  and end at  $W(K) = (n, m)$ .
- **Continuity and monotonic condition:** Every point in the two time-series must be used in the warping path, and both  $i$  and  $j$  indexes can only increase by 0 or 1 on each step along the path. In other words, if we take a point  $(i, j)$  from the matrix, the next point must be either of  $(i, j + 1)$ ,  $(i + 1, j + 1)$  and  $(i + 1, j)$ .
- **Warping Window condition:** The path as we know intuitively, should not wander too far from the diagonal.

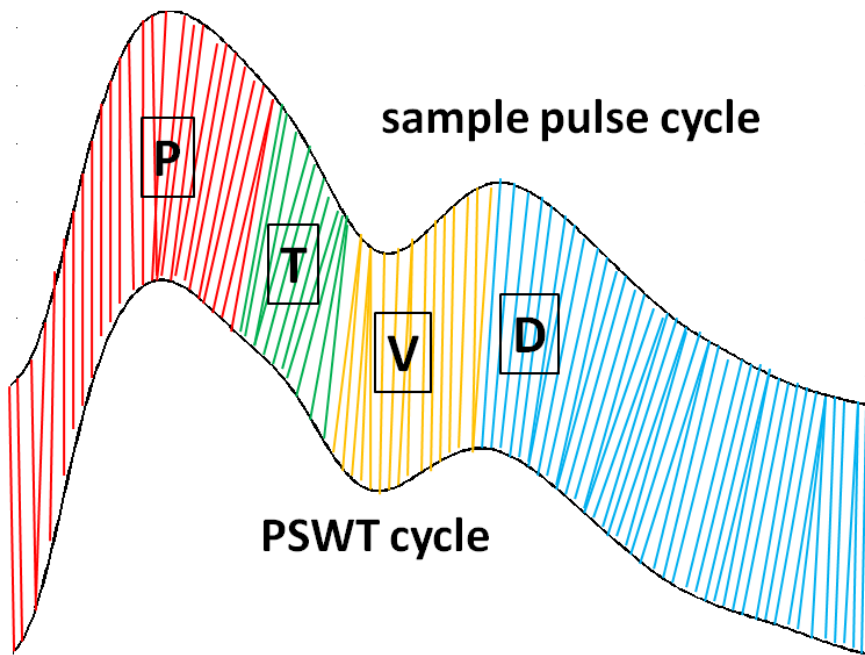
The warping path for maximum coincidence between the PSWT cycle (P1) and the pulse cycles (P2) under consideration is shown in Figure 5.17. The two pulse cycles P1 and P2 get reorganized for the optimum matching. We utilized the DTW concept for the purpose of online pattern matching. The result of the algorithm is illustrated in Figure 5.18 for warping the PSWT cycle (P1) to the sample pulse cycle (P2). Different colors are used to indicate the corresponding locations of the subwaves.

### 5.4.3 Observations

Our novel PSWT-DTW approach is summarized in Figure 5.19. We now present the results of our approach. First, we show that the PSWT cycle captures the trend. Then



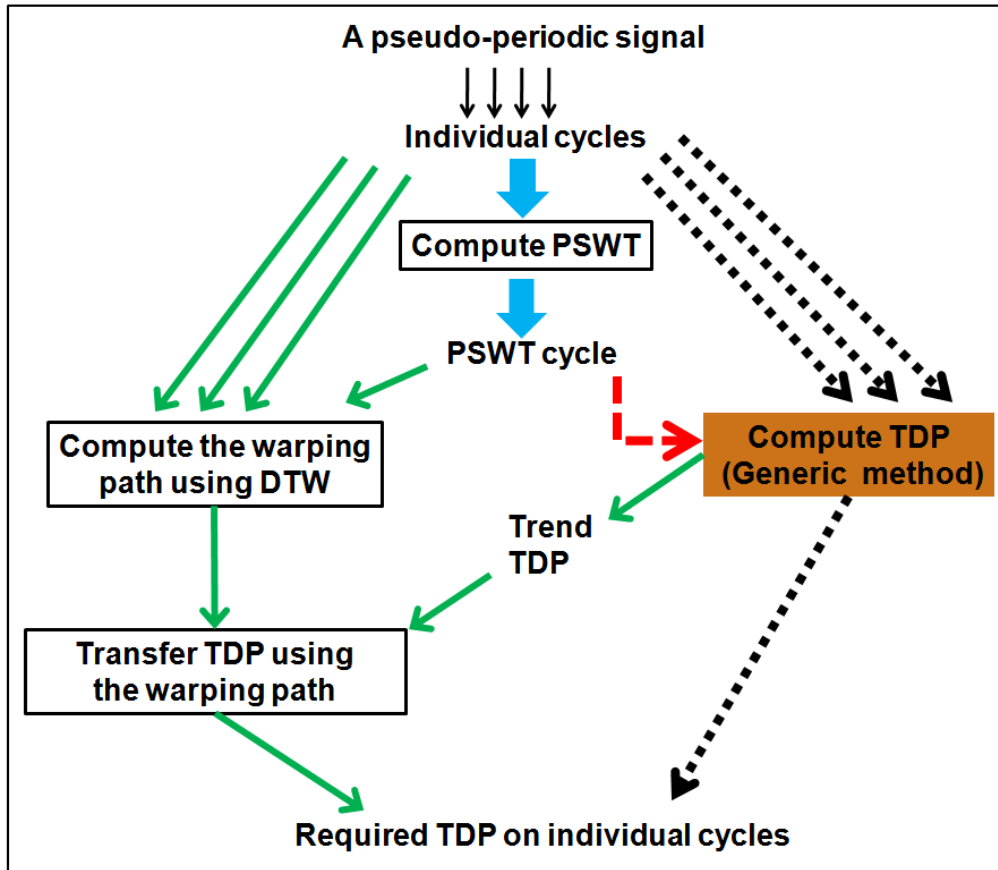
**Figure 5.17:** The PSWT cycle (P1) is on x-axis, while a sample pulse cycle (P2) is on y-axis. The warping path indicates the maximum coincidence between the two.



**Figure 5.18:** Cycle on top is one of the cycles of pulse signal, while PSWT cycle is at bottom. Using DTW, we achieve nonlinear elastic matching between the two cycles.

we validate our novel PSWT-DTW approach both in a summarized quantitative form, and in a manual ground truth-based qualitative form.

The classification results, for detecting 10 different behaviors of the pulse signals, are presented later in Section 7.3. We use two feature sets as the input vectors to the classifier, namely (a) TDP individually computed on the pulse cycles using our maxima-minima based approach, and (b) TDP computed using our PSWT-DTW approach. We show that the classification performance is improved with the second feature set.

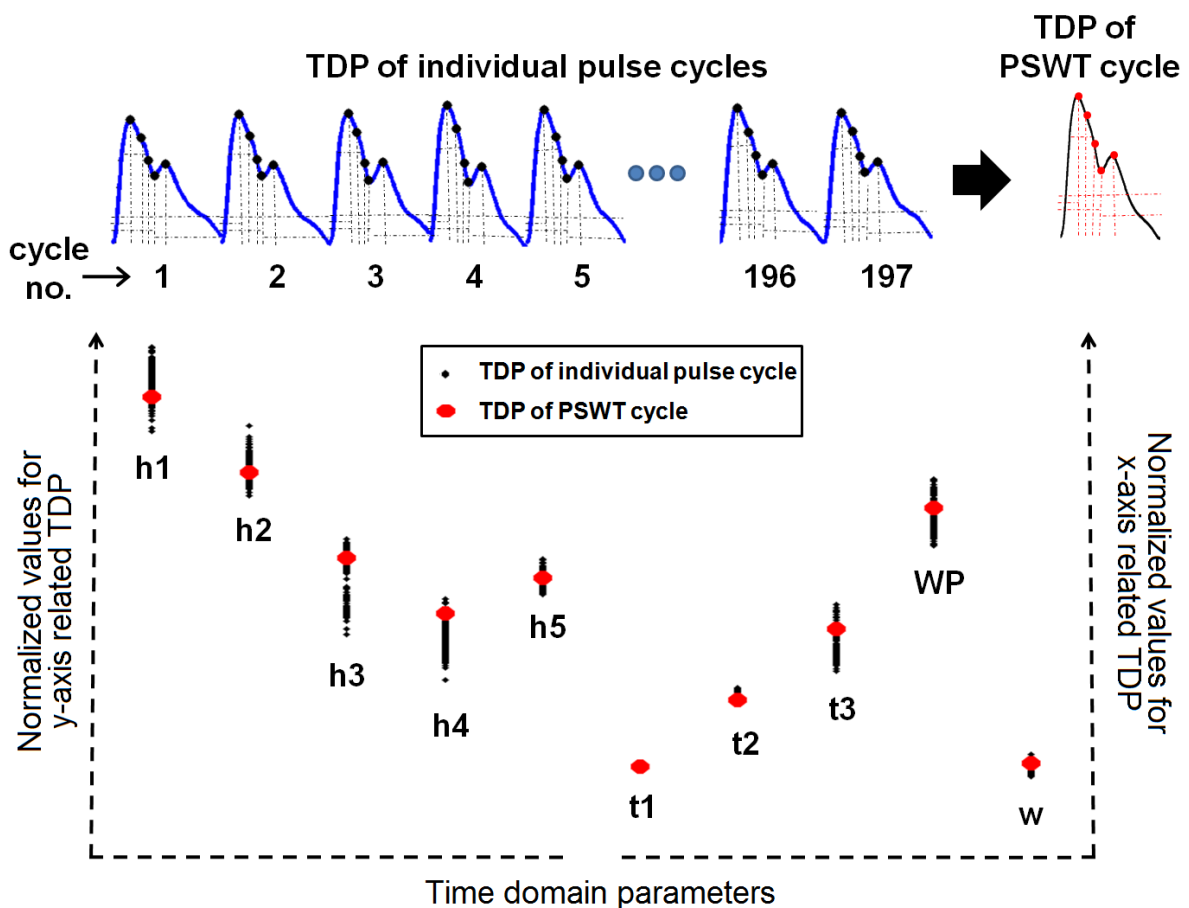


**Figure 5.19:** Method overview. We wish to extract the TDP of individual cycles in a pseudo-periodic signal. The dotted, black path displays the conventional approach which employs a generic method. On the other hand, we use the PSWT (bold blue arrow) to capture the trend values (dashed, red arrow). Now this trend information is further used along with the dynamic time warping (green arrows), to obtain the TDP of all cycles.

### 5.4.3.1 Capturing the Trend TDP on the PSWT Cycle

To show that the PSWT cycle captures the overall trend values, we take an example of a long pulse signal with 197 pulse cycles. First, we compute the TDP on the individual 197 pulse cycles using our maxima-minima based (Section 5.3.4) methodology. Then, we obtain the PSWT cycle of the signal (Section 5.2.2) and compute the TDP on this cycle using our maxima-minima based methodology.

The statistics of the TDP of the signal are shown with black values (197 values for each parameter) in Figure 5.20. The TDP of the PSWT cycle are shown in red circles. It is clear that the trend TDP is a good approximation to any individual parameter. We summarize these values in Table 5.2.



**Figure 5.20:** TDP of a pulse signal with 197 pulse cycles. The black points indicate the TDP computed from individual pulse cycles. The TDP of the PSWT cycle is indicated by red circles. (y-axis values are normalized.)

TDP	Individual 197 pulse cycles (mean $\pm$ std.dev.)	PSWT cycle
h1	0.10 $\pm$ 0.0037	0.0974
h2	0.08 $\pm$ 0.0029	0.08
h3	0.058 $\pm$ 0.0054	0.06
h4	0.044 $\pm$ 0.0036	0.048
h5	0.056 $\pm$ 0.0017	0.056
t1	63.48 $\pm$ 1.48	62
t2	144.89 $\pm$ 4.36	139
t3	204.62 $\pm$ 15.63	220
WP	349.51 $\pm$ 15.89	359
w	61.27 $\pm$ 4.72	66

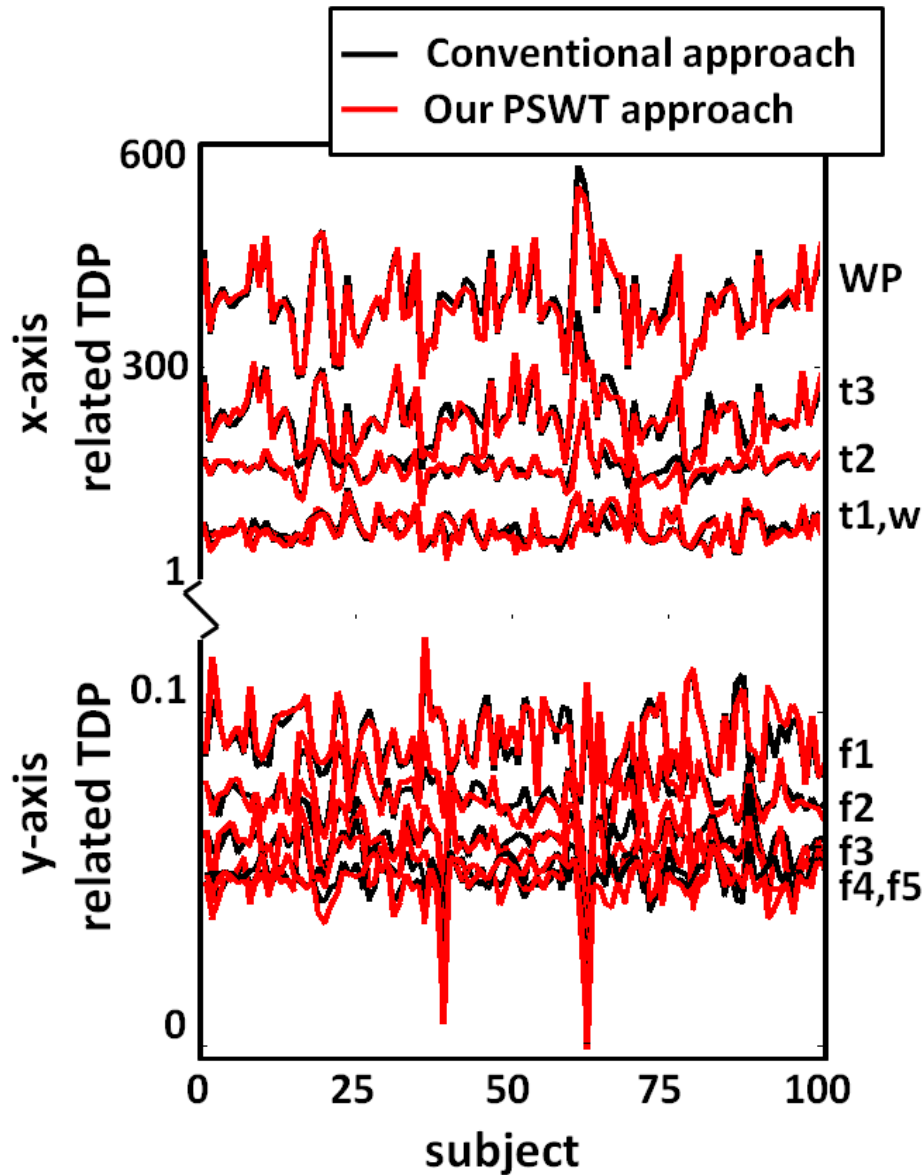
**Table 5.2:** Comparison of statistics in the computation of TDP. We observe that the PSWT cycle captures the trend values.

So far, summary statistics appear meaningful. For further validation, we present the comparison for randomly chosen 100 subjects (from a total 263) in Figure 5.21. All the TDP are on the y-axis. For each parameter, we display 100 values of both the conventional approach and our approach. The black line corresponds to the second column (conventional approach) in Table 5.2. On the other hand, the red line represents the TDP computation using our PSWT-based approach.

#### 5.4.3.2 Effectiveness of Our Novel PSWT-DTW Approach

So far we have shown that if the pseudo-periodic signals are well behaved (“typical”) then both the conventional approach and our PSWT trend approach results in values that are virtually distinguishable in a macro sense. The advantage of our approach becomes more clear when we consider poorly behaved signals. We compare the TDP using two approaches:

- Approach 1: TDP obtained for each pulse cycle using the maxima-minima based

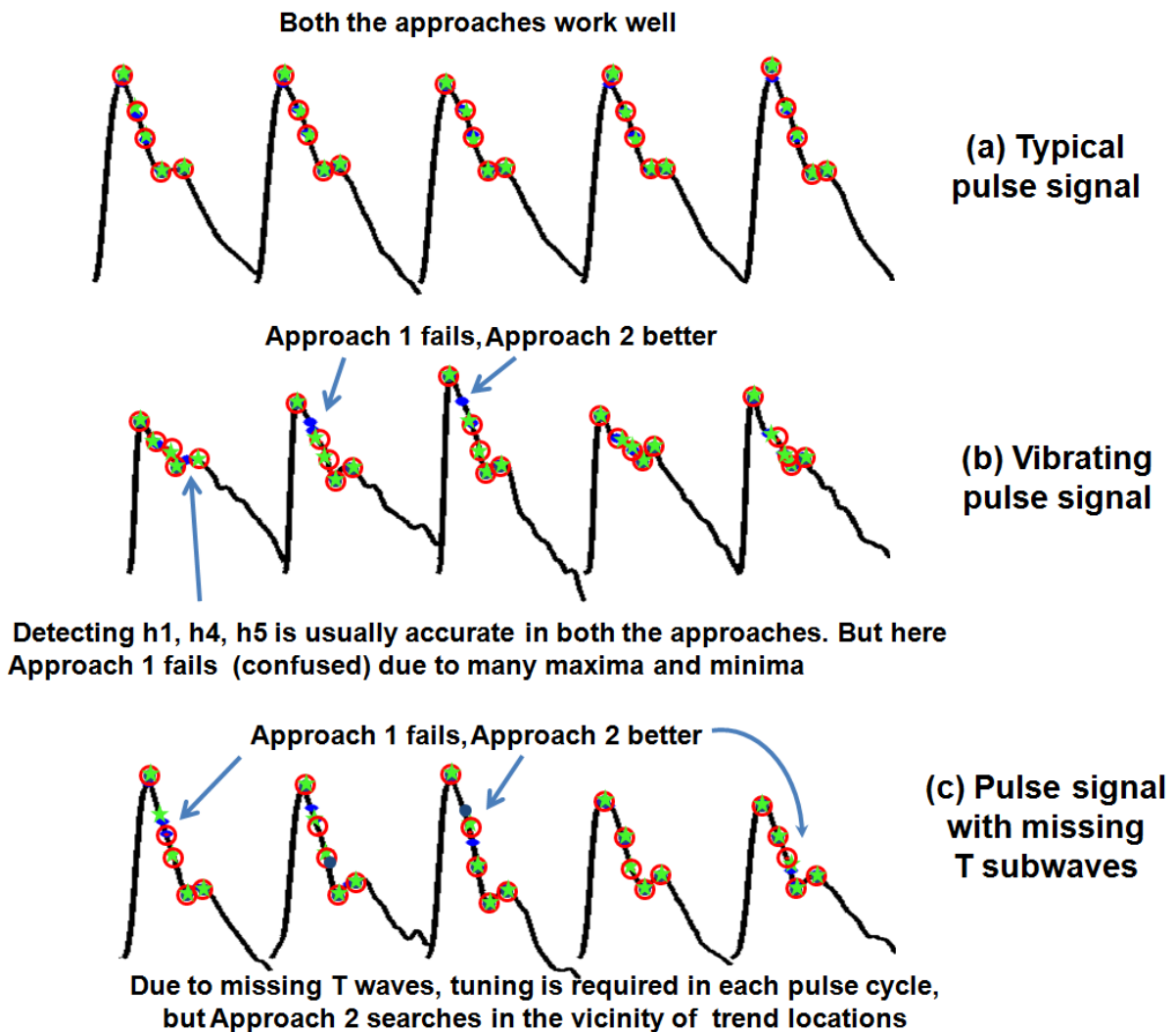


**Figure 5.21:** Comparison between the two approaches of computing TDP. Black points correspond to the second column (mean values) of Table 5.2 and red points to the third column (PSWT-based approach). The two colors follow each other, which makes PSWT cycle acceptable for mapping in the next step. (Figure best viewed in color.)

method (Section 5.3.4).

- Approach 2: TDP obtained for each pulse cycle using our novel approach by first computing the trend TDP locations on the PSWT cycle, and then transferring through DTW.

We show three examples indicating the comparison in Figure 5.22. To facilitate comparisons, we also mark the TDP locations manually. In our experiments, the TDP of a signal computed using Approach 1 requires less than a second, while Approach 2 requires more than 10 seconds due to extra steps (computation of the PSWT and DTW paths).



**Figure 5.22:** Comparison showing the effectiveness of our novel PSWT-DTW approach (Approach 2). The green star, blue filled circles, and red circles (respectively) indicate the ground truth, results from Approach 1 (black path in Figure 5.19) and Approach 2 (red path in Figure 5.19) respectively. Approach 2 detects the TDP locations accurately even for ill-behaved pulse signals.

When the pulse signal shows a typical behavior (example a), both approaches work equally well in computing the TDP. However, in the case of ill-behaved pulse signals,

our novel approach (Approach 2) works better. Specifically,

- In the case of vibrating pulse signal (example b), Approach 1 gets confused among many maxima and minima points and thus sometimes provides false TDP locations. On the other hand, in Approach 2, the trend locations are already computed on the *smooth PSWT cycle* and then are transferred in their vicinity for every pulse cycle.
- In the case of the pulse signal with the missing T-subwave (example c), corresponding locations involve a search for h2 and h3 locations for every cycle in Approach 1. On the other hand, in Approach 2, the search is done only for one PSWT cycle (assuming that the T-subwave is also negligible in the PSWT cycle) and are then transferred to each cycle.

In terms of limitations, our approach requires accurate computation of the trend TDP of the PSWT cycle, as an error in this computation will result in replication on all pulse cycles through the DTW mapping. However, as the computation of the PSWT always results in a smooth trend cycle as explained in Section 5.2.2, the chances of false detection in our approach are less.

When a doctor in traditional medicine diagnoses a subject by feeling the pulse, he or she considers both the average as well as varying values of the pulse parameters such as amplitude, frequency, energy and so on [104] (which can be easily expressed in terms of the TDP). We have accomplished both through our PSWT-DTW approach. The PSWT stage helps to analyze the average behavior (essence) of the entire pulse signal, and furthermore, final results take account of the period-to-period variabilities on several scales.

### **Other Application: Detection of P-wave in the ECG signal**

Our novel approach can be suitably applied in other applications whenever we see signals with pseudo-periodic nature. There could be different sets of X-axis related parameters for a signal depending upon the property under study. For example, two different sets of parameters for an electrocardiogram (ECG) [9] signal



could be (i) Locations and amplitudes of P-wave, QRS-complex and T-wave (for purposes of segmentation) and (ii) Widths of P-wave, QRS-complex (for studying blood circulation).

Consider an example of detecting important P-waves [47, 212] in the P-QRS-T patterns of an ECG (refer Section 2.1). This detection can be tedious due to the varying slopes and amplitudes of P-waves. Our approach can be suitably used as follows:

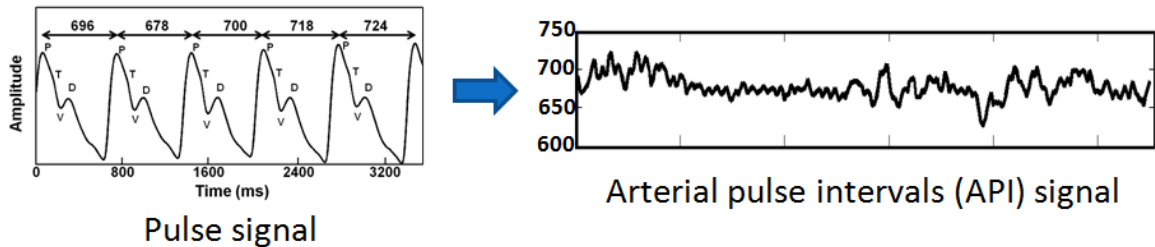
- First, peak detection step (Section 5.1.1) will provide the peaks at the QRS-complex in the ECG signal.
- In obtaining the PS representation, the conventional peak-to-peak segments will result in splitting the important QRS-complex. Instead, our approach will consider portions on both sides of the peak, such that the P-wave, QRS-complex and T-wave will have significant durations in the segments of pseudo-periods.
- The PSWT of the ECG signal is now computed to obtain the trend (ECG) cycle.
- Any preferred algorithm for detecting P-wave locations (such as [47, 212]) is applied on this trend (ECG) cycle.
- Finally, nonlinear DTW-based mapping is used to transfer these detected P-wave trend locations back to each of the ECG cycles.

## 5.5 Pulse Rate Variability Measures

As said earlier, the clinical importance of the heart rate variability became apparent in the late 1980s, to evaluate the equilibrium between the sympathetic system (which accelerates the heart rhythm), and parasympathetic system (which decelerates the heart rhythm). It has been shown that the properties of the HRV play an important role in the better prognosis in patients. Related, but inexpensive, physiological signals derived from the finger plethysmogram [40, 91] has also been preliminarily studied for the variability.

In this section, we establish the efficacy of the pulse rate variability (PRV). We show

that various time-domain, frequency-domain and nonlinear measures are useful in studying the variations in the arterial pulse intervals (API) signals. Note the example of a typical pulse signal in Figure 5.1. The API signal is obtained by computing the peak-to-peak distances as shown in Figure 5.23.



**Figure 5.23:** Typical behavior of a pulse signal. The API signal is the sequence of its peak-to-peak distances (696–678–700–718–724 milliseconds and so on).

### 5.5.1 Time-domain Measures

These are different than explained in Section 5.3, which were related to the important locations on a pulse cycle. Time-domain measures (TDM) may be divided into two classes, (a) those derived from direct measurements of the API (e.g. SDPP), and (b) those derived from the differences between API (e.g. SDAPP). The time-domain measures of PRV are summarized in Table 5.3. Each measure extracts particular significance of the API signal.

**Statistical measures:** The standard deviation (SDPP) reflects all components responsible for the variability. The measures PP50 and pPP50 estimate high frequency variations. In the experiments, we standardize the length of the API signal in the computation of measures which depend upon the length (such as SDPP).

**Geometric measures:** The API are converted into a geometric pattern, such as the sample density distribution of API, or the differences between adjacent API. Then, the measures are computed on these patterns. E.g. the PP triangular index measurement is the integral of the density distribution (i.e. the number of all API) divided by the maximum of the density distribution. The major advantage of geometric methods (PP triangular index & TIPP) lie in their relative insensitivity to the quality of the API signal

[148].

Variable	Units	Description
AVPP & SDPP	ms	Mean and standard deviation of the API
SDAPP		Standard deviation of the mean of API in segments
ASDAPP		Mean of the standard deviation of API in segments
RMSSD	ms	The root mean square of differences of successive API
PP50 count		Number of consecutive PP intervals that differ more than 50 ms
pPP50	%	The percentage value of consecutive API that differ more than 50 ms
PP triangular index		The integral of the sample density distribution (SDD) of API divided by the maximum of the density distribution
TIPP	ms	Baseline width of the minimum square difference triangular interpolation of the maximum of the SDD of API

**Table 5.3:** *Time-domain measures of API.*

## 5.5.2 Frequency-domain Measures

It has been widely accepted that the very low frequency oscillations are related to the thermoregulatory system, the low frequency oscillations to the vasomotor activity, and the high frequency oscillations to the respiratory activity [189]. We have already mentioned that the sympathetic and parasympathetic branches of the autonomic nervous system have differing characteristic time constants. The high frequency power is a measure of parasympathetic tone, while the low frequency power reflects both sympathetic and parasympathetic activity [214]. The LF/HF ratio, when high, is considered to reflect dominance of sympathetic activity, and when low, is thought to reflect dominance of vagal activity. We consider three frequency bands as very low frequency (VLF, 0–0.04 Hz), low frequency (LF, 0.04–0.15 Hz), and high frequency (HF, 0.15–0.4 Hz).

Though the pulse signal is a regularly sampled (at 500Hz) data, the API is inherently irregularly spaced in time. Therefore in our experiments, we firstly resample this irregularly time-sampled API signal before the computation of the frequency-domain measures (FDM).

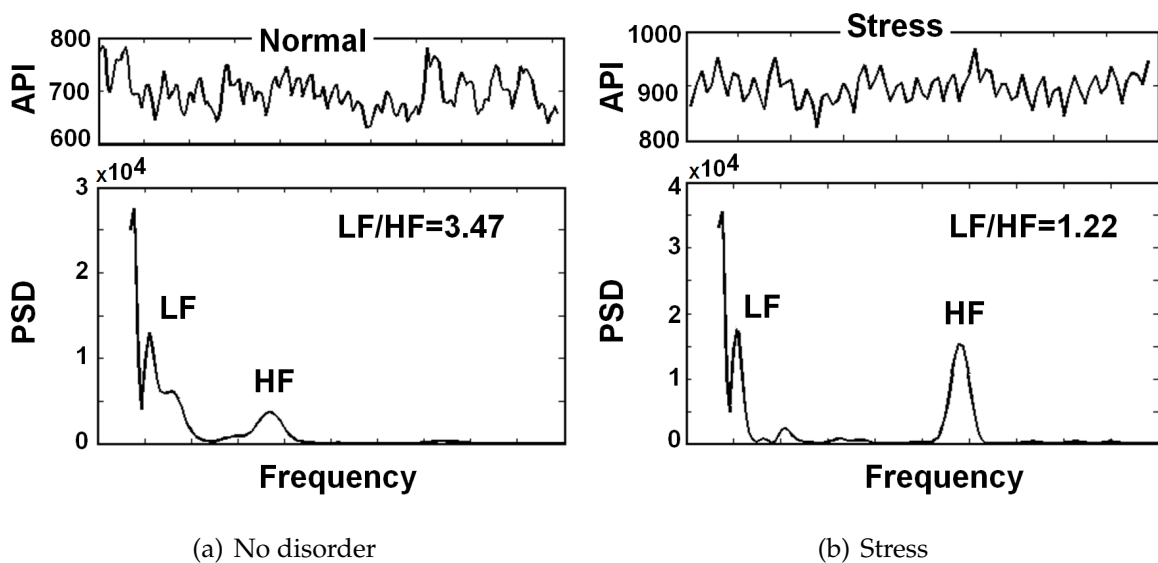
We compute the power spectral density (PSD) of the API using parametric and nonparametric approaches and then choose the appropriate powers from the above mentioned ranges. The PSD analysis provides the basic information of how the power (i.e. variance) is distributed as a function of frequency. We compute the PSD using the fast Fourier transform (FFT) [nonparametric] model, and the autoregressive (AR) [parametric] model. The representation of LF and HF in normalized units emphasizes the controlled and balanced behavior of the two branches of the autonomic nervous system [148].

Table 5.4 lists the FDM computed from PSD using both the parametric and nonparametric approaches.

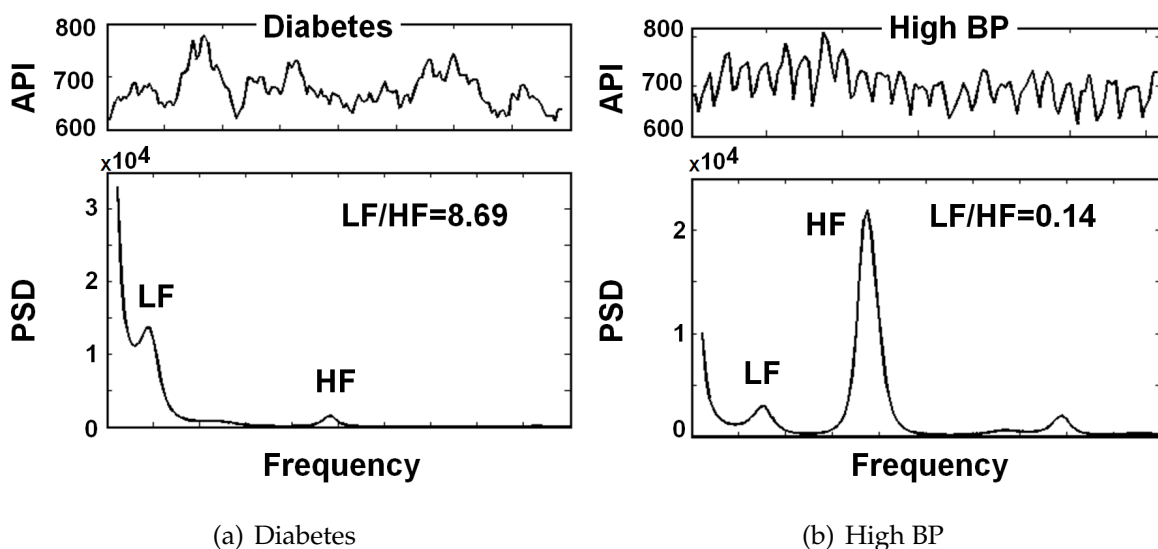
Variable	Units	Description
Peak frequency	Hz	Peak frequencies of the PSD estimate for VLF, LF, and HF frequency bands (see §5.5.2)
Power	$ms^2, \%$	Respectively for VLF, LF, and HF frequency bands
Power	n.u.	Respectively for LF, and HF frequency bands in normalized units
LF/HF	%	Ratio of LF and HF frequency band power

**Table 5.4:** *Frequency-domain measures.*

Representative examples of power spectrums of the PRV for various disorders are shown in Figure 5.24[a-b] and Figure 5.25[a-b]. Only two frequency bands are studied here, LF and HF. We applied our PRV method, an alternative to the HRV, to many subjects and observed that most of the results of LF/HF measure are consistent with the HRV results obtained earlier in the literature.



**Figure 5.24:** Power Spectrums of pulse signals acquired from different subjects, computed using the nonparametric FFT spectrum approach.



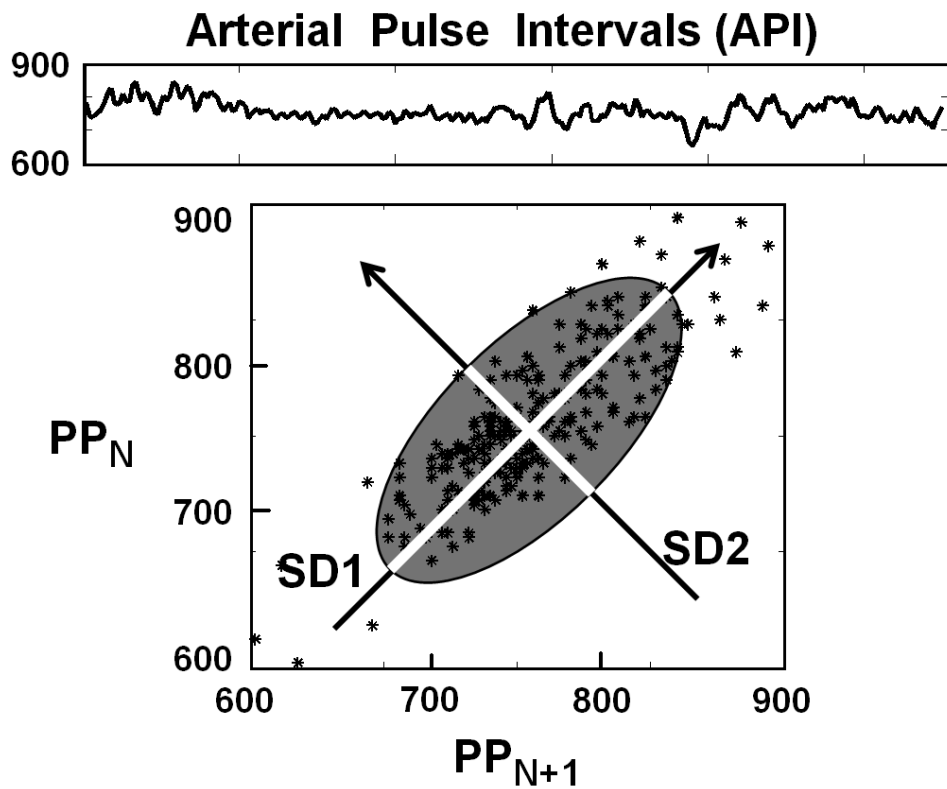
**Figure 5.25:** Power Spectrums of pulse signals acquired from different subjects, computed using the parametric AR spectrum approach.

### 5.5.3 Nonlinear Poincaré-based Measures

It has been speculated that analysis of the HRV based on methods of nonlinear dynamics might elicit valuable information for the physiological interpretation of the HRV and for the assessment of the risk of sudden death [174]. We use the easy to comprehend nonlinear method of the Poincaré plot. The method summarizes the

entire recording, at the same time making it possible to extract the information on both short as well as long time behavior [170] of the system (API in this case). It is both: A visual technique which can make use of the human eye's ability to recognize patterns, and a quantitative one, in the sense that it introduces various measures to quantify the information contained in the Poincaré plot.

Poincaré plot is a graphical presentation of the correlation between consecutive values (PP) in API signal. Thus, each PP is plotted against the next one. For an API signal with healthy behavior, we obtain a comet-like shape, similar to the one shown in Figure 5.26. The reason is that for healthy behavior, the consecutive PP values usually contain very close values, even though the overall variability is high.



**Figure 5.26:** A sample API signal and its Poincaré plot.  $SD1$  and  $SD2$  are the dispersions of points perpendicular and along the axis of line-of-identity.

As shown, an ellipse is fitted on the line-of-identity at  $45^\circ$  to the normal axis [148] so that maximum density is achieved.  $SD2$  is defined as the standard deviation of the projection of the PP on the line of identity ( $y = x$ ), and  $SD1$  is the standard deviation of projection of the PP on the line perpendicular to the line of identity ( $y = -x$ ).

**Observations.** In the Poincaré plot, the points towards the top right indicate two consecutive increasing intervals (implying that the pulse rate is decreasing), while the bottom-left points indicate two consecutive decreasing intervals (implying that the pulse rate is increasing). The Poincaré plot summarizes the entire data as  $SD2$  indicates the long-term variability, and  $SD1$  indicates the short-term variability (mainly caused by respiratory sinus arrhythmia). Table 5.5 lists all the nonlinear Poincaré-based measures. We observe variations in the plots for different API signals.

Variable	Units	Description
centroid		The centroid of the ellipse (see §5.5.3)
SD1	ms	The standard deviation perpendicular to the line-of-identity
SD2	ms	The standard deviation along the line-of-identity
SD1/SD2		SD1 to SD2 ratio

**Table 5.5:** *Nonlinear Poincaré-based measures.*

**Variations with age.** As the age increases, the heart rate and the pulse rate decreases. In other words, it results in increase of the pulse intervals. This can be verified in the Poincaré representation as shown in Figure 5.27. The elliptical area shifts along the line-of-identity as the age increases from 2 months to 45 years.

**Capturing arrhythmic behavior.** In Figure 5.28, we display the Poincaré plot of an API signal of arrhythmic behavior, where the third beat is missing (every now and then). The Poincaré plot in this case shows a different coverage of points. The region of the small durations followed by large durations and that of the large durations followed by small durations get easily separated. Thus, such abnormality in a pulse signal can be easily visualized (detected) using this technique.

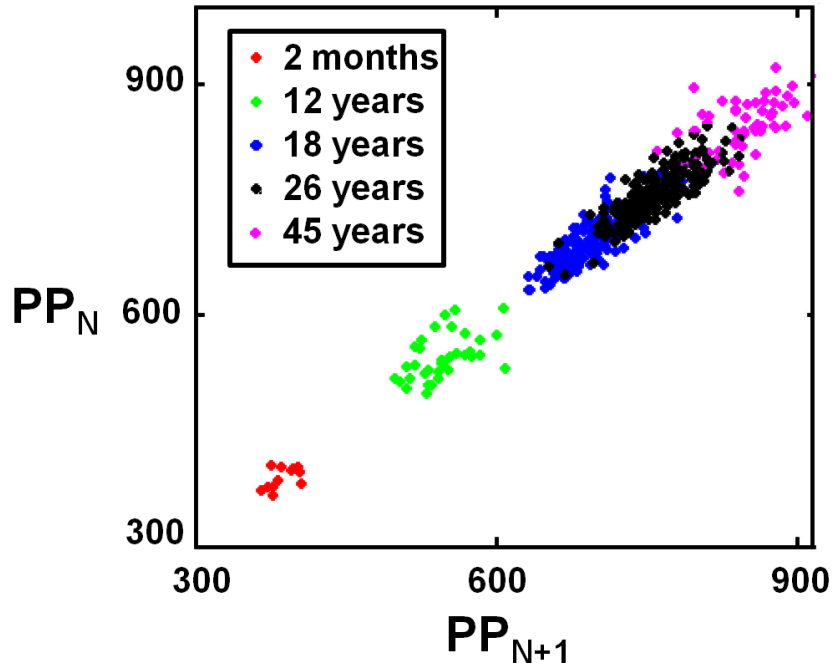


Figure 5.27: The elliptical area on the Poincaré plot shifts as age increases.

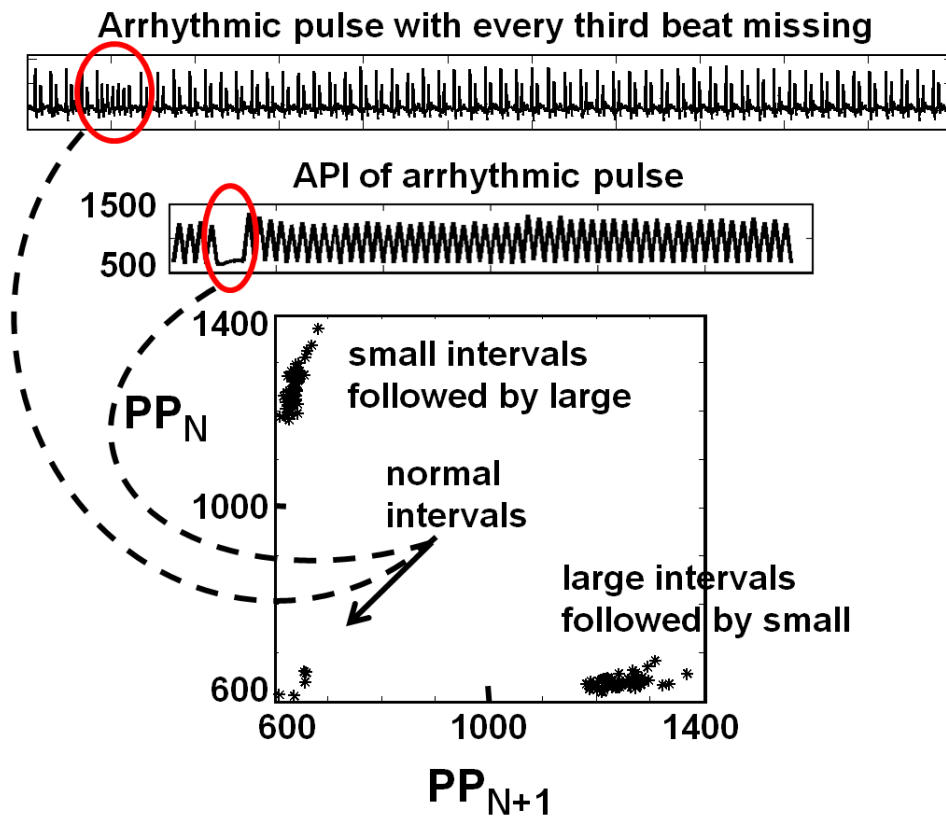


Figure 5.28: Poincaré plot of an arrhythmic pulse shows two distinct areas instead of a typical elliptical behavior.



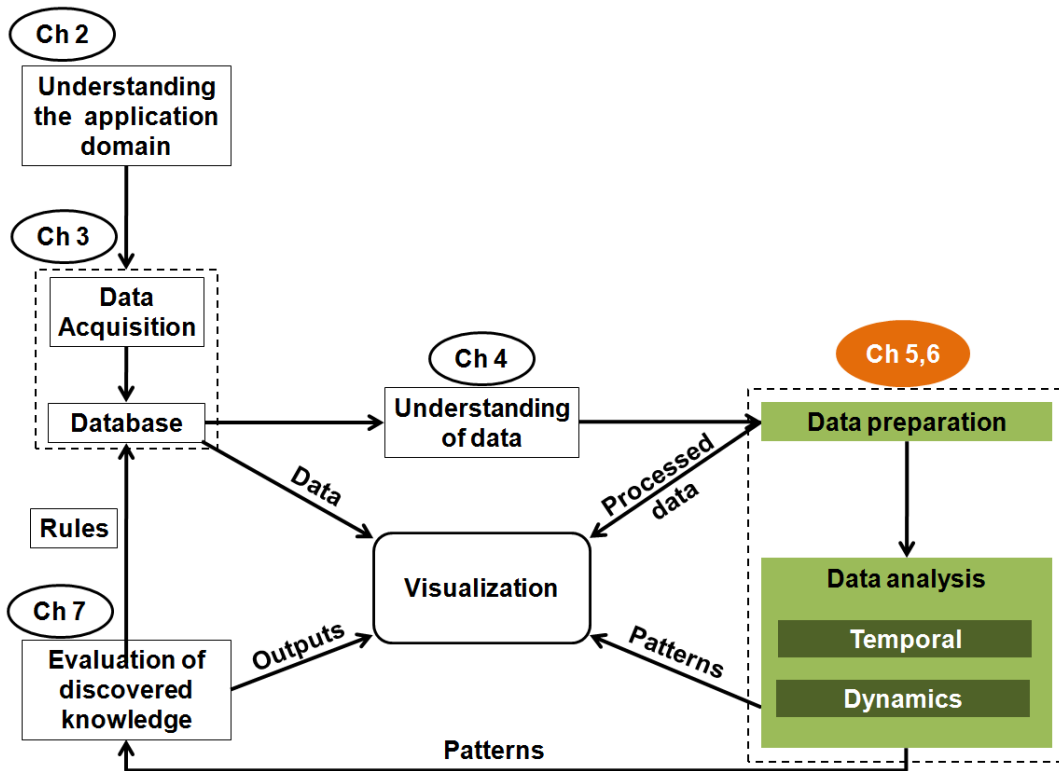
## Conclusions

In this chapter, we explained the physiologically important pseudo-periodic nature of the pulse signals. Though adjacent pulse cycles look similar in shape (usually composed of P-T-V-D subwaves), they show different patterns. These evolutionary (period-to-period) changes in the behavior are medically significant.

A novel PSWT-DTW approach was established to compute various time-axis related parameters taking advantage of the pseudo-periodicity of the signals. Our novel approach reduces the efforts of searching the pseudo-locations when they have either confusing or overlapping values. Specifically, the tedious efforts of searching for the parameters in the ill-behaviors were reduced at two stages. In the computation of TDP, we have developed a maxima-minima based approach to avoid the searching of pseudo parameters arising due to the false negatives of missing zeros. Furthermore, our novel PSWT-DTW approach helps for two reasons. First, the resulting trend (PSWT) cycle is smooth in nature and second, the efforts (if any) of tuning the TDP are only for one PSWT cycle.

We established the pulse rate variability (PRV) by computing various time-domain (TDM), frequency-domain (FDM), and nonlinear Poincaré based measures. We showed their efficacy in capturing variations with age, and in detecting various disorders including high BP, diabetes, skin disorder and arrhythmic behavior. All our methods find scope in future clinical use.





---

## Chapter 6

# Capturing Nonlinear Dynamics

---

*Extensive research has been done to show that heartbeats are composed of the interaction of many physiological components operating on different time scales, with nonlinear and self-regulating nature [8, 62, 64, 70]. The beat-to-beat variations of the heart are members of a special class of complex processes – those which are inhomogeneous, non-stationary, and display chaotic properties [25, 82, 83].*

*The chaotic behavior is highly sensitive to the initial conditions. So, a small change in a parameter can lead to sudden and dramatic changes in the behavior of the system. Nonlinear dynamics is concerned with the study of such systems whose time evolution equations are nonlinear. Poincaré showed that no matter how clever we are, we won't be able to write down the equations that solve many nonlinear systems.*

*A chaotic system with strong dependance on the initial conditions is better visualized through a phase space, whose coordinates describe the dynamical state at any instant. A dynamical rule is used to specify the immediate future trend of all state variables, given only the present values of those same state variables.*

*In this chapter, we study the nonlinear dynamics of the arterial pulse signals. We have shown – for the first time – that the pulse (API signal) exhibits a chaotic self-similar nature, and requires a large number of exponents to characterize their scaling properties. We systematically study the chaotic properties, fractal nature, and structures of the pulse signals, respectively through the recurrence plot-based analysis, local Hölder exponents & multifractal spectrum-based analysis, and the topological analysis. We also analyze their variations for different subjects.*

## **6.1 Fractal Nature**

The fractal structures are resulted from an iterative process (at every scale of observation) such that when their pieces are compared to larger pieces or to those of the whole, they prove similar to each other. However, in practical applications, such similarity has some limitations, and minute variations are observed.

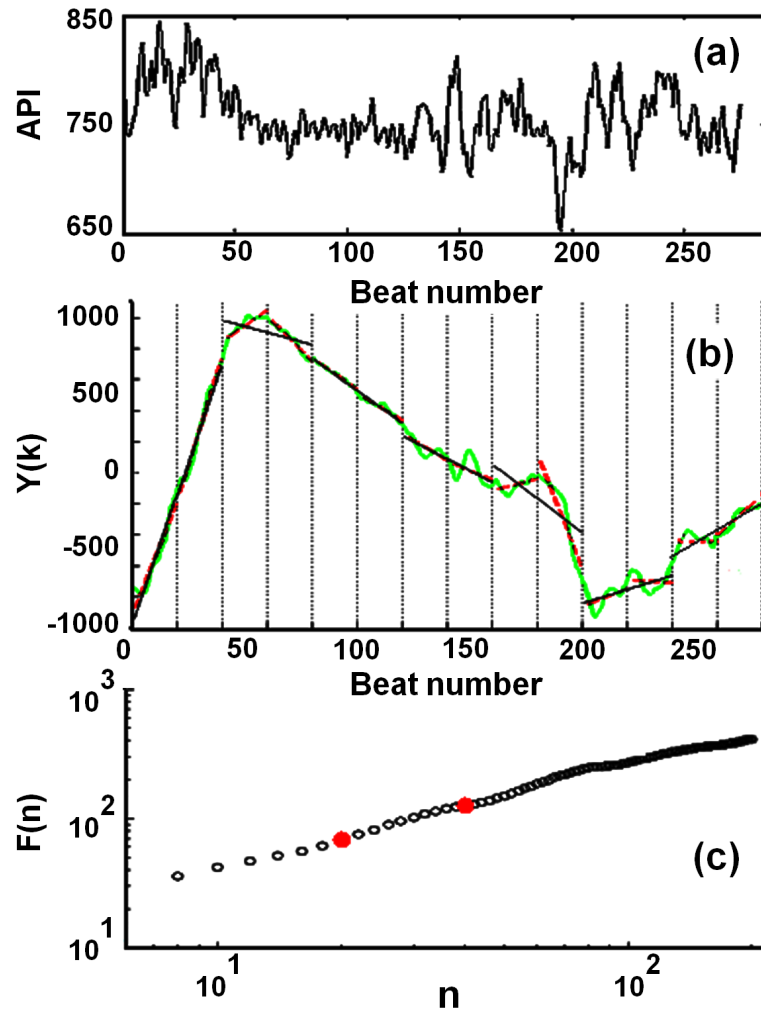
The pulse signals show a complex nature as a result of the interaction of many physiological components operating on different time scales. These interactions are nonlinear and self-regulating (rhythmic heart action) through feedback control, leading to the nonlinear character. Due to these actions, the pulse signal is also part of such complex system. This indicates that different parts of the signal (both pulse and API) have different scaling properties, and thus it is natural to expect the need for many exponents to characterize them.

### **6.1.1 Self Similarity of the API Signal**

*Detrended fluctuation analysis (DFA)* was introduced in [161] to improve on root mean square analysis of highly non-stationary data, by removing non-stationary trends from a long-range-correlated time series [43].

Figure 6.1 illustrates the DFA algorithm briefly. A key issue of the DFA-analysis is

that the extrinsic fluctuations from uncorrelated stimuli can be interpreted as “trends”, and decomposed (by detrending across time scales) from the intrinsic dynamics of the system itself [129].



**Figure 6.1:** (a) An example of an API signal. (b)  $y(k)$  (in green), and local trends for  $n = 20$  (in red) and  $n = 40$  (in black). (c) Log-log plot of  $F(n)$  versus  $n$  shows a linear trend indicating the presence of self-similar nature.

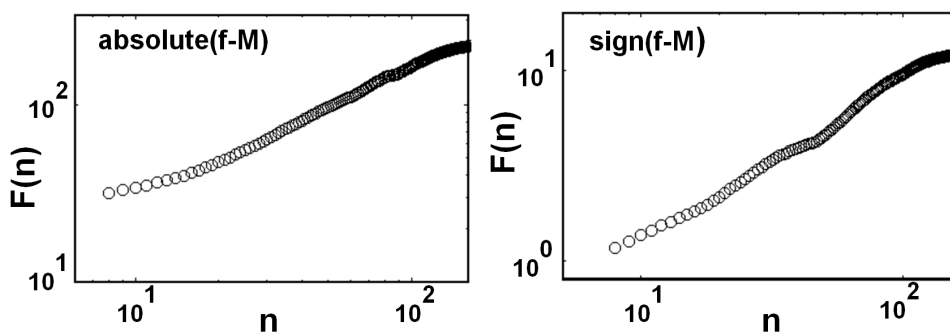
Consider a peak-to-peak (API) signal  $f(i)$  [ $i = 0, \dots, N - 1$ ] shown in Figure 6.1 (a). First,  $f$  is suitably integrated  $y(k) = \sum_{i=0}^{k-1} (f(i) - M)$  where  $M$  is the mean. (Notice that values of  $f(i) - M$  are signed.)  $y(k)$  is now divided into boxes of equal width  $n$  and the least-square line fitting the data in each box,  $y_{p,n}(j)$ ,  $j = 0 \dots (N/n) - 1$ ,  $p = 1 \dots n$ , is computed. The integrated time series is then detrended by subtracting the local trend  $y_{p,n}(j)$ , and the root-mean square fluctuation of the entire detrended series,  $F(n)$  is computed [81]:

$$F(n) = \sqrt{\left(\frac{1}{N} \sum_{k=0}^N [y(k) - y_{p,n}(\lfloor k/n \rfloor)]^2\right)} \quad (6.1)$$

As shown in Figure 6.1 (b),  $y(k)$  is the solid green curve and the vertical dotted lines indicate boxes of width  $n = 20$ . The red and black straight line segments represent the *trends* estimated in each box of sizes  $n = 20$  &  $n = 40$  respectively by a linear least-squares-fit. Note that the typical deviations for the two lines is different.  $F(n)$  is now plotted against the box width  $n$  for various values in Figure 6.1 (c). The two highlighted points in red are for  $n = 20$  &  $n = 40$ .

Typically,  $F(n)$  increases with the box size  $n$ . The linear relationship on a log-log graph (Figure 6.1 (c)) indicates the presence of power-law scaling (self-similarity), such that fluctuations in small boxes are related to the fluctuations in larger boxes in a power-law fashion. The slope of the line relating  $\log(F(n))$  to  $\log(n)$  determines the fractal scaling exponent.

**DFA on absolute and signed values:** In another experiment, self scaling is observed if we take absolute values (i.e.,  $|f - M|$ ). A similar nature is found if  $f - M$  is approximated to be a binary number (1 or -1). These situations (Figure 6.2) have been previously explored for beat-to-beat patterns of ECG [82].



**Figure 6.2:** Linear DFA plots for absolute and signed values of an API signal.

## 6.1.2 Hierarchical Organization of Singularities

### 6.1.2.1 Wavelet Action

In the early nineties, a statistical approach based on the continuous wavelet transform was proposed for singular measures and multi-affine functions [11, 144]. The (continuous) wavelet transform, with its ability to remove local trends, and to extract inter-beat variations on different time scales, enables us to identify fractal patterns in the pulse fluctuations.

As already explained in Section 5.1.1, the wavelet transform (WT),  $W_{s,x_0}(f)$ , provides a way to analyze the local behavior of a signal  $f$ , which is a convolution product of the signal with the scaled( $s$ ) and translated( $x_0$ ) kernel.

$$W_{s,x_0}(f) = \int_{-\infty}^{+\infty} \frac{1}{s} \Psi\left(\frac{x-x_0}{s}\right) f(x) dx \quad (6.2)$$

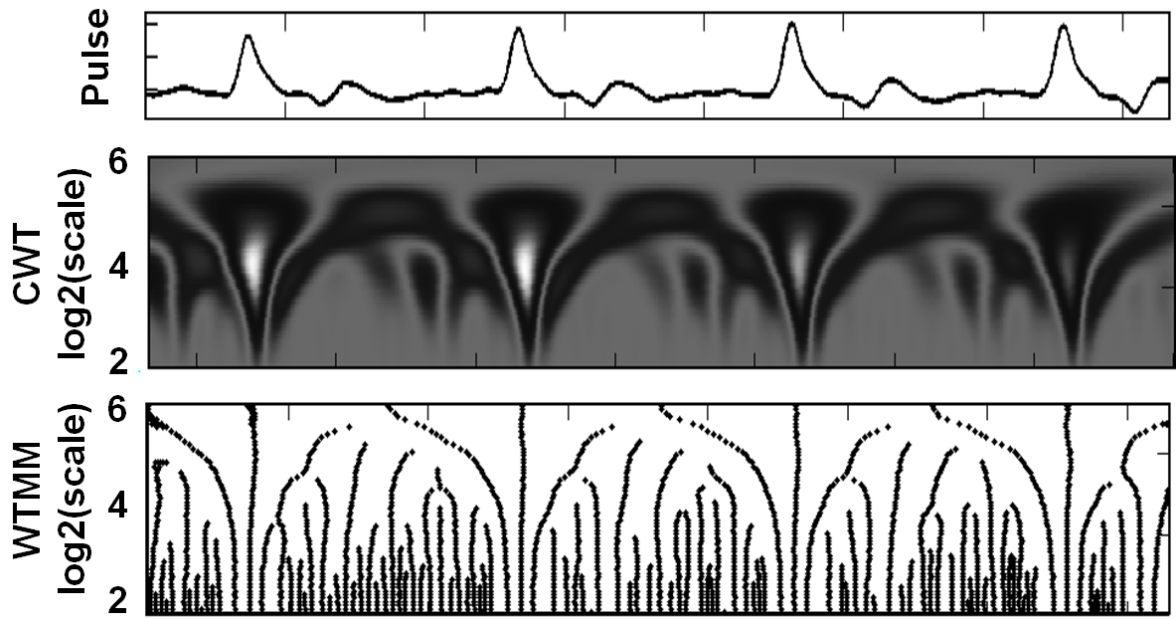
where  $s$  is the scale parameter and  $x_0$  is the translation parameter.

### 6.1.2.2 Wavelet Transform Modulus Maxima

The continuous wavelet transform (CWT) in Equation 6.2 is an extremely redundant and a computationally expensive representation. Therefore, we have employed wavelet transform modulus maxima (WTMM) method [79, 167, 174, 203], which changes the continuous sum over space into a discrete sum by taking only the local maxima. Figure 6.3 illustrates the branching structure of the WTMM skeleton in the (scale,time). It indicates the hierarchical organization of the singularities.

The WTMM particularly takes advantage of the space-scale partitioning given by this skeleton to define a series of exponents  $\tau(q)$  through the partition functions  $Z(s,q)$  of  $q^{th}$  moment [11, 144].  $Z(s,q)$  is defined as the sum of the  $q^{th}$  powers of the local WTMM coefficients at scale  $s$ .





**Figure 6.3:** A sample pulse signal, its CWT, and corresponding WTMM tree. The branching structure of the WTMM tree is useful in extracting the hierarchical organization of the singularities.

$$Z(s, q) = \sum_{\Omega(s)} |W(f)|^q \propto s^{\tau(q)} \quad (6.3)$$

where  $\Omega(s)$  is the set of all maxima at the scale  $s$ . Here,  $q$  has the ability to select a desired range of values [83]:

- For large negative  $q$ , weak exponents are addressed, while strong exponents are filtered out.
- For large positive  $q$ , strong exponents are addressed, while weak exponents are effectively filtered out.

### 6.1.3 Local Hölder Exponents

For a time series  $f$ , if there exists a polynomial  $P_n$  of degree  $n < h$  and constant  $C$ , such that

$$|f(x) - P_n(x - x_0)| \leq C|x - x_0|^h \quad (6.4)$$

the supremum of all exponents  $h(x_0) \in (n, n + 1)$  is termed the Hölder exponent, which characterizes the singularity strength. It is evident [185] that the Hölder exponent describes the local regularity of the function (or distribution)  $f$ . The higher the value of  $h$ , more regular is the local behavior the function  $f$ . Thus it characterizes the scaling of the function locally and the distinct scaling behavior of different signals can be exploited to characterize and classify time series.

The ability of WT (refer Equation 6.2) to reveal even the weaker singularities within the time series by adjusting  $s$  makes it an indispensable tool for singularity analysis. This is formalized by its relation [144] with the Hölder exponent:

$$W_{s,x_0}(f) \propto |s|^{h(x_0)}, \quad s \rightarrow 0 \quad (6.5)$$

### 6.1.3.1 Stable Computation

The WTMM-based formalism (Section 6.1.2.2) provides global estimates of scaling properties of a time series (pulse signal). Recently, it has been found that even though such global estimates of scaling is often a required property, local analysis may provide more useful information.

In the traditional form, the estimation of local singularity strengths and their spectra may not be possible due to the fact that in real life data, the singularities are not isolated but densely packed. This causes the logarithmic rate of increase or decrease of the corresponding wavelet transform maximum line to fluctuate. But very recently, Struzik [202, 204] has provided a stable methodology for estimating the local exponents, in which he has modeled the singularities as if they were created through a multiplicative cascading process. This method has been successfully applied for classification of human gait [185].

The method [185, 204] is as explained below. The mean Hölder exponent  $\bar{h}$  is given by

$$\log[M(s)] = \bar{h} \log(s) + C \quad (6.6)$$

where

$$M(s) = \sqrt{\frac{Z(s, 2)}{Z(s, 0)}} \quad (6.7)$$

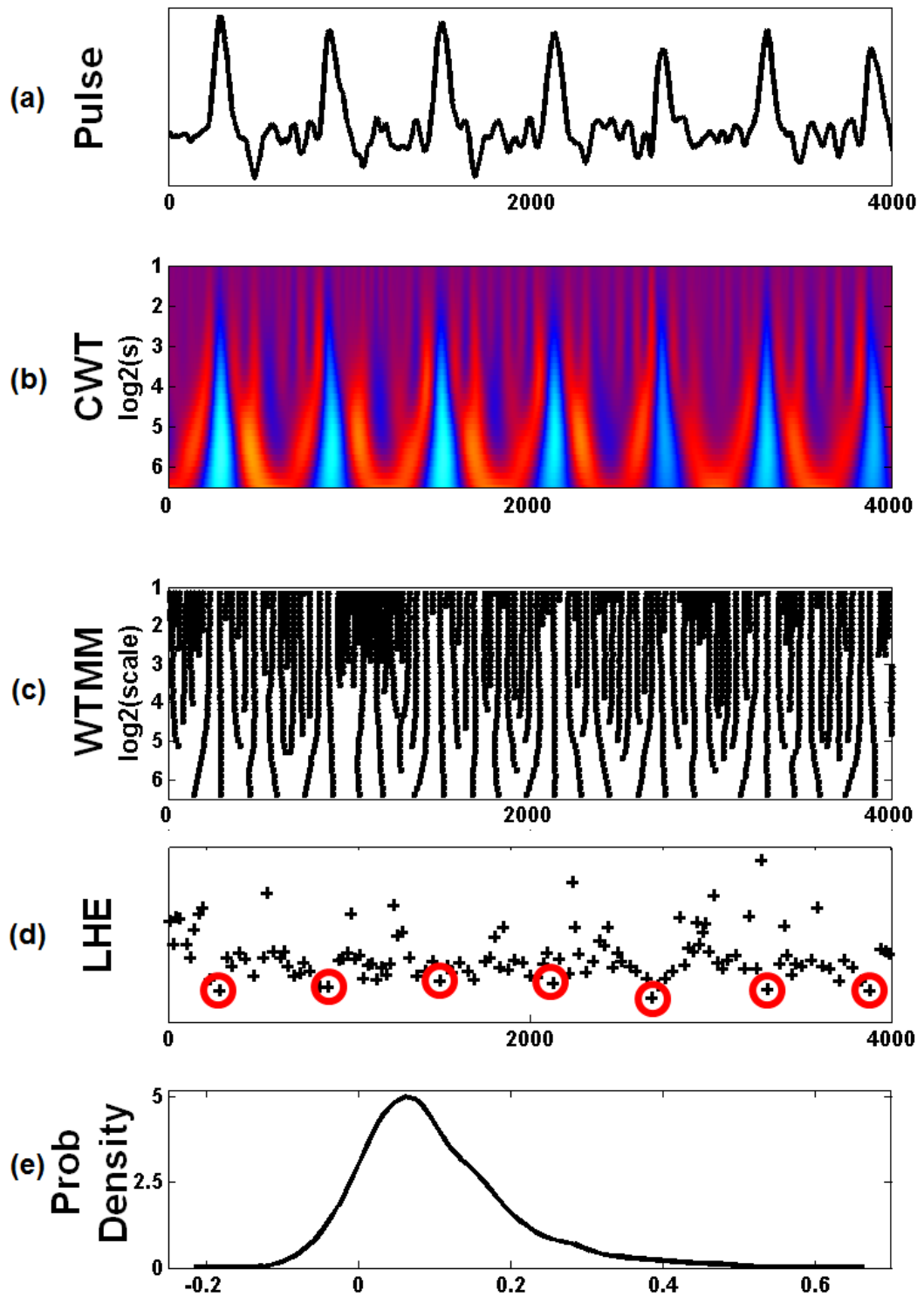
Employing the multiplicative cascade model, the approximate local Hölder exponent  $\hat{h}(x_0, s)$  at the singularity  $x_0$  can now be evaluated as the slope:

$$\hat{h}(x_0, s) = \frac{\log(|W_{s, x_0}(f)|) - (\bar{h} \log(s) + C)}{\log(s) - \log(s_N)} \quad (6.8)$$

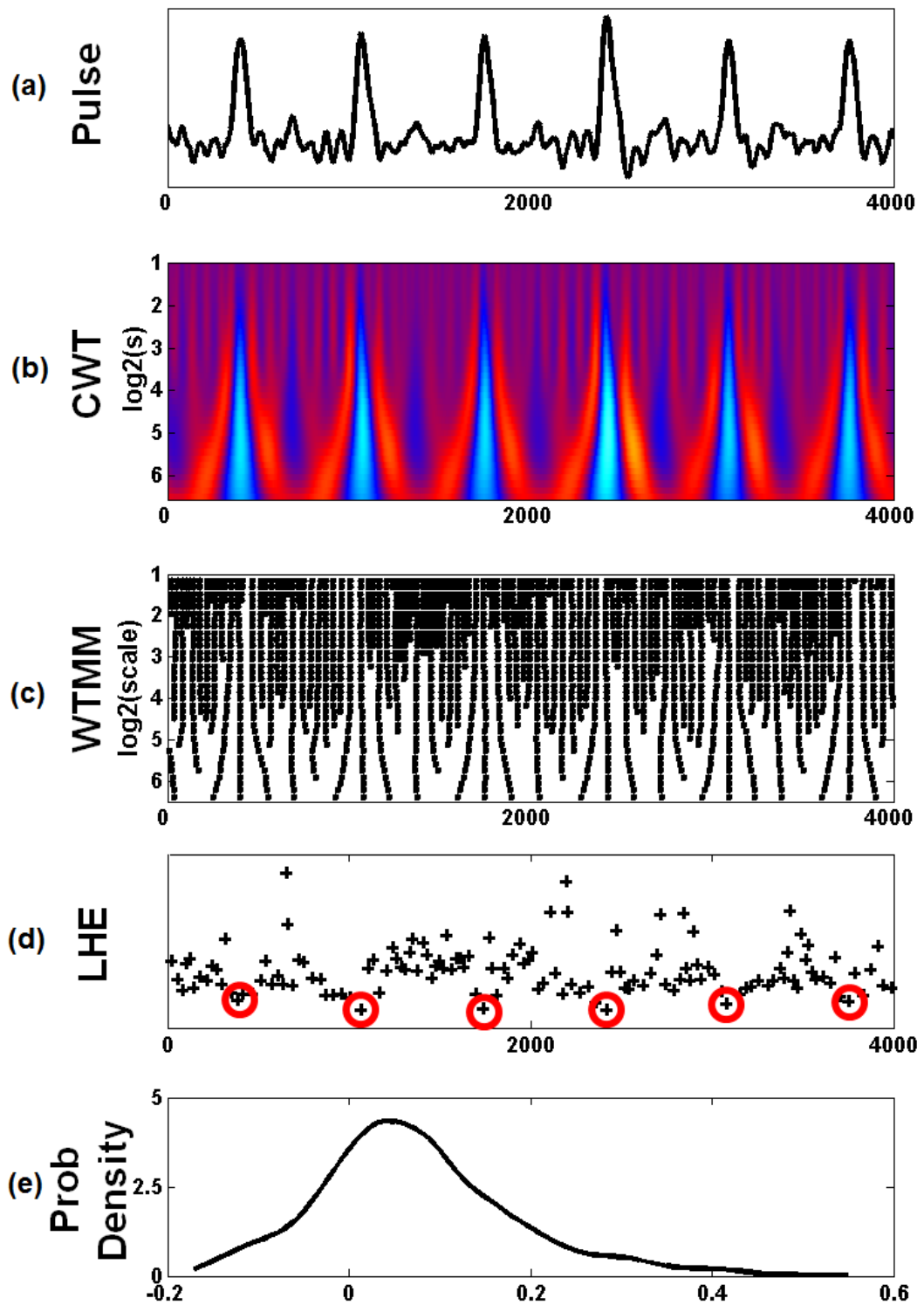
where  $s_N$  is the length of the entire wavelet maxima line tree, that is, the maximum available scale that coincides with the sample length  $s_N = N$ , and  $x_0$  belongs to the set  $\Omega(s)$  of all wavelet maxima at the scale  $s$  that assume the value  $W_{s, x_0}(f)$ .

**Methodology:** Figure 6.4 and 6.5 shows two examples of the computation of the WT and local Hölder exponents. The procedure is as follows. We first compute the continuous wavelet transform of a pulse signal. We have used the second derivative of the Gaussian function, i.e. Mexican hat, as the analyzing wavelet. This wavelet is extensively used, as it possesses good localization capabilities in both position and frequency. Then we compute the local Hölder exponents as explained above. We then compute the probability density of these local Hölder exponents, and then fit this density with a Gaussian kernel.

The local Hölder exponents capturing the singularity strengths at the peaks of the signal, e.g. , are highlighted. Note that the exponents are observed in the vicinity of the peaks on the time-axis. Therefore, it is clear from the examples that, the time-scale localization properties of the wavelets reveal the underlying hierarchy of exponents defining the singularities of the pulse signal.



**Figure 6.4:** (a) Sample pulse signal of diabetes behavior, (b) Continuous wavelet transform (CWT), (c) Wavelet transform modulus maxima (WTMM) tree, (d) Local Hölder exponents (LHE) at scale one, and (e) Probability density of LHE at scale 1. Note that the local Hölder exponents at red circles capture, e.g., the singularity strengths at the peaks in the pulse signal.



**Figure 6.5:** (a) Sample pulse signal of skin disorder behavior, (b) Continuous wavelet transform (CWT), (c) Wavelet transform modulus maxima (WTMM) tree, (d) Local Hölder exponents (LHE) at scale one, and (e) Probability density of LHE at scale 1. Note that the local Hölder exponents at red circles capture, e.g., the singularity strengths at the peaks in the pulse signal.

### 6.1.3.2 Combination of Wavelet Scales

Prior works, such as [185, 205] have used only the lowest scale in the computation of LHE, for different applications. But, we make use of the properties of the wavelet transform that it captures different information at different scales (or resolutions). Figure 6.4 and 6.5 displays two cases of pulse signals and their local Hölder exponents.

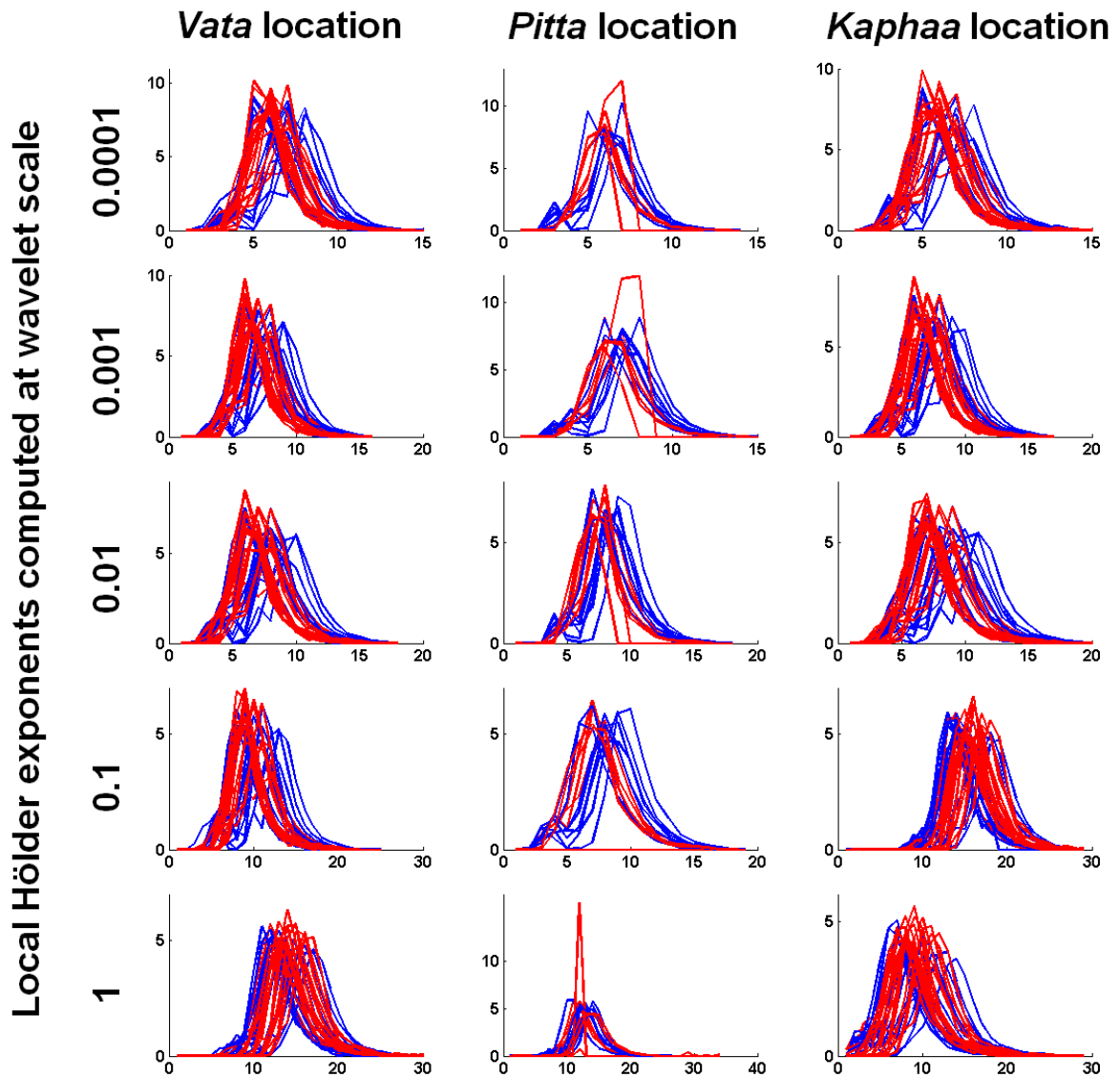
Now, we present more examples, in Figure 6.6, of the computation of local Hölder exponents at various wavelet scales. This dataset contains pulse signals from 22 subjects of diabetes behavior and 13 subjects of skin abnormalities. The results for *Tridosha* locations (refer Section 4.2.2) are provided.

The probability densities of the exponents for the two behaviors in the figure show different patterns for each of the scales, and hence there is an opportunity of the classification. Furthermore, the densities show different patterns, for the same behavior, across different scales. Actually, keeping Equation 6.5 as a reference, the local Hölder exponents capture different singularity strengths at different wavelet scales. Therefore in future, selection of *lower* wavelet scales, computing more and more informative local singularity features, and selecting the best features would provide better classification results.

### 6.1.4 Multifractal Spectrum

The finding of power-law correlations through detrended fluctuation analysis (Section 6.1.1) indicates the presence of scale-invariant, fractal structures in the pulse. Multifractal time series are self-similar only in local ranges of the structure, their fractal measure does vary in time; hence they can be characterized by a set of local fractal measures. We used the wavelet transform modulus maxima-based formalism [144] to capture the spectrum of these fractal measures of the API signals.

In Equation 6.3, we observe that  $\tau(q)$  is a nonlinear function, which indicates that the pulse is a multifractal signal. The local tangent slope to  $\tau(q)$  gives the corresponding exponent  $h(q)$ . The related dimensions  $D(h)$  for each value of  $h$  form the spectrum of



**Figure 6.6:** Comparison of probability densities of the local Hölder exponents of various pulse signals. The blue color is for the skin disorder behavior and the red color is for the diabetes behavior. Note the differences in the patterns of the densities for different wavelet scales.

singularities of the signal. Formally, the transformation from  $\tau(q)$  to  $D(h)$  is referred to as the Legendre transformation [11]:

$$h(q) = d\tau(q)/dq \quad (6.9)$$

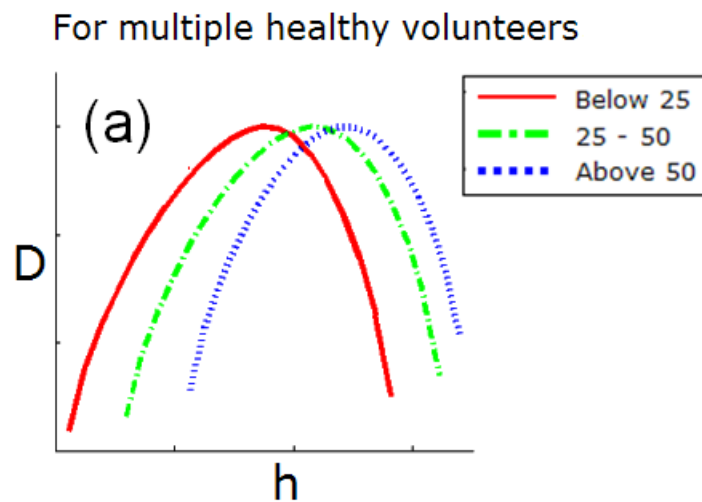
$$D[h(q)] = q h(q) - \tau(q) \quad (6.10)$$

## Observations

In our experiments, we first extracted peaks in a pulse signal using a complex frequency b-spline wavelet (Section 5.1.1) to obtain the arterial pulse intervals (API) signal. The length of each API was at least 250. We obtained the multifractal spectra for all the API signal in the database as explained above.

### Variations with Age

Figure 6.7 shows the results of the *vata* pulse signals from the left hand of the subjects of normal behavior. The spectra are colored according to the subject's age as – 'below 25 years' (G1), '25–50 years' (G2), and 'above 50 years' (G3) respectively. We observe that the multifractal spectrum drifts to higher values of  $h$ , as the age increases.



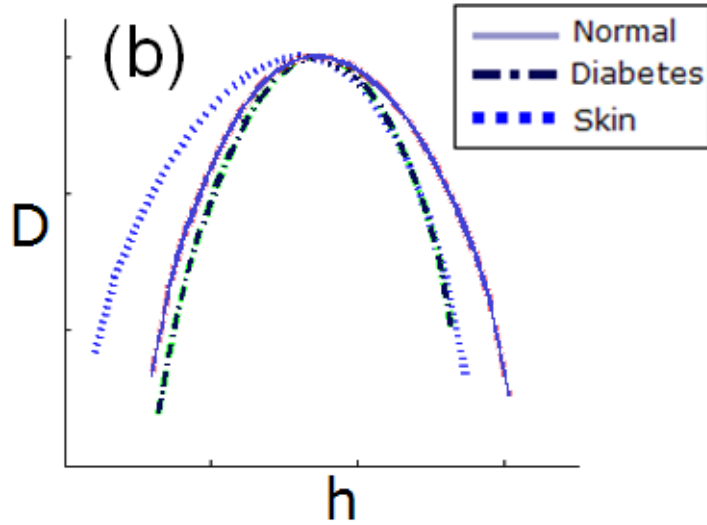
**Figure 6.7:** Variations in multifractal spectra for three age-groups with normal behavior.

### Variations with Disorder

Figure 6.8 shows the results. In this case, the multifractal spectra of volunteers in the age group G3 have been chosen. It can be seen that the spectra vary, and hence there is an opportunity for classification. These spectra were for the pulse signals obtained from the *vata* location on the left hand; but again, similar behavior was observed for all pick-up positions.



For volunteers of age above 50



**Figure 6.8:** Variations in multifractal spectra for subjects in group G3 with normal, diabetes and skin abnormalities behaviors.

The varying ranges of the  $D(h)$  spectrum (Figure 6.7 and Figure 6.8) for various disorders may be useful in detection of various types of *nadi*. These results are consistent with the heart-beat results from [234], which state that a person’s multifractal spectrum is controlled mainly by his/her neurosystem. With advancing age, the neuroautonomic control of a person’s body on the ECG decreases and tends from multifractality to monofractality. We provide the ratios of areas under the spectra in the Table 6.1. We observe that the areas under G2 have higher values as compared to those under G3. This reinforces the result in [234] that the areas under multifractal spectrum of young adult subject is larger than the older ones.

	Left Hand			Right Hand		
	G1	G2	G3	G1	G2	G3
Normal	14.95	15.23	14.82	15.04	15.40	13.71
Diabetes	-	-	14.36	-	-	15.27
Skin Disorder	13.67	-	14.25	14.82	-	13.47

**Table 6.1:** Ratios of areas under the multifractal spectra for multiple age groups, and varying disorders.

## 6.2 Chaotic Nature

One of the conventional approaches of studying the complex chaotic system is to compute its Lyapunov exponents. The computation of Lyapunov exponents on our pulse signals resulted in positive numbers, which indicates that the pulse signals contain chaotic properties, such that the behavior depends upon the initial conditions. However, this computation (or computation of fractal dimensions, correlation dimensions) suffers from the curse of dimensionality, and requires rather a long data series. Therefore, we have studied the chaotic behavior through the recurrence plot (RP) based analysis, as RPs make no such demands [122]. We felt that RPs could be useful in the pulse-based analysis, where the dynamics is changing.

### 6.2.1 Recurrence plot

Small disturbances in parameters in a dissipative dynamical system causes exponential divergence of its state, but the system (pulse signal in our case) comes back to a state that is arbitrarily close to a former state, and passes through a similar evolution. This property is called *recurrence* [122].

Recurrence plot (RP) [42] is a graphical tool which helps visualize the recurrent behavior. RPs have found utility in a wide range of scientific explorations, including geophysics [121], dynamics [217], neuronal spike trains [92], electromyographic EMG data [241], and intracranial EEG recordings [164].

The collection of all possible states of a recurrent system is called the phase space of the system. However, for a chaotic dynamical systems, the phase space, and a mathematical description of the system are often unknown. Attractor reconstruction methods have been developed as a means to reconstruct the phase space. One or more signals from the system must be observed as a function of time.

An embedding procedure is commonly used to obtain the phase space from a one-dimensional time series (pulse signal in our case). Using the method of time delays, the points on the attractor are constructed from a time series  $x$  as follows:

$$\vec{x}_i = x(i), x(i + \tau), x(i + 2\tau), \dots, x(i + (m - 1)\tau) \quad (6.11)$$

where  $\vec{x}_i$  are the new vector points,  $x(i)$  is a sampled data point at time  $i$ ,  $\tau$  is a constant time delay, and  $m$  is the embedding dimension. Refer Section 6.3.1.1 for more details on the embedding procedure, to obtain the phase space of a time series.

## 6.2.2 Visualization

The recurrence of the embedded time series can be noted by observing all the situations, where a state at position  $i$  is close (recurrent) to that at a later position  $j$ . In the recurrence plot, such situation is represented by a black dot at  $i$  versus  $j$  location. Therefore, a high-dimensional phase space trajectory is conveniently visualized in the recurrence plot as a matrix of black dots (close recurrent situations) and white dots (apart non-recurrent situations).

Mathematically, whenever the Euclidian distance between these two vectors  $\vec{x}_i$  and  $\vec{x}_j$  is less than a fixed radius  $\varepsilon$ , the two states are said to be close. It is expressed as:

$$R_{i,j} = \Theta(\varepsilon - \|(\vec{x}_i) - (\vec{x}_j)\|); \quad \vec{x}_i, \vec{x}_j \in R^m, \quad i, j = 1, \dots, N \quad (6.12)$$

where  $N$  is the number of states  $\vec{x}_i$  considered for the analysis,  $\varepsilon$  is the radius,  $\| \cdot \|$  is the norm, and  $\Theta$  is the Heavyside function [122, 242].

The radius  $\varepsilon$  provides a cut-off limit (for Heavyside function) that transforms the distance matrix (Euclidian distance) into the recurrence matrix (RM). The points falling within the circle with radius equal to  $\varepsilon$  distance units are the nearest neighbors of (and recurrent with) the reference point.

Proper care is needed in the selection of all the parameters (embedding dimension  $m$ , time delay  $\tau$  and radius  $\varepsilon$ ) to obtain the best results.

- **Time Delay ( $\tau$ ):** Too large time delay results in the loss of information between axes, while very small values make the axes closely related.
- **Embedding Dimension ( $m$ ):** If  $m$  is selected too small, the delayed phase space cannot completely unfold [89] the attractor and thus false nearest neighbors will occur.
- **Radius ( $\epsilon$ ):** As the radius increases, the number of neighbors (and recurrences) increases; some of which might be the false neighbors.

### 6.2.3 Observations

In our experiments, to capture the hidden information of the pulse signals, we performed a careful selection of the parameters, along with grid search for the optimum values.

By sequentially increasing the embedding dimension  $m$  and noting the corresponding percentage of false neighbors, the minimal  $m$  was selected. It ensures that the phase space is fully unfolded in the state space with no crossing of trajectories of the attractor. For various pulse signals in our database, embedding dimension of 3 to 7 was decided. Time delay  $\tau$  of 8 was decided based on the Takens time delay theorem [211]<sup>1</sup>. Finally, radius  $\epsilon$  was selected such that it would not exceed 10% of the mean of the phase space diameter [270], and such that the informative values of DET (refer Section 6.2.4) would remain between 1–5%.

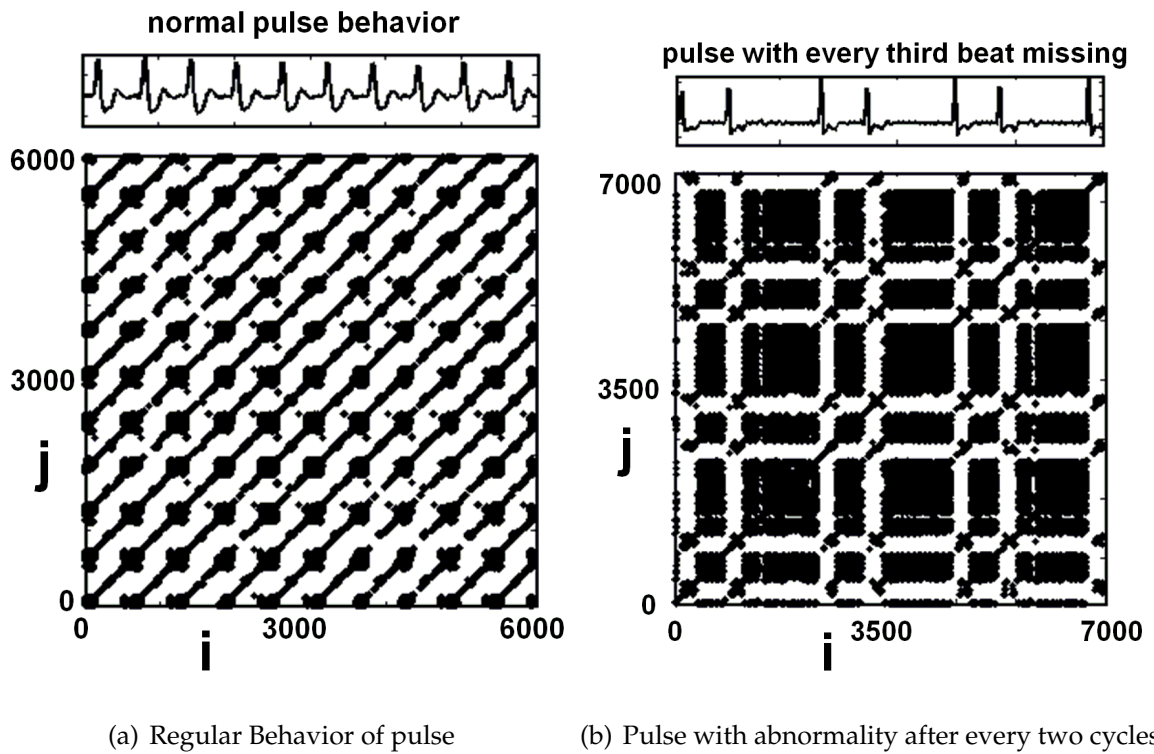
Figure 6.9 displays RPs of two different pulse signals. In each plot, as the same pulse signal is compared to itself,  $R$  is a square, symmetrical matrix.

RPs show different patterns and structural changes in the pulse signals. Both the small-scale and large-scale structures are observed in the plots such as:

- **Single dots:** They represent the states at which the two pulse signals come closer than  $\epsilon$ , but the next states do not.

---

<sup>1</sup>Takens theorem states that in the general case, the dynamics of the system recovered by delay coordinate embedding are the same as the dynamics of the original system.



**Figure 6.9:** Recurrence plots of two pulse signals with different behaviors. We used embedding dimension 5, time delay 8, and radius 0.3. The structures of black points show the recurrence in the pulse signal, and is subsequently captured using RQA descriptors.

- **Diagonals:** They represent the lines parallel to the main diagonal of the square matrix. They represent similar local evolution of different parts of the trajectory. Thus, the main diagonal is always black.
- **Vertical lines:** They represent the states, where the pulse signal on the y-axis evolves locally along the trajectory, but the signal on the x-axis does not change.
- **Horizontal lines:** They represent the states, where the pulse signal on x-axis evolves locally along the trajectory, but the signal on the y-axis does not change.

## 6.2.4 Recurrence Quantification Analysis

To study the structures in the recurrence plot, various descriptors of quantification of RPs were introduced in [56, 57, 242]. They make use of the recurrence point density, and the diagonal structures in the RP. Along with the RP, the recurrence quantification

analysis (RQA) has been successfully applied in exploring cardiac signals [122, 130, 245, 268, 269]. We make use of these descriptors for our pulse database. A feature vector is obtained using RQA descriptors that captures the hidden information.

Important descriptors namely recurrence, determinism, entropy, and laminarity are explained in details below.

- **Recurrence (REC):** REC is the percentage of recurrent points per total triangular area, excluding the central diagonal (which represents the distance between each embedded vector and itself). As mentioned above, single (isolated) points in the RP signify states that do not fluctuate or persist consistently. However, these points do not necessarily imply noise in the data. REC is a count of all such points.

Periodic embedded processes have higher (percent) REC than processes that exhibit aperiodic dynamics. REC ranges from 0% (no recurrent points) to 100% (all points recurrent).

- **Determinism (DET):** DET measures the proportion of recurrent points forming diagonal line (of length  $l$ ) structures. The diagonal length measures how long the time series will be close to each other.

Periodic signals have very long diagonal lines, chaotic signals have short diagonal lines, while stochastic signals do not give any diagonal lines (unless radius is set too high). Diagonal line segments must have a minimum length (usually  $l_{\min} = 2$ ) defined by the line parameter.

- **Entropy (ENT):** ENT is the Shannon information entropy of all diagonal line lengths distributed over integer bins in a histogram. ENT is a measure of signal complexity and is calibrated in units of bits/bin. Individual histogram bin probabilities ( $p(l)$ ) are computed for each non-zero bin and then summed according to Shannon's equation. As the logarithms in the equation are to the base 2, ENT can be interpreted as number of bits.

For simple periodic systems in which all diagonal lines are of equal length,

the ENT would be expected to be 0 bits/bin (narrow diversity in diagonal line lengths), but relatively high within chaotic windows (wide diversity in diagonal line lengths).

- **Laminarity (LAM):** LAM measures the percentage of recurrent points comprising vertical line structures. A vertical/horizontal line marks a length in which a state does not change or changes very slowly, i.e. it represents a state in which the series is “trapped”.  $P(v)$  is the number of vertical lines of length  $v$ .

Table 6.2 lists all the descriptors that we have used with their descriptions and equations used for computation.

For the two cases of pulse with normal (& abnormal) behaviors in Figure 6.9, the RQA descriptors LMAX, REC, ENT, and TRND vary as 0.062 (& 0.116), 96 (& 179), 3.428 (& 2.330) and -0.059 (& -0.034) respectively. Therefore, the numbers show possibility of classification using all the descriptors. The quantification converts the pulse signal into just a 10-dimensional output, resulting in less storage memory as well as effective feature vector for classification.

### 6.3 Topological Structures

Many physical phenomena (such as reactions) and inventions (such as the NMR laser) have been analyzed using concepts from nonlinear dynamics. The chaotic behavior inherent in pseudo-periodic time series data consists of stretching (signals start drifting), and squeezing (signals start coming closer) in a somewhat unpredictable manner.

As we discussed in the previous chapter, the pulse signal follows a P-T-V-D rhythm in an pseudo-periodic fashion as shown in Figure 6.10.

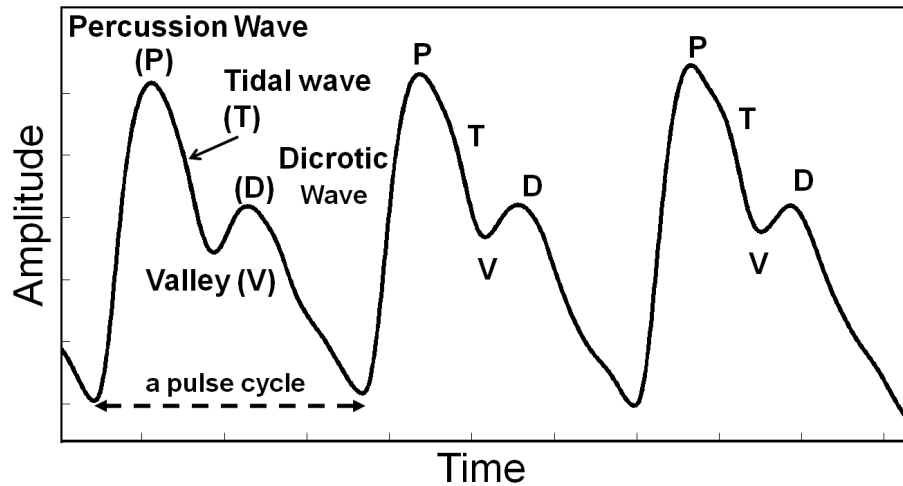
There are two broad approaches to study such chaotic behavior in the pulse signal. First approach is metric-based, where the distances between points in a (strange) attractor are particularly studied, such as Lyapunov exponents [231, 265]. The second

Descriptor	Description	Formula
START	First point for each epoch calculation	
LMAX	The length of the longest diagonal line	$LMAX = \max(l_i; i = 1, \dots, N_l)$
MEAN	Mean of input points	
Entropy, ENT	Shannon's entropy of distribution of the diagonal lines	$ENT = -\sum_{l=l_{min}}^N p(l) \ln(p(l))$
STDEV	Standard deviation of input points	
Trend, TRND	Paling of the recurrence plot towards the edges	$TRND = \frac{\sum_{i=1}^{N-2} [i-(N-2)(RR_i - \langle RR \rangle)]}{\sum_{i=1}^{N-2} [i-(N-2)/2]^2}$
Recurrence, REC	The percentage of points in the recurrence matrix	$REC = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}$
Laminarity, LAM	The percentage of recurrence points that form vertical lines	$LAM = \frac{\sum_{v=v_{min}}^N vP(v)}{\sum_{v=1}^N vP(v)}$
Determinism, DET	The percentage of recurrence points that form diagonal lines	$DET = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{i,j=1}^N R_{i,j}}$
Trapping time, TRAPT	Average length of vertical lines	$TRAP = \frac{\sum_{v=v_{min}}^N vP(v)}{\sum_{v=v_{min}}^N P(v)}$

**Table 6.2:** RQA descriptors with their meaning and the equations for computation.

and newer approach is studying the topological properties. Not much work has been done in both the approaches.





**Figure 6.10:** *Sample pulse with typical behavior. It follows a continuous P-T-V-D rhythm.*

### 6.3.1 Nonlinear Dynamics in Pulse

A century earlier Poincaré had observed that the key to understand a dynamical system, like arterial pulse, lay in identifying its unstable periodic orbits [136]. These periodic orbits are formed in the geometric object created as a result of embedding of the time series in a higher dimension (say 3D), also called as an attractor. It consists of orbits of different pseudo-periods. Topological invariants quantify the topological interaction of these orbits with different periods, and thus are useful in capturing the information about the hidden structures.

The pulse signal, although driven by the heart, is transmitted by blood flow through arteries and travels a considerable distance from the heart. Due to the momentary heart rate, and constant interaction between sympathetic and parasympathetic branches of the autonomic nervous system, nonlinearity [166] is suspected. In Section 6.1 we successfully establish this self-similar chaotic nature.

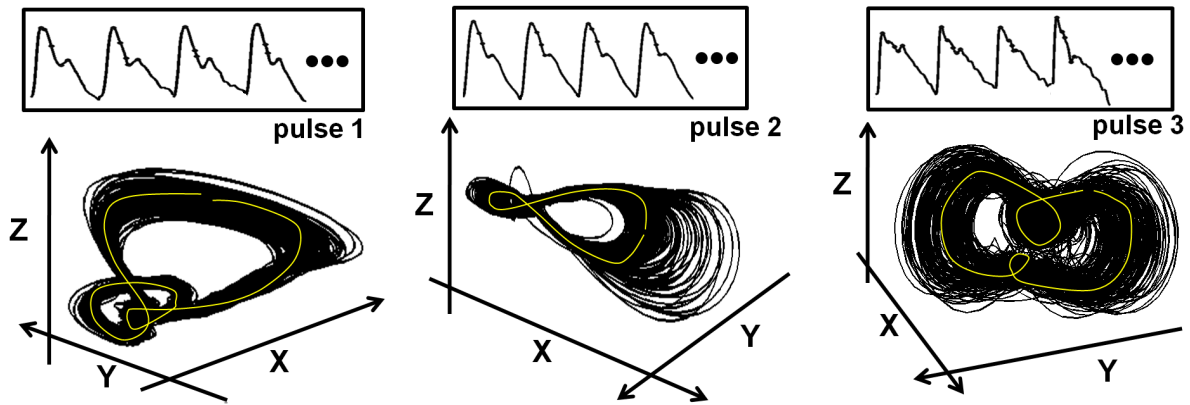
We first computed the Lyapunov exponents [265] and Hausdorff dimensions ( $d_H$ ). The positive Lyapunov exponents (close to 2) indicate the presence of chaotic behavior (strange attractor) [219], and the value of  $d_H$  less than 3 makes the use of topological analysis feasible [136]. However, the geometrical information necessary for computing the topological invariants is not present in the time series format. Subsequent study thus requires embedding, most commonly, in three dimensions.

### 6.3.1.1 Pulse Embedding

Embeddings are the primary tool used by experimentalists to study scalar time series generated by chaotic dynamical systems [220]. There are various ways of embedding a time series into a higher dimensional space, such as time-delay embedding. We have used the *differential phase space* (DPS) embedding to embed a pulse signal  $x(t)$  into a 3D space as  $x(t) \rightarrow (y_1(t), y_2(t), y_3(t))$  where

$$y_1(t) = x(t), \quad y_2(t) = dy_1/dt, \quad y_3(t) = dy_2/dt \quad (6.13)$$

The advantages of using this embedding over the other embeddings will be clear in section 6.3.3.2. Explicitly, differential phase space embedding ensures the existence of a Poincaré section in the resulting attractor [136], and therefore enables easy and trivial computation of the linking numbers of two orbits. Figure 6.11 shows examples of embedding of three different pulse signals.



**Figure 6.11:** *Attractors of three different pulse signals using the DPS embedding. The viewpoint of the 3D rendering is such that the holes in the attractors are clearly visible. Though all the signals have P-T-V-D nature, the attractors show distinctive patterns. We want to capture these patterns quantitatively using topological invariants.*

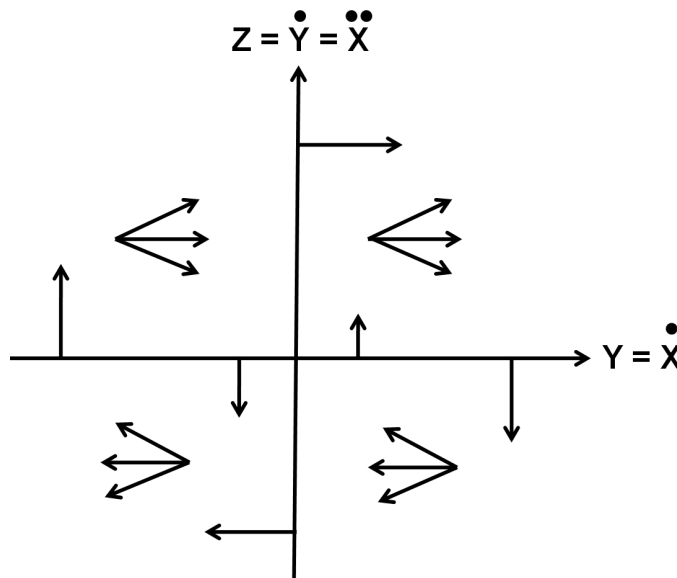
The 3D structure, termed attractor, is a set of points in the phase space to which a dynamical system (the pulse signal in our case) evolves after a long enough time [220]. In Section 6.1, we have already shown that the pulse signal has a fractal nature. Thus, the attractor of the pulse shows strangeness, and dependence on the initial conditions.

The key concept here is that an approximate circle in three dimensional space is traced, possibly after several self intersections.

### 6.3.1.2 Poincaré Section

Poincaré map is the intersection of a periodic orbit in the state space with a certain lower dimensional subspace, called the Poincaré section, transversal to the flow of the system [170]. The intersection defines the period of the orbit (or circle) based on the number of intersections with the section. In this section, we show how DPS embedding makes the further computation of the period of the orbit trivial. Consider a projection of the attractor in Figure 6.12.

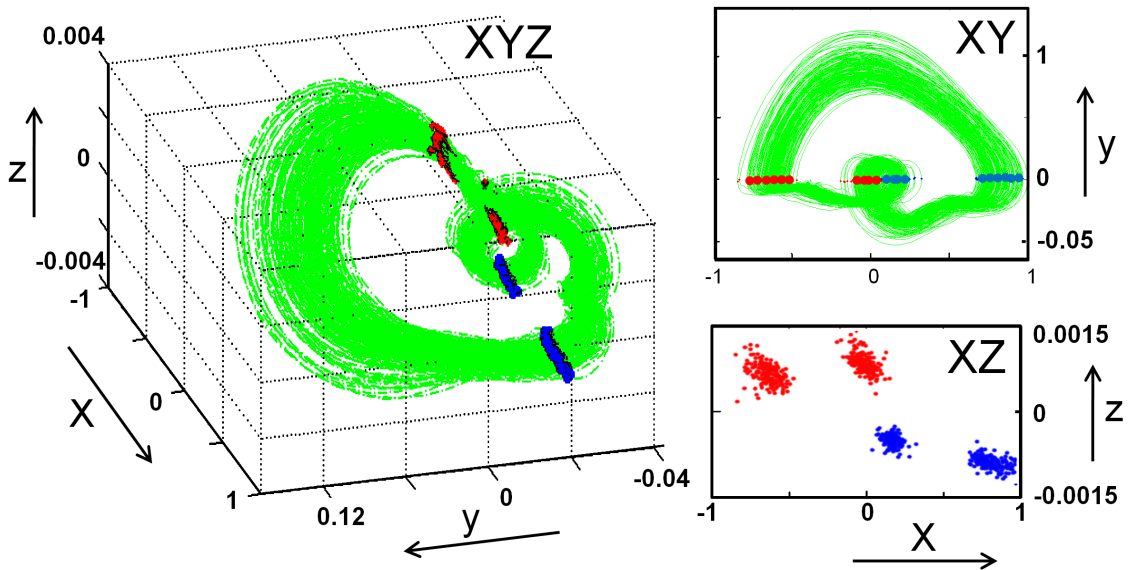
In the first quadrant, on the  $Z$  axis (where  $Y$  is 0),  $dY/dt = Z > 0$ . Hence the flow direction, i.e., the direction with which the orbit is being traced will be towards the right. Similarly for  $Z < 0$  it will be towards the left. Hence the plane consisting of  $X$  axis and  $Y = 0$  can be considered as a Poincaré section enabling a trivial existence of a Poincaré section.



**Figure 6.12:** The projection of the DPS embedding onto the  $Y$ - $Z$  plane clearly shows the flow directions. The plane bounded by  $X$  axis with  $Y = 0$  serves as a Poincaré section.

Figure 6.13 clarifies the Poincaré section for a pulse signal through three view points. We use the  $Y=0$  plane to observe the *going in* and *coming out* properties of the orbits.

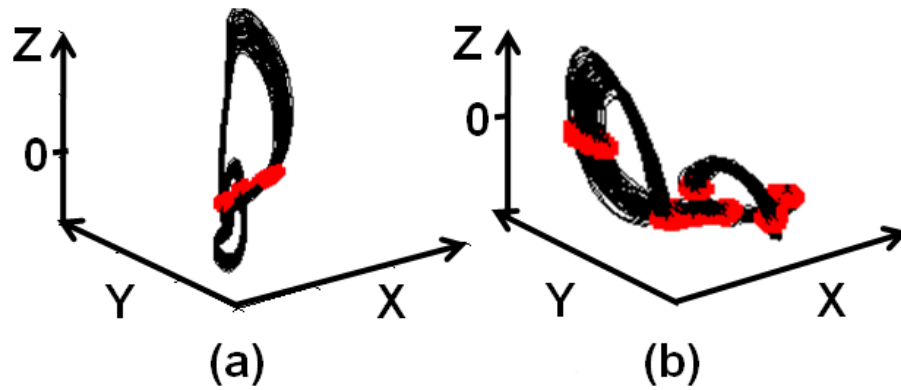
The XZ plane shows four clusters of red points (going in) and blue points (coming out) indicating the pseudo-periodic nature of the pulse signal.



**Figure 6.13:** A sectional cut on the attractor is shown through three views – XYZ, XY and XZ. Red dots indicate orbits going into the plane, while blue dots indicate orbits coming out of the plane.

The Poincaré section helps studying the periodic orbits starting on the subspace flow through it and not parallel to it [61]. More precisely, one considers a pseudo-periodic orbit with initial conditions on the Poincaré section and observes the point at which this orbit first returns to the section. The orbits which comprise the attractor cross the plane many times. We want to study the structures in the attractor, by taking advantage of these intersecting locations on the Poincaré section.

Suppose we use an alternate embedding (Figure 6.14) such as the time-delay embedding. We again highlight (in red) the Poincaré sections which has been manually acquired by laboriously studying the 3D structure using a visualization tool. The resulting pseudo-periodic patterns are not as clear in these two embeddings when compared to Figure 6.13.



**Figure 6.14:** Embedding of a pulse using (a) time-delay embedding with delay 6. (b)  $x(t) \rightarrow (\int x, x, dx/dt)$ . Refer Figure 6.13 for comparison.

### 6.3.2 Determining Period of a Pulse cycle

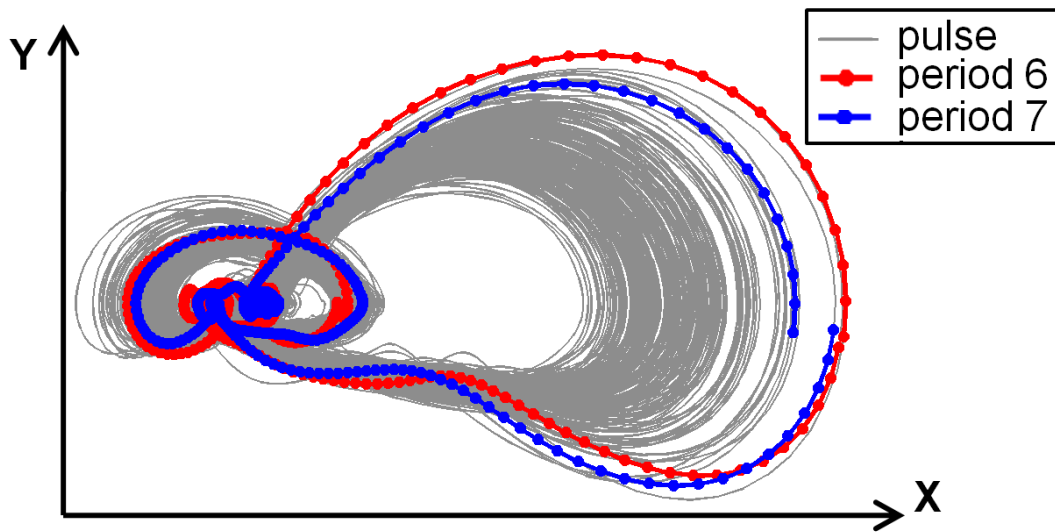
A pulse signal is comprised of many beats one after the other (three such beats are shown in Figure 6.10). As noted earlier, the period of each beat can be trivially computed using the Poincaré section. Specifically, we note the number of times the flow cuts the section before it comes back to the original group of points (one out of four in the XZ plane in Figure 6.13).

However, this is true if the data capture process is not corrupted by noise. For example, we observed that some orbits of signal in the database failed to cut the section (false negative) due to a vibrating behavior around  $Y=0$ . Therefore for accurate computations, we have adopted a slope-based methodology of counting the number of orbits. In summary, the idea is to decide when the curve folds back on itself in the projection based on the Poincaré section. Specifically the algorithm is:

1. We first find the start and the end of each pulse beat using the Poincaré points. This corresponds to the P-subwaves in Figure 6.10 (the extreme right blue points on X-Z plane in Figure 6.13).
2. We start with this first embedded point and follow to the next point in the time series and note the angle with which we need to rotate using the tangent at that point. As we continue through the subsequent points, we note when the flow has rotated through  $180^\circ$ , and then later  $360^\circ$ .

3. One such complete rotation through  $360^\circ$  signifies a loop. The number of loops in a pulse beat is the period of the orbit.

We catch all the loops in each of the pulse cycles using the above algorithm. It is observed that a pulse signal contains pulse cycles (or beats) of varying period (pseudo-period), arising due to inherent nonlinearity and variability of the pulse. Figure 6.15 indicates an example of two different periods (6 & 7) in the attractor of sample pulse signal in Figure 6.10.

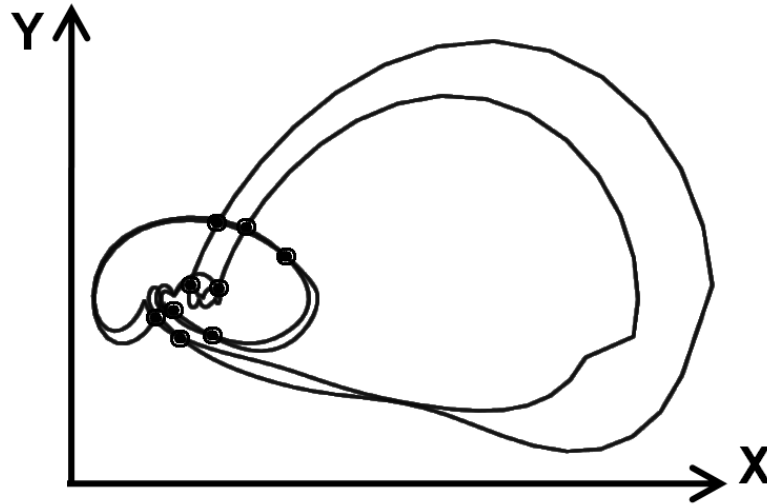


**Figure 6.15:** *Pulse beats with different periods follow different trajectories in the attractor. Two pulse beats with period of 6 and 7 are highlighted.*

We observed that the pulse cycles with different periods follow different trajectories in the attractor. Furthermore, they interact with each other in a particular fashion as shown in Figure 6.16.

### 6.3.3 Topological Invariants

Topological invariants study these interactions in a systematic way to reveal the underlying stretching and squeezing mechanisms [61, 137, 219]. The stretching mechanism causes nearby points in the phase space to diverge from each other and is responsible for “sensitive dependence on initial conditions”. On the other hand, the squeezing mechanism prevents phase space points from escaping from a



**Figure 6.16:** *Intersections between pulse beats of period 3 and 4. Systematic study of them reveals underlying patterns and information.*

compact domain and thus is responsible for the “recurrence” characteristics. Both the mechanisms act to organize the strange attractor in phase space in a unique way. The next goal is to study these interactions of pulse cycles with different pseudo-periods *taken pairwise*.

### 6.3.3.1 Choosing a Pulse Cycle

It is obvious that there will be many pulse beats with the same value of the period. Though the basic structure or the properties of each pulse beat are the same, there are nonlinear variations from one beat to next one, known as pulse rate variability (refer Chapter 5). Thus there is a conundrum of choosing “the” pulse beat from a long pulse signal. We have employed four ways to choose a representative beat (of any fixed period). Similar strategies can be employed for any time series.

1. **Random selection:** If the modeling assumes that each pulse beat contains the basic properties of the input one may obviously choose any one.
2. **Average beat:** Earlier literature [222] has suggested creation of a mean beat from the average of all beats.

3. **Novel minimum distance paradigm:** In this approach, distances of each beat with all the other beats of the same period is computed. The beat with the least sum, which indicates that this beat is the closest to all beats, is chosen.
4. **Novel big orbit paradigm:** In the computation of the period (Section 6.3.2), the number of holes is available. Each hole, however, covers a different area. In this min-max approach reminiscent of the Hausdorff distance, the shortest hole is found in each orbit (of the same period), and then the beat with the largest area (amongst the minima) is chosen.

As a representative example, we now provide our methodology for computing the topological invariants for one sample pulse signal using the ‘minimum distance’ paradigm. In our experiments, the same procedure was repeated considering all the above four ways for all the pulse signals in our database.

### 6.3.3.2 Linking Numbers

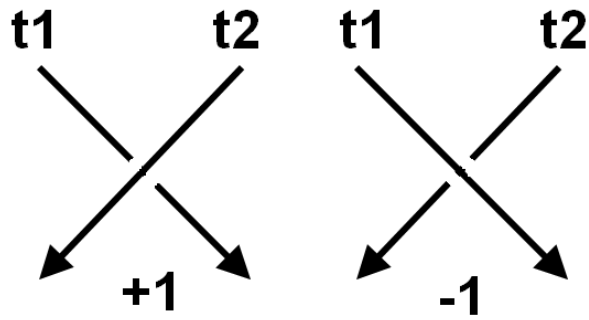
Linking numbers (LN) were introduced in [61] to describe the organization of vortex tubes in ether. LN describe the interaction between two beats quantitatively by counting the crossings between them. For a pair of beats of different periods, LN is the number of times they rotate around each other. If only one beat is considered, then the self-linking number is computed.

Though the beats deform themselves continuously as the initial conditions are changed, their intersections (shown in Figure 6.16) and crossing information (as explained in Section 6.3.3.3) does not vary. Thus the linking information of the unstable periodic orbits is *invariant* and useful. It is always an integer, the sign however depends on the orientation of the flow.

### 6.3.3.3 Computation of LN

The conventional steps for computing the LN of a time series (pulse signal in this case) are [61] as follow. The procedure is repeated for (representative beats of) all pairs of





**Figure 6.17:** *Two possibilities of intersection of the orbits.*

differing periods.

1. Initialize a crossing number ( $CN$ ) variable to 0. Given the periodic beats, project all the beats onto a two dimensional subspace. Find out all the crossing over points.
2. Find out tangent vectors at these points. They will be of one of the formats shown in Figure 6.17.
3. At each of the crossing points, locate the tangent vector with the higher  $z$  coordinate. Let this vector be  $t_1$  and the lower vector be  $t_2$ .
4. Compute  $t_1 \times t_2$ .
5. If this vector has +ve  $z$  co-ordinate then  $CN = CN + 1$  else  $CN = CN - 1$
6. When all the crossing points are over, return  $CN/2$  as the linking number.

**Our approach.** In our experiments, we used an alternate method for steps 3, 4 and 5 above, by taking the advantage of DPS embedding (Section 6.3.1.1). The slope at any point is given by

$$slope = \frac{d\dot{X}}{dX} = \frac{d\dot{X}/dt}{dX/dt} = \frac{Z}{Y} \quad (6.14)$$

Hence,  $Z = slope \times Y$ . Now in the  $Y \geq 0$  upper half plane, the trajectory with greater slope at the crossing point implies that this trajectory is nearer to the observer than the

other. In 3D co-ordinates, it resembles the case of positive  $Z$  co-ordinate in step 5 of the above algorithm. And the opposite situation takes place for  $Y < 0$  lower half plane.

The linking numbers between beats with periods 3,4,5 & 6 (derived using methodology explained in Section 6.3.2) are provided in Figure 6.18.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4 & -2 & -7 \\ 0 & 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 & -7 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

**Figure 6.18:** *LN matrix for periods 3,4,5,6.*

The LN matrix is symmetric. We can also find out the self-linking number of an orbit. It involves calculating the number of times an orbit winds around itself. More precisely it is the sum of the signed crossings of an orbit with itself.

#### 6.3.3.4 Relative Rotation Rates

Relative rotation rates (RRR) were originally introduced [196] to help describe periodically driven two-dimensional systems, such as oscillators. They describe roughly how often one beat rotates around another on the average in a more fine-grained manner as compared to LN.

Consider a dynamical system with beats  $A$  and  $B$  of period  $p_A$  and  $p_B$  respectively. Let  $A$  intersect the Poincaré section  $p_A$  number of times in points  $a_1, a_2, \dots, a_{p_A}$ , while  $B$  intersects it at  $p_B$  points  $b_1, b_2, \dots, b_{p_B}$ . Let  $(a_i, b_j)$  be any pair of points (initial conditions) on the Poincaré section. Let the vector connecting these points be  $\Delta r$ . As the flow is evaluated across time, the vector will come back to its original position (nearby location) and direction on the section after  $p_A \times p_B$  periods. Thus it rotates an

integer number of  $2\Pi$  radians in the plane perpendicular to the flow in  $p_A \times p_B$  periods. RRR ( $R_{ij}$ ) is the average rotation per period during these  $p_A \times p_B$  periods [136]:

$$R_{ij}(A,B) = \frac{1}{2\Pi p_A p_B} \oint \frac{n \cdot (\Delta r \times d(\Delta r))}{\Delta r \cdot \Delta r} \quad (6.15)$$

This integral depends on the initial points  $(a_i, b_j)$  in the Poincaré section. Thus for the beats  $A$  and  $B$ , total  $p_A \times p_B$  relative rotation rates can be defined.

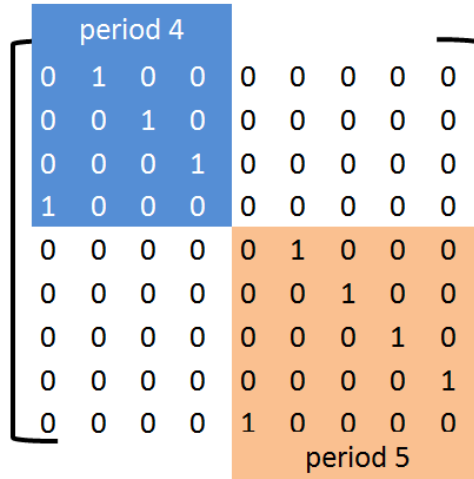
### 6.3.3.5 Computation of RRR

Computation of RRR matrix involves four steps. Consider a case of interactions between beats with period 4 and period 5.

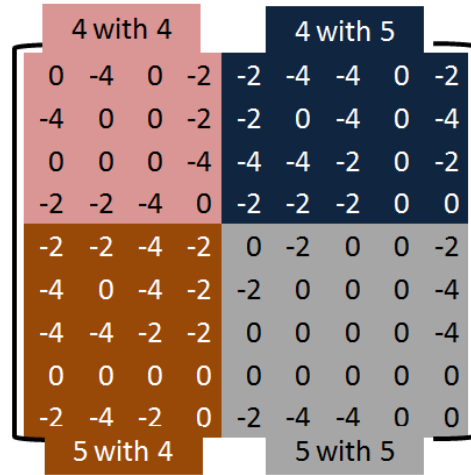
1. **Naming the orbits.** The pulse beats of period 4 and 5 contain four and five orbits respectively. For period 4, the four orbits are named as  $(a, b, c, d)$  in temporal order, assuming the flow direction is anticlockwise. Further, the five orbits for the period 5 beat are named  $(e, f, g, h, i)$ .
2. **Permutation matrix (P).** For each beat, a permutation matrix  $P$  is constructed which depicts how motion occurs along the beat under forward time evolution for the orbits (e.g. for period 4,  $a \rightarrow b \rightarrow c \rightarrow d \rightarrow a$ ).

Therefore if the system is in some orbit  $i$ , we can find out its next orbit by checking out which column in the  $i^{\text{th}}$  row is 1. The direct sum of these matrices gives us how the nine orbits are permuted under forward time evolution. Concatenating these matrices, the final  $P$  is as shown in Figure 6.19. The permutation matrices for beats of periods 4 and 5 are highlighted.

3. **Crossing matrix (C).** The crossing matrix ( $C$ ) is computed by superimposing the two beats (of period 4 and 5). The element  $C(i, j)$  is the signed number of times the beat orbit  $i$  has crossed orbit  $j$ . E.g. if the orbit  $b$  crosses in succession orbits  $c(-1)$ ,  $d(-1)$ ,  $a(-1)$ ,  $a(+1)$ , and  $d(+1)$ , then  $C(b,a) = -1+1 = 0$ ,  $C(b,c) = -1$ ,  $C(b,d) = -1+1 = 0$ , and for the rest orbits  $k$ ,  $C(b,k) = 0$ . The crossing matrix is symmetric as shown in Figure 6.20.



**Figure 6.19:** *Permutation matrix.*



**Figure 6.20:** *Crossing matrix.*

4. Finally we compute the RRR matrix using above matrices. The initial conditions residing on beats  $i$  and  $j$  cross each other  $C(i, j)$  times. After one loop (or flow in an orbit),  $i \rightarrow P_i, j \rightarrow P_j$ , the initial conditions will cross  $(P^{-1}CP)(i, j)$  times. The matrix  $P$  is orthogonal. After  $p_A \times p_B$  iterations, the total number of crossings is the sum of such transformations. Thus we compute a symmetric matrix for relative rotation rates  $RRR$  with periods  $p_A$  and  $p_B$ , for all possible initial conditions [61]:

$$p_A \times p_B \times RRR = \sum_{k=0}^{p_A \times p_B - 1} (P^T)^k C P^k \quad (6.16)$$

RRR for all the pairs of the periodic orbits are computed through the above steps and the final symmetric *RRR* matrix is as shown below.

0.00	-18.27	-12.87	-21.17	-14.62	-16.58	-16.72	-16.32	-16.72	-12.98	-10.45	-4.98	-2.80	-2.40
-18.27	0.00	-18.27	-12.87	-18.05	-16.05	-16.32	-16.72	-16.32	-13.38	-10.05	-5.38	-2.40	-2.80
-12.87	-18.27	0.00	-18.27	-16.98	-17.12	-16.72	-16.32	-16.72	-12.98	-10.45	-4.98	-2.80	-2.40
-21.17	-12.87	-18.27	0.00	-15.15	-16.05	-16.32	-16.72	-16.32	-13.38	-10.05	-5.38	-2.40	-2.80
-14.62	-18.05	-16.98	-15.15	0.00	-13.43	-14.36	-16.09	-13.03	-11.83	-9.56	-4.98	-2.80	-2.40
-16.58	-16.05	-17.12	-16.05	-13.43	0.00	-10.84	-11.71	-18.03	-11.13	-6.01	-4.47	-2.40	-2.80
-16.72	-16.32	-16.72	-16.32	-14.36	-10.84	0.00	-10.04	-9.71	-15.36	-7.79	-3.23	-0.80	-1.60
-16.32	-16.72	-16.32	-16.72	-16.09	-11.71	-10.04	0.00	-10.04	-7.04	-11.09	-4.52	-2.40	-0.80
-16.72	-16.32	-16.72	-16.32	-13.03	-18.03	-9.71	-10.04	0.00	-8.71	-4.91	-5.32	-1.60	-2.40
-12.98	-13.38	-12.98	-13.38	-11.83	-11.13	-15.36	-7.04	-8.71	0.00	-6.04	-1.63	-2.40	-1.60
-10.45	-10.05	-10.45	-10.05	-9.56	-6.01	-7.79	-11.09	-4.91	-6.04	0.00	-3.27	-0.80	-2.40
-4.98	-5.38	-4.98	-5.38	-4.98	-4.47	-3.23	-4.52	-5.32	-1.63	-3.27	0.00	-1.60	-0.80
-2.80	-2.40	-2.80	-2.40	-2.80	-2.40	-0.80	-2.40	-1.60	-2.40	-0.80	-1.60	0.00	-1.60
-2.40	-2.80	-2.40	-2.80	-2.40	-2.80	-1.60	-0.80	-2.40	-1.60	-2.40	-0.80	-1.60	0.00

## Templates

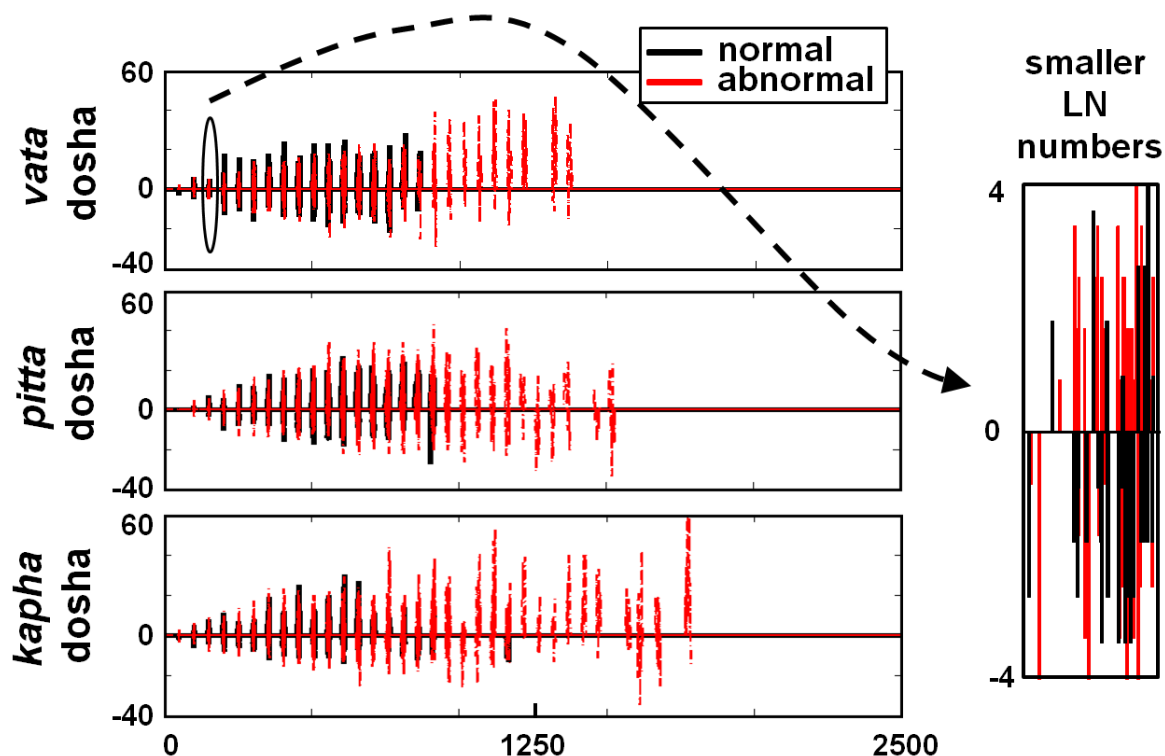
Templates (also known as knot holders or branched manifolds) were constructed in [15] to describe the ensemble of unstable periodic orbits in a strange attractor, as well as the topological organization of those periodic orbits. Templates summarize the stretching and squeezing mechanisms that generate strange attractors. They contain all the crossing information required in order to construct the two previously discussed topological invariants: LN and RRR [16]. Similar to LN and RRR, a template is also invariant under smooth change of variables and changes in control-parameter.

### 6.3.4 Observations

From the LN and RRR matrices, it is clear that different pulse signals will have different periods (LN) and orbits (RRR) (always keeping in mind that these are computed for the representative beat). Furthermore, they will provide different signed intersections (amplitudes in the matrix) between these periods and orbits.

## Classification

Figure 6.21 shows how LN are different for pulse signals with normal (black color) and abnormal (red color) behaviors. We found that the periods (*minimum distance* way) were spanning a wide range (approximately up to 40) for pulse with vibrating nature (e.g. pulse showing fever behavior). Thus, we entered the LN matrix into a  $50 \times 50$  matrix at appropriate locations for each pulse and finally converted the matrix into a single vector of length 2500. These vectors were used as inputs to the classifier.



**Figure 6.21:** Varying LN numbers for pulse signals with normal (black) and abnormal (red) behaviors indicate possibility for classification.

For each pulse signal, we find the representative beats using the four ways (Section 6.3.3.1) including our novel *minimum distance* and *big orbit* approaches. Table 6.3 shows the classification results for pulse signals of healthy and disorder behaviors. We perform the experiments with three sets of features as (i) only linking numbers (ii) only relative rotation rates and (iii) both LN and RRR numbers combined.

Our ways of selecting the representative beats perform well in the classification. Currently, we have not considered other dimensions such as age, gender and so on.

Features	Selection of representative beat			
	Random	Average	Minimum distance	Big orbit
LN	83.4	84.4	87.6	89
RRR	82.6	76.9	88.4	85.8
LN + RRR	80.8	82.3	85.9	87.2

**Table 6.3:** Comparison of binary (healthy-disorder) classification accuracies.

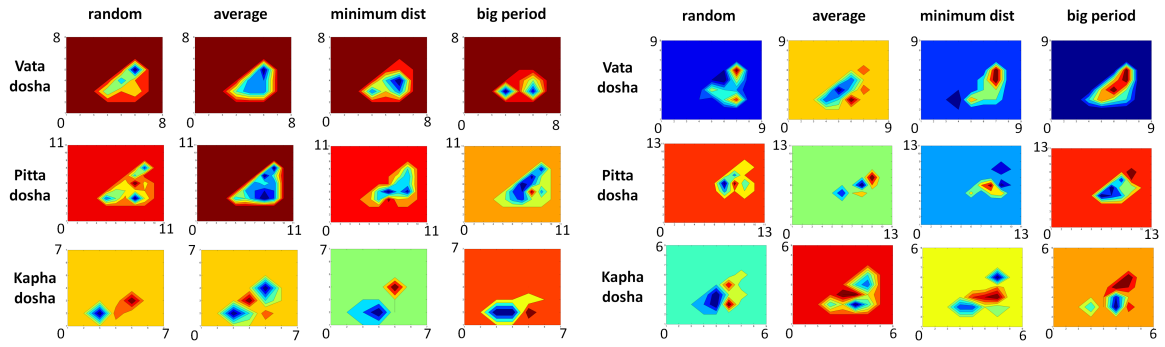
In future, such dimensions could be incorporated with larger dataset for multiclass detection.

### 6.3.5 Visualization

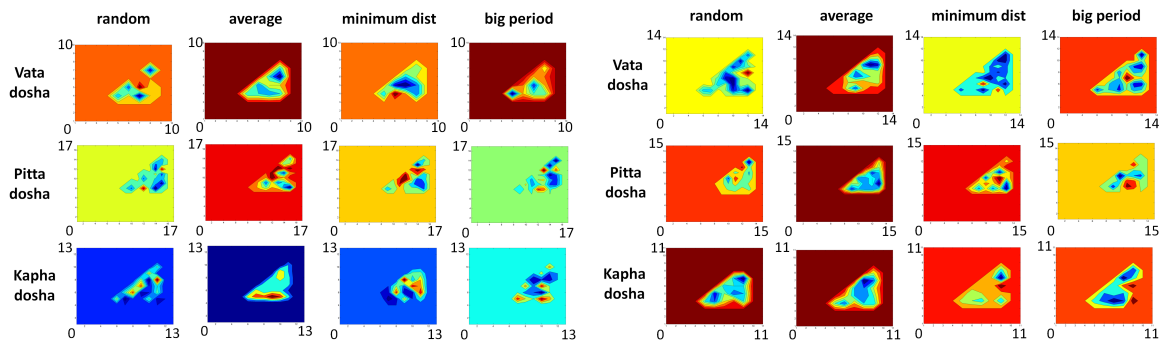
We observe that the LN and RRR matrices show some patterns, which we visualize as contours in the form of images. Figure 6.22 and 6.23 provide the contour diagrams and our observations for four different subjects in our database. The pulse signals number mentioned correspond to our numbering sequence in database and have no special significance.

The contour diagrams of LN and RRR for healthy and disorder behaviors show different patterns. We observed that the LN diagrams for healthy behavior have fewer number of periods (lower base of triangular contours), and their values are also lower. Also, the number of contours and colors are fewer in number indicating more regularity in the intersections between the orbits of different periods. The opposite observations are noted for the disorder behavior with lesser regularity among the beats. Similarly in the case of RRR contour diagrams, we observed lesser number of orbits for healthy behavior. Also, the widths of the contours at the diagonal are smaller as compared to the disorder behavior.

It is known that the sets of LN & RRR numbers for the periodic beats provide a unique signature for the underlying structure of the attractor [219]. We captured such patterns through the contour diagrams of the LN and RRR matrices. The four ways of selecting



(a) Healthy behavior (pulse signals no. 30 & 135). Usually fewer periods with lower values are observed. Also the number of colors and contours are fewer.

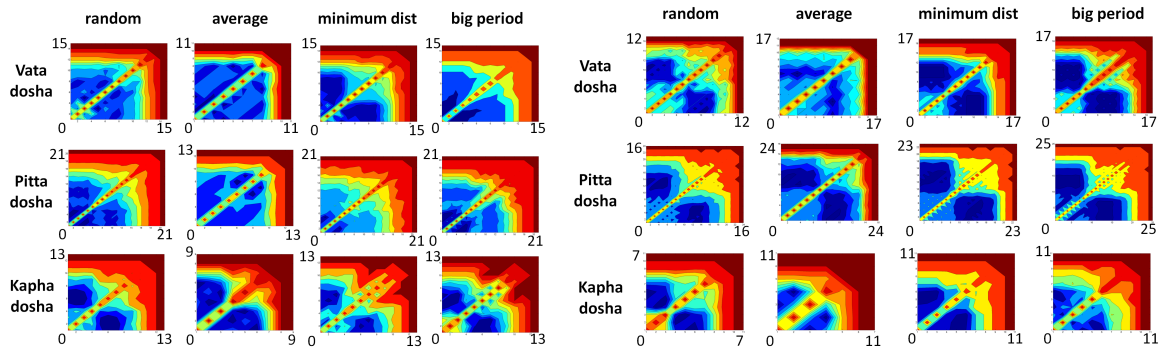


(b) Disorder behavior (pulse signals no. 71 & 86). Usually more periods with higher values are observed. Also the number of colors and contours are higher.

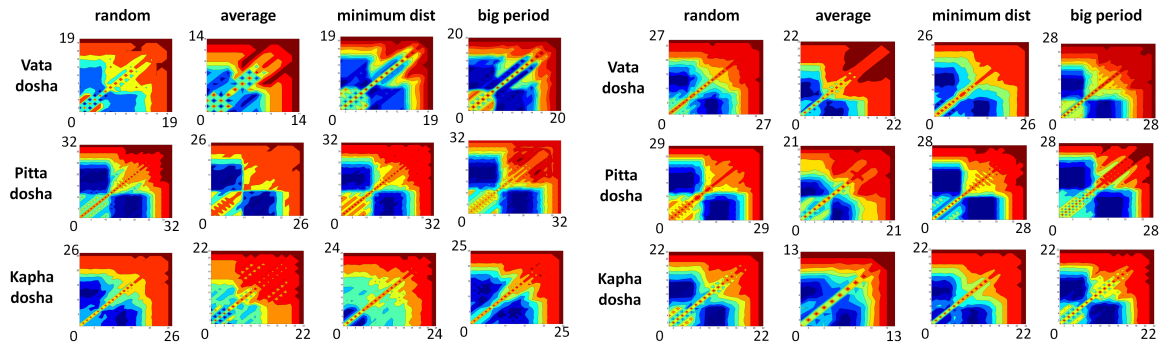
**Figure 6.22:** Contour diagrams indicating LN of pulse signals for respective periods. The colors in the contours indicate the amplitudes of LN, and the background color depends upon the range of the amplitudes.

the representative beat for each period provides different contours. The ‘average beat’ (previously considered in the literature) does not work well – random contours are exhibited in the diagram when the individual beats are vibrating and thus should be avoided. Our suggested ways of selection perform better.





(a) Healthy behavior (pulse signals no. 30 & 135). Usually lesser number of orbits are observed. Also the width of the diagonal is small.



(b) Disorder behavior (pulse signals no. 71 & 86). Usually more orbits are observed. Also the width of the diagonal (and its parallel contours) is large.

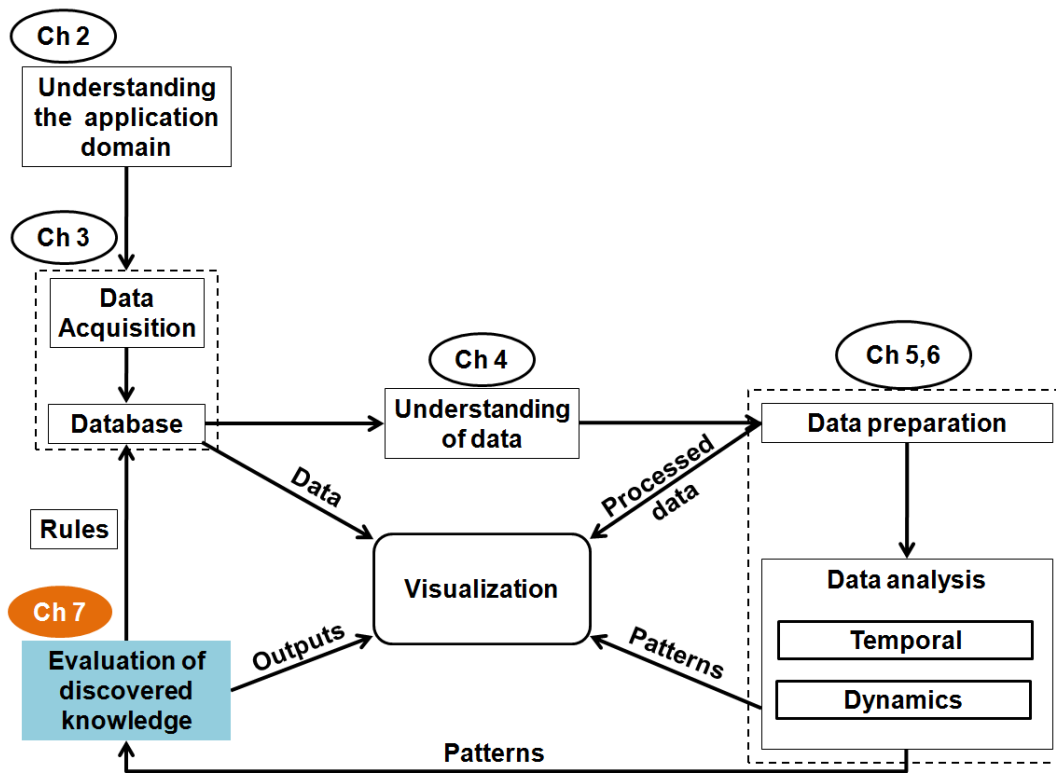
**Figure 6.23:** Contour diagrams indicating RRR of pulse signals for respective orbits. The colors in the contours indicate the amplitudes of RRR. The symmetric RRR matrices form diagonal and parallel contours.

## Conclusions

We showed that the pulse signals has a chaotic nature due to its pseudo-periodic P-T-V-D rhythm. For extracting hidden patterns, we implemented four useful feature sets namely, recurrence plot-based descriptors, local Hölder exponents, multifractal spectrum-based exponents, and topological invariants.

Linear log-log graph of the detrended fluctuation analysis, and the hump-like shape of the multifractal spectrum confirmed – for the first time – that the pulse signals has a self-similar nature. We showed that the combination of wavelet scales in the computation of local Hölder exponents, provides an additional information at different scales (or resolutions). In the computation of linking numbers and relative rotation rates, we suggested novel *minimum distance* paradigm and *big orbit* paradigm. Intuitively the selection has the representative behavior of the entire input signal. We also introduced the contour diagrams for the topological invariants, which help in the visualization of the invariants.

We showed the efficacy of our approaches in the classification. The results in each case were visually displayed as well as numbers were provided whenever possible. There has not been much prior work to extract the nonlinear dynamics of the pulse signals. Therefore, our results show capability in developing a pulse-based automated tool. Also in future, the topological analysis shows a wide scope in detecting various types of the pulse signals using the ‘template’ information.



---

## Chapter 7

# Classification

---

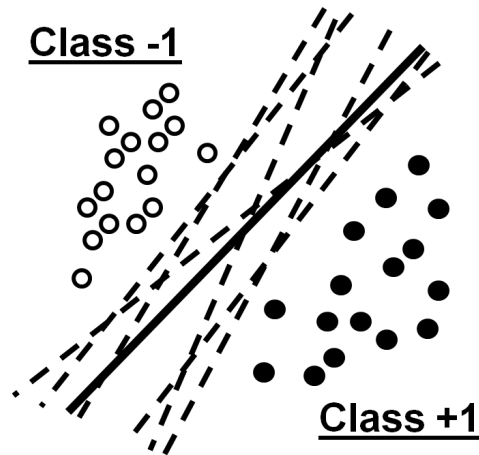
*In Chapters 5 and 6, we have proved the usefulness of the pulse signal through our observations. To further show its efficacy in a quantitative form, in this chapter, we present classification results using our PSWT-based features.*

*We describe our novel classifier using a hybrid arrangement of binary and multiclass SVMs, expressly designed to take care of class imbalance rampant in biomedical signals. As a result, we significantly reduce the number of false negatives – patients falsely classified as normal.*

### 7.1 Support Vector Machine

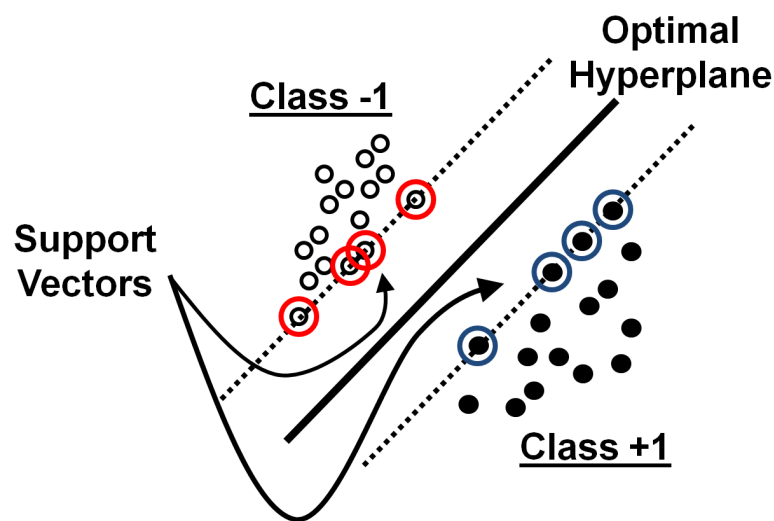
Traditional neural network approaches have suffered difficulties with generalization, thus producing models that can overfit the data. To overcome these drawbacks, Vapnik (1995) has developed support vector machine (SVM) [227], a classifier that works well on unseen examples, and converges to a single globally optimal solution. Because of its advantages and easy implementation, SVM is being extensively used for several classification and regression applications. In this section, we briefly explain classification using SVM.

Consider a 2-class classification model as shown in Figure 7.1. The goal is to produce a classifier that will work well on unseen examples (generalization). There are many possible linear classifiers that can separate the data (dashed hyperplanes), but there is only one that maximizes the margin (bold hyperplane).



**Figure 7.1:** Possible classifying hyperplanes for a two-class problem. The bold line is the best hyperplane exactly at the center from both the classes.

The two classes (+1 and -1) are said to be *optimally separated* by the hyperplane (Figure 7.2), if the distance between the closest data vector in either of the classes to the hyperplane is maximal. The data vectors defining these optimum margins are termed *support vectors* (SV).

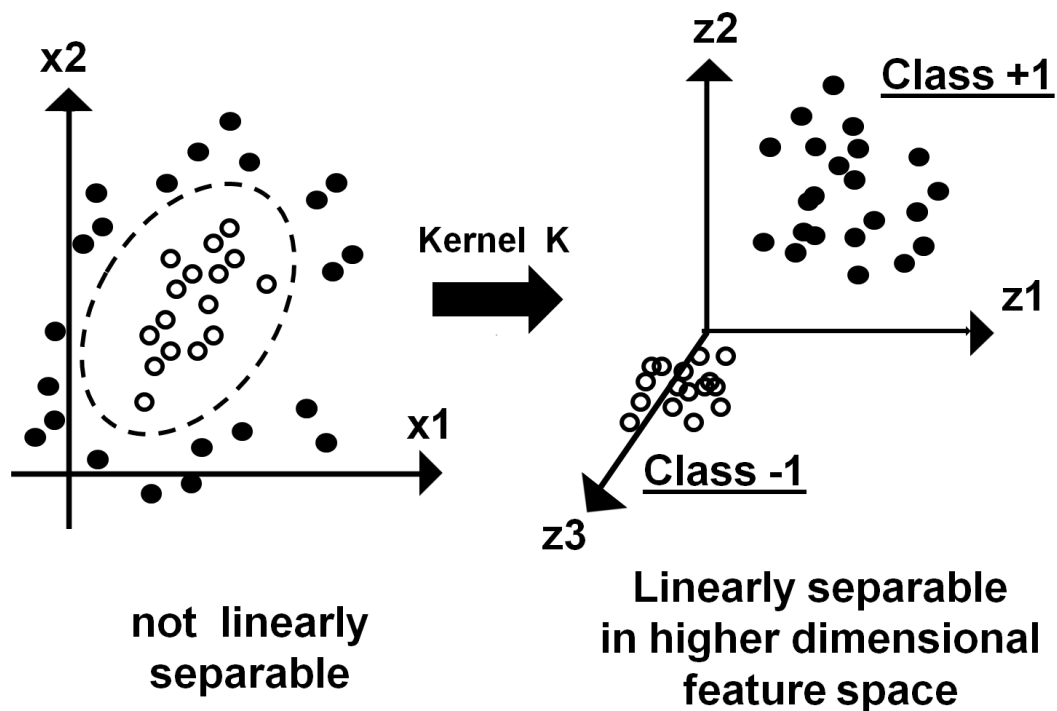


**Figure 7.2:** Optimal hyperplane for two classes with labels +1 & -1. The two dashed lines are the margins and the points on them are the support vectors (SV).

In general, the two classes are not linearly separable. In such situations, the support vectors will no longer be on the margins, but are in fact on both sides of the classifying hyperplane. Also, the orientation of the hyperplane and/or the width of the margin might be different.

A solution to linearly non-separable case is to map the input data vectors into a high dimensional feature space. By choosing a nonlinear mapping a priori, the SVM can construct an optimal separating hyperplane in this higher dimensional space.

For constructing the high dimensional feature space, kernel function  $K$  is conveniently used as shown in Figure 7.3. It performs the nonlinear mapping into feature space with the constraints being unchanged. Among acceptable mappings are polynomials, radial basis functions, and certain sigmoid functions.



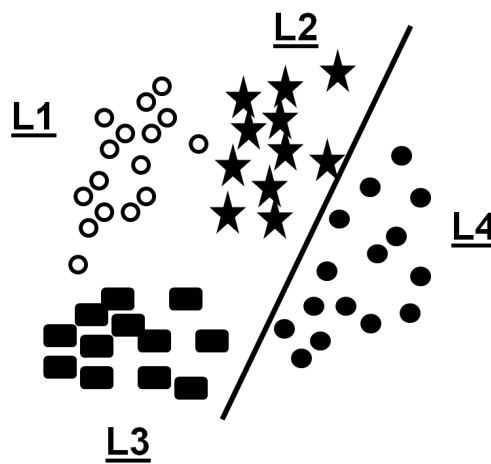
**Figure 7.3:** Kernel action: Transformation of 2D space into 3D feature space

The idea of the kernel function [227] is to enable operations to be performed in the input space rather than the potentially high dimensional feature space. Finally, a classifier implementing the optimal linearly separating hyperplane in the feature space is achieved.

### 7.1.1 Multiclass Support Vector Machine

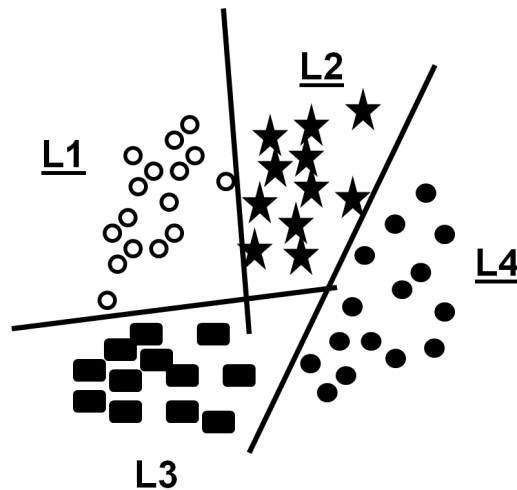
SVM for pattern classification are particularly based on two-class classification problems. As said earlier, the generalization ability is maximized while minimizing the training error. Because of this formulation, extension to multiclass problems is not unique. The  $N$ -class problem is generally solved by considering the problem as a collection of binary classification problems. The most commonly used multiclass SVMs are as follows [76, 98]:

- **One-Against-All approach:** This approach works by constructing a set of  $N$  binary classifiers. The  $i^{th}$  classifier is trained with all of the examples in the  $i^{th}$  class with negative labels (-1) and all other examples with positive labels (+1) as shown in Figure 7.4. The final output is the class that corresponds to the classifier with the highest output value. An important limitation of this approach is that some region in space remains unclassified.



**Figure 7.4:** One-against-all approach for multiclass classification. Here, class with label  $L4$  is separated from all  $L1$ ,  $L2$  and  $L3$ .

- **One-Against-One approach:** This approach simply constructs all possible binary classifiers from a training set of  $N$  classes. Each class is trained on only two out of  $N$  classes as shown in Figure 7.5. Thus, there will be  $N(N-1)/2$  classifiers. Now, In the Max Wins Algorithm, a test example is classified by all of classifiers. Each classifier provides one vote for its preferred class and the majority vote is used to make the final output.

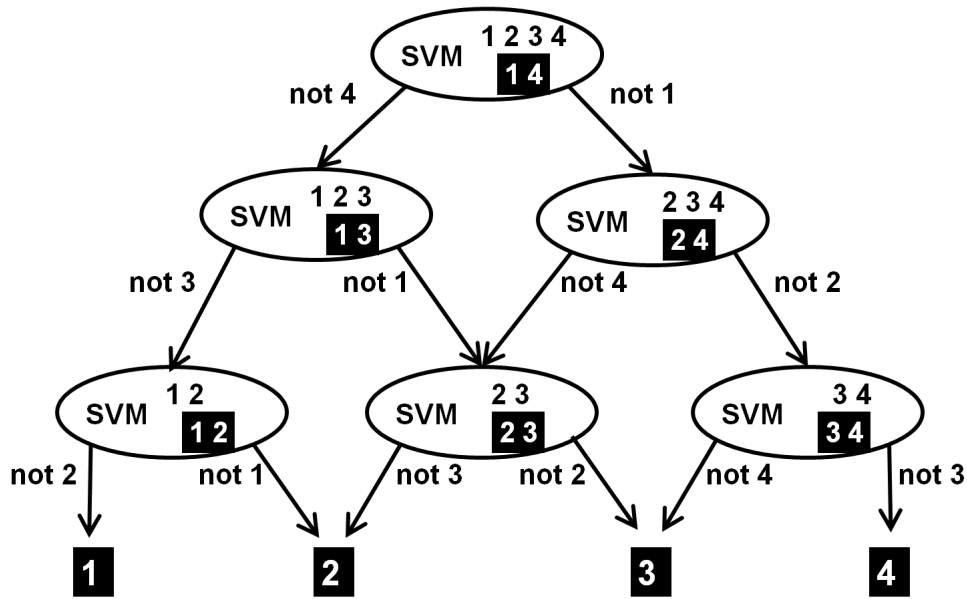


**Figure 7.5:** *One-against-one approach for multiclass classification. All the four classes L1, L2, L3 and L4 are separable (may not be the best hyperplane shown here for better visibility) from each other.*

- Directed Acyclic Graph:** In a problem with  $N$  classes, a rooted binary DAG has  $N$  leaves labeled by the classes where each of the  $N(N-1)/2$  internal nodes is labeled with an element of a boolean function. To evaluate a Decision DAG (DDAG), starting at the root node, the binary function at a node is evaluated. The node is then exited via the Left edge, if the binary function is -1; or the Right edge, if the binary function is +1 as shown in Figure 7.6. The next node's binary function is then evaluated. The value of the decision function is the value associated with the final leaf node [77]. Therefore, we require  $(N-1)$  decision nodes in order to derive the answer.

The DAG-SVM algorithm creates a DDAG whose nodes are maximum margin classifiers over a kernel-induced feature space. Such a DDAG is obtained by training each  $i$ - $j$  node only on the subset of training points labeled by  $i$  or  $j$ . The final class decision is derived by using the DDAG architecture, as described above. Now, the DAG-SVM separates the individual classes with large margin. It is safe to discard the losing class at each 1-v-1 decision because all of the examples of the losing class are far away from the decision surface.





**Figure 7.6:** DDAG approach for classification of 4 classes, which are at the leaves. At each node of SVM binary classification, 2 classes (highlighted bottom numbers of each node) are selected of all the classes available at that node (top numbers of each node).

## 7.2 Novel Hybrid Support Vector Machine

In general we observe that biomedical signals of normal behavior are more common and thus are widely available as compared to signals with abnormalities. Thus, in the multiclass classification of such unbalanced dataset, the classifier gets trained with more bias towards huge *Normal* class. Thus in any diagnostic system, it is critical to reduce the ‘false negatives;’ that is the decisions predicting *Normal* when there is some abnormality present.

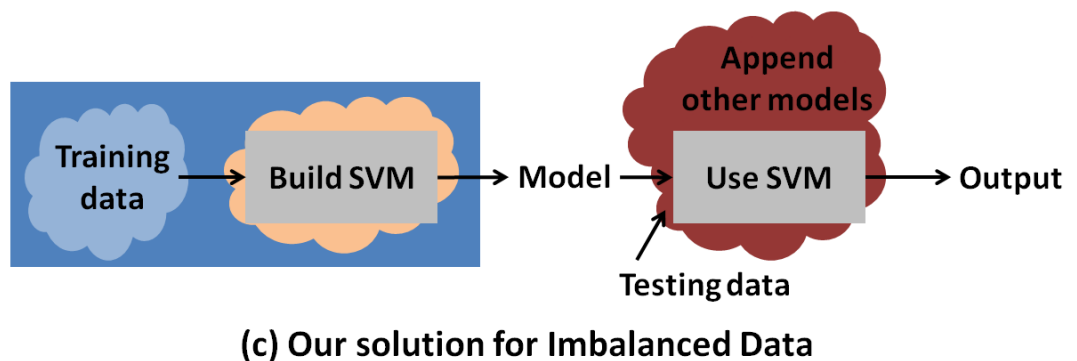
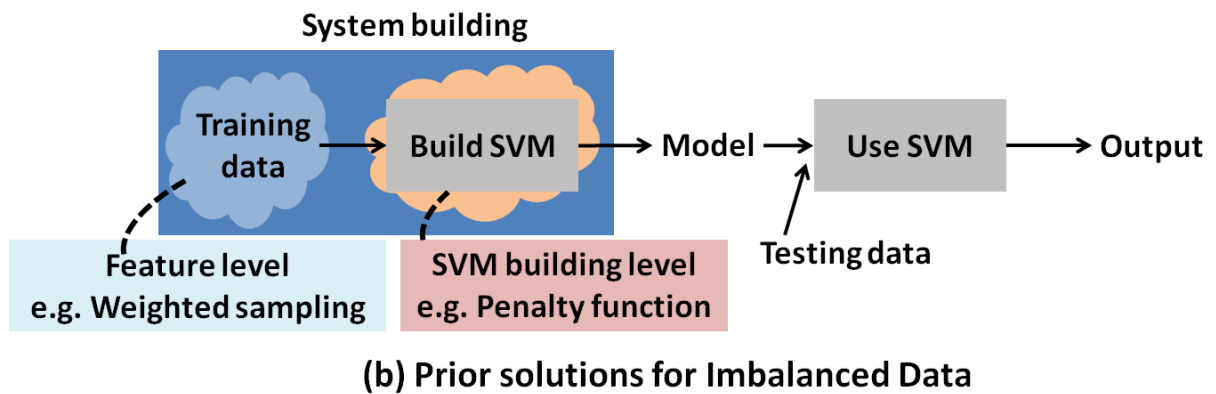
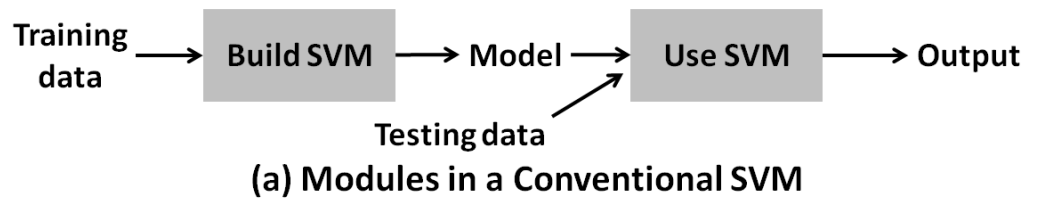
### 7.2.1 Prior Work

The treatment of unbalanced data in the classification process is not new but has been undertaken at different levels in the machine learning process. For example at the feature extraction level, skewed datasets are evaluated in [85] for over- and under-sampling. At the SVM building level, an algorithm is proposed in [32] that changes the kernel function to modify the bias, while the kernel matrix is adjusted in [32] to fit the training data. Different penalty constants for different classes of data can

also be employed [228] making errors on positive instances costlier than errors on negative instances. At the system building level [4, 213], the penalizing method can be combined with SMOTE [22] to develop a classifier that performs better than the use of individual algorithms.

## 7.2.2 Our Approach

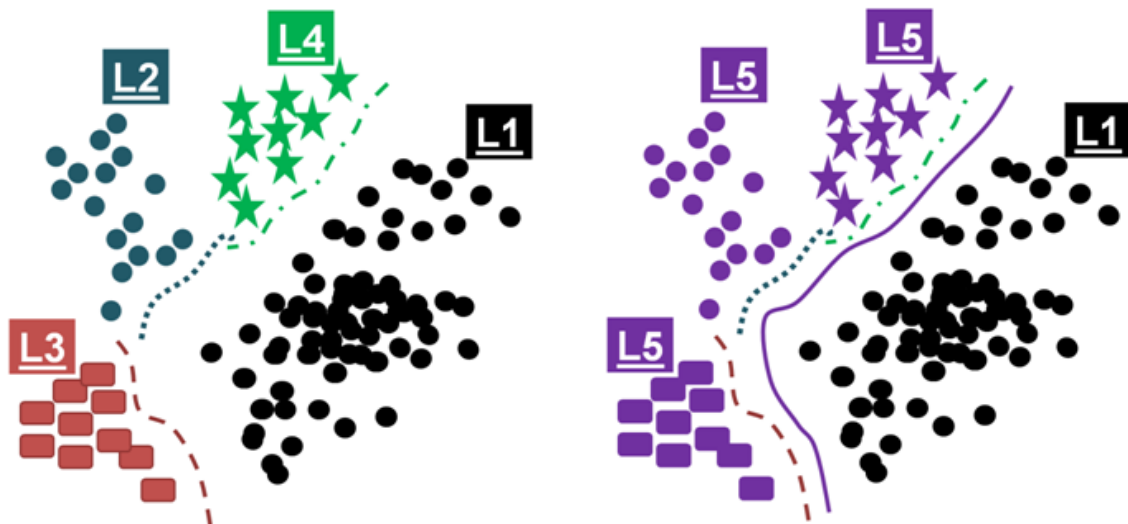
Figure 7.7(a) shows typical blocks in a conventional SVM. The training data is trained to produce a model, which predicts the outputs for the testing data.



**Figure 7.7:** Handling imbalanced data in (a) conventional SVM. Prior methods (b) emphasize the creation of the model while building the SVM. In contrast, our method (c), while compatible with earlier methods, emphasizes the testing phase.

Figure 7.7(b) shows prior solutions at various levels for classifying the imbalanced data. Our solution is shown in Figure 7.7(c), and works independent of, or along with, prior solutions. Thus, the particular advantage of our method is that it is at a “black box” level and is generic enough to work with a variety of SVMs, with or without penalties.

As said earlier, the system usually gets biased toward the larger *Normal* class in unbalanced datasets. Our scheme takes advantage of the point that, when all the abnormal rhythms are combined into one *Abnormal* class, its number becomes comparable to that of the *Normal* class and therefore there is a higher likelihood to get a system trained without bias as explained in Figure 7.8.



(a) The classifying hyperplanes look biased for the huge L1 class. (b) The classifying hyperplane is not biased for binary.

**Figure 7.8:** *Intuition for Hybrid-SVM. (a) A representation of multiclass classification for four classes L1, L2, L3 and L4. L1 is huge as compared to the other three classes, and thus the classifying hyperplanes are biased. (b) When the three smaller classes are combined into one L5 class, the new hyperplane is not biased for L1 class.*

## Hybrid-SVM

Our Hybrid-SVM approach<sup>1</sup> is as explained in Figure 7.9. Consider that we have a Normal class and ' $n$ ' different abnormal classes in the classification system. First, using a multiclass SVM, we train a system with all  $(n+1)$  classes, each class being individually considered. Let the predicted output coming out of this classifier be  $opM$  (here 'M' stands for multiclass). Next, we use a binary SVM to train the *Normal* class versus the *Abnormal* class (considering all  $n$  abnormal classes put together). Let  $opB$  be its predicted output (here 'B' stands for binary). Finally, we use another multiclass SVM to train only the  $n$ -abnormals and let  $opN$  be its predicted output (here 'N' stands for the considered  $n$  abnormalities). *An important point to note is that all the mentioned SVM blocks are trained offline.*

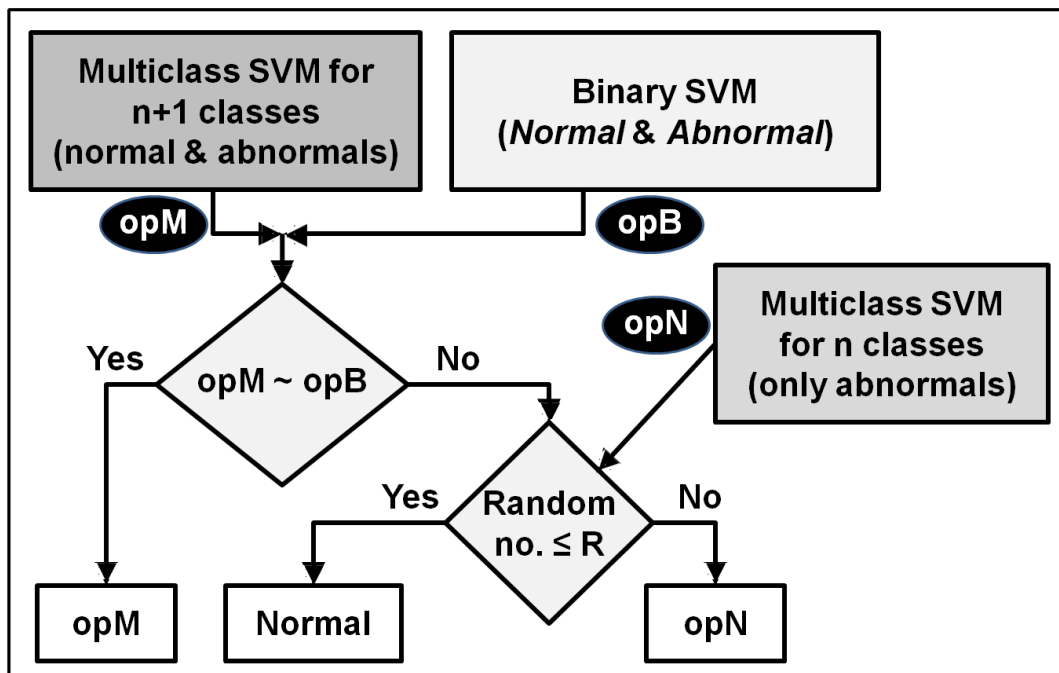


Figure 7.9: Our Hybrid-SVM method. See text for details.

Now, when a test case enters the system, we first compare  $opM$  and  $opB$  coming from the above mentioned two sub-systems.

<sup>1</sup>Our novel Hybrid-SVM approach is different from prior hybrid SVMs, which are hybrid combinations of a support vector machine with a pattern recognition tool. Alternatively, in our novel approach, we have used a hybrid combination of three SVM blocks (one binary-SVM and two multiclass-SVMs) as explained in Figure 7.9, and is unique.

- If the predicted outputs are consistent (that is both are *Normal*, or *opM* declares an abnormality and *opB* declares the *Abnormal* class), then we use *opM* as the final predicted output.
- On the other hand, if they are not consistent, then we speculate that the incompatibility is due to the bias in the trained SVM (conventional multiclass). Therefore, we devote more care to the  $n$  abnormal classes. We choose a random number between  $[0,1]$ . If the random number is below a threshold value ( $R$ ), then we consider the *Normal* class as the final answer; and if the random number is above threshold, then we consider *opN*, declared by the  $n$ -class SVM block, as the final predicted output.

## Threshold Choice

At first glance, it might look a bit strange to use a random number generator. Indeed, if we set  $R = 0$ , as in our initial experiments, then the random number is essentially useless, and we will minimize false negatives by pushing the output to an abnormality. However, this leaves less control for reducing false positives, which is why our Hybrid-SVM has the random number based “knob” for the control. For concreteness, we consider the following three cases in the second decision box in Figure 7.9: same weightage is given to both the *Normal* & *Abnormal* classes ( $R= 0.5$ ), more weightage to the *Normal* class ( $R= 0.9$ ), and more weightage to the *Abnormal* class ( $R= 0$ ).

(In the next section, we use our Hybrid-SVM approach for the unbalanced dataset of pulse signals. Please refer to Appendix B, where we show the efficacy of novel Hybrid-SVM for a variety of feature vectors, and the effect of different values of threshold  $R$  on classification.)

## 7.3 Classification Results

In Chapters 5 and 6, our results in each case were visually displayed, and numbers were provided whenever possible. In Section 6.3.4, we provided the classification

results for the topological invariants-based features. We showed that our ways of selecting the representative beats of each pseudo-period perform well in the classification.

Now we present the results for PSWT-based features. In Section 5.3.4, we showed our approach of computing the useful time domain parameters (TDP) (Section 5.3) of a pulse cycle. We further proposed, and proved the improvement of this computation through our novel PSWT-DTW based approach (Section 5.4). Now, we classify the pulse signals of various subjects, with TDP as input vectors to the classifiers (conventional SVM and our Hybrid-SVM).

The steps in the classification procedure are as follow.

- For each pulse signal, we first detect the peaks (Section 5.1.1). Using these peaks, we segment the pulse signal into individual cycles of different pseudo-periods.
- **Feature set 1 (FS1):** We compute the TDP on all the cycles individually using our maxima-minima based approach (Section 5.3.4). We repeat the steps for all the pulse signals.
- **Feature Set 2 (FS2):** This feature set is obtained using our novel PSWT-DTW approach. The cycles of a pulse signal are arranged in our pitch-synchronous representation (Section 5.2.1.2). Its wavelet transform results in a smooth PSWT cycle (Section 5.2.2). TDP are computed on this trend cycle and are then transferred on each of the pulse cycles using nonlinear dynamic time warping path. We repeat these steps for all the pulse signals.

**Dataset.** The data was generated using the above mentioned steps. For this experiment, the complete dataset includes 1264 pulse cycles of normal (N) behavior, 437 pulse cycles of muscular pain (MP) behavior, 397 pulse cycles of high BP (HBP) behavior, 426 pulse cycles of diabetes (D) behavior, 401 pulse cycles of slipped disk (SD) behavior, 448 pulse cycles of Fits (F) behavior, 427 pulse cycles of skin disorder (S) behavior, 421 pulse cycles of cough (C) behavior, 451 pulse cycles of Fever (FV) behavior, and 315 pulse cycles of other (O) disorders. Note that number of data instances in N is high as compared to the other abnormal labels. Out of the whole

dataset, approximately 66% randomly selected data constituted the training set, and the remaining constituted the testing set.

### 7.3.1 Efficacy of Our Hybrid-SVM

The classification results for FS1 appear in Table 7.1. It displays confusion matrices of both the conventional SVM (*opM* case in Figure 7.9) and the Hybrid-SVM with threshold  $R=0$ .

In a confusion matrix, diagonal elements represent the correct identification of the respective class. We observe that there are very few non-diagonal entries for both the confusion matrices. However, in case of the conventional SVM, there were many misclassification cases for the first row and the first column (which are due to the large number of 'N' class). Our Hybrid-SVM reduces few of these false negatives, by sending them on the right path of the second decision box in Figure 7.9.

Table 7.2 displays the efficacy of our novel Hybrid-SVM approach using the second feature set FS2. Note the improvement in the diagonal entries after applying our Hybrid-SVM approach.

### 7.3.2 Efficacy of Our PSWT-DTW approach

The results in Table 7.2 (TDP using our novel PSWT-DTW approach) appear better to those of Table 7.1 (TDP algorithm on individual pulse cycles). Now, we provide the sensitivity and specificity numbers for all the above cases in Table 7.3.

We observe:

- The sensitivity and specificity numbers indicate that the classification was successfully achieved in all the cases, regardless of the feature set and the nature of the SVM.
- The numbers in the diagonal of the confusion matrix (Table 7.1 and Table 7.2) in the case of the Hybrid-SVM are higher as compared to the corresponding

Testing		Predicted classification									
Data	Classes	N	MP	HBP	D	SD	F	S	C	FV	O
418	N	337 ( <b>337</b> )	10 ( <b>10</b> )	3 ( <b>3</b> )	14 ( <b>14</b> )	12 ( <b>12</b> )	8 ( <b>8</b> )	6 ( <b>6</b> )	11 ( <b>11</b> )	13 ( <b>13</b> )	4 ( <b>4</b> )
142	MP	24 ( <b>15</b> )	101 ( <b>107</b> )	0 ( <b>2</b> )	4 ( <b>4</b> )	2 ( <b>2</b> )	1 ( <b>2</b> )	4 ( <b>4</b> )	4 ( <b>4</b> )	2 ( <b>2</b> )	0 ( <b>0</b> )
145	HBP	17 ( <b>13</b> )	1 ( <b>1</b> )	80 ( <b>83</b> )	5 ( <b>5</b> )	8 ( <b>8</b> )	16 ( <b>17</b> )	12 ( <b>12</b> )	0 ( <b>0</b> )	5 ( <b>5</b> )	1 ( <b>1</b> )
150	D	23 ( <b>11</b> )	13 ( <b>13</b> )	5 ( <b>6</b> )	81 ( <b>87</b> )	5 ( <b>6</b> )	1 ( <b>1</b> )	7 ( <b>10</b> )	4 ( <b>4</b> )	4 ( <b>4</b> )	7 ( <b>8</b> )
141	SD	17 ( <b>9</b> )	4 ( <b>4</b> )	2 ( <b>2</b> )	12 ( <b>12</b> )	60 ( <b>65</b> )	2 ( <b>2</b> )	13 ( <b>15</b> )	24 ( <b>24</b> )	6 ( <b>7</b> )	1 ( <b>1</b> )
141	F	23 ( <b>15</b> )	1 ( <b>2</b> )	2 ( <b>5</b> )	2 ( <b>2</b> )	3 ( <b>3</b> )	107 ( <b>109</b> )	1 ( <b>1</b> )	0 ( <b>0</b> )	2 ( <b>4</b> )	0 ( <b>0</b> )
128	S	13 ( <b>5</b> )	6 ( <b>6</b> )	0 ( <b>2</b> )	9 ( <b>11</b> )	4 ( <b>4</b> )	10 ( <b>10</b> )	70 ( <b>74</b> )	10 ( <b>10</b> )	3 ( <b>3</b> )	3 ( <b>3</b> )
129	C	16 ( <b>10</b> )	3 ( <b>4</b> )	1 ( <b>1</b> )	3 ( <b>4</b> )	13 ( <b>14</b> )	0 ( <b>0</b> )	8 ( <b>9</b> )	83 ( <b>85</b> )	2 ( <b>2</b> )	0 ( <b>0</b> )
125	FV	21 ( <b>10</b> )	0 ( <b>1</b> )	4 ( <b>5</b> )	6 ( <b>9</b> )	6 ( <b>8</b> )	0 ( <b>0</b> )	6 ( <b>6</b> )	8 ( <b>8</b> )	70 ( <b>73</b> )	4 ( <b>5</b> )
102	O	14 ( <b>4</b> )	1 ( <b>2</b> )	1 ( <b>2</b> )	15 ( <b>18</b> )	2 ( <b>2</b> )	0 ( <b>1</b> )	1 ( <b>2</b> )	0 ( <b>0</b> )	1 ( <b>1</b> )	67 ( <b>70</b> )

**Table 7.1:** Confusion matrix for the multiclass classification using feature set FS1. Each entry corresponds to the conventional case (*opM*), and the Hybrid-SVM ( $R=0$ ) case (*in bold*).



Testing		Predicted classification												
Data	Classes	N	MP	HBP	D	SD	F	S	C	FV	O			
418	N	405 ( <b>405</b> )	5 ( <b>5</b> )	0 ( <b>0</b> )	2 ( <b>2</b> )	1 ( <b>1</b> )	3 ( <b>3</b> )	1 ( <b>1</b> )	0 ( <b>0</b> )	0 ( <b>0</b> )	1 ( <b>1</b> )			
142	MP	3 ( <b>1</b> )	135 ( <b>137</b> )	0 ( <b>0</b> )	1 ( <b>1</b> )	0 ( <b>0</b> )	0 ( <b>0</b> )	3 ( <b>3</b> )	0 ( <b>0</b> )	0 ( <b>0</b> )	0 ( <b>0</b> )			
145	HBP	1 ( <b>0</b> )	2 ( <b>2</b> )	132 ( <b>132</b> )	0 ( <b>0</b> )	0 ( <b>0</b> )	3 ( <b>3</b> )	6 ( <b>7</b> )	0 ( <b>0</b> )	1 ( <b>1</b> )	0 ( <b>0</b> )			
150	D	7 ( <b>3</b> )	2 ( <b>2</b> )	1 ( <b>1</b> )	127 ( <b>130</b> )	3 ( <b>3</b> )	5 ( <b>5</b> )	0 ( <b>0</b> )	2 ( <b>2</b> )	3 ( <b>4</b> )	0 ( <b>0</b> )			
141	SD	0 ( <b>0</b> )	1 ( <b>1</b> )	0 ( <b>0</b> )	7 ( <b>7</b> )	125 ( <b>125</b> )	2 ( <b>2</b> )	0 ( <b>0</b> )	5 ( <b>5</b> )	1 ( <b>1</b> )	0 ( <b>0</b> )			
141	F	8 ( <b>6</b> )	2 ( <b>2</b> )	0 ( <b>0</b> )	3 ( <b>3</b> )	0 ( <b>0</b> )	126 ( <b>128</b> )	2 ( <b>2</b> )	0 ( <b>0</b> )	0 ( <b>0</b> )	0 ( <b>0</b> )			
128	S	7 ( <b>5</b> )	0 ( <b>0</b> )	3 ( <b>3</b> )	0 ( <b>1</b> )	2 ( <b>2</b> )	6 ( <b>6</b> )	110 ( <b>110</b> )	0 ( <b>0</b> )	0 ( <b>0</b> )	0 ( <b>1</b> )			
129	C	3 ( <b>0</b> )	0 ( <b>0</b> )	0 ( <b>0</b> )	0 ( <b>0</b> )	3 ( <b>3</b> )	0 ( <b>0</b> )	1 ( <b>1</b> )	121 ( <b>122</b> )	1 ( <b>1</b> )	0 ( <b>2</b> )			
125	FV	3 ( <b>2</b> )	2 ( <b>2</b> )	1 ( <b>1</b> )	3 ( <b>3</b> )	1 ( <b>1</b> )	0 ( <b>0</b> )	1 ( <b>1</b> )	1 ( <b>1</b> )	113 ( <b>114</b> )	0 ( <b>0</b> )			
102	O	1 ( <b>0</b> )	0 ( <b>0</b> )	0 ( <b>0</b> )	2 ( <b>3</b> )	0 ( <b>0</b> )	1 ( <b>1</b> )	1 ( <b>1</b> )	2 ( <b>2</b> )	0 ( <b>0</b> )	95 ( <b>95</b> )			

**Table 7.2:** Confusion matrix for the multiclass classification using feature set FS2. Each entry corresponds to the conventional case (*opM*), and the Hybrid-SVM (*R=0*) case (in bold).

Classes	Sensitivity				Specificity			
	Conventional SVM		Hybrid-SVM		Conventional SVM		Hybrid-SVM	
	FS1	FS2	FS1	FS2	FS1	FS2	FS1	FS2
N	66.73	92.47	78.55	95.97	92.74	98.90	93.20	98.92
MP	72.14	90.60	71.33	90.73	97.23	99.52	97.62	99.66
HBP	81.63	96.35	74.77	96.35	95.73	99.12	95.89	99.12
D	53.64	87.59	52.41	86.67	95.31	98.44	95.67	98.64
SD	52.17	92.59	52.42	92.59	94.62	98.92	94.92	98.92
F	73.79	86.30	72.67	86.49	97.70	98.98	97.82	99.12
S	54.69	88.00	53.24	87.30	96.12	98.80	96.36	98.80
C	57.64	92.37	58.22	92.42	96.89	99.46	97.02	99.53
FV	64.81	94.96	64.04	94.21	96.36	99.20	96.55	99.27
O	77.01	98.96	76.09	95.96	97.72	99.54	97.91	99.54

**Table 7.3:** Comparison of the classification performances of our novel approaches with the conventional approach.

conventional SVM numbers. This proves the efficacy of our Hybrid-SVM approach in reducing false negatives.

- If we look at the sensitivity numbers in Table 7.3, we note the Hybrid-SVM are sometimes lower or comparable to the corresponding conventional SVM numbers. We clarify that we do not see this as a serious problem because the purpose of the Hybrid-SVM was to reduce false negatives. If a disorder is present, then that subject should be treated. Next, the reason for this is as follows. Consider the *MP–HBP* combination in Table 7.1. In conventional SVM, this number was zero; but in Hybrid-SVM, it becomes two due to the application of the “NO” path in the first decision box of Figure 7.9. This means that, the false negatives are reduced at the expense of increase in false positives (zero becomes two). The lower numbers in the table are due to such false positive numbers.
- The numbers for FS2 are much better as compared to those of FS1. This proves that our novel PSWT-DTW approach helps (Section 5.4.3.2) for two reasons. First, the resulting trend cycle is smooth in nature and second, the efforts (if any) of tuning the TDP are only for one PSWT cycle.

Our investigations show the immense potential benefit of both the novel PSWT-DTW and Hybrid-SVM approaches respectively for the pseudo-periodic signals, and the imbalanced datasets. We believe the two methods can be suitably extended to other biomedical as well as non-biomedical applications.

---

## Chapter 8

# Conclusions

---

*In Chapter 3, we presented Nadi Tarangini a system that acquires noise-free pulse signals. After understanding the nature of pulse signals in Chapter 4, we systematically studied the pseudo-periodic nature, and nonlinear dynamics of pulse signals in Chapters 5 and 6 respectively.*

*In this chapter, we describe applications and future directions of our work. Finally, with the pictorial summary of our approaches, we conclude the thesis on a positive note that we have successfully established the methodology of acquisition and quantitative analysis of pulse signals.*

### 8.1 Applications

Our system will work as a computer-aided diagnostic system to offer a second opinion to the doctors. When a patient comes, his or her *nadi* will be recorded using *Nadi Tarangini*, and based on the outputs of our approaches, a decision will be made about his or her physiological conditions. Our system can never replace a doctor, but will assist in many ways. Furthermore, *Nadi Tarangini* will be helpful to the society in various other ways including:

- *Nadi Tarangini* will particularly help when a doctor is not available, or distantly available. In such situations, professional nurse can record the pulse signal, and after ‘understanding’ the output of our approaches, a required action could be taken.
- Our system is portable and inexpensive. A common man can use it at home, to check the pulse (and thus physiological conditions) on frequent basis. He or she can observe whether the important parameters (outputs) are in the required range or not. Thus, this could help to detect deviations, and work as a watchdog for the continuous monitoring.
- *Nadi Tarangini* can further be used to reproduce, compress and then transmit the measured pulse to a student or *Ayurvedic* practitioner for diagnosis.
- Just as simulators are used to train pilots learning to fly, *Nadi Tarangini* can be used for training students or practitioners, which is to be mastered, otherwise, with actual pulse feeling with concentration and experience.
  - ◇ The *nadi* or pulse signal changes from person to person, for different age-groups, gender, habits, time of acquisition and so on. But with an assumption of retaining the basic structure in such disorders, different types of reference-pulse-waveforms can be visualized using *Nadi Tarangini*.
  - ◇ In addition, various frequency, amplitude, slope and rhythm features will be visibly available to understand and learn the patterns in various pulse types, similar to a reference guide.
  - ◇ For higher studies, the accuracy of the student’s pulse determinations could be verified using *Nadi Tarangini*, which will help in their learning process.
- Pulse-based diagnosis when developed in its finest form will help in reducing the cost of detecting the disorders (and may be treatments), which involve different scans and biopsies. We need to prove in future – the point in literature that the *Ayurvedic* pulse has all the important information.

## 8.2 Limitations and Future Work

In our work, we have established the efficacy of pulse signals through results of various approaches. However, the domain of the pulse-based methodology (especially in the traditional medicine) is very huge. The enormous hidden information in pulse signals and its complexity itself, though important, is the largest limitation in the analysis of the pulse signal. Very high attention is required to develop the pulse-based analysis, similar to current status of the advanced automated ECG tools.

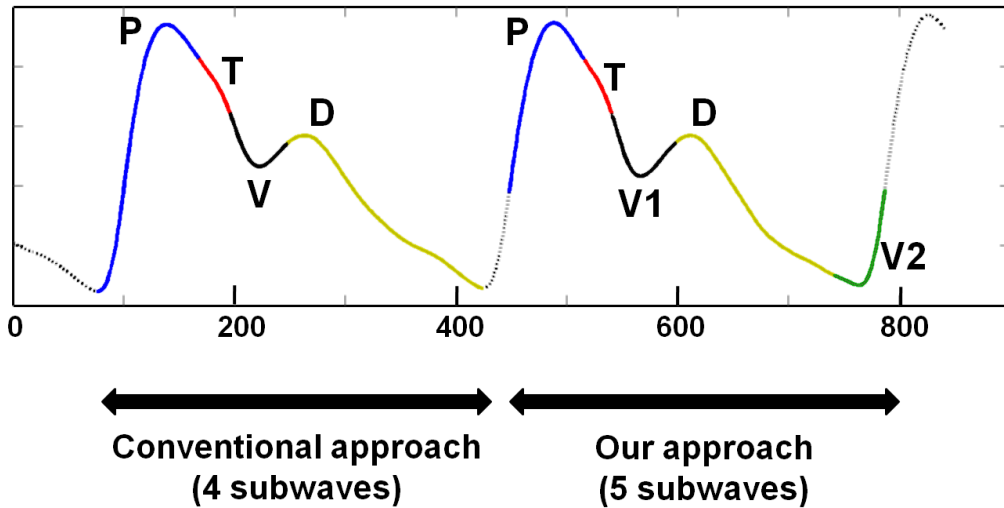
There are still a number of ways in which the pulse signals can be further studied to understand the hidden dynamics. We are currently examining the pulse signals through its segmentation using hidden Markov models (HMM), and its symbolic formation (with the known fact that sometimes symbolic representation captures the dynamics which are invisible to the usual time series format). We provide some preliminary observations and results here.

### 1. Segmentation of the Pulse Using Hidden Markov Model

As already explained in Figure 4.18, a typical behavior of an arterial pulse follows a P-T-V-D rhythm. When such underlying behavior of a given signal is known (or assumed to be known), an efficient hidden Markov model (HMM) [173] can be formed.

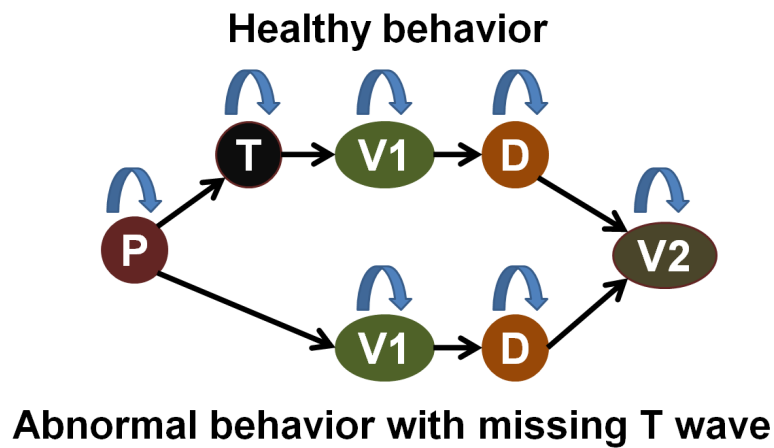
HMM is a statistical model in which a state (e.g. P-subwave) is not directly visible, but the variables influenced by the state are visible. Each state has a probability distribution over the possible output tokens (the pulse signal numbers). Also, the transitions among the states are governed by a set of probabilities called transition probabilities [28, 224].

After observing the variations of different pulse signals through HMM, we have introduced a notion of valley-2 (V2) on the pulse cycle. Thus, each pulse cycle is now divided into five (instead of conventional four) segments as shown in Figure 8.1. The new set of subwaves (as compared to Section 4.2.2) are percussion subwave (P), tidal subwave (T), valley-1 (V1), dicrotic subwave (D), and valley-2 (V2).



**Figure 8.1:** Our novel approach of splitting a pulse cycle into five subwaves (instead of conventional four subwaves) for capturing the variations at the V2 location using HMM segmentation.

The sequence of tokens generated by an HMM gives some information about the sequence of states as shown in Figure 8.2. The extracted model parameters can subsequently be used for classification.



**Figure 8.2:** The top sequence shows HMM model for normal behavior, modeled using 5-state HMM, while the bottom sequence shows that for a disorder with missing T-subwave, modeled using 4-state HMM.

The advantage of using HMMs is that they can model the P-T-V1-D-V2 behavior, by sequentially taking into consideration the intra-individual variability of the pseudo-period. Furthermore, It is possible to carry out, at the same, time segmentation and classification [9].

## 2. Extraction of Short Patterns Using Symbolic Dynamics

In this work, we focus on dynamic features of short API signals using symbolic dynamics (SD) based analysis [244]. Our intuition is that the analysis will take into account the *short patterns* distributed in the API symbolic signals, and will be appropriate for studying the instabilities in various disorders. We carry out the quantitative analysis of the nonlinear behavior of the symbolic API signal using different measures, namely 'plvars' & 'phvars' [246], forbidden words [246] and Shannon & Renyi entropies [33]. We hope that they will quantify different kinds of regularity (or order) of the signals.

### Future Work

We now elaborate the future work in the acquisition and analysis of the pulse signals.

- **Tactile sensors:** As far as a system with three-sensors (*Tridosha* locations) is concerned, then *Nadi Tarangini* provides noise-free signals with high details. *Ayurvedic* practitioners believe that the pulse is available at many minute locations on each fingertip. Though these locations (on each fingertip) will contain almost the same information, the practitioners are able to sense the differences with high concentration.

Currently our system is limited to just one location (center) for each fingertip, which is a good approximation to start with. But to completely mimic the *Ayurvedic* methodology, we will require a data acquisition system to capture the pulse signals at many more locations.

A *tactile-sensing array* can be used for this purpose, which is a coordinated group of touch sensors. It involves the detection and measurement of the spatial distribution of forces perpendicular to a predetermined sensory area, and the subsequent interpretation of the spatial information. When this tactile sensor will be exposed to wrist region (where the pulse is available), it will help to acquire the pulse signals from many locations.



- **Massive datasets and high dimensionality:** As explained, the pulse signal is very complex, and it depends upon many dimensions such as age, gender, habits, profession, stress, family background and so on. Therefore, enormous dataset has to be collected in order to increase the chances that a data mining algorithm will find the valid patterns. In our study, we have recorded the other dimensions in our database, but have not used due to high complexity of inter-relations. In future, all the dimensions have to be taken into consideration before coming to the conclusion about the subject.
- **Nonstandard and incomplete data:** As the literature of the quantitative description of the pulse signals is missing in the traditional methodology, the standardization of each pulse type is very important. In our work, we have provided an efficient way to capture the overall essence of the pulse signal using PSWT (Section 5.2). This essence pulse cycle can further be used, in future, for standardizing particular pulse types and sub-types.

The other limitation is the outliers in the data due to sudden movements of hand during the acquisition, or other noise inducing factors. Also in few cases, data for a particular sensor (*dosha* location) is missing due to its negligible amplitude, or an improper placement of the sensor on wrist. Currently, we do not handle these issues, as our scope was to establish the importance of the pulse signals. In future, the missing or noisy data should be properly handled with care.

- **Mixed data types:** Currently our database contains the pulse signals (time series format), the patient's information, and the doctor's diagnosis. We also plan to incorporate patient's fingerprints in future. Thus, the database will be a combination of various media – numeric (data), symbolic (gender, habits), text (doctor's diagnosis), and images (fingerprints). The handling of all these data types should be properly done for better results.
- **Compression:** Storage of large databases is often required to be in compressed form [24]. Hence the development of compression technology, particularly suitable for data mining, is required. The pseudo-periodic pulse signals have not only a temporal intrabeat correlation between successive samples, but also

a strong temporal interbeat similarity between successive beats. These facts could be used in future to compress the huge amount of data based on pattern matching.

- **Hemodynamics:** Hemodynamics is the study of physical factors that govern blood flow. In future, signals can be obtained simultaneously from various body locations, and studied to understand the hemodynamics. Few examples of these signals are the electrocardiogram (from the heart), the *arterial pulse* (on the wrist), the *carotid pulse* (on both sides on the neck), the *brachial pulse* (inside of arm), and the *pedal pulse* (on top of the foot). The signal from each body location will provide different information in terms of patterns, to infer the pressure, resistance and flow inside human body. Further, correlation between these signals can be derived, e.g. between the arterial pulse and the ECG to understand whether heart actions can be captured using just the pulse signal.
- **Fuzzyness:** Fuzzy logic is a computational method tolerant to sub-optimality and impreciseness (vagueness) and thus gives quick, simple and sufficiently good solutions. This is the most important future work to incorporate. This will help to remove the limitation of inherent subjectiveness in the traditional treatment.

A new database of pulse signals is to be prepared in the following manner. Same set of patients will go to a number of *Ayurvedic* practitioners, and their diagnosis will be noted down. It will consider the diagnosis of all the practitioners in respective proportions, and will provide the summarized output.

- **Networked Multimedia Communications:** Wireless networking technology (WNT) is rapidly coming of age with the recent passage of initial standards, widespread performance improvements, and the introduction of personal communications networks (PCNs) [254]. Therefore different future directions in WNT are open such as:
  - ◇ A pocket-sized portable “smart” terminal combined with wireless communication will help in long-term monitoring and processing of pulse signals for timely intervention.

- ◇ Pulse signals can be acquired with *Nadi Tarangini* at a patient's home, and transmitted over internet to a doctor or for further processing and/or storage. It will also be possible to provide a quick diagnostic feedback if the patient has an unusual rhythm while at home.
- ◇ Problems might arise while handling extremely large heterogeneous databases. Development of algorithms for mining data from distributed databases will be helpful in the future.
- **Improvement in our PSWT-DTW approach:** In the computation of the pitch synchronous wavelet transform (Section 5.2.2), wavelet packet transform (instead of the discrete wavelet transform) may provide approximations at various scales, which can further be utilized as trends of the pulse signal at different scales. These trends can be individually used to study the overall behavior, or can be utilized in the searching of TDP in ill-behaved pulse signals.

In our experiments, we have used dynamic time warping in its simplest form, making use of the *warping window constraints*. For speeding up the DTW computations further, advanced approaches of lower bounding measures and warping windows [177, 180, 181] can be utilized.

- **Improvement in our Hybrid-SVM approach:** In our novel Hybrid-SVM (Section 7.2), we use a threshold parameter ( $R$ ). For purposes of illustration, we use 3 different values of  $R$  in Appendix B to prove its efficacy. A particular value of  $R$  might reduce the number of false negatives at the cost of increasing the number of false positives, and the balance is best chosen by the user. A systematic way of building the Hybrid-SVM, in future, would use a feedback system with various values of  $R$ , and pick the value that is most suitable for the problem at hand.

As our Hybrid-SVM is an ensemble kind of arrangement, different classifiers can be used offline. In the first decision box of Figure 7.9, outputs of two different classifiers (support vector machine for *opM* case and a neural network-based classifier for *opB* case) can be utilized. The two classifiers will get trained with different set of rules, and thus could provide better information for checking the

bias in the unbalanced datasets.

Before we summarize, we would like to mention that – throughout the world, patients are going outside of modern medicine to look for help. Now, the visits to alternative practitioners (*Ayurveda*, TCM, etc.) exceed visits to primary care providers [34]. Many alternative therapies are even moving to the hospital sector [44]. The review in [12] suggests that a large number of physicians are either referring to, or practising some of the more prominent and wellknown forms of the alternative medicine, and that many physicians believe in these useful therapies such as Yoga [58, 119, 188]. Our system *Nadi Tarangini* shows lots of promise in the pulse-based diagnosis (alternative medicine) such that a patient will get treated just by analyzing the arterial pulse.

### 8.3 Pictorial Summary of Our Work

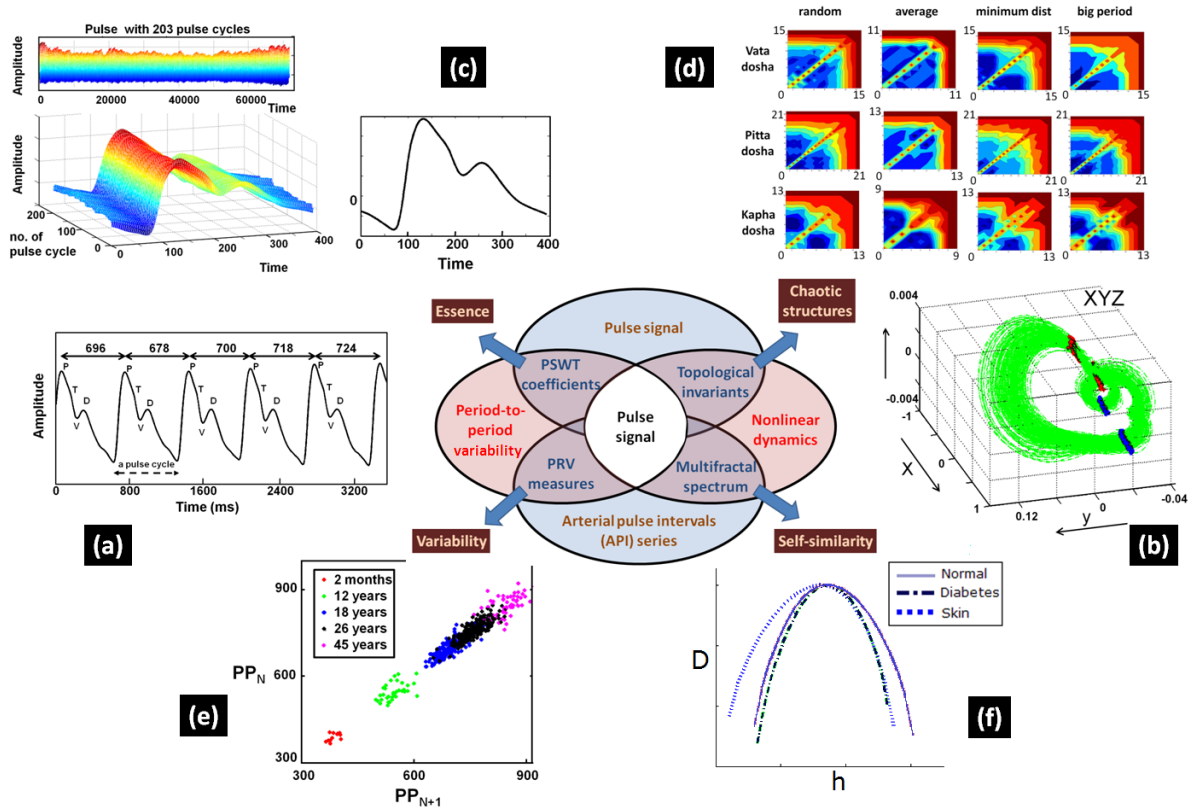
In this thesis, we established the efficacy of the pulse signals with a scientific basis. The experience, knowledge, and vision of the *Ayurvedic* practitioners were successfully incorporated with our expertise of visualization and modern pattern recognition methodologies as shown in Figure 8.3.

With reference to above figure, our approaches can be classified as the combination of:

- **Pulse signal and period-to-period variability:** PSWT-based analysis
- **API signal and period-to-period variability:** PRV-based analysis
- **Pulse signal and Nonlinear dynamics:** Topological invariants-based analysis, local Hölder exponents-based analysis
- **API signal and nonlinear dynamics:** Multifractal spectrum-based analysis

We summarize our work in Figure 8.4.

We developed a portable and inexpensive system *Nadi Tarangini* to acquire noise-free pulse signals. The design of our system reflects the “feeling” of fingertips. We showed that our data contains typical physiological properties and patterns, and is



**Figure 8.3:** The systematic methodology of our work to study different properties and patterns of the pulse signals.

consistent with the *Ayurvedic* literature. We systematically analyzed the pulse signals to reveal important properties, and hidden information:

- **Trend cycle:** We used the approximate coefficients in the pitch synchronous wavelet transform to extract the P-T-V-D rhythm (essence) of the pulse signal.
- **Time domain parameters:** We utilized the pseudo-periodic nature of the pulse signal to compute the time domain parameters (TDP) using our novel PSWT-DTW approach. Our novel approach reduces the efforts of searching the pseudo-locations; particularly when they have either confusing or overlapping values. Our approach can be suitably extended to other signals with pseudo-periodic behavior, to study average values as well as cycle-to-cycle variations.
- **Period-to-period variability:** This is a characteristic property of the pulse signal. We studied the variability using various time-domain, frequency-domain, and nonlinear Poincaré based measures. Our results showed the variations in the

measures for different age-groups and disorders; akin to the heart rate variability in clinical use.

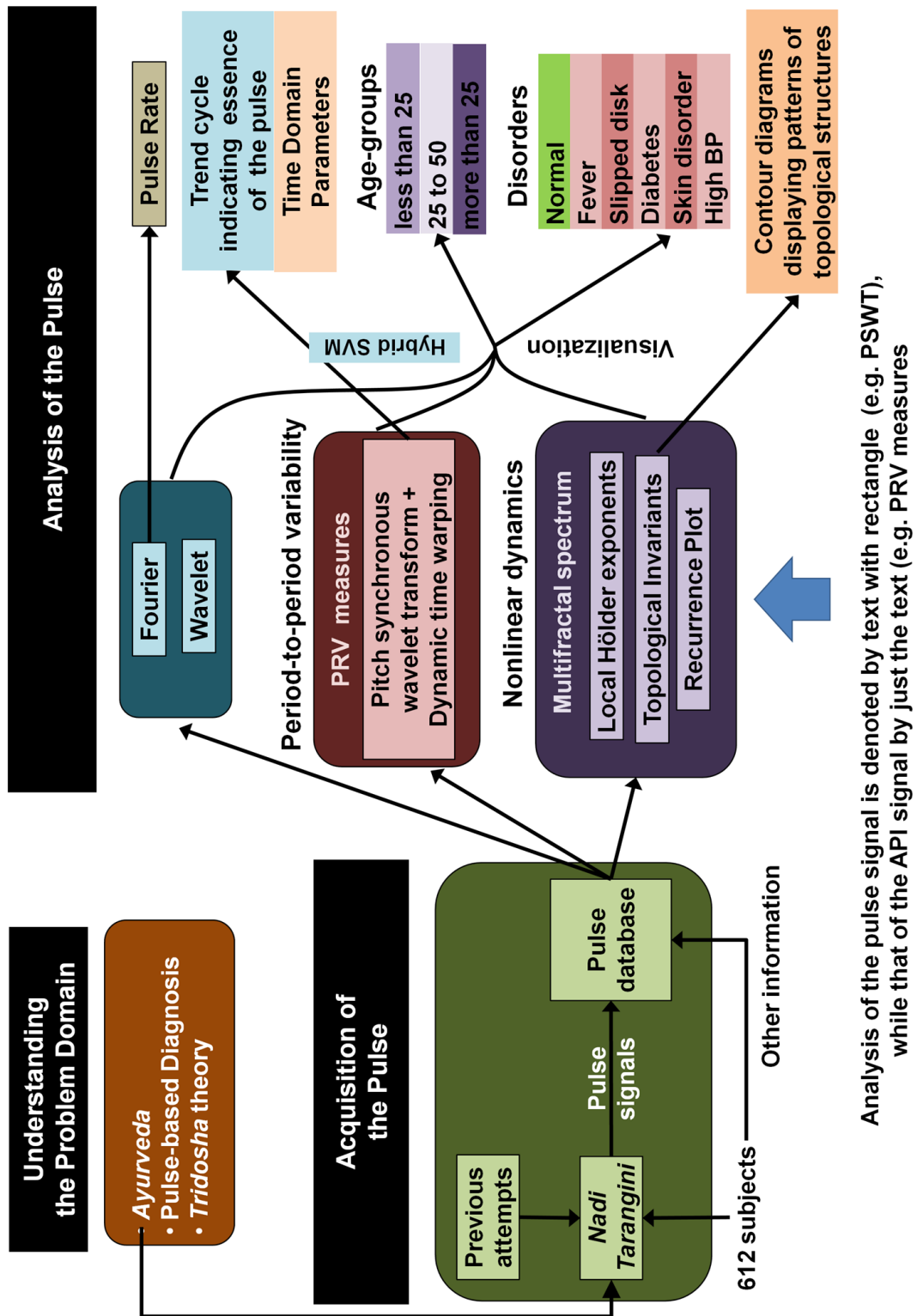


Figure 8.4: Summary of our work.

- **Nonlinear dynamics:** This is a characteristic property of the pulse signal. We studied the complex chaotic nature through four approaches, namely multifractal spectrum based analysis, recurrence plot based analysis, local Hölder exponents based analysis and topological analysis.

We showed that a combination of wavelet scales in the computation of local Hölder exponents provides an additional information. Also, in the computation of topological invariants, we suggested novel ‘minimum distance’ and ‘big orbit’ paradigms for selecting the representative pulse beats of each pseudo-period.

- **Structures:** We introduced contour diagrams for visualizing the topological structures in the embedded pulse signals. We showed the varying patterns for different pulse signals based on the number of contours, base of triangular contours, and their colors.
- **Results:** The results in each methodology were visually displayed as well as numbers were provided whenever possible. We showed the efficacy of our novel ways of selecting representative beats of each pseudo-period in the computation of topological invariants through the improvements in the classification results. Also, the classification of pulse cycles into ten different behaviors was successfully achieved using the feature sets of our maxima-minima based time domain parameters.

Further in the classification procedure, we introduced the notion of Hybrid-SVM – a combination of binary and multiclass SVM – to increase the testing accuracy in unbalanced datasets. This becomes particularly important in reducing the false negative (patients falsely classified as normal) predictions.

The classification results with time domain parameters as the input feature vectors, displayed the improvements due to both novel PSWT-DTW and Hybrid-SVM approaches. We believe the two methods can suitably be extended to other biomedical as well as non-biomedical applications.

As a matter of fact, electrocardiogram (ECG) for cardiovascular system witness the development of home health monitoring. In future, the pulse signals can be equally

effective in the quantitative study. The large amount of information of the pulse signals will lead us to an improved understanding of the disease process affecting the patient. Establishment of this correlation in this thesis is just the beginning ...





---

## Appendix A

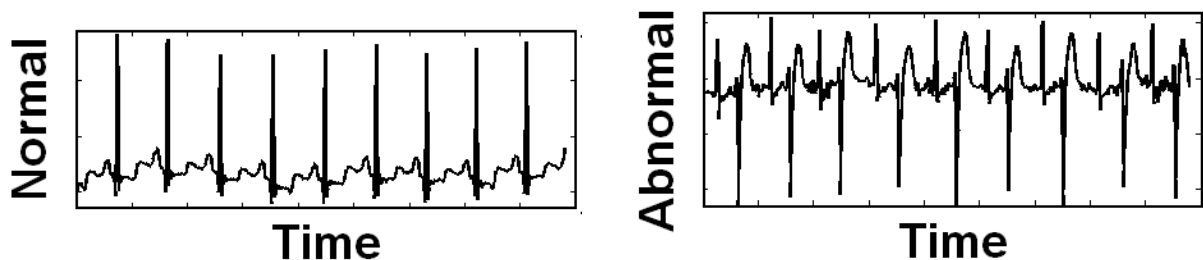
# ECG-based Arrhythmia Classification

---

*We propose a novel Hölder-SVM detection algorithm for arrhythmia classification. The Hölder exponents are computed efficiently using the wavelet transform modulus maxima (WTMM)-based method.*

## A.1 Introduction

Arrhythmia is a common term for any cardiac rhythm which deviates from normal sinus rhythm. Characterization and classification of such arrhythmia signals is an important step in developing devices for monitoring the health of individuals. Typical and abnormal signals are shown in Figure A.1.



**Figure A.1:** A normal consistent P-QRS-T rhythm is exhibited on the left. Abnormal rhythm for a patient in every alternate beat appears on the right.

Arrhythmias are of different kinds, and exhibit long-term non-stationary patterns. Concepts and techniques including Fourier [140], wavelets [140], chaos parameters [156] have been employed to extract information present in such physiologic time series. All these methods exhibit different degrees of advantages and disadvantages, the main concern being low specificity and accuracy.

Our interest is in the recently developed analytic tools based on nonlinear dynamics theory and fractals. These are attractive because they have the ability to perform a reliable local singularity analysis. For example, gait analysis [185] and localization of outliers [206] have been performed using this approach. This approach offers a new and potentially promising avenue for quantifying features of a range of physiologic signals that differ [205] in health and disease.

Features detected using such approaches exhibit *local* hidden information in time series and thus are suitable for classification. Support vector machine (SVM) rigorously based on statistical learning theory simultaneously minimizes the training and test errors. Apart from that, SVM produces a unique globally optimal solution and hence is extensively used in diverse applications including medical diagnosis.

## A.2 Local Hölder Exponents

For a time series  $f$ , if there exists a polynomial  $P_n$  of degree  $n < h$  and constant  $C$ , such that:

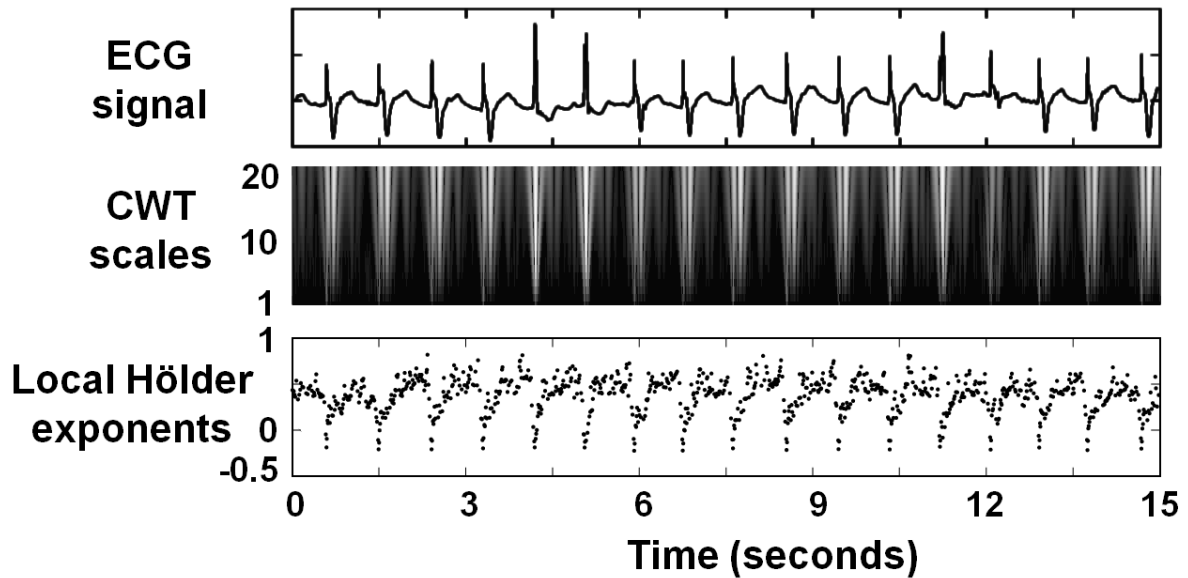
$$|f(x) - P_n(x - x_0)| \leq C|x - x_0|^h \quad (\text{A.1})$$

the supremum of all exponents  $h(x_0) \in (n, n + 1)$  is termed the Hölder exponent, which characterizes the singularity strength. It is evident [185] that the Hölder exponent describes the local regularity of the function (or distribution)  $f$ . The higher the value of  $h$ , more regular is the local behavior the function  $f$ . Thus it characterizes the scaling of the function locally and the distinct scaling behavior of different signals can be exploited to characterize and classify time series.

The wavelet transformation (WT) i.e.  $W_{s,x_0}(f)$  provides a way to analyze the local behavior of a signal  $f$ , which is a convolution product of the signal with the scaled( $s$ ) and translated( $x_0$ ) kernel. One of the main aspects of the WT is the ability to reveal the hierarchy of (singular) features, including the scaling behavior. This is formalized by its relation [144] with the Hölder exponent:

$$W_{s,x_0}(f) \propto |s|^{h(x_0)}, \quad s \rightarrow 0 \quad (\text{A.2})$$

Figure A.2 shows an example of the WT and the Hölder exponents. We used the second derivative of the Gaussian function, i.e. Mexican hat, as the analyzing wavelet. This wavelet is also extensively used by the other authors, since it possesses good localization capabilities in both position and frequency.



**Figure A.2:** Sample ECG waveform, its continuous wavelet transform (CWT) and the corresponding local Hölder exponents at first scale.

### A.2.1 Efficiency Considerations

As continuous WT in its original form is an extremely redundant representation, Mallat and Hwang [79] have come up with an alternative approach called wavelet transform modulus maxima (WTMM). Here, the computation of the hierarchical distribution of

local behavior, as explained above, can be effectively computed by considering the space-scale partitions. The dependence of the scaling function on the moments  $q$  can be captured using WTMM tree as:

$$Z(s, q) = \sum_{\Omega(s)} |W_{s, x_0}(f)|^q \propto s^{\tau(q)} \quad (\text{A.3})$$

where  $Z(s, q)$  is the partition function and  $\Omega(s)$  is the set of all maxima at the scale  $s$ . The computation of singularity strength and the transformation from  $\tau(q)$  to spectrum of singularities  $D(h)$  is given by the Legendre transformation [11]:

$$h(q) = d\tau(q)/dq; \quad D[h(q)] = qh(q) - \tau(q) \quad (\text{A.4})$$

## A.2.2 Stable Computation

The WTMM based formalism developed by Muzy [144] as described above provides global estimates of scaling properties of time series. Recently, it has been found that even though such global estimates of scaling is often a required property, local analysis may provide more useful information.

In the traditional form, the estimation of local singularity strengths and their spectra may not be possible due to the fact that in real life data, the singularities are not isolated but densely packed. This causes the logarithmic rate of increase or decrease of the corresponding wavelet transform maximum line to fluctuate. But very recently, Struzik [202, 204] has provided a stable methodology for estimating the local exponents, in which he has modeled the singularities as if they were created through a multiplicative cascading process. This method has been successfully applied for classification of human gait [185].

The method [185, 204] is as explained below. The mean Hölder exponent  $\bar{h}$  is given by  $\log[M(s)] = \bar{h}\log(s) + C$ , where  $M(s) = \sqrt{\frac{Z(s,2)}{Z(s,0)}}$ . Employing the multiplicative cascade model, the approximate local Hölder exponent  $\hat{h}(x_0, s)$  at the singularity  $x_0$  can now be evaluated as the slope:

$$\hat{h}(x_0, s) = \frac{\log(|W_{s,x_0}(f)|) - (\bar{h} \log(s) + C)}{\log(s) - \log(s_N)} \quad (\text{A.5})$$

where  $s_N$  is the length of the entire wavelet maxima line tree, that is, the maximum available scale that coincides with the sample length  $s_N = N$ , and  $x_0$  belongs to the set  $\Omega(s)$  of all wavelet maxima at the scale  $s$  that assume the value  $W_{s,x_0}(f)$ . (In our calculations we used  $s=1$  in the WT.)

## Support Vector Machines

The local Hölder exponents are appropriate as the informative features for classification using the support vector machines (SVM)[227]. SVM is being extensively used for several classification and regression applications. As the theory is well developed, we provide only the basic ideas [102, 138] involved in binary classification.

1. Transform the input data into a higher dimensional feature space to enable linear classification; specifically define an appropriate kernel in the input space in place of the dot product in the high dimensional feature space.
2. Maximize the margin of the linear hyperplane separating the instances belonging to the two classes by solving the dual formulation of the convex quadratic programming problem to obtain the unique global solution for the classifier.

For multiclass classification, we used popular One-Against-One method [77].

## A.3 Results and Discussion

**Data Set.** The data set used was extracted from ECG recordings of MIT-BIH Arrhythmia Database according to the beat and rhythm annotations. Each record of these rhythms is at least 10 seconds long. Here, the complete dataset includes 160 of Normal sinus rhythm (NSR), 25 of Ventricular bigeminy (VB), 155 of Atrial fibrillation (AF) and 146 of Nodal (A-V junctional) rhythm (NR) records. Out of the whole dataset,  $2/3^{rd}$  of

randomly selected data was used as the training set and the remaining  $1/3^{rd}$  was used as the testing set.

For each of the extracted rhythms, we computed the features to be used by SVM for classification in the following manner. First, we de-noised the data series using soft threshold wavelet method [150]. Then we computed the wavelet coefficients using WTMM approach, which were subsequently used for the computation of local Hölder exponents as explained in Section A.2. We then computed the probability density of these local Hölder exponents and then fitted this density with Gaussian kernel. For all the rhythms belonging to different classes the local Hölder exponents were in the range  $[-0.5 : 1.5]$ . We divided this range into 12 equal sub-parts and chose 12 points (as shown in Figure A.3(a)) on the fitted probability density curve corresponding to the mid-points of the 12 sub-ranges.

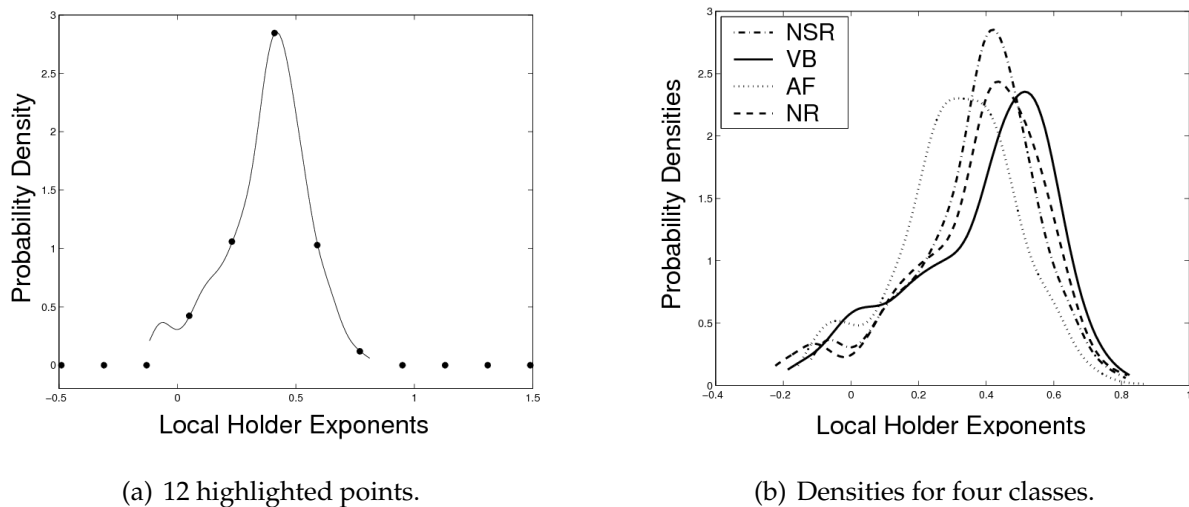
### **A.3.1 Binary Classification**

We considered NSR as the ‘normal class’ and all other arrhythmias as ‘arrhythmia class’. We employed LIBSVM [21] toolbox for the classification purpose and used radial basis function (RBF) as the kernel. Selection of other SVM parameters was done using cross-validation on training data. Using multiple runs of 5-fold cross validation with chosen parameters, the overall accuracy of classifier model combinations ranged between 96% to 98%, and which show average (and best respectively) classification for 96.3% (and 98.48%) of normal class and 98.25%(and 98.89%) of arrhythmia class on the test set.

### **A.3.2 Multiclass Classification**

For classification of four types of rhythms (for which the probability densities vary as shown in Figure A.3(b)), again, the parameters were tuned in the same way as explained in binary classification. The results show overall 96.94% accuracy, and average (and best respectively) classification for 95.98% (and 98.89%) of NSR, 82.05% (and 100%) of VB, 98.51% (and 99.23%) of AF and 98.92% (and 99.20%) of NR rhythms.

The results for multiclass classification can be summarized as the confusion matrix of four classes and sensitivity & specificity of each class as given in Table 1.



**Figure A.3:** *probability density of local Hölder exponents.*

		NSR	VB	AF	NR
NSR		31.93	0.70	0.51	0.11
VB		0.62	3.69	0.18	0
AF		0.84	0	31.75	0
NR		0.48	0	0.15	29.01
Sensitivity	Avg	96.04	82.11	97.42	97.9
	Best	98.29	100	99.23	99.20
Specificity	Avg	97.1	99.27	98.75	99.84
	Best	98.89	100	99.63	99.64

**Table A.1:** *Confusion matrix (average case).*

We also used the features derived from Fourier analysis, as mentioned in introduction, with SVM classification for comparison with our method. Fourier analysis (by selecting best 300 features) gives average correct classifications respectively 86.48%, 89.33%, 95.7% and 96.74% for the above four classes. Hölder-SVM methodology was found to provide superior performance. An interesting point to note is that in both the binary and multiclass classification, we used data provided by both sensing leads. It was observed that even if we use just a single lead data, classification gives results with



comparable accuracy. Arrhythmia being a disorder in the normal rhythm, can thus be captured in any of the two leads.

## **Conclusion**

In this study, it is demonstrated that support vector machine in conjunction with wavelet transform modulus maxima based local singularity feature extraction provides an excellent combination in arrhythmia classification. The recognition of the normal and different types of rhythms representing the arrhythmias has been done with a good accuracy. These investigations show that the presented method may find practical application in the recognition of many more types of arrhythmias.

---

## Appendix B

# Hybrid-SVM for Imbalanced Data

---

*We propose a novel hybrid arrangement of binary and multiclass SVMs designed to take care of class imbalance rampant in biomedical signals. As a result, we significantly reduce the number of false negatives – patients falsely classified as normal. Our concept of Hybrid-SVM is applicable to a wide variety of multiclass classification problems. We update the relevant singularity-driven Hölder features proposed earlier, and compare our Hybrid-SVM with a suitable conventional SVM. While the use of the updated features results in higher accuracy, using the Hybrid-SVM shows even more gains.*

## B.1 Introduction

Changes in the normal rhythm of a human heart may result in different cardiac arrhythmias which cause damage to the heart when sustained over long periods of time. The Electrocardiogram (ECG) is an important tool in diagnosing the condition of the heart. In modern medicine, the ability to automatically classify arrhythmias from ECG recordings is important for clinical diagnosis and treatment, as well as for understanding the electrophysiological mechanisms of the arrhythmias.

Arrhythmias are of different kinds and *machine learning* methods are frequently applied with considerable amount of success in their classification [135, 169]. One major problem for an autonomous system is the seemingly erratic variation of the ECG signals of different patients and patient groups, leading to a significant risk of inconsistent classification. Therefore, characterization and classification of such arrhythmia signals is an important step in developing devices for monitoring the health of individuals. Our interest is in the recently developed analytic tools based on nonlinear dynamics theory and fractals. These are attractive because they have the ability to perform a reliable local singularity analysis. For example, gait analysis [185], flow regime identification [84], and localization of outliers [206] have been performed using this approach. We make use of the property of *local Hölder exponents* in that they capture the hidden information in a time series [205], and thus are suitable for classification. In Appendix A, a Hölder-SVM based classification scheme was presented on a relatively small number of classes. In this chapter, we show the efficacy of a modified scheme on a larger dataset with an increase in number (seven) of arrhythmia (abnormal) classes.

## Our contributions

Our results show considerable improvements over the previous schemes as can be seen in Table B.13. Besides the use of the Hölder-SVM scheme, our improvements include the following.

- Unlike prior methods, we compute local Hölder exponents, using a *combination of wavelet scales*. Our results (Section B.4.1) show that additional scales are significant in classification.
- Further, we observe that ECG recordings of the normal behavior are more common and thus are widely available as compared to ECG beats with some abnormality. But when all the abnormal beats are combined into one ‘Arrhythmia’ class, the Normal-Arrhythmia (binary) classification is unbiased, when compared to individual (multiclass) classification inherent in the SVM paradigm. We exploit this observation and introduce a novel *Hybrid-SVM* approach to consider

such class imbalance. This combinations reduces the false negatives (situations in which abnormal ECG signals are erroneously classified as normal), resulting in increased testing accuracy (Section B.5). Our Hybrid-SVM scheme can be combined (Section B.5.4) with other class imbalance solution methods.

## Prior Work

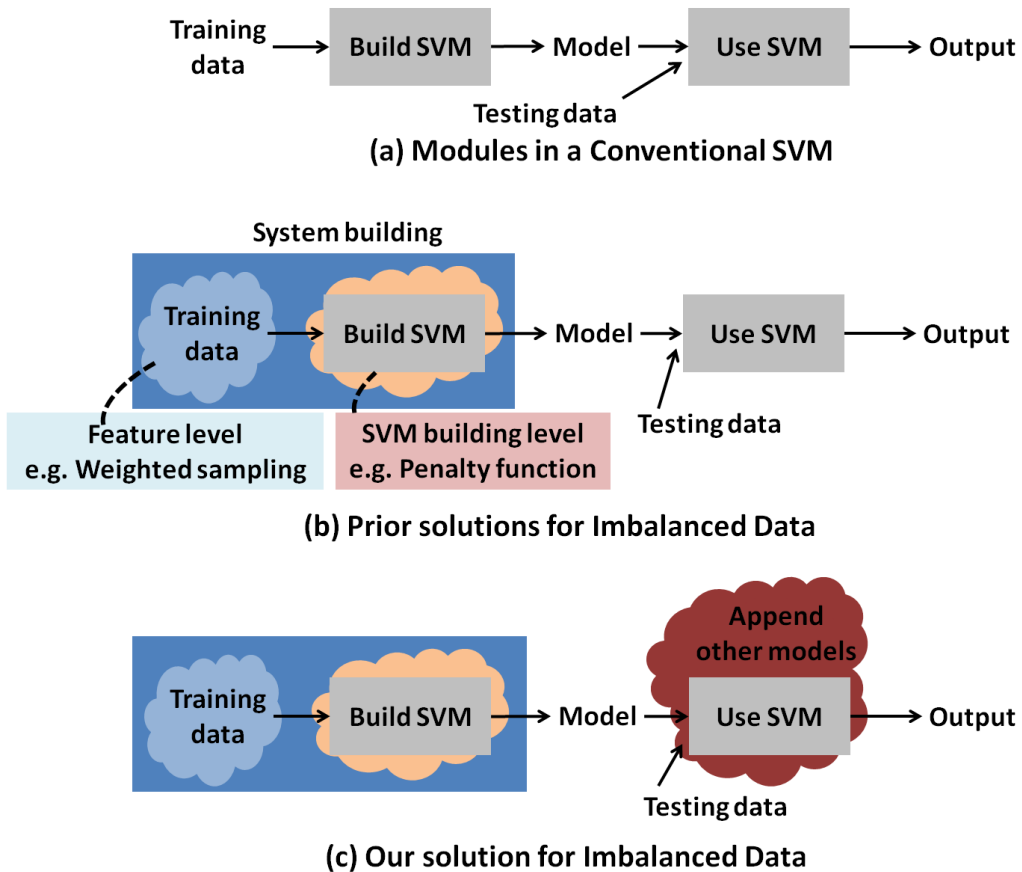
The treatment of unbalanced data in the classification process is not new but has been undertaken at different levels in the machine learning process. For example at the feature extraction level, skewed datasets are evaluated in [85] for over- and under-sampling. At the SVM building level, an algorithm is proposed in [32] that changes the kernel function to modify the bias, while the kernel matrix is adjusted in [32] to fit the training data. Different penalty constants for different classes of data can also be employed [228] making errors on positive instances costlier than errors on negative instances. At the system building level [4, 213], the penalizing method can be combined with SMOTE [22] to develop a classifier that performs better than the use of individual algorithms.

Figure B.1(a) shows typical blocks in a conventional SVM. The training data is trained to produce a model, which predicts the outputs for the testing data.

Figure B.1(b) shows prior solutions at various levels for classifying the imbalanced data. Our solution is shown in Figure B.1(c), and works independent of, or along with, prior solutions. Thus, the particular advantage of our method is that it is at a “black box” level and is generic enough to work with a variety of SVMs, with or without penalties. The details of our enhanced model that feeds into the testing phase appears in Section B.2.1.

## Local Hölder Features

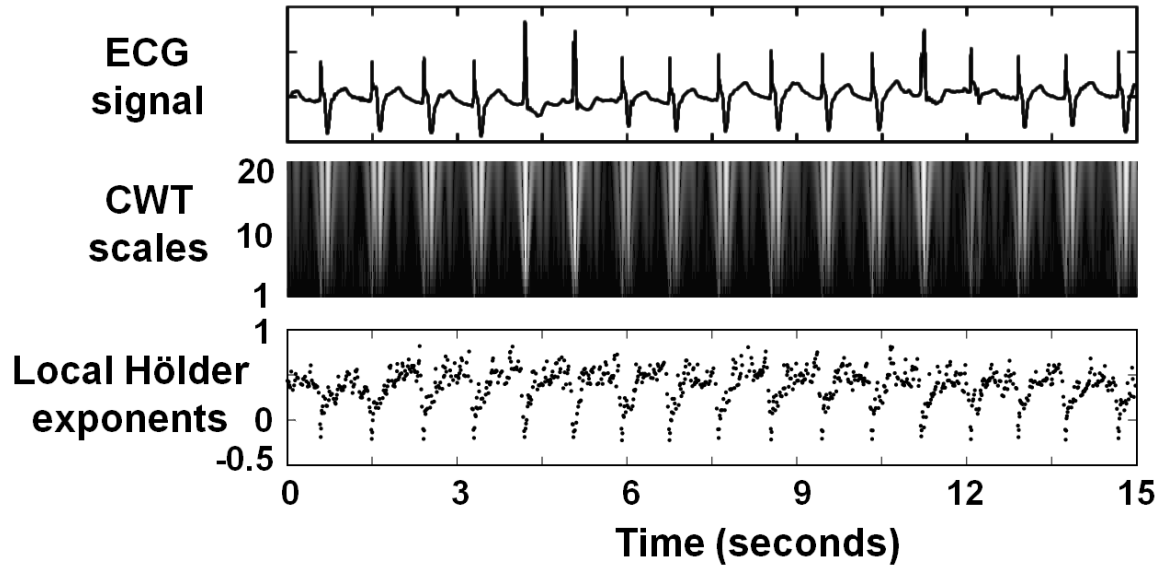
In the work of Appendix A, a wavelet-based local Hölder exponents spectrum approach was applied for feature extraction. In this chapter, we process these spectra



**Figure B.1:** Handling imbalanced data in (a) conventional SVM methodology. Prior methods (b) emphasize the creation of the model while building the SVM. In contrast, our method (c), while compatible with earlier methods, emphasizes the testing phase.

further so as to feed them into Hybrid-SVM-based classification system.

Many physiological signals contain step or cusp-like singularities, which are the rapid changes in the variable values for a very small change in time. The strength of such singularity is usually defined by an exponent called the Hölder exponent [205]. We use a stable methodology for estimating the local Hölder exponents provided in [204], where the singularities are modeled as if they were created through a multiplicative cascading process. Please refer to [204, 205, 185] for further details. Figure B.2 summarizes important steps.



**Figure B.2:** Sample ECG waveform, its continuous wavelet transform (CWT) with scales 1 to 20 in which the maxima lines are visible, and corresponding local Hölder exponents at scale 1.

## B.2 Classification Method

Support vector machines (SVM) are being used [106, 155] extensively in the classification and regression applications in the biomedical area. The advantages of using SVM are that it produces a classifier that works well on unseen examples and also converges to a single globally optimal solution [227]. In the classification problem of two classes, the job is to find an optimal hyperplane that maximizes the distance between itself and the nearest data points of each class. For linearly non-separable classes, a suitable kernel ( $K$ ) is introduced in the feature space. Please refer to [20, 67] for further details.

### Multiclass Support Vector Machine

The SVM for pattern classification is based on two-class classification problems. Thus, the  $n$ -class problem is generally solved by considering the problem as a collection of binary classification problems [7]. One of the most commonly used approaches for multiclass SVMs is the *One-Against-One approach* [76]. This approach simply constructs all possible binary classifiers from a training set of  $n$  classes. Each class is trained on only two out of  $n$  classes. Thus, there will be  $n(n-1)/2$  classifiers. Now, in the Max

Wins Algorithm [98], a test example is classified by all of classifiers. Each classifier provides one vote for its preferred class and the majority vote is used to decide the final winner. Other popular multiclass approaches include one-against-all, and the directed acyclic graph SVM [77, 210].

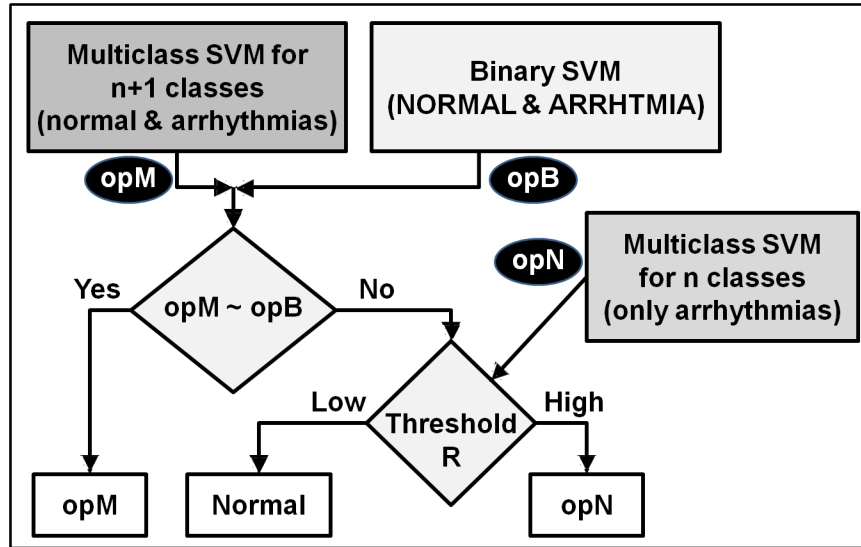
## Reducing False Abnormality Predictions

In any diagnostic system, it is critical to reduce the ‘false negatives’; that is the decisions predicting ‘normal’ when there is some abnormality. We now describe an arrangement of binary and multiclass classifiers so as to improve the final accuracy on testing data. Our scheme takes advantage of the point that in a diagnostic system, we get an enormous number of normal signals as compared to the abnormal signals. Thus, when the system is trained with the data consisting of all the individual classes, it gets biased toward the larger *Normal* class. However, when the abnormal rhythms are combined into one *Arrhythmia* class, its number becomes comparable to that of the *Normal* class and therefore there is a higher likelihood to get a system trained without bias.

### B.2.1 Hybrid-SVM

Our approach is as explained in Figure B.3. Consider that we have a Normal class and ‘ $n$ ’ different abnormal classes (arrhythmias in our case) in the classification system. First, using a multiclass SVM, we train a system with all  $(n+1)$  classes, each class being individually considered. Let the predicted output coming out of this classifier be  $opM$  (here ‘M’ stands for multiclass). Next, we use a binary SVM to train the *Normal* class versus the *Arrhythmia* class (considering all  $n$  abnormal classes put together). Let  $opB$  be its predicted output (here ‘B’ stands for binary). Finally, we use another multiclass SVM to train only the  $n$ -abnormals and let  $opN$  be its predicted output (here ‘N’ stands for the considered  $n$  abnormalities). *An important point to note is that all the mentioned SVM blocks are trained **offline**.*

Now, when a test case enters the system, we first compare  $opM$  and  $opB$  coming from the above mentioned two sub-systems.



**Figure B.3:** Our Hybrid-SVM method. See text for details.

- If the predicted outputs are consistent (that is both are *Normal*, or *opM* declares an abnormality and *opB* declares the *Arrhythmia* class), then we use *opM* as the final predicted output.
- On the other hand, if they are not consistent, then we speculate that the incompatibility is due to the bias in the trained SVM (conventional multiclass). Therefore, we devote more care to the *n* abnormal classes. We choose a random number between  $[0,1]$ . If the random number is below a threshold value ( $R$ ), then we consider the *Normal* class as the final answer; and if the random number is above threshold, then we consider *opN*, declared by the *n*-class SVM block, as the final predicted output.

### B.2.1.1 Threshold Choice

At first glance, it might look a bit strange to use a random number generator. Indeed, if we set  $R = 0$ , as in our initial experiments, then the random number is essentially useless, and we will minimize false negatives by pushing the output to an abnormality (refer Table B.4). However, this leaves less control for reducing false positives, which is why our Hybrid-SVM has the random number based “knob” for the control. For concreteness, we consider the following three cases in the second decision box



in Figure B.3: same weightage is given to both the *Normal* & *Arrhythmia* classes ( $R= 0.5$ ), more weightage to the *Normal* class ( $R= 0.9$ ), and more weightage to the *Arrhythmia* class ( $R= 0.1$ ). The comparison of the new Hybrid-SVM approach with the conventional one-against-one multiclass approach is discussed next in Section B.5.

## B.3 Experimental Setup

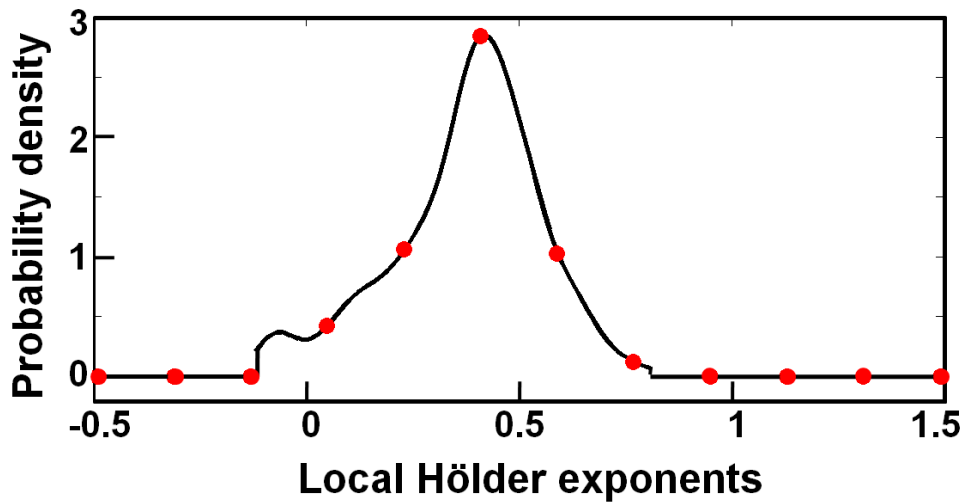
We first explain our arrhythmia dataset and the methodology to extract the local Hölder features. Next, we show that these features are indeed useful in the multiclass arrhythmia classification. Further, we show the improvement in the testing accuracy by employing our Hybrid-SVM approach. We show that our approach is compatible with other types of features (e.g. Fourier coefficients) also by improving performance. Finally, we show that Hybrid-SVM can also be used with other prior methods for imbalanced dataset.

### Data Set

The data set used was extracted from the ECG recordings of the ‘MIT-BIH Arrhythmia Database’ and ‘MIT-BIH Malignant Ventricular Arrhythmia Database’ [63] according to the beat and rhythm annotations. Each record of these rhythms is 10 seconds long. The complete dataset includes 1273 Normal sinus (N) rhythms, 91 Paced (P) rhythms, 372 Atrial fibrillation (AFIB) rhythms, 123 Nodal (A-V junctional) rhythms (NOD), 108 Ventricular fibrillation (VF) rhythms, 278 First degree heart block (BI) rhythms and 34 Ventricular bigeminy (B) rhythms. It can be seen that number of data instances in N is very high as compared to the other abnormalities.

The extraction of features from each rhythm, for classification, was done in the following manner. First, we de-noised the data series using the soft threshold wavelet method [150]. Then we computed WTMM-based local Hölder exponents as explained in Section B.1. We then computed the probability density of these local Hölder exponents and observed that for all considered rhythms (irrespective of the classes),

the local Hölder exponents were in the range  $[-0.5, 1.5]$ . We chose 12 equidistant points (as shown in Figure B.4) on this range.



**Figure B.4:** *Highlighted points on the probability density function of the local Hölder exponents are used as features.*

We employed LIBSVM [21] toolbox for the classification purpose and used the radial basis functions (RBF) for the kernel. The above-mentioned methodology was employed for all possible ranges of the other SVM parameters, and the combination which gives the best 10-fold cross validation accuracy (explained in section B.3) was selected as the final set of parameters for further classification. Out of the whole dataset, approximately 66% randomly selected data constituted the training set, and the remaining constituted the testing set.

## Performance Measures

In subsequent sections, we present the implementation of our Hybrid-SVM approach. For every possibility, we give

- The *10-fold cross validation accuracy* as a percentage value, which is calculated as follows. The training dataset (excluding 30% testing data of the whole data) is divided into  $k$  ( $k=10$  here) subsets. Of the  $k$  subsets, a single subset is retained as the validation data for testing the model, and the remaining  $k - 1$  subsets are used as the training data. Then the average accuracy across all  $k$  trials is computed.

The advantage of this method is that now it does not matter how training and testing data get divided. Every data point gets to be in a test set exactly once, and gets to be in a training set  $k - 1$  times.

- The *confusion matrix* of predicted classification values along with their sensitivity-specificity numbers in the tabular form.
- The *Matthews correlation coefficients (MCC)* [123], defined as:

$$MCC = \frac{[(TP)(TN) - (FP)(FN)]}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where  $TP$  denotes true positives,  $TN$  denotes true negatives,  $FP$  denotes false positives, and  $FN$  denotes false negatives. If an approach has a higher MCC value in the range [-1:1], its predictive ability is better. Since MCC takes account of both over- and under-predictions, it is one robust and frequently used measure.

- The areas under *receiver operating curve (ROC plot)* for showing the efficacy of our Hybrid-SVM approach. The ROC plot is a graph of true positive rate (TPR) versus false positive rate (FPR).

## B.4 Validity of Features

### Conventional Multiclass Classification

This corresponds to the *opM* case in Section B.2.1. Here, we consider all the classes individually. The 10-fold cross validation accuracy is  $90.89 \pm 2.1\%$ ; details of the 7-class confusion matrix appear in Table B.1. (Notation: The 10-fold cross validation accuracy is written as [mean  $\pm$  standard deviation]). In such a matrix, diagonal elements represent the correct identification of the respective rhythm. We observe that there are very few non-diagonal entries, except for the first row and the first column (which are due to the large number of rhythms present in the ‘Normal’ class). The high sensitivity

Testing Data	Classes	Predicted classification						
		N	P	AFIB	NOD	VF	BI	B
369	N	348	1	3	8	2	7	0
26	P	5	20	1	0	0	0	0
101	AFIB	5	1	95	0	0	0	0
33	NOD	4	0	1	28	0	0	0
31	VF	5	0	0	0	26	0	0
81	BI	1	0	0	0	0	80	0
11	B	4	0	0	0	1	1	5
	Sensitivity	93.55	90.91	95	77.78	89.66	90.91	100
	Specificity	92.5	99.05	98.91	99.19	99.2	99.82	99.07

**Table B.1:** Confusion matrix for the conventional multiclass classification using local Hölder exponents (*opM* case in Figure B.3). Six abnormal and one normal class were considered.

and specificity numbers indicate that the classification was successfully achieved in all the cases.

## Binary Classification

An important observation to note is that all the abnormal rhythms (i.e., excluding numbers in the first row and the first column in Table B.1) are very well separable. In the right-bottom  $6 \times 6$  matrix, all the non-diagonal entries are negligible. As we have already explained in Section B.2, *we take advantage of these observations to increase the testing-accuracy using our Hybrid-SVM classification approach.*

This combination corresponds to the *opB* case in Section B.2.1. Here, we considered N as the ‘Normal’ class and all other abnormalities as the ‘Arrhythmia’ class. The 10-fold cross validation accuracy of the classifier was  $91.52 \pm 2.3\%$ . It is higher than that of the conventional multiclass SVM. Table B.2 gives the binary class confusion matrix (one hopes for small values in the off-diagonal entries).

Testing Data	Classes	Predicted classification	
		Normal	Arrhythmia
369	Normal	340	29
283	Arrhythmia	17	266

**Table B.2:** Confusion matrix for the binary classification using local Hölder exponents (*opB* case in Figure B.3).

### B.4.1 Combination of Wavelet Scales

Prior work in the computation of local Hölder exponents for different applications such as [185, 205] have used only the lowest scale (scale 1). Instead, we also make use of the properties of wavelet transform (WT) that it captures different information at different scales (or resolutions). Thus, this extra information can also be used in the process of classification. In fact, selection of other *lower* wavelet scales, computing more and more informative local singularity features and selecting the best features can provide better classification results. For example, we provide the 7-class confusion matrix in Table B.8, in which the local Hölder exponents were computed using scales 1&2. The 10-fold cross validation accuracy is  $93.73\pm 3.3\%$ , which is much higher than the previous case of computation using only scale 1.

### B.4.2 ECG Lead Selection

In an another experiment, we computed the classification results for data of only the first ECG lead (out of the two). The local Hölder exponents were computed using scale 1 only. The 10-fold cross validation accuracy is  $84.44\pm 2.8\%$ . Even though only half the number of features were considered (coming out of just first lead), the classification was still successful.

## Summary

We observe that all the confusion matrices for the above approaches are clean and the MCC values (Table B.9) are higher positive numbers, which indicate that the classification was successful. Thus, the probability densities of the local singularity measures (or patterns) differ in these arrhythmia classes and are useful in the classification.

## B.5 Validity of Hybrid-SVM

Table B.3 gives the 6-class confusion matrix (only six abnormalities, excluding N from Table B.1). This corresponds to the *opN* case in Section B.2.1. One of the key points to note is that the confusion matrix of only arrhythmias is very clean. All the classes are observed to be very well separated from each other.

Testing Data	Classes	Predicted classification					
		P	AFIB	NOD	VF	BI	B
26	P	23	2	0	1	0	0
101	AFIB	2	98	0	0	1	0
33	NOD	2	1	30	0	0	0
31	VF	0	0	0	30	1	0
81	BI	0	0	0	0	81	0
11	B	0	0	0	1	1	9
	Sensitivity	85.19	97.03	100	93.75	96.43	100
	Specificity	98.83	98.35	98.81	99.6	100	99.27

**Table B.3:** Confusion matrix for six abnormal classes, using local Hölder exponents (*opN* case in Figure B.3). Only abnormal rhythms were considered.

### B.5.1 Comparison with Conventional Multiclass SVM

We implement the scheme of Section B.2.1 and establish (Table B.4) the efficacy of the Hybrid-SVM scheme.

Test Data	Classes	Predicted classification (threshold 0.1)						
		N	P	AFIB	NOD	VF	BI	B
369	N	348	1	3	8	2	7	0
26	P	4	21	1	0	0	0	0
101	AFIB	4	1	96	0	0	0	0
33	NOD	3	0	1	29	0	0	0
31	VF	3	0	0	0	28	0	0
81	BI	1	0	0	0	0	80	0
11	B	1	0	0	0	1	1	8
	Sensitivity	95.6	91.3	95.05	78.38	90.32	90.91	100
	Specificity	92.71	99.21	99.09	99.35	99.52	99.82	99.53

**Table B.4:** *Confusion matrix for the multiclass classification using local Hölder exponents with our Hybrid-SVM approach considering threshold 0.1. See Table B.1 to contrast with the earlier approach.*

Recall the issue of threshold in Section B.2.1. To show the effect of the threshold (Figure B.3), we display the results for two different threshold values (0.5 and 0.9). The confusion matrices are respectively shown in Tables B.5, and B.6.

Test Data	Classes	Predicted classification (threshold 0.5)						
		N	P	AFIB	NOD	VF	BI	B
369	N	350	1	3	7	2	6	0
26	P	4	21	1	0	0	0	0
101	AFIB	5	1	95	0	0	0	0
33	NOD	3	0	1	29	0	0	0
31	VF	4	0	0	0	27	0	0
81	BI	2	0	0	0	0	79	0
11	B	3	0	0	0	1	1	6
	Sensitivity	94.34	91.3	95	80.56	90	91.86	100
	Specificity	93.24	99.21	98.91	99.35	99.36	99.65	99.23

**Table B.5:** *Confusion matrix for multiclass classification using local Hölder exponents with our Hybrid-SVM approach considering threshold 0.5. See Table B.1 and Table B.4 to contrast.*

Test Data	Classes	Predicted classification (threshold 0.9)						
		N	P	AFIB	NOD	VF	BI	B
369	N	350	1	3	7	2	6	0
26	P	5	20	1	0	0	0	0
101	AFIB	5	1	95	0	0	0	0
33	NOD	4	0	1	28	0	0	0
31	VF	5	0	0	0	26	0	0
81	BI	3	0	0	0	0	78	0
11	B	4	0	0	0	1	1	5
	Sensitivity	93.09	90.91	95	80	89.66	91.76	100
	Specificity	93.12	99.05	98.91	99.19	99.2	99.47	99.07

**Table B.6:** Confusion matrix for multiclass classification using local Hölder exponents with our Hybrid-SVM approach considering threshold 0.9. See Table B.1 and Table B.4 to contrast.

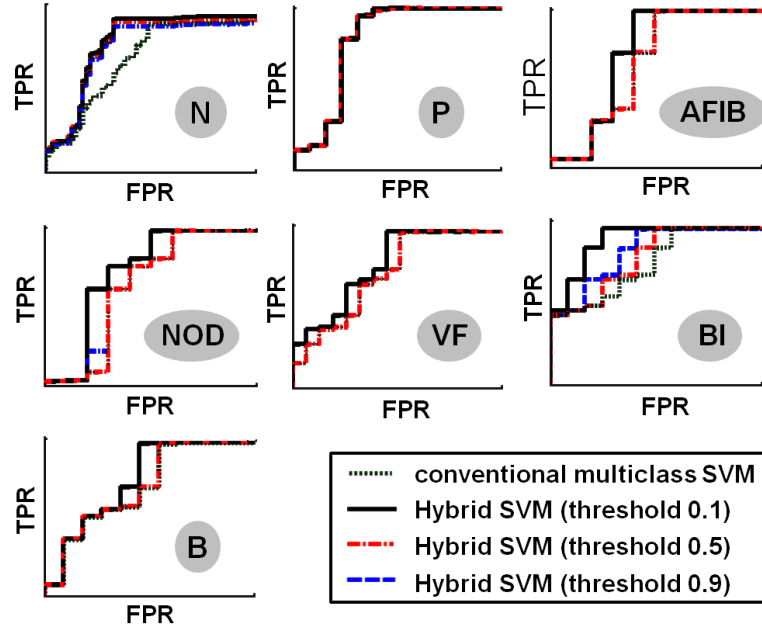
The values in the first column (which are due to the larger Normal class) are smaller after implementing our Hybrid-SVM approach. Furthermore, as the threshold ( $R$ ) reduces, the performance of our approach increases. Reduction in such false negatives (FN) is very important, especially in medicine, where a subject with an abnormality should be considered for appropriate the treatment.

The performance of our Hybrid-SVM approach can be further observed in the ROC plots shown in Figure B.5. The solid lines (our Hybrid-SVM with threshold 0.1) are above the dashed lines (our Hybrid-SVM with thresholds 0.5 and 0.9), which in turn are above the dotted lines (conventional multiclass SVM approach). The areas under the curves are as indicated in Table B.7, where we observe that the areas are higher in the case of Hybrid-SVM approach (particularly when the threshold value is 0.1).

## B.5.2 Performance of Hybrid-SVM for Enhanced Features

In Section B.4.1, we proposed the increased use of wavelet scales for the computation of the Local Hölder exponents. The 10-fold cross validation accuracy is  $93.73 \pm 3.3\%$ , which is much higher than the previous case (Section B.4) of computation using only scale 1. We provide a 7-class confusion matrix in Table B.8 where we also provide the





**Figure B.5:** Receiver operating curves for the considered seven classes (N, P, AFIB, NOD, VF, BI and B) in our database. The abscissa in the subfigures are different for better clarity. The curve for the Hybrid-SVM case (with threshold 0.1) is above all others, and hence significantly better.

	conventional	Hybrid-SVM with threshold		
	multiclass SVM	0.1	0.5	0.9
N	0.90	0.86	0.97	0.92
P	0.88	0.97	0.70	0.92
AFIB	0.86	0.98	0.92	0.88
NOD	0.98	0.79	0.92	0.86
VF	0.97	0.93	0.87	0.98
BI	0.79	0.91	0.83	0.97
B	0.91	0.87	0.98	0.70

**Table B.7:** The comparison of areas under ROC plots for the multiclass classification using local Hölder exponents.

results for classification using the Hybrid-SVM (numbers given in brackets). This table can be contrasted with Table B.1. We observe that our Hybrid-SVM approach further

increases the testing accuracy for the enhanced features. Finally we provide the MCC values in Table B.9.

Testing		Predicted classification						
Data	Classes	N	P	AFIB	NOD	VF	BI	B
369	N	354(354)	1(1)	2(2)	5(5)	2(2)	4(4)	1(1)
26	P	6(5)	20(21)	0(0)	0(0)	0(0)	0(0)	0(0)
101	AFIF	3(2)	0(0)	98(98)	0(0)	0(1)	0(0)	0(0)
33	NOD	4(3)	0(1)	0(0)	29(29)	0(0)	0(0)	0(0)
31	VF	6(5)	0(0)	0(0)	0(0)	25(26)	0(0)	0(0)
81	BI	1(1)	0(0)	0(0)	0(0)	0(0)	80(80)	0(0)
11	B	3(0)	0(0)	0(1)	0(0)	0(1)	1(1)	7(8)
	Sensitivity	93.9 (95.68)	95.24 (91.3)	98 (97.03)	85.29 (85.29)	92.59 (86.67)	94.12 (94.12)	87.5 (88.89)
	Specificity	94.55 (94.68)	99.05 (99.21)	99.46 (99.46)	99.35 (99.35)	99.04 (99.2)	99.82 (99.82)	99.38 (99.53)

**Table B.8:** *Confusion matrix for the multiclass classification using local Hölder exponents for scales '1 & 2'. Each entry corresponds to the conventional case, and the Hybrid-SVM (R=0.1) case (in brackets).*

The MCC values for our Hybrid-SVM approach are higher as compared to the corresponding conventional multiclass SVM (column) in all the combinations.

### B.5.3 Performance of Hybrid-SVM for Other Features

Using prior published results, we extracted features such as Fourier coefficients [135, 140], wavelet coefficients [140, 169], and auto-regressive (AR) coefficients [36]. These features were then subsequently used in the conventional multiclass SVM, and our Hybrid-SVM approach.

With respect to the training process, as mentioned earlier, the 10-fold cross validation accuracy using local Hölder exponents is  $90.89 \pm 2.1\%$ . In contrast, the 10-fold cross validation accuracies for the three methods were found to be  $88.93 \pm 2.8\%$ ,  $82.21 \pm 2.9\%$  and  $91.07 \pm 1.8\%$  respectively. These numbers vary from the ones reported earlier due

Classes	1S2L		2S2L		1S1L	
	Normal	Hybrid	Normal	Hybrid	Normal	Hybrid
N	0.86	<b>0.88</b>	0.88	<b>0.91</b>	0.75	<b>0.79</b>
P	0.83	<b>0.85</b>	0.85	<b>0.85</b>	0.76	<b>0.76</b>
AFIB	0.94	<b>0.94</b>	0.97	0.96	0.85	<b>0.88</b>
NOD	0.80	<b>0.82</b>	0.86	<b>0.86</b>	0.76	<b>0.80</b>
VF	0.86	<b>0.90</b>	0.86	<b>0.85</b>	0.87	<b>0.88</b>
BI	0.94	<b>0.94</b>	0.96	<b>0.96</b>	0.84	<b>0.86</b>
B	0.67	<b>0.85</b>	0.74	<b>0.80</b>	0.34	<b>0.34</b>

**Table B.9:** MCC values for the considered seven classes in different scenarios. ‘S’ indicates the number of scales used (1 or 2) for the computation of local Hölder exponents ‘L’ indicates the number of ECG leads used (1 or 2) in the feature extraction. The Hybrid-SVM used the threshold value of 0.1.

to the overlap between classes as seen in the confusion matrix. With respect to the testing process, we provide the confusion matrices in Tables B.10, B.11 and B.12.

Testing		Predicted classification						
Data	Classes	N	P	AFIB	NOD	VF	BI	B
369	N	358(361)	7(6)	4(2)	0(0)	0(0)	0(0)	0(0)
26	P	6(2)	18(20)	2(4)	0(0)	0(0)	0(0)	0(0)
101	AFIF	13(8)	1(3)	86(87)	1(1)	0(1)	0(1)	0(0)
33	NOD	0(0)	0(0)	1(1)	26(26)	4(4)	2(2)	0(0)
31	VF	1(1)	0(0)	0(0)	2(2)	22(22)	6(6)	0(0)
81	BI	3(2)	0(0)	1(1)	7(7)	1(2)	67(67)	2(2)
11	B	0(0)	0(0)	0(0)	2(2)	0(0)	3(3)	8(8)
Sensitivity		93.96 (96.52)	69.23 (68.97)	91.49 (91.58)	72.22 (72.22)	81.48 (75.86)	85.9 (84.81)	80 (80)
Specificity		95.94 (97.12)	98.72 (99.04)	97.31 (97.49)	98.86 (98.86)	98.56 (98.56)	97.56 (97.56)	99.53 (99.53)

**Table B.10:** Confusion matrix for the multiclass classification using Fourier coefficients [140]. Each entry corresponds to the conventional case, and the Hybrid-SVM (R=0) case (in brackets).

Testing		Predicted classification						
Data	Classes	N	P	AFIB	NOD	VF	BI	B
369	N	355(355)	5(5)	4(4)	1(1)	1(1)	1(1)	2(2)
26	P	15(13)	10(12)	1(1)	0(0)	0(0)	0(0)	0(0)
101	AFIF	7(3)	0(1)	90(93)	1(1)	0(0)	3(3)	0(0)
33	NOD	0(0)	0(0)	2(2)	24(24)	5(5)	2(2)	0(0)
31	VF	0(0)	0(0)	0(0)	5(3)	22(24)	4(4)	0(0)
81	BI	1(0)	0(0)	1(1)	11(11)	2(2)	64(65)	2(2)
11	B	0(0)	0(0)	0(0)	2(2)	0(0)	4(4)	5(5)
	Sensitivity	93.92 (95.69)	66.67 (66.67)	91.84 (92.08)	54.55 (57.14)	73.33 (75)	82.05 (82.28)	55.56 (55.56)
	Specificity	94.89 (95.02)	97.49 (97.79)	98.01 (98.55)	98.52 (98.52)	98.55 (98.87)	97.04 (97.21)	99.07 (99.07)

**Table B.11:** Confusion matrix for the multiclass classification using wavelet coefficients [140]. Each entry corresponds to the conventional case, and the Hybrid-SVM ( $R=0$ ) case (in brackets).

Testing		Predicted classification						
Data	Classes	N	P	AFIB	NOD	VF	BI	B
369	N	339(341)	10(10)	4(4)	0(0)	1(1)	15(13)	0(0)
26	P	2(2)	23(23)	1(1)	0(0)	0(0)	0(0)	0(0)
101	AFIF	3(3)	2(2)	96(96)	0(0)	0(0)	0(0)	0(0)
33	NOD	2(2)	0(0)	0(0)	31(31)	0(0)	0(0)	0(0)
31	VF	1(0)	0(0)	0(0)	0(0)	21(22)	9(9)	0(0)
81	BI	1(0)	0(0)	0(0)	0(0)	0(0)	80(81)	0(0)
11	B	11(2)	0(0)	0(0)	0(0)	0(0)	0(1)	0(8)
	Sensitivity	94.43 (97.43)	65.71 (65.71)	95.05 (95.05)	100 (100)	95.45 (95.65)	76.92 (77.88)	NaN (100)
	Specificity	89.76 (90.73)	99.51 (99.51)	99.09 (99.09)	99.68 (99.68)	98.41 (98.57)	99.82 (100)	98.31 (99.53)

**Table B.12:** Confusion matrix for the multiclass classification using AR coefficients [36]. Each entry corresponds to the conventional case, and the Hybrid-SVM ( $R=0$ ) case (in brackets).

Two points are worth observing. First, for each of these methods including the method with AR coefficients, the results obtained using our Hybrid-SVM approach (given in

brackets) are superior. Second, the success in our method is independent of which classes confuse the methods. For example, the classes (N & AFIB) cause confusion in the Fourier method, and the classes (N & AFIF), (VF & BI) cause confusion in the Wavelet method.

We also provide the comparison of MCC values in Table B.13 which compares the classification using all the above mentioned methods. It can be observed that irrespective of the method, the classification performance further increases when used with our Hybrid-SVM approach.

Classes	Hölder		Fourier		wavelet		AR coeff	
	Normal	Hybrid	Normal	Hybrid	Normal	Hybrid	Normal	Hybrid
N	0.80	<b>0.86</b>	0.89	<b>0.93</b>	0.88	<b>0.89</b>	0.84	<b>0.88</b>
P	0.87	0.85	0.67	<b>0.71</b>	0.49	<b>0.53</b>	0.75	<b>0.75</b>
AFIB	0.91	<b>0.93</b>	0.86	<b>0.86</b>	0.88	<b>0.90</b>	0.94	<b>0.94</b>
NOD	0.78	<b>0.79</b>	0.74	<b>0.74</b>	0.60	<b>0.63</b>	0.96	<b>0.96</b>
VF	0.80	<b>0.80</b>	0.74	0.72	0.70	0.68	0.79	<b>0.81</b>
BI	0.92	<b>0.94</b>	0.82	0.81	0.77	<b>0.77</b>	0.85	<b>0.86</b>
B	0.49	<b>0.67</b>	0.75	<b>0.75</b>	0.49	<b>0.49</b>	NaN	<b>0.85</b>

**Table B.13:** *MCC values for the considered seven classes in different scenarios.*

## B.5.4 Compatibility of Hybrid-SVM with Prior Solutions

As explained in Figure B.1(b), there are various levels at which the effect of imbalance in the data can be reduced for better classification. Our solution can be combined with any of these solutions for even better performance. In Table B.14 and B.15, we provide the combination of our Hybrid-SVM with one of the commonly used techniques of under- and over-sampling (respectively). Here, the data is preprocessed by under-sampling the majority class and over-sampling the minority classes (respectively) in order to create a balanced dataset (refer Figure B.1(b)). Further, as shown in Figure B.1(c), our novel Hybrid-SVM methodology can be used to improve the classification performance.

Testing		Predicted classification						
Data	Classes	N	P	AFIB	NOD	VF	BI	B
171	N	156 (156)	1 (1)	2 (2)	3 (3)	1 (1)	5 (5)	3 (3)
29	P	4 (2)	24 (26)	0 (0)	0 (0)	0 (0)	1 (1)	0 (0)
93	AFIF	4 (4)	0 (0)	89 (89)	0 (0)	0 (0)	0 (0)	0 (0)
51	NOD	7 (5)	0 (1)	3 (3)	41 (42)	0 (0)	0 (0)	0 (0)
21	VF	2 (2)	0 (0)	0 (0)	0 (0)	19 (19)	0 (0)	0 (0)
84	BI	4 (3)	0 (0)	0 (0)	0 (0)	0 (0)	80 (81)	0 (0)
15	B	2 (1)	0 (0)	1 (1)	0 (0)	1 (1)	3 (3)	8 (9)

**Table B.14:** Confusion matrix using the principle of under-sampling data and then applying the Hybrid-SVM approach. Each entry corresponds to the conventional case, and the Hybrid-SVM ( $R=0$ ) case (in brackets).

Testing		Predicted classification						
Data	Classes	N	P	AFIB	NOD	VF	BI	B
363	N	331 (331)	2 (2)	8 (8)	7 (7)	3 (3)	9 (9)	3 (3)
73	P	0 (0)	73 (73)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
105	AFIF	7 (6)	1 (1)	97 (97)	0 (1)	0 (0)	0 (0)	0 (0)
106	NOD	5 (5)	0 (0)	0 (0)	101 (101)	0 (0)	0 (0)	0 (0)
99	VF	0 (0)	0 (0)	0 (0)	0 (0)	99 (99)	0 (0)	0 (0)
147	BI	2 (0)	0 (0)	0 (0)	0 (0)	0 (0)	145 (147)	0 (0)
58	B	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	58 (58)

**Table B.15:** Confusion matrix using the principle of over-sampling data, and then applying Hybrid-SVM approach. Each entry corresponds to the conventional case, and the Hybrid-SVM ( $R=0$ ) case (in brackets).

In both the tables, the classification performance using the sampling technique is better than Table B.1. By combining it with our Hybrid-SVM approach, the false negatives further reduce. Therefore, our approach can improve the performance of the multiclass classifier independent of, or in combination with, other solutions on the imbalanced data.

## Conclusion

Our experiments show that different arrhythmias inherently carry distinct patterns. For classification, the method of the wavelet transform modulus maxima based local singularity features forms a potent combination with SVMs. Further, we showed that the combination of scales for the computation of local Hölder exponents increases the classification accuracy.

We have introduced the notion of Hybrid-SVM – a combination of binary and multiclass SVM – to increase the testing accuracy. This becomes particularly important in reducing the false negative predictions. Our black box approach is suitable whenever there is a class imbalance – a phenomenon commonly seen in biomedical signals, but also found elsewhere. In addition, the approach is also compatible with other solutions to the imbalanced data problem. Finally, while we have used three offline SVMs, classifier ensembles with a mix-match approach can also be used.

In building our Hybrid-SVM offline, our method uses a threshold parameter ( $R$ ). For purposes of illustration, we use 3 different values of  $R$  to prove the efficacy of our work. A systematic way of building the Hybrid-SVM would use a feedback system with various values of  $R$ , and pick the value that is most suitable for the problem at hand. We do not describe this here, except to comment that a particular value of  $R$  might reduce the number of false negatives at the cost of increasing the number of false positives, and the balance is best chosen by the user. In conjunction with  $R$ , and during the online phase, the system we describe uses a “knob” – namely a random number generator. This conjunction can, for instance, be completely eliminated if the user want to minimize the false negatives. On the flip side, the random number generator aids the use of the knob.

These investigations show immense potential benefit of the proposed classifier. We believe it can be easily extended to other biomedical as well as non-biomedical applications.

---

## Appendix C

# Example Pulse Waveforms

---

We present few examples of pulse signals from our database. In each example, we show the *Tridosha* pulse of 5 second long. The three *dosha*, namely *vata*, *pitta*, and *kapha* are displayed in red, green, and blue colors respectively. If an example shows only two colors, then it indicates negligible strength of the signal at the third *dosha* location.

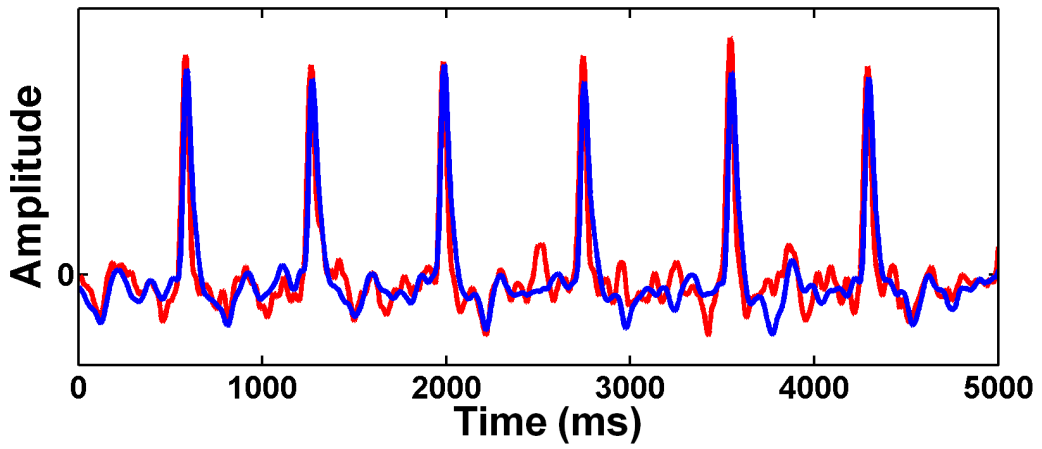
For each pulse signal, we provide the following details of the subject:

- **Behavior:** Healthy or unhealthy
- **Age group:** 1 (below 25 years), 2 (25–50 years), and 3 (above 50 years)
- **Gender:** Male or female

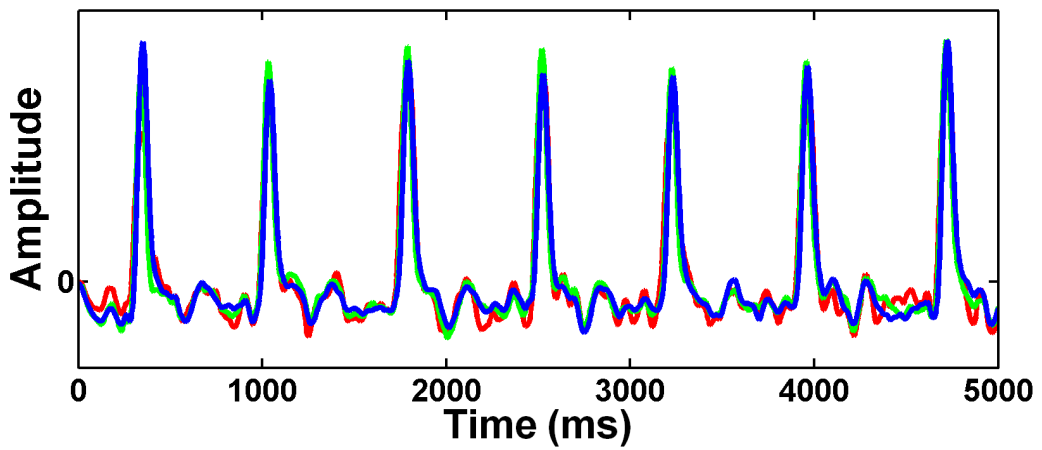
The examples (no specific order) are provided to visualize the varying patterns across many pulse signals. The variations in these patterns are captured using our approaches in Chapters 5 and 6.



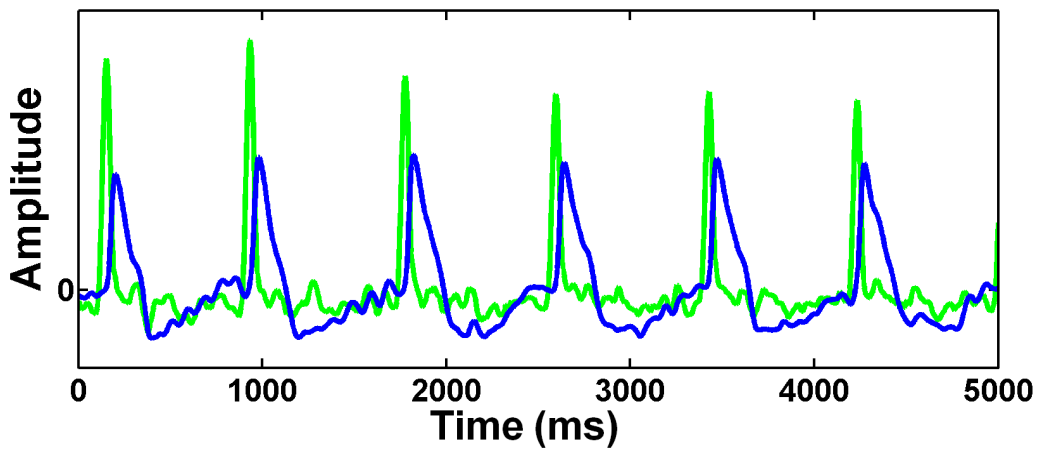
Behavior = Unhealthy, Age group = 2, Gender = Female.



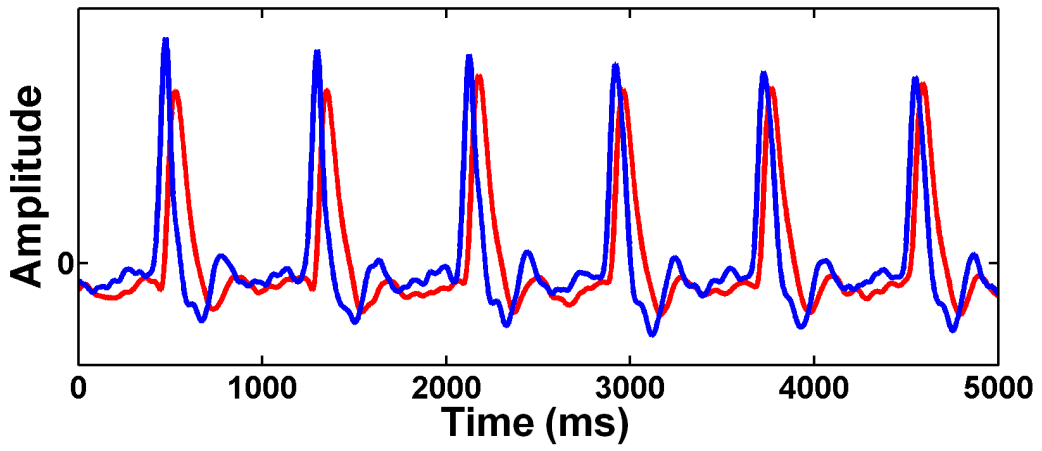
Behavior = Healthy, Age group = 2, Gender = Female.



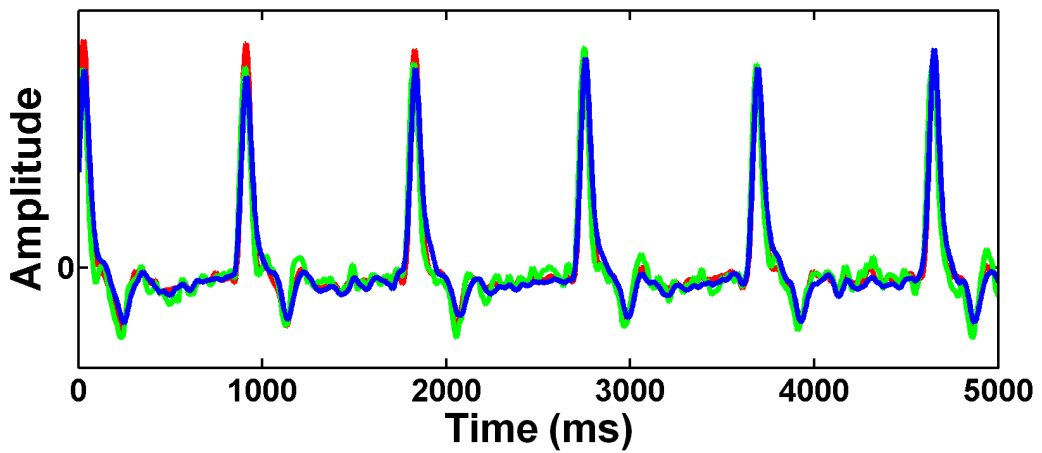
Behavior = Unhealthy, Age group = 2, Gender = Female.



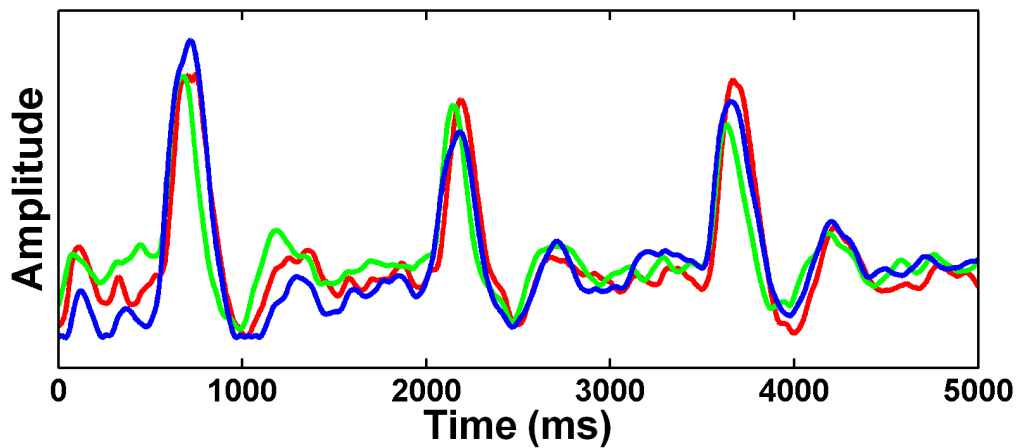
Behavior = Healthy, Age group = 2, Gender = Male.



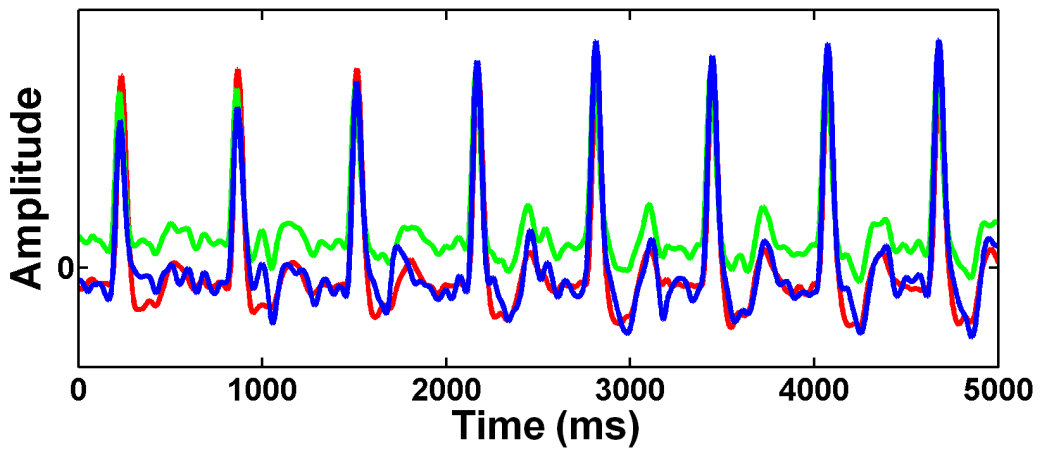
Behavior = Unhealthy, Age group = 3, Gender = Male.



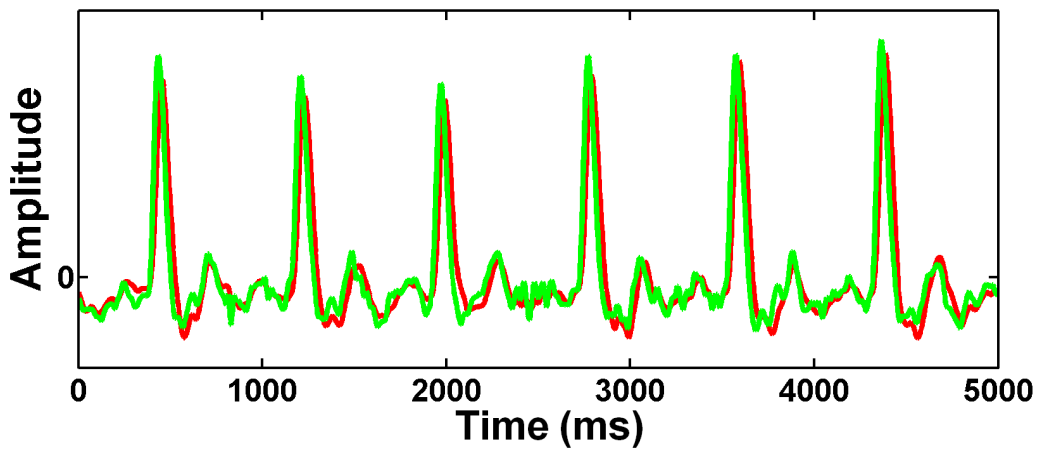
Behavior = Healthy, Age group = 2, Gender = Male.



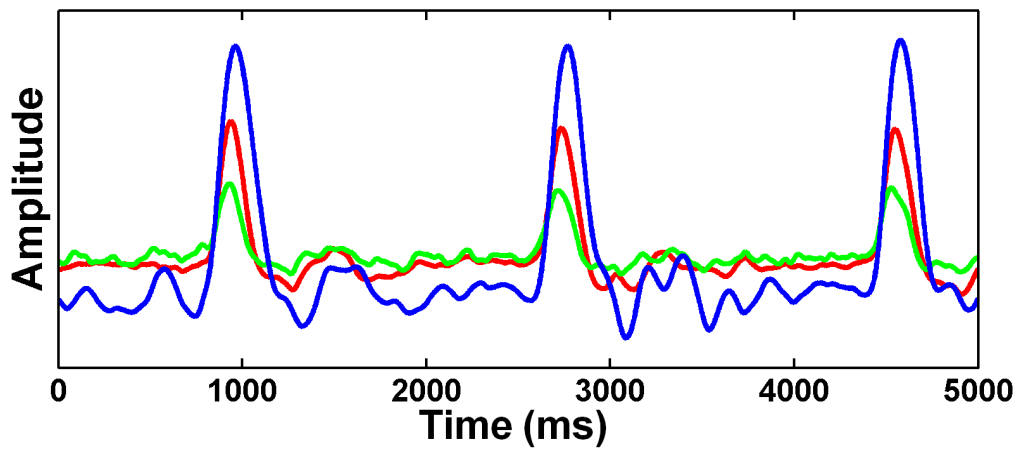
Behavior = Unhealthy, Age group = 2, Gender = Male.



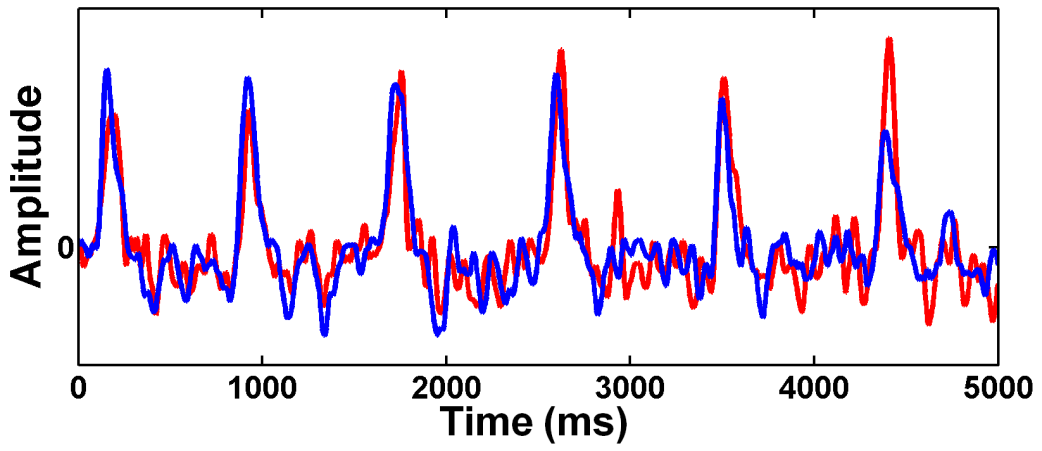
Behavior = Healthy, Age group = 2, Gender = Male.



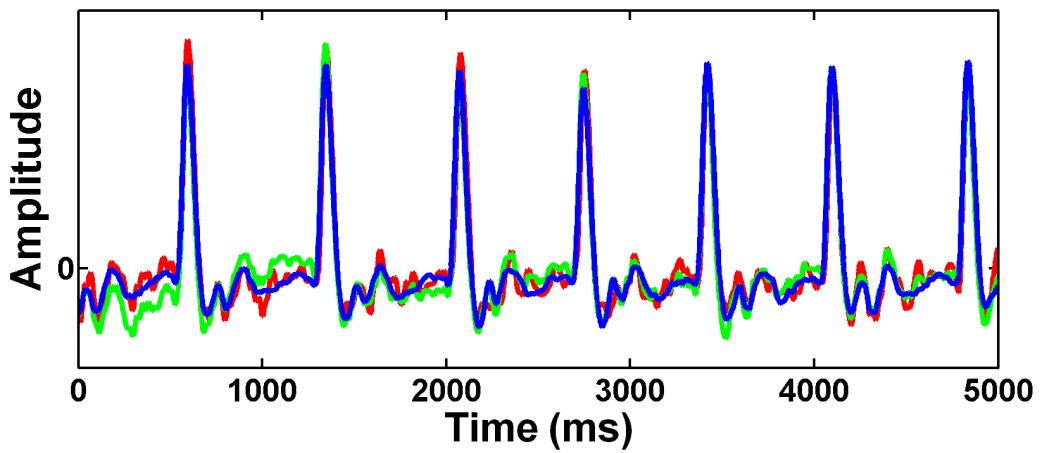
Behavior = Healthy, Age group = 2, Gender = Male.



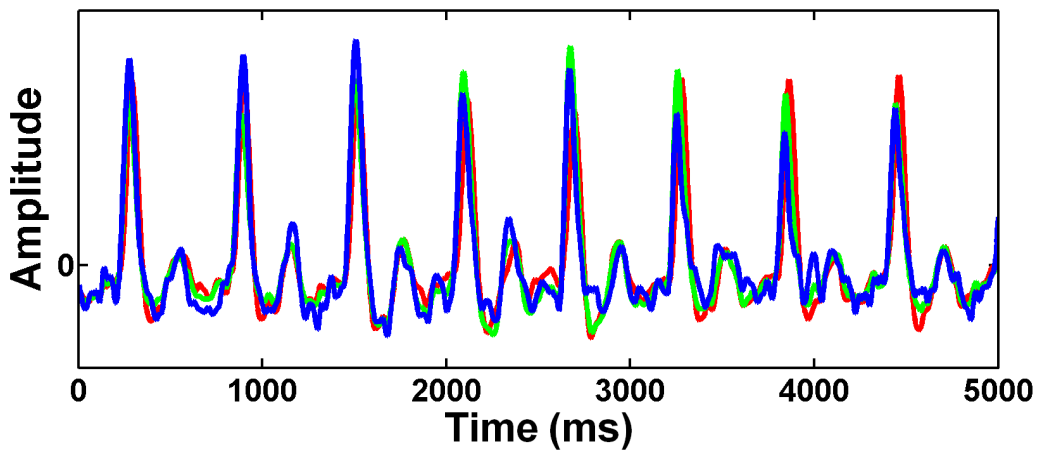
Behavior = Unhealthy, Age group = 3, Gender = Female.



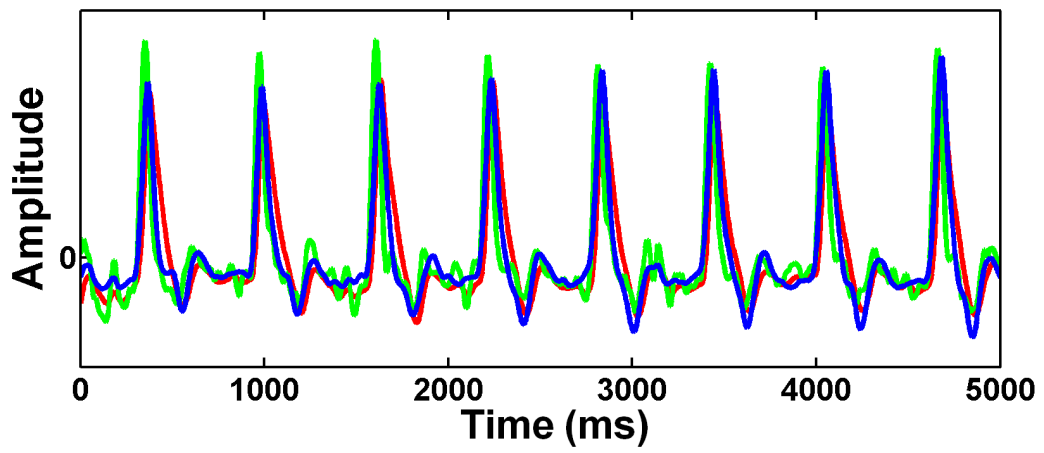
Behavior = Unhealthy, Age group = 2, Gender = Male.



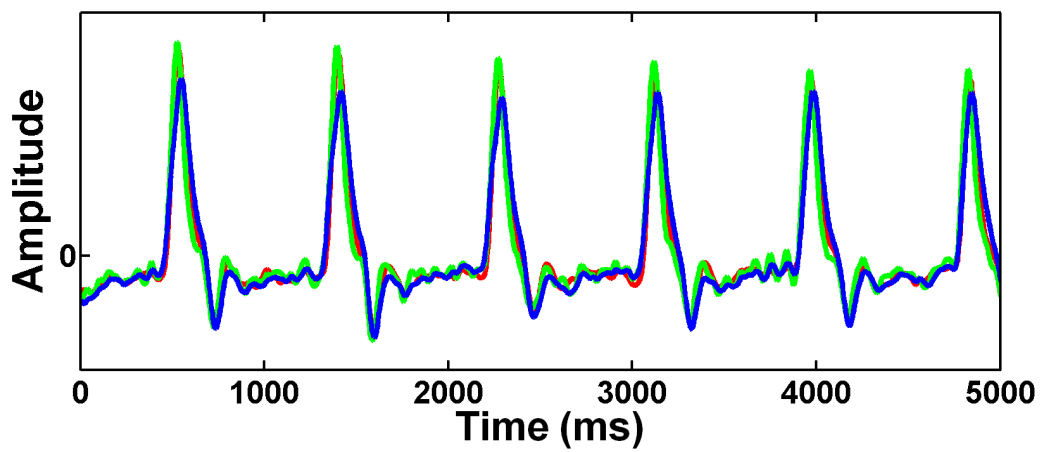
Behavior = Unhealthy, Age group = 2, Gender = Male.



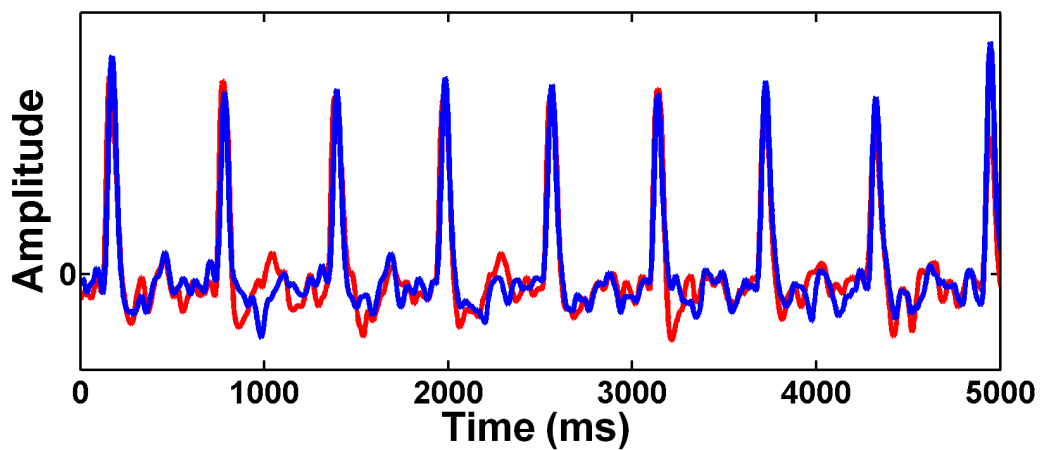
Behavior = Unhealthy, Age group = 3, Gender = Male.



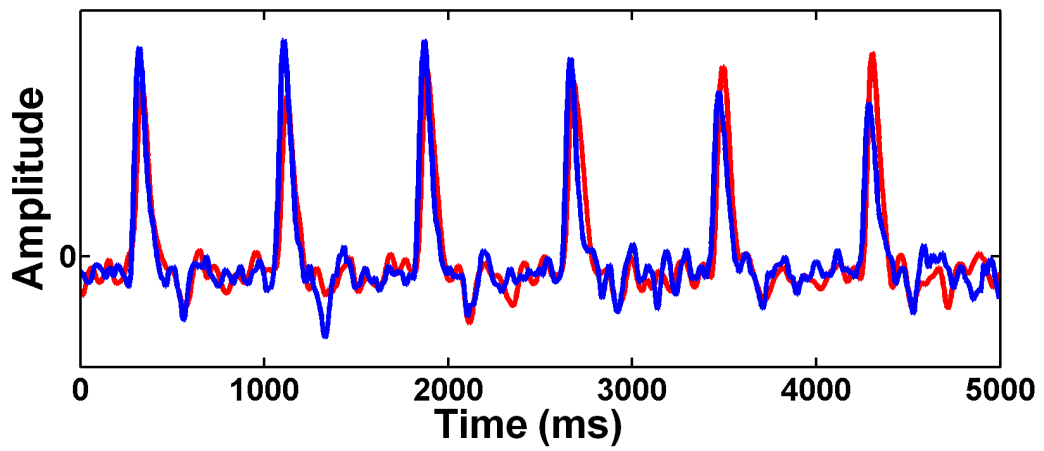
Behavior = Unhealthy, Age group = 3, Gender = Male.



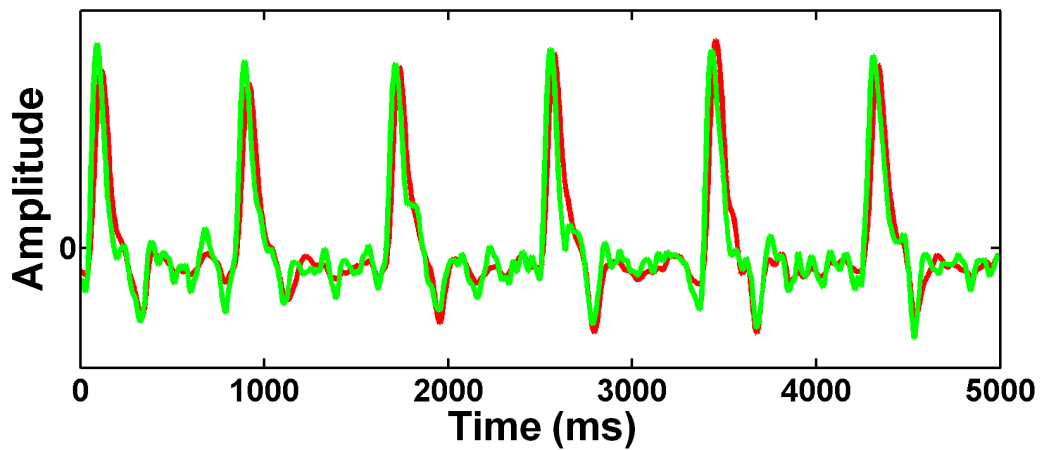
Behavior = Unhealthy, Age group = 1, Gender = Female.



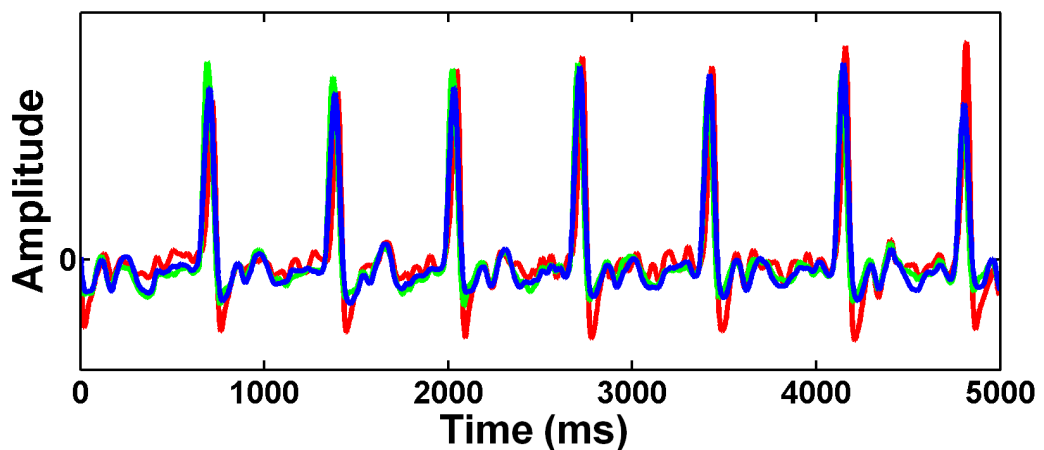
Behavior = Unhealthy, Age group = 3, Gender = Female.



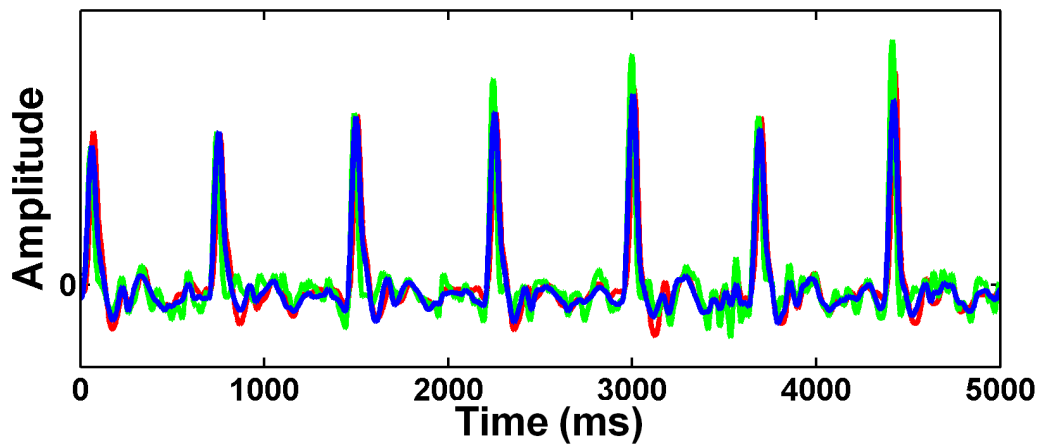
Behavior = Unhealthy, Age group = 3, Gender = Female.



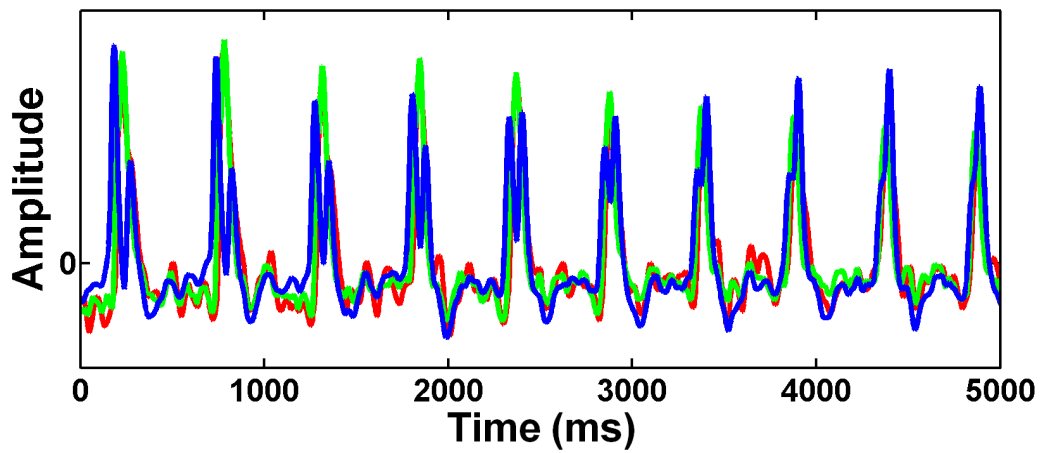
Behavior = Unhealthy, Age group = 1, Gender = Male.



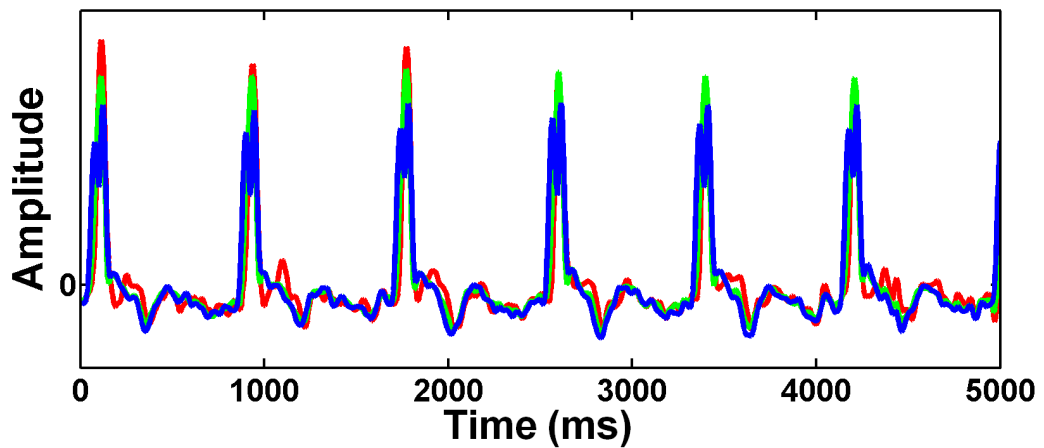
Behavior = Unhealthy, Age group = 1, Gender = Male.



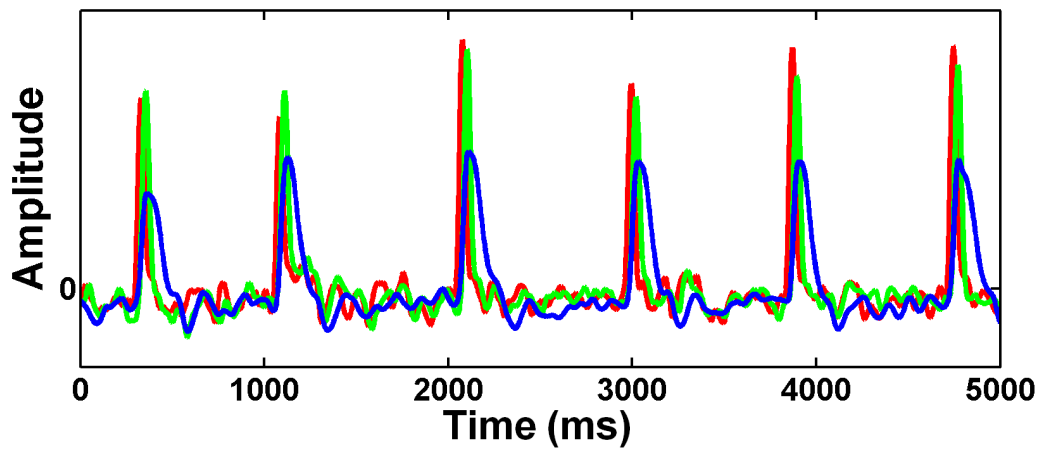
Behavior = Unhealthy, Age group = 1, Gender = Female.



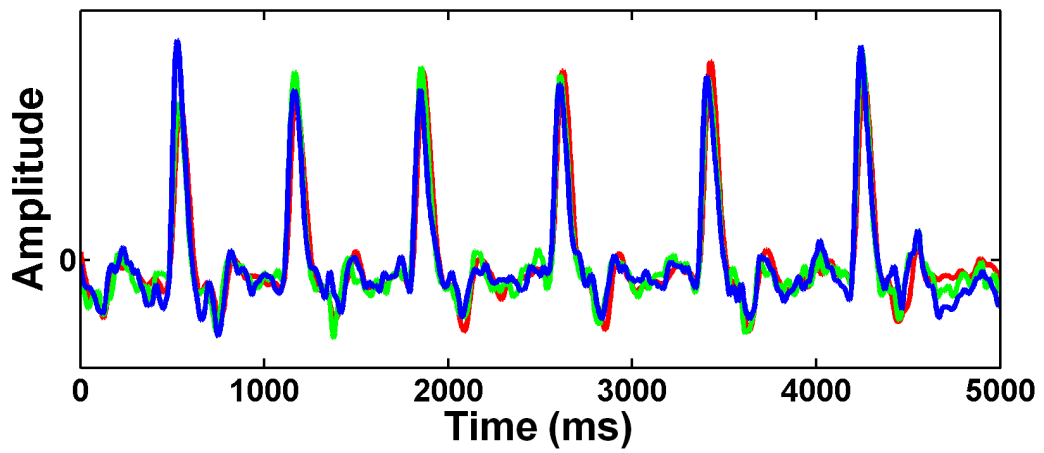
Behavior = Unhealthy, Age group = 3, Gender = Female.



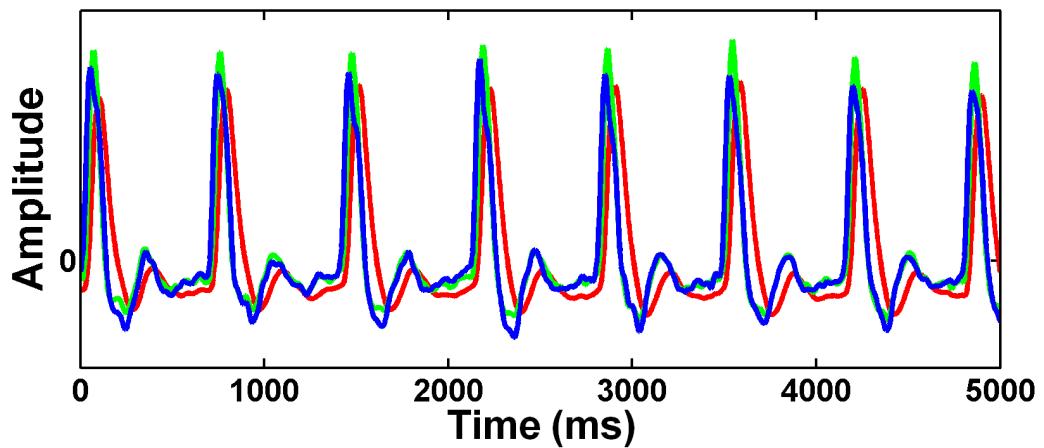
Behavior = Normal, Age group = 1, Gender = Male.



Behavior = Unhealthy, Age group = 3, Gender = Female.

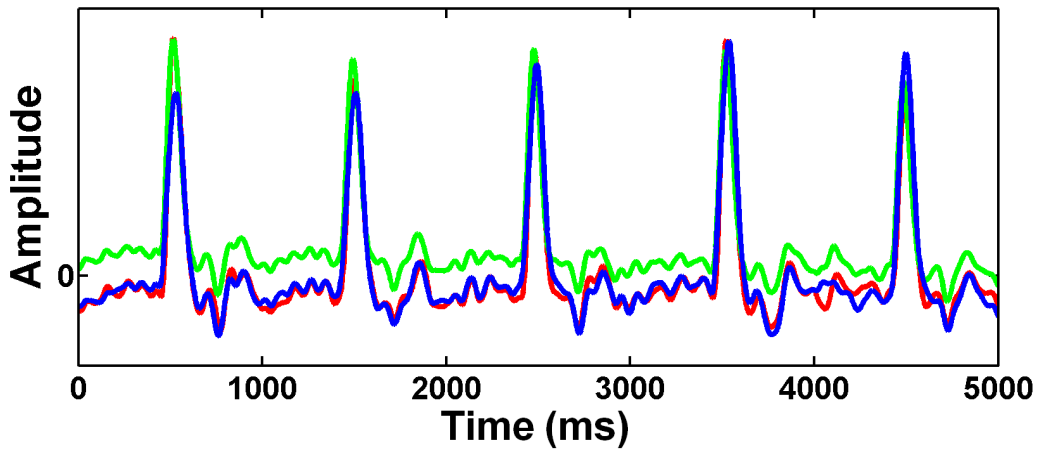


Behavior = Normal, Age group = 2, Gender = Male.

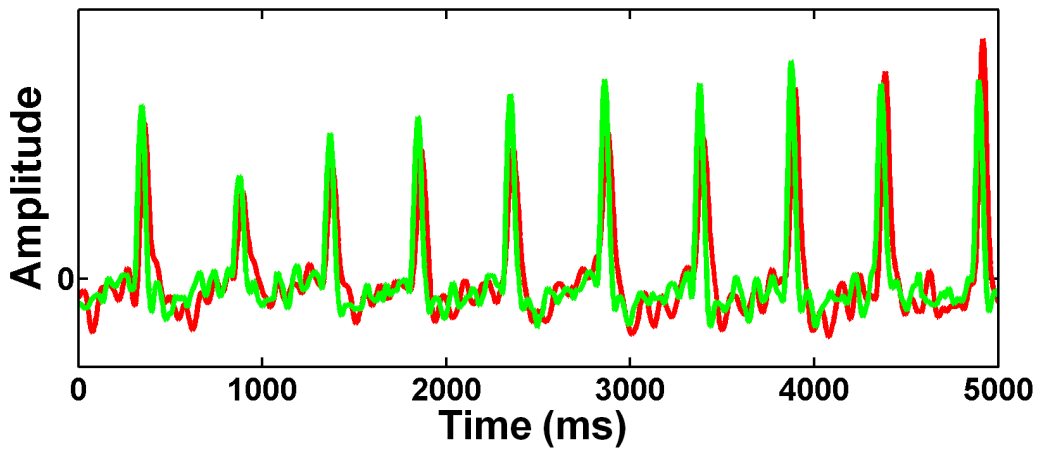




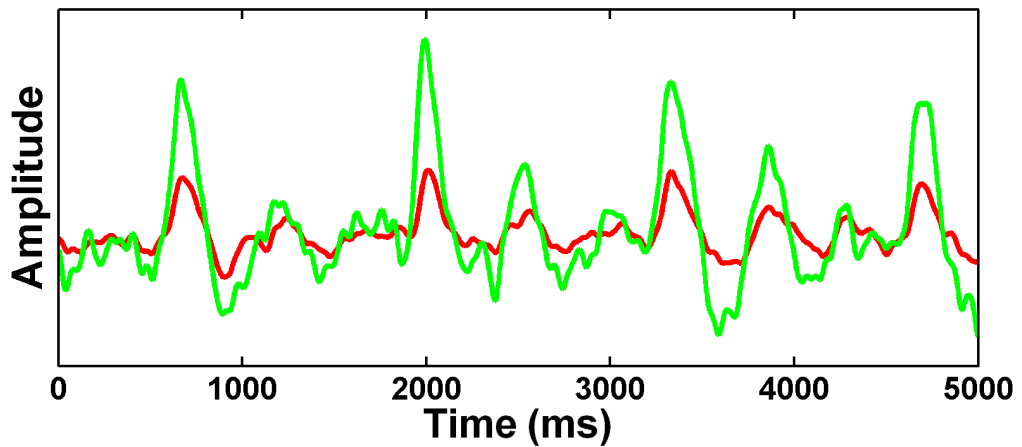
Behavior = Unhealthy, Age group = 2, Gender = Male.



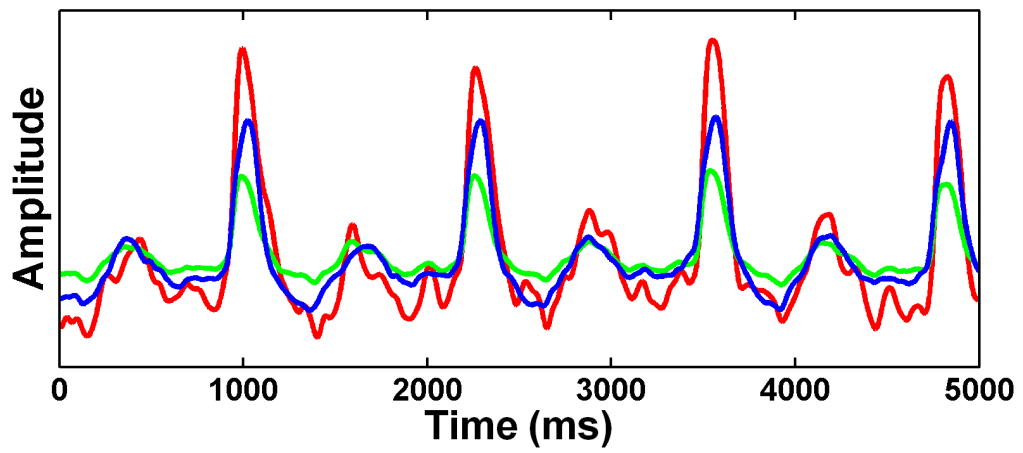
Behavior = Unhealthy, Age group = 1, Gender = Male.



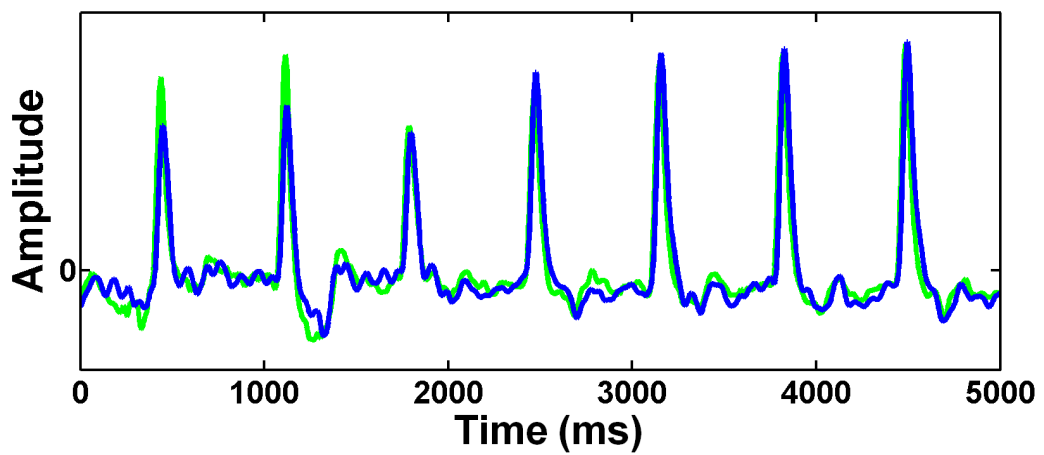
Behavior = Normal, Age group = 1, Gender = Male.



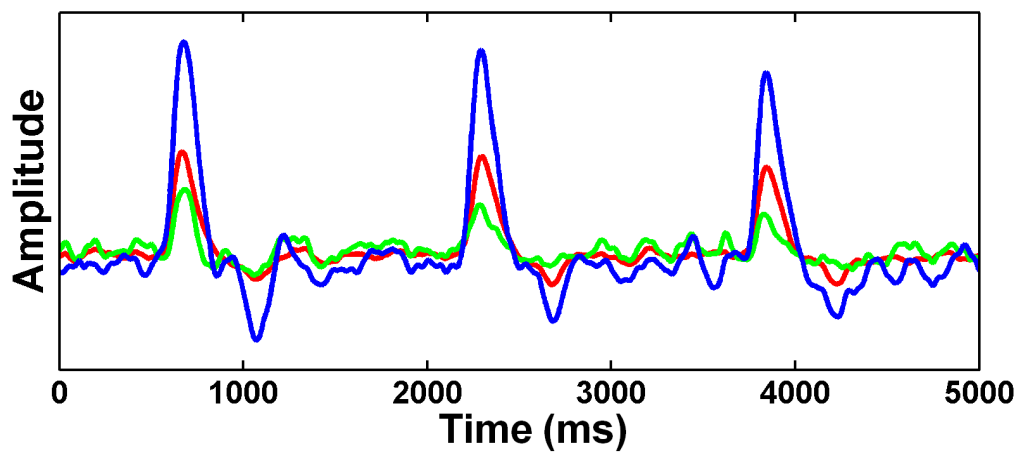
Behavior = Unhealthy, Age group = 1, Gender = Male.



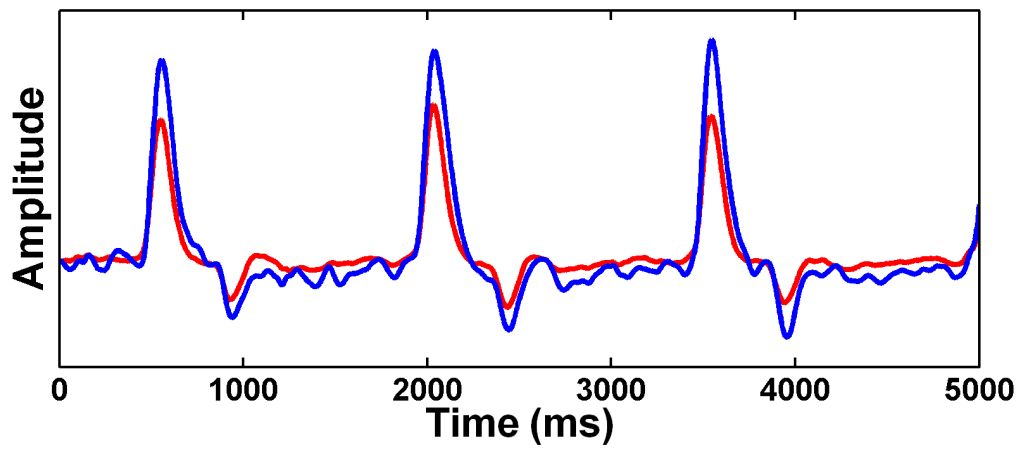
Behavior = Unhealthy, Age group = 2, Gender = Female.



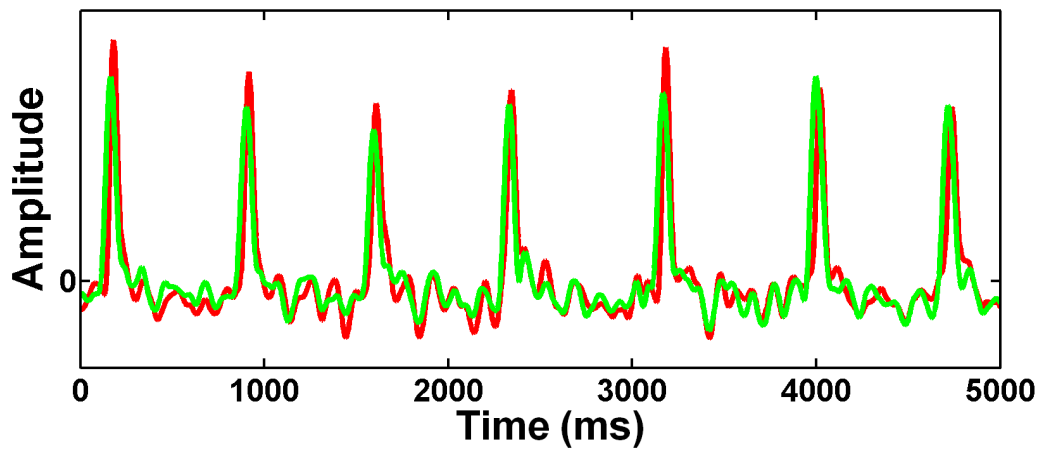
Behavior = Unhealthy, Age group = 3, Gender = Male.



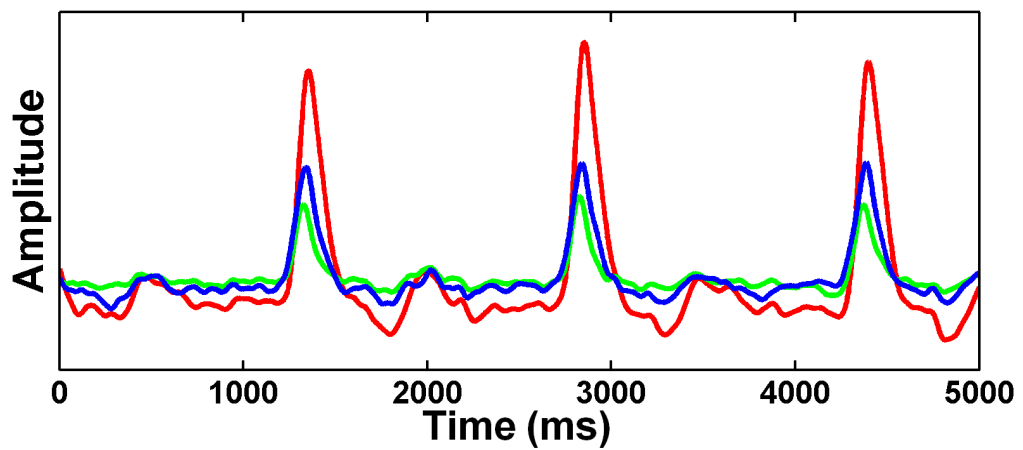
Behavior = Unhealthy, Age group = 3, Gender = Male.



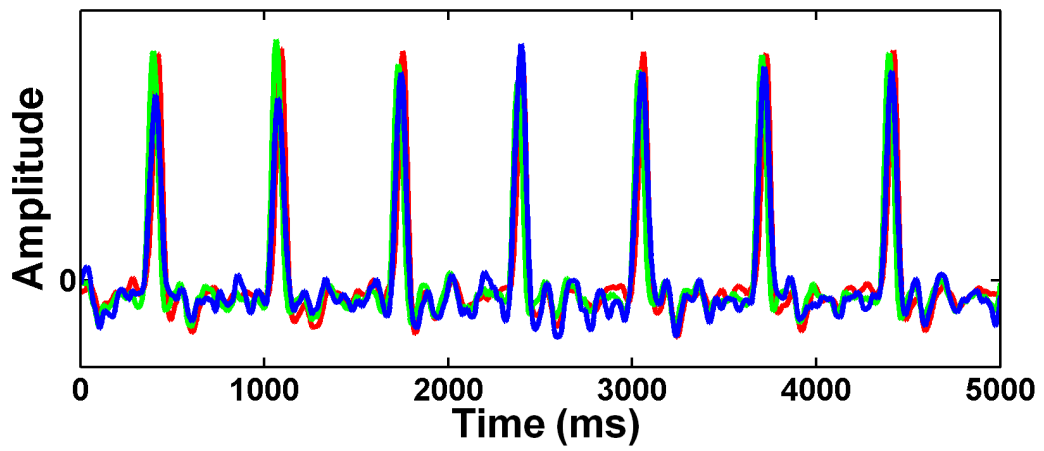
Behavior = Unhealthy, Age group = 1, Gender = Female.



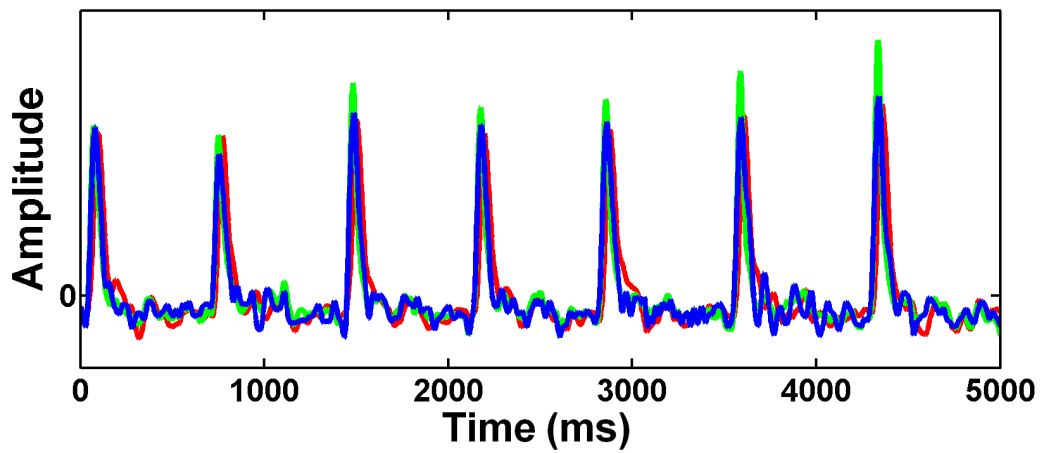
Behavior = Unhealthy, Age group = 2, Gender = Female.



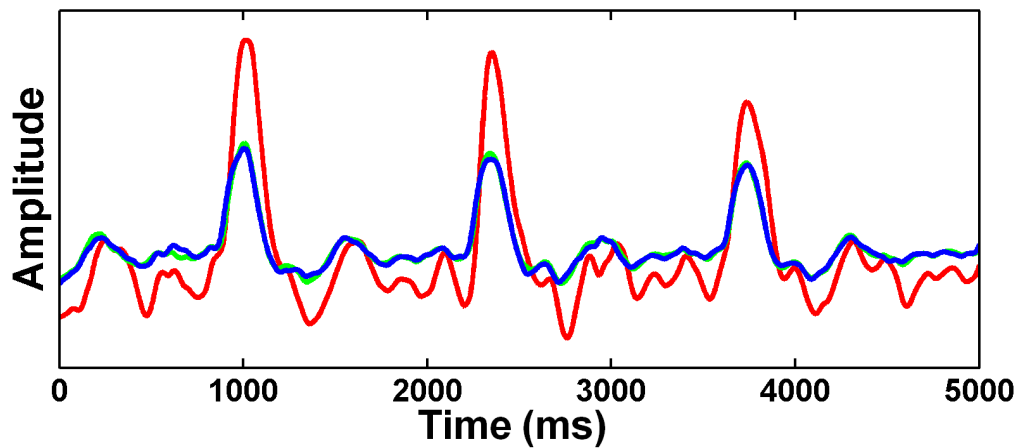
Behavior = Unhealthy, Age group = 2, Gender = Female.



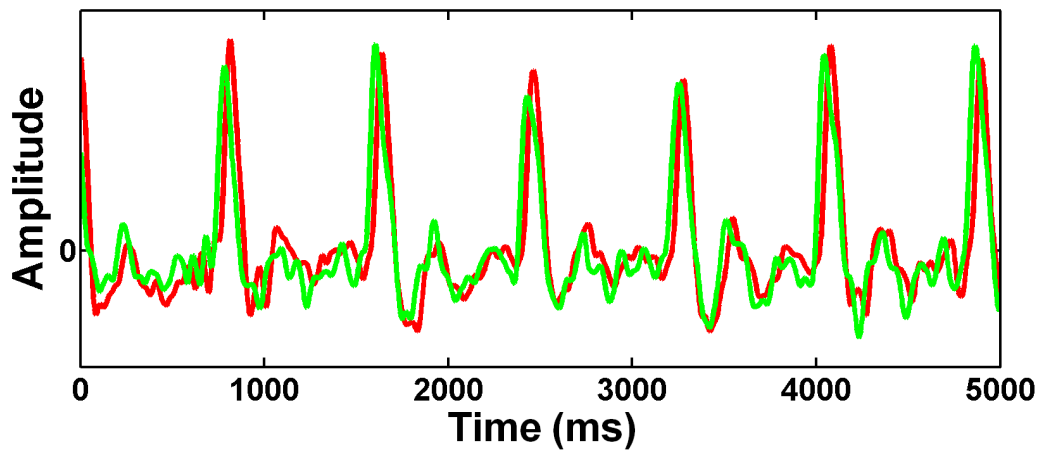
Behavior = Unhealthy, Age group = 2, Gender = Female.



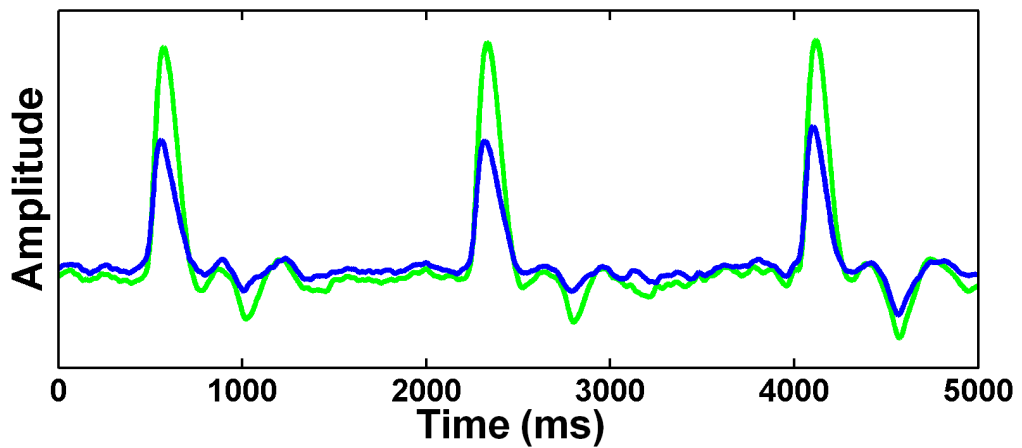
Behavior = Unhealthy, Age group = 2, Gender = Male.



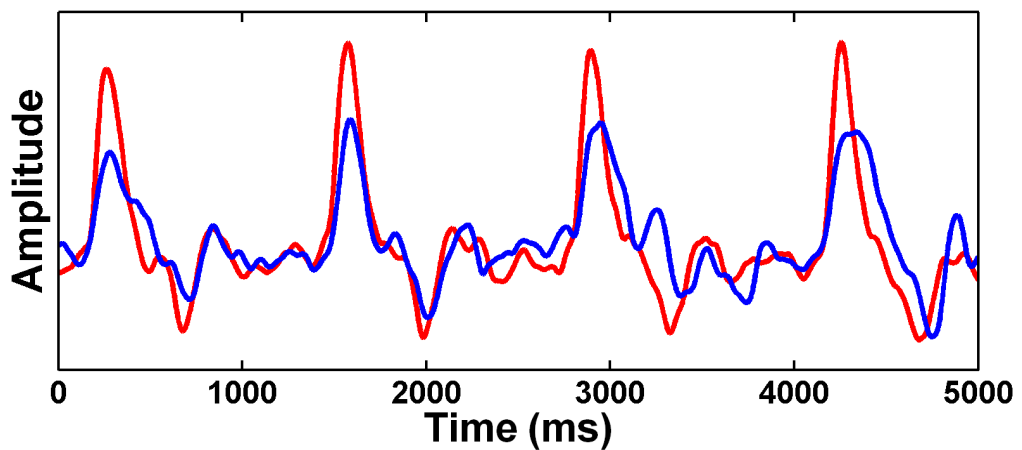
Behavior = Healthy, Age group = 2, Gender = Male.



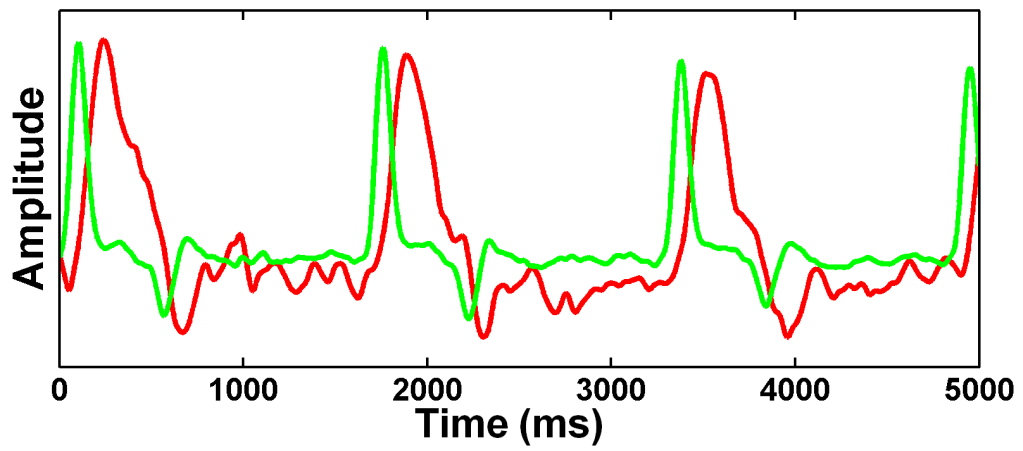
Behavior = Unhealthy, Age group = 2, Gender = Male.



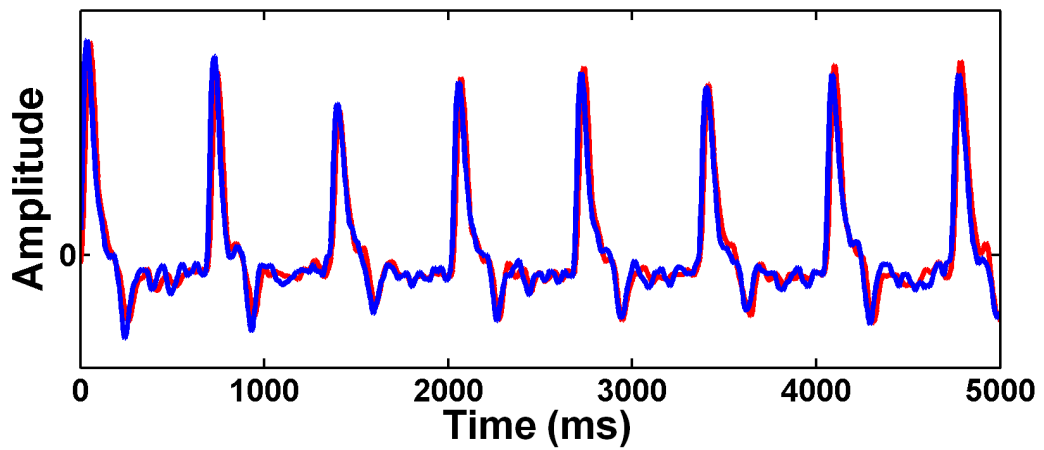
Behavior = Unhealthy, Age group = 2, Gender = Female.



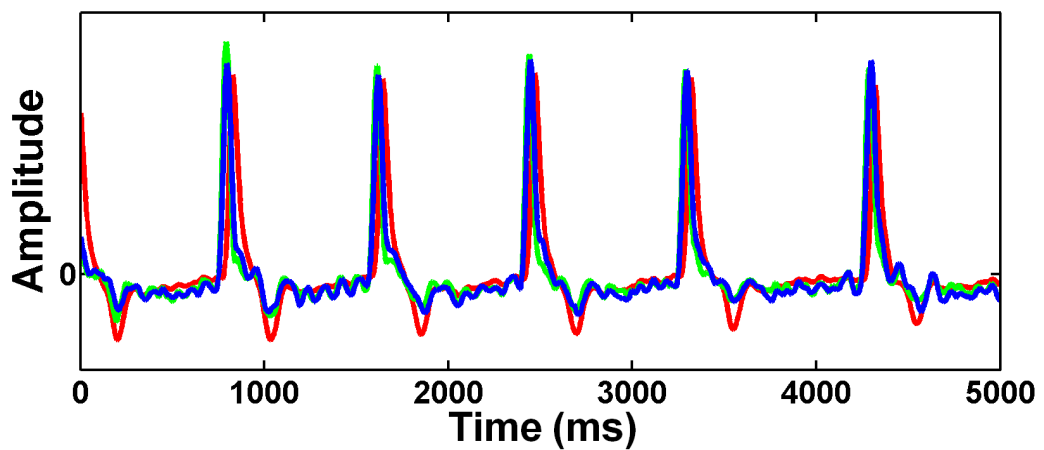
Behavior = Unhealthy, Age group = 2, Gender = Female.



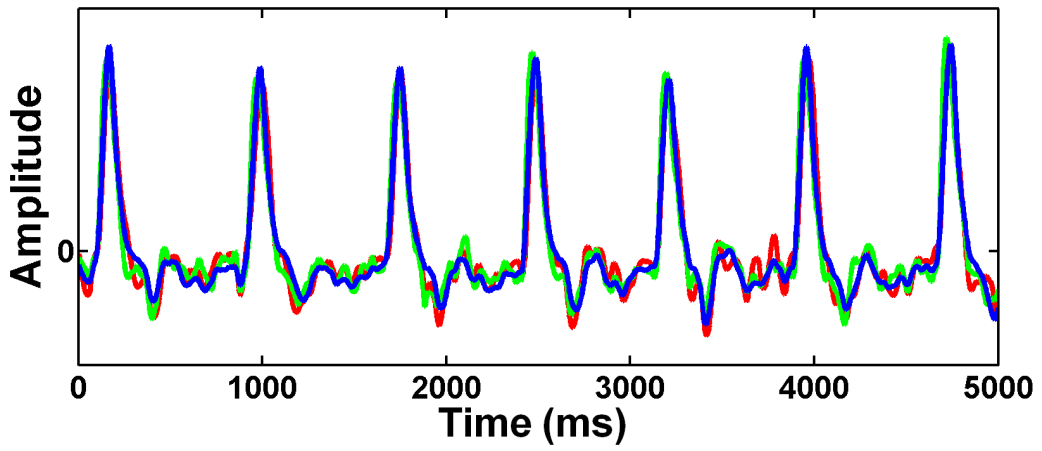
Behavior = Unhealthy, Age group = 3, Gender = Female.



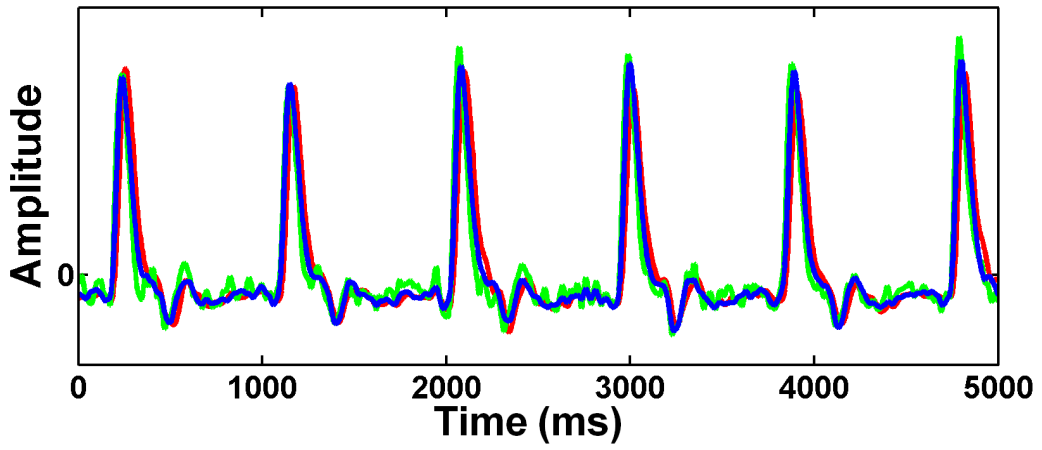
Behavior = Unhealthy, Age group = 3, Gender = Female.



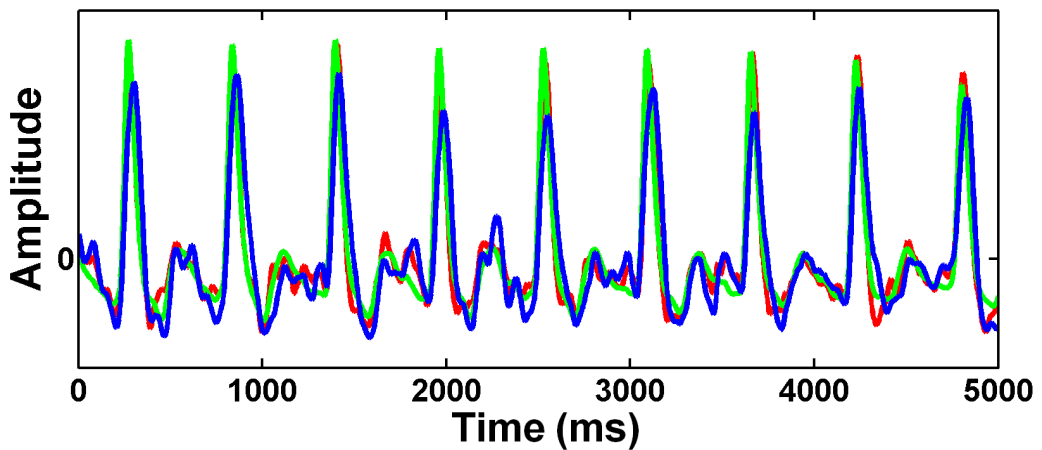
Behavior = Unhealthy, Age group = 3, Gender = Male.



Behavior = Unhealthy, Age group = 3, Gender = Male.



Behavior = Unhealthy, Age group = 2, Gender = Male.



---

## References

---

- [1] J. Aach and G. Church. Aligning gene expression time series with time warping algorithms. *Bioinformatics*, 17:495–508, 2001.
- [2] Abhinav, M. Sareen, M. Kumar, S. Anand, A. Salhan, and J. Santhosh. Nadi yantra : A robust system design to capture the signals from the radial artery for non-invasive diagnosis. In *2nd International Conference on Bioinformatics and Biomedical Engineering*, pages 1387–1390, 2008.
- [3] U. Acharya, N. Kannathal, L. Hua, and L. Yi. Study of heart rate variability signals at sitting and lying postures. *Journal of Bodywork and Movement Therapies*, 9:134–141, 2005.
- [4] R. Akbani, S. Kwek, and N. Japkowicz. Applying support vector machines to imbalanced datasets. *15th European Conference on Machine Learning*, 3201:39–50, 2004.
- [5] J. Allen and A. Murray. Age-related changes in peripheral pulse timing characteristics at the ears, fingers and toes. *Human Hypertension*, 16:711–717, 2002.
- [6] J. Allen and A. Murray. Age-related changes in the characteristics of the



- photoplethysmographic pulse shape at various body sites. *Physiological measurement*, 24:297–307, 2003.
- [7] E. Allwein, R. Schapire, and Y. Singer. Reducing multiclass to binary: A unifying approach for margin classifiers. In *17th International Conference on Machine Learning*, pages 9–16, 2000.
- [8] L. Amaral, P. Ivanov, N. Aoyagi, I. Hidaka, S. Tomono, A. Goldberger, H. Stanley, and Y. Yamamoto. Behavioral-independent features of complex heartbeat dynamics. *Physical Review Letters*, 86:6026–6029, 2001.
- [9] R. Andreao, B. Dorizzi, and J. Boudy. ECG signal analysis through hidden Markov models. *IEEE Transactions on Biomedical Engineering*, 53:1541–1549, 2006.
- [10] M. Aritomo, Y. Yonezawa, and W. Caldwell. A wrist-mounted activity and pulse recording system. In *1st Joint BMES/EMBS Conference*, volume 2, page 693, 1999.
- [11] A. Arneodo, E. Bacry, and J. Muzy. The thermodynamics of fractals revisited with wavelets. *Physica A*, 213:232–275, 1994.
- [12] J. Austin, A. Marie, and K. Pelletier. A review of the incorporation of complementary and alternative medicine by mainstream physicians. *Archives of Internal Medicine*, 158:2303–2310, 1998.
- [13] K. Belani, J. Buckley, and M. Poliac. Accuracy of radial artery blood pressure determination with the vasotrac. *Canadian Journal of Anesthesia*, 46:488–496, 1999.
- [14] J. Bigger, J. Fleiss, R. Steinman, L. Rolnitzky, R. Kleiger, and J. Rottman. Frequency domain measures of heart period variability and mortality after myocardial infarction. *Circulation*, 85:164–171, 1992.
- [15] J. Birman and R. Williams. Knotted periodic orbits in dynamical systems I: Lorenz’s equations. *Topology*, 22:47–82, 1983.
- [16] J. Birman and R. Williams. Knotted periodic orbits in dynamical systems II: Knot holders for fibered knots. *Contemporary Math*, 20:1–60, 1983.

- [17] C. Bratteli and S. Finkelstein. Analysis of aging effects on the arterial pulse contour using an artificial neural network. In *20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 20, pages 1333–1336, 1998.
- [18] E. Brien, B. Waeber, G. Parati, J. Staessen, and M. Myers. Blood pressure measuring devices: Recommendations of the european society of hypertension. *British Medical Journal*, 322:531–536, 2001.
- [19] A. Brumfield and M. Andrew. Digital pulse contour analysis: Investigating age-dependent indices of arterial compliance. *Physiological Measurement*, 26:599–608, 2005.
- [20] C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.
- [21] C.-C. Chang and C.-J. Lin. *LIBSVM: A library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [22] N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.
- [23] C. Chen, E. Nevo, B. Fetics, P. Pak, F. Yin, W. Maughan, and D. Kass. Estimation of central aortic pressure waveform by mathematical transformation of radial tonometry pressure: Validation of generalized transfer function. *Circulation*, 95:1827–1836, 1997.
- [24] W.-S. Chena, L. Hsiehb, and S.-Y. Yuan. High performance data compression method with pattern matching for biomedical ECG and arterial pulse waveforms. *Computer Methods and Programs in Biomedicine*, 74:11–27, 2004.
- [25] A. Chhabra, C. Meneveau, R. Jensen, and K. Sreenivasan. Direct determination of the  $f(\alpha)$  singularity spectrum and its application to fully developed turbulence. *Physical Review A*, 40:5284–5294, 1989.

- [26] N. Chong, I. Bumett, J. Chicharo, and M. Thomson. Use of the pitch synchronous wavelet transform as a new decomposition method for WI. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 513–516, 1998.
- [27] N. Chong, I. Burnett, and J. Chicharo. A new waveform interpolation coding scheme based on pitch synchronous wavelet transform decomposition. *IEEE Transactions on Speech and Audio Processing*, 8:345–348, 2000.
- [28] A. Cohen. Hidden Markov models in biomedical signal processing. In *20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 20, pages 1145–1150, 1998.
- [29] C.Jr. Contant, C. Robertson, J. Crouch, S. Gopinath, R. Narayan, and R. Grossman. Intracranial pressure waveform indices in transient and refractory intracranial hypertension. *Journal of Neuroscience Methods*, 57:15–25, 1995.
- [30] V. Corinoa, M. Matteuccib, L. Cravelloc, E. Ferraric, A. Ferrarid, and L. Mainardi. Long-term heart rate variability as a predictor of patient age. *Computer Methods and Programs in Biomedicine*, 82:248–257, 2006.
- [31] V. Creigen, L. Ferracina, A. Hlod, S. Mourik, K. Sjauw, V. Rottschfer, M. Vellekoop, and P. Zegeling. Modeling a heart pump. In *58th European Study Group with Industry*, pages 7–25, 2007.
- [32] N. Cristianini, A. Elisseeff, J. Shawe-Taylor, and J. Kandola. On kernel target alignment. In *Proceedings of Neural Information Processing Systems*, 2001.
- [33] D. Cysarz, S. Lange, P. Matthiessen, and P. Leeuwen. Regular heartbeat dynamics are associated with cardiac health. *American Journal of Physiology – Regulatory, Integrative and Comparative Physiology*, 292:R368–R372, 2007.
- [34] J. Dalen. Is integrative medicine the medicine of the future? *Archives of Internal Medicine.*, 159:2122–2126, 1999.
- [35] M. Dasrao, Y. Hock, and E. Sim. Diagnostic blood pressure wave analysis and ambulatory monitoring using a novel non-invasive portable device. *International*

- Conference on Biomedical Engineering Biovision*, pages 267–272, 2001.
- [36] G. De, N. Srinivasan, and S. Krishnan. Cardiac arrhythmia classification using autoregressive modeling. *BioMedical Engineering OnLine*, 1:1585–1588, 2002.
- [37] N. Dearden. Intracranial pressure monitoring. *Critical Care III*, pages 8–13, 1985.
- [38] A. Deary, A. Schumann, H. Murfet, S. Haydock, R. Foo, and M. Brown. Influence of drugs and gender on the arterial pulse wave and natriuretic peptide secretion in untreated patients with essential hypertension. *Clinical Science*, 103:493–499, 2002.
- [39] S. Dharmananda. <http://www.itmonline.org/arts/pulse.htm>. Last viewed on May 22, 2008.
- [40] M. Di Rienzo, P. Castiglioni, G. Mancia, G. Parati, and A. Pedotti. 24h sequential spectral analysis of arterial blood pressure and pulse interval in free-moving subjects. *IEEE Transactions on Biomedical Engineering*, 36:1066–1075, 1989.
- [41] D. Duprez, D. Kaiser, W. Whitwam, S. Finkelstein, A. Belalcazar, R. Patterson, S. Glasser, and J. Cohn. Determinants of radial artery pulse wave analysis in asymptomatic individuals. *American Journal of Hypertension*, 17:647–653, 2004.
- [42] J. Eckmann, S. Kamphorst, and D. Ruelle. Recurrence plots of dynamical systems. *Europhysics letters*, 4:973–977, 1987.
- [43] A. Eke, P. Herman, L. Kocsis, and L. Kozak. Fractal characterization of complexity in temporal physiological signals. *Physiological Measurement*, 23:R1–R38, 2002.
- [44] A. Elash. Move into hospital sector another sign of complementary medicine’s growing popularity. *Canadian Medical Association Journal*, 157:1589–1592, 1997.
- [45] A. Elloumi, Z. Lachiri, and N. Ellouze. Denoising ECG contaminated with EMG components based on pitch synchronous wavelet analysis. In *IEEE International Conference on Industrial Technology*, pages 1660–1663, 2004.

- [46] A. Elloumi, Z. Lachiri, and N. Ellouze. Pitch synchronous wavelet based fetal ECG extraction. In *1st International Symposium on Control, Communications and Signal Processing*, pages 239–242, 2004.
- [47] O. Escalona, L. Gavidia, and G. Wright. A robust procedure for P-wave detection and segmentation in highresolution 12-lead ECG. In *18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1365–1366, 1996.
- [48] G. Evangelista. Pitch-synchronous wavelet representations of speech and music signals. *IEEE Transactions on Signal Processing*, 41:3313–3329, 1993.
- [49] B. Fetics, E Nevo, C. Chen, and D. Kass. Parametric model derivation of transfer function for noninvasive estimation of aortic pressure by radial tonometry. *IEEE Transactions on Biomedical Engineering*, 46:698–706, 1999.
- [50] B. Flaws. *The secret of Chinese pulse diagnosis*. Blue Poppy Press, 1995.
- [51] E. Freis, W. Heath, P. Luhsinger, and R. Snell. Changes in the carotid pulse which occur with age and hypertension. *American Heart Journal*, 71:757–765, 1966.
- [52] E. Freis and M. Kyle. Computer analysis of carotid and brachial pulse waves: Effects of age in normal subjects. *American Journal of Cardiology*, 22:691–695, 1968.
- [53] H. Friedberg. Syncope during standby cardiac pacing. *British Heart Journal*, 31:281–284, 1969.
- [54] S. Fu and S. Lai. A system for pulse measurement and analysis of Chinese medicine. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 5, pages 1695–1696, 1989.
- [55] M. Fukui, Y. Kitagawa, N. Nakamura, S. Mogami, M. Ohnishi, C. Hirata, N. Ichio, K. Wada, K. Kamiuchi, M. Shigeta, M. Sawada, G. Hasegawa, and T. Yoshikawa. Augmentation of central arterial pressure as a marker of atherosclerosis in patients with type 2 diabetes. *Diabetes research and clinical practice*, 59:153–161, 2003.

- [56] J. Gao. Recurrence time statistics for chaotic systems and their applications. *Physical Review Letters*, 83:3178–3181, 1999.
- [57] J. Gao and H. Cai. On the structures and quantification of recurrence plots. *Physics Letters A*, 270:75–87, 2000.
- [58] M. Garfinkel, A. Singhal, W. Katz, D. Allan, R. Reshetar, and H. Schumacher. Yoga-based intervention for carpal tunnel syndrome: A randomized clinical trial. *The Journal of the American Medical Association*, 280:1601–1603, 1998.
- [59] C. Gatzka, B. Kingwell, and J. Cameron. Gender difference in the timing of arterial wave reflection beyond differences in body height. *Journal of Hypertension*, 19:2197–2203, 2001.
- [60] N. Gilardino, P. Lethrer, and R. Edelberg. Comparison of finger plethysmograph to ECG in the measurement of heart rate variability. *Psychophysiology*, 39:246–253, 2002.
- [61] R. Gilmore. Topological analysis of chaotic dynamical systems. *Reviews of Modern Physics*, 70:1455–1529, 1998.
- [62] A. Goldberger. Fractal variability versus pathologic periodicity: Complexity loss and stereotypy in disease. *Perspectives in Biology and Medicine*, 40:543–561, 1997.
- [63] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C.-K. Peng, and H. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101:e215–e220, 2000.
- [64] A. Goldberger, L. Amaral, J. Hausdorff, P. Ivanov, C. Peng, and H. Stanley. Fractal dynamics in physiology: Alterations with disease and aging. In *National Academy of Sciences, USA*, volume 99, pages 2466–2472, 2002.
- [65] G. Gratze, J. Fortin, and A. Holler. A software package for non-invasive, real-time beat-to-beat monitoring of stroke volume, blood pressure, total peripheral

- resistance and for assessment of autonomic function. *Computers in Biology and Medicine*, 28:121–142, 1998.
- [66] J. Guerrero, P. Lopez, J. Chorro, M. Martinez, G. Camps, R. Magdalena, A. Rosado, and J. Ampudia. Analysis of heart rate variability in diabetic patients affected by autonomic cardiovascular neuropathy. *Computers in Cardiology*, pages 241–244, 1999.
- [67] S. Gunn. Support vector machines for classification and regression. Technical Report ISIS-1-98, Department of Electronics and Computer Science, University of Southampton, 1998.
- [68] A. Guyton and J. Hall. *Textbook of Medical Physiology*. W.B. Saunders Company, 2005.
- [69] L. Hammer. *Chinese pulse diagnosis: A contemporary approach*. Eastland Press, 2001.
- [70] S. Havlin, L. Amaral, Y. Ashkenazy, A. Goldberger, P. Ivanov, C. Peng, and H. Stanley. Application of statistical physics to heartbeat diagnosis. *Physica A*, 274:99–110, 1999.
- [71] J. Hayano, A. Barros, A. Kamiya, N. Ohte, and F. Yasuma. Assessment of pulse rate variability by the method of pulse frequency demodulation. *BioMedical Engineering OnLine*, 4, 2005.
- [72] C. Hayward and R. Kelly. Gender-related differences in the central arterial pressure waveform. *American College of Cardiology*, 30:1863–1871, 1997.
- [73] R. Hermes, D. Geselowitz, and G. Oliver. Development, distribution, and use of the American Heart Association database for ventricular arrhythmia detector evaluation. *Computers in Cardiology*, pages 263–266, 1980.
- [74] A. Hertzman. Photoelectric plethysmograph of the finger and toes in man. In *The Society for Experimental Biology and Medicine*, volume 37, pages 1633–1637, 1937.

- [75] E. Hon and S. Lee. Electronic evaluations of the fetal heart rate patterns preceding fetal death, further observations. *American Journal of Obstetrics and Gynecology*, 87:814–826, 1965.
- [76] C. W. Hsu and C. J. Lin. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 13:415–425, 2002.
- [77] C.-W. Hsu and C.-J. Lin. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 13:415–425, 2002.
- [78] H. Huikuri, T. Makikallio, K. Airaksinen, R. Mitrani, A. Castellanos, and Myerburg R. Measurement of heart rate variability: A clinical tool or a research toy? *American College of Cardiology*, 34:1878–1883, 1999.
- [79] W.-L. Hwang and S. Mallat. Characterization of self-similar multifractals with wavelet maxima. *Applied and Computational Harmonic Analysis*, 1:316–328, 1994.
- [80] F. Itakura. Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 23:52–72, 1975.
- [81] P. Ivanov, L. Amaral, A. Goldberger, S. Havlin, M. Rosenblum, H. Stanley, and Z. Struzik. From  $1/f$  noise to multifractal cascades in heartbeat dynamics. *Chaos*, 11:641–652, 2001.
- [82] P. Ivanov, Z. Chen, K. Hu, and H. Stanley. Multiscale aspects of cardiac control. *Physica A*, 344:685–704, 2004.
- [83] P. Ivanov, M. Rosenblum, L. Amaral, Z. Struzik, S. Havlin, A. Goldberger, and H. Stanley. Multifractality in human heartbeat dynamics. *Nature*, 399:461–465, 1999.
- [84] A. Jade, V. Jayaraman, B. Kulkarni, A. Khopkar, V. Ranade, and A. Sharma. A novel local singularity distribution based method for flow regime identification: Gasliquid stirred vessel with rushton turbine. *Chemical Engineering Science*, 61:688–697, 2006.



- [85] N. Japkowicz. The class imbalance problem: Significance and strategies. In *International Conference on Artificial Intelligence*, volume 1, pages 111–117, 2000.
- [86] G. Jindal, S. Nerurkar, S. Pednekar, J. Babu, M. Kelkar, and A. Deshpande. Corrected formula for estimating peripheral blood flow by impedance plethysmography. *Computers in Biology and Medicine*, 32:625, 1994.
- [87] V. Johneff. Complex valued wavelet analysis for QRS detection in ECG signals. *Bioinformatics, Images, and Wavelets*, pages 134–136, 2004.
- [88] C. Jones, J. Goodfellow, R. Bleasdale, and M. Frenneaux. Modulation of interaction between left ventricular ejection and the arterial compartment: Assessment of aortic wave travel. *Heart Vessels*, 15:247–255, 2000.
- [89] P. Joseph, R. Acharya, D. Poo, J. Chee, L. Min, S. Iyengar, and H. Wei. Effect of reflexological stimulation on heart rate variability. *Innovations and Technology in Biology and Medicine (ITBM-RBM)*, 25:40–45, 2004.
- [90] R. Joshi. A biostatistical approach to Ayurveda: Quantifying the tridosha. *The Journal of Alternative and Complementary Medicine*, 10:879–889, 2004.
- [91] T. Kageyama, M. Kabuto, T. Kaneko, and N. Nishikido. Accuracy of pulse rate variability parameters obtained from finger plethysmogram: A comparison with heart rate variability parameters obtained from ECG. *Journal of Occupational Health*, 39:154–155, 1997.
- [92] P. Kaluzny and R. Tarnecki. Recurrence plots of neuronal spike trains. *Biological Cybernetics*, 68:527–534, 1993.
- [93] M. Karamanoglu, M. O'Rourke, A. Avolio, and R. Kelly. An analysis of the relationship between central aortic and peripheral upper limb pressure waves in man. *European Heart Journal*, 14:160–167, 1993.
- [94] R. Kelly, C. Hayward, A. Avolio, and M. O'Rourke. Noninvasive determination of age-related changes in the human arterial pulse. *Circulation*, 80:1652–1659, 1989.

- [95] E. Keogh and M. Pazzani. Scaling up dynamic time warping for data mining applications. In *6th International Conference on Knowledge Discovery and Data Mining*, pages 285–289, 2000.
- [96] E. Keogh and M. Pazzani. Derivative dynamic time warping. In *1st SIAM International Conference on Data Mining*, 2001.
- [97] A. Khir and K. Parker. Measurements of wave speed and reflected waves in elastic tubes and bifurcations. *Journal of Biomechanics*, 35:775–783, 2002.
- [98] B. Kijssirikul and N. Ussivakul. Multiclass support vector machines using adaptive directed acyclic graph. In *IEEE/INNS International Joint Conference on Neural Networks*, volume 1, pages 980–985, 2002.
- [99] Y. Kim, R. Britto, and T. Kesavadas. Diagnostics of arterial pressure pulse using haptic kymograph: Remote diagnosis of vital signs through a telehaptic device. In *1st Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 539–540, 2005.
- [100] R. Kleiger, J. Miller, J. Bigger, and A. Moss. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *American Journal of Cardiology*, 59:256–262, 1987.
- [101] S. Krimi, K. Ouni, and N. Ello. Pitch synchronous wavelet and hidden semi-Markov models based method for ECG analysis. In *European Signal Processing Conference*, pages 1407–1411, 2007.
- [102] A. Kulkarni, V. Jayaraman, and B. Kulkarni. Support vector classification with parameter tuning assisted by agent based techniques. *Computers and Chemical Engineering*, 28:311–318, 2004.
- [103] P. Kumar, A. Gupta, Rajshekhar, V. Jayaraman, and B. Kulkarni. Aligning time series with genetically tuned dynamic time warping algorithm. *Advances in Metaheuristics for Hard Optimization*, pages 251–261, 2008.

- [104] V. Lad. *Secrets of the pulse: The ancient art of Ayurvedic pulse diagnosis*. Motilal Banarsidass, Delhi, 2005.
- [105] O. Lau and A. Chwang. Relationship between wrist-pulse characteristics and body conditions. *14th Engineering Mechanics Conference*, 2000.
- [106] H. Lee and S. Choi. PCA+HMM+SVM for EEG pattern classification. In *7th International Symposium on Signal Processing and Its Applications*, volume 1, pages 541–544, 2003.
- [107] H. Lee, S. Suzuki, Y. Adachi, and M. Umeno. Fuzzy theory in traditional Chinese pulse diagnosis. *International Joint Conference on Neural Networks*, 1:774–777, 1993.
- [108] Y. Lee, H. Han, and J. Kim. Influence of motion artifacts on photoplethysmographic signals for measuring pulse rates. In *International Conference on Control, Automation and Systems*, pages 962–965, 2008.
- [109] P. Leonard, T. Beattie, P. Addison, and J. Watson. Wavelet analysis of pulse oximeter waveform permits identification of unwell children. *Journal of Emergency Medicine*, 21:59–60, 2004.
- [110] B. Li, M. Dong, and M. Vai. Pulse signal monitoring and analysis for home healthcare. In *International Summer School and Symposium on Medical Devices and Biosensors*, pages 53–56, 2006.
- [111] N. Linton and R. Linton. Estimation of changes in cardiac output from the arterial blood pressure waveform in the upper limb. *British Journal of Anaesthesia*, 86:486–496, 2001.
- [112] W. Lu, Y. Wang, and W. Wang. Pulse analysis of patients with severe liver problems. *IEEE Engineering in Medicine and Biology Magazine*, 18:73–75, 1999.
- [113] S. Lukman, Y. He, and S. Hui. Computational methods for traditional Chinese medicine: A survey. *Computer Methods and Programs in Biomedicine*, 88:283–294, 2007.

- [114] J. Lygouras and P. Tsalides. Optical-fiber finger photo-plethysmograph using digital techniques. *IEEE Sensors Journal*, 2:20–25, 2002.
- [115] S. Mahesh, M. Manivannan, and T. Anandan. Three radial artery pulse sensor design for Siddha based disease diagnosis. *International Journal of Systemics, Cybernetics and Informatics*, pages 19–23, 2008.
- [116] S. Mahesh, M. Manivannan, and T. Anandan. Design of pulse sensor for Siddha based disease diagnosis. In *International Conference of Research Into Design*, pages 419–426, 2009.
- [117] M. Malik, T. Farrell, T. Cripps, and A. Camm. Heart rate variability in relation to prognosis after myocardial infarction: Selection of optimal processing techniques. *European Heart Journal*, 10:1060–1074, 1989.
- [118] J. Malmivuo and R. Plonsey. *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, USA, 1995.
- [119] S. Manchanda, R. Narang, K. Reddy, U. Sachdeva, D. Prabhakaran, S. Dharmanand, M. Rajani, and R. Bijlani. Retardation of coronary atherosclerosis with yoga lifestyle intervention. *Journal of Association of Physicians of India*, 48:687–694, 2000.
- [120] J. Martin, L. Volfson, V. Kirzon-Zolin, and V. Schukin. Application of pattern recognition and image classification techniques to determine continuous cardiac output from the arterial pressure waveform. *IEEE Transactions on Biomedical Engineering.*, 41:913–920, 1994.
- [121] N. Marwan and J. Kurths. Nonlinear analysis of bivariate data with cross recurrence plots. *Physics Letters A*, 302:299–307, 2002.
- [122] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths. Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Physical Review E*, 66:026702, 2002.

- [123] B. Mathews. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta*, 405:442–451, 1975.
- [124] M. Mathews, J. Miller, and E. Jr. David. Pitch synchronous analysis of voiced sound. *The Journal of the Acoustical Society of America*, 33:179–186, 1961.
- [125] D. McDonald and M. Taylor. Hydrodynamics of the arterial circulation. *Progress in Biophysics and Biophysical Chemistry*, 9:107–173, 1959.
- [126] G. McVeigh. Pulse waveform analysis and arterial wall properties. *Hypertension*, 41:1010–1011, 2003.
- [127] G. McVeigh, C. Bratteli, D. Morgan, C. Alinder, S. Glasser, S. Finkelstein, and J. Cohn. Age-related abnormalities in arterial compliance identified by pressure pulse contour analysis: Aging and arterial compliance. *Hypertension*, 33:1392–1398, 1999.
- [128] G. McVeigh, D. Morgan, S. Finkelstein, L. Lemay, and J. Cohn. Vascular abnormalities associated with long-term cigarette smoking identified by arterial waveform analysis. *American Journal of Medicine*, 102:227–231, 1997.
- [129] M. Meyer and O. Stiedl. Self-affine fractal variability of human heartbeat interval dynamics in health and disease. *European Journal of Applied Physiology*, 90:305–316, 2003.
- [130] U. Meyerfeldt, N. Wessel, H. Schtt, D. Selbig, A. Schumann, A. Voss, J. Kurths, Ch. Ziehmann, R. Dietz, and A. Schirdewan. Heart rate variability before the onset of ventricular tachycardia: Differences between slow and fast arrhythmias. *International journal of cardiology*, 84:141–151, 2002.
- [131] S. Millasseau, F. Guigui, R. Kelly, K. Prasad, J. Cockcroft, and J. Ritter. Noninvasive assessment of the digital volume pulse: Comparison with the peripheral pressure pulse. *Hypertension*, 36:952–956, 2000.
- [132] S. Millasseau, S. Patel, S. Redwood, J. Ritter, and P. Chowienczyk. Pressure wave reflection assessed from the peripheral pulse is a transfer function necessary?

- Hypertension*, 41:1016–1020, 2003.
- [133] W. Milnor. *Hemodynamics*. Williams and Wilkins, 1982.
- [134] W.R. Milnor. *Cardiovascular Physiology*. Oxford Press, 1990.
- [135] K. Minami, Y. Ohkuma, H. Nakajima, and T. Toyoshima. Real-time ventricular arrhythmia detection with Fourier analysis and neural network. *Computers in Cardiology*, pages 545–548, 1996.
- [136] G. Mindlin and R. Gilmore. Topological analysis and synthesis of chaotic time series. *Physica D*, 58:229–242, 1992.
- [137] G. Mindlin, X. Hou, H. Solari, R. Gilmore, and N. Tufillaro. Classification of strange attractors by integers. *Physical Review Letters*, 64:2350–2353, 1990.
- [138] P. Mitra, U. Shankar, and S. Pal. Segmentation of multispectral remote sensing images using active support vector machines. *Pattern recognition letters*, 25:1067–1074, 2004.
- [139] G. Moody and R. Mark. The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20:45–50, 2001.
- [140] F. Morchen. Time series feature extraction for data mining using DWT and DFT. Technical Report 33, Dept. of Mathematics and Computer Science, Philipps-University Marburg., Germany, 2003.
- [141] Mouser Electronics. <http://www.mouser.com/>. Last viewed on August 1, 2007.
- [142] M. Munich and P. Perona. Continuous dynamic time warping for translation invariant curve alignment with applications to signature verification. In *International Conference on Computer Vision*, pages 108–115, 1999.
- [143] J. Murgo, N. Westerhof, J. Giolma, and S. Altobelli. Aortic impedance in normal man: Relationship to pressure waveforms. *Circulation*, 62:105–116, 1980.
- [144] J. Muzy, E. Bacry, and A. Arneodo. The multifractal formalism revisited with wavelets. *International Journal of Bifurcation and Chaos*, 4:245–302, 1994.

- [145] P. Nations-Weissman. <http://www.naturalhealthweb.com/articles/nations-weissman3.html>. Last viewed on March 19, 2008.
- [146] M. Navakatikyan, C. Barrett, G. Head, J. Ricketts, and S. Malpas. A real-time algorithm for the quantification of blood pressure waveforms. *IEEE Transactions on Biomedical Engineering*, 49:662–670, 2002.
- [147] W. Nichols and M. O'Rourke. *McDonald's blood flow in arteries: Theoretical, experimental and clinical principles*. Arnold, 1990.
- [148] J. Niskanen, M. Tarvainen, P. Ranta-aho, and P. Karjalainen. Software for advanced HRV analysis. *Computer Methods and Programs in Biomedicine*, 76:73–81, 2004.
- [149] A. Noordergraaf. *Circulatory System Dynamics*. Academic Press, 1978.
- [150] D. Novak, D. Cuesta-Frau, V. Eck, J.C. Perez-Cortes, and G. Andreu-Garcia. Denoising electrocardiographic signals using adaptive wavelets. *BIOSIGNAL*, pages 18–20, 2000.
- [151] Panel on Definition and CAM Research Methodology Conference Description. A wrist-mounted activity and pulse recording system. *Alternative Therapies*, 3:49–57, 1997.
- [152] M. O'Rourke. Arterial hemodynamics in hypertension. *Circulation*, pages 123–133, 1970.
- [153] M. O'Rourke. Influence of ventricular ejection on the relationship between central aortic and brachial pressure pulse in man. *Circulation Research*, 4:291–300, 1970.
- [154] M. O'Rourke and R. Kelly. Wave reflection in the systemic circulation and its implications in ventricular function. *Journal of Hypertension*, 11:327–337, 1993.
- [155] S. Osowski, L. T. Hoai, and T. Markiewicz. Support vector machine-based expert system for reliable heartbeat recognition. *IEEE Transactions on Biomedical Engineering*, 51:582–589, 2004.

- [156] M. Owis, A. Abou-Zied, A.-B. Youssef, and Y. Kadah. Robust feature extraction from ECG signals based on nonlinear dynamical modeling. In *23th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2, pages 1585–1588, 2001.
- [157] S.-Y. Park and J.-J. Lee. Self-diagnosis device using wrist pulse. In *33rd Annual Conference of the IEEE Industrial Electronics Society*, pages 139–142, 2007.
- [158] K. Parker, C. Jones, J. Dawson, and D. Gibson. What stops the flow of blood from the heart? *Heart Vessels*, 4:241–245, 1988.
- [159] G. Parulkar, G. Jindal, R. Padmashree, G. Haridasan, and J. Dharani. Impedance cardiology in mitral valve disease. *Journal of Post-graduate Medicine*, 26:155–161, 1980.
- [160] G. Parulkar, R. Padmashree, R. Bapat, R. Rege, K. Bhagtani, and G. Jindal. A new electrical plethysmograph: Observations in peripheral arterial occlusive diseases. *Journal of Post-graduate Medicine*, 27:66–72, 1981.
- [161] C-K. Peng, S. Buldyrev, S. Havlin, M. Simons, H. Stanley, and A. Goldberger. Mosaic organization of DNA nucleotids. *Physical Review E*, 49:1685–1689, 1994.
- [162] C. Perkins. <http://faculty.sgc.edu/cperkins/2211/ch21.ppt>. Last viewed on July 21, 2008.
- [163] Pico Technology. <http://www.picotech.com/>. Last viewed on June 9, 2007.
- [164] J. Pijn, D. Velis, M. van der Heyden, J. DeGoede, C. van Veelen, and F. Lopes da Silva. Nonlinear dynamics of epileptic seizures on basis of intracranial EEG recordings. *Brain Topography*, 94:249–270, 1997.
- [165] S. Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 88:2297–2301, 1991.
- [166] J. Piskorski and P. Guzik. Filtering poincaré plots. *Computational Methods in Science and Technology*, 11:39–48, 2005.



- [167] M. Potter and W. Kinsner. Multifractal characterization of synthetic ECG in the presence of coloured noise. In *Canadian Conference on Electrical and Computer Engineering*, volume 4, pages 2221–2225, 2004.
- [168] P. Prakash, M. Sareen, Abhinav, and S. Anand. Application of wavelets based multi-resolution analysis to detect relevant points of interest from finger-tip photoplethysmography and pressure signal from the radial artery. In *4th Cairo International Biomedical Engineering Conference*, pages 1–4, 2008.
- [169] G. Prasad and J. Sahambi. Classification of ECG arrhythmias using multi-resolution analysis and neural networks. In *Conference on Convergent Technologies for Asia-Pacific Region*, volume 1, pages 227–231, 2003.
- [170] Y. Qiu, Y. Cai, Y. Zhu, P Cheang, V. Crabtree, P Smith, and S. Hu. Poincaré plot analysis for pulse interval extracted from non-contact photoplethysmography. In *27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1964–1967, 2005.
- [171] C. Quick, D. Berger, and A. Noordergraaf. Apparent arterial compliance. *American Journal of Physiology*, 274:H1393–H1403, 1998.
- [172] C. Quick, M. Mohiuddin, G. Laine, and A. Noordergraaf. The arterial system pressure-volume loop. *Physiological measurement*, 26:N29–N35, 2005.
- [173] L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 1989.
- [174] S. Ramchurn and A. Murray. Multifractal analysis of the day and night characteristics of heart rate variability. *Computers in Cardiology*, 29:421–424, 2002.
- [175] M. Ramsey and M. Sugawara. Arterial wave intensity and ventriculoarterial interaction. *Heart and Vessels Supplement*, 12:128–134, 1997.
- [176] C. Ratanamahatana and E. Keogh. Three myths about dynamic time warping. In *SIAM International Conference on Data Mining*, pages 506–510, 2005.

- [177] T. Rath and R. Manmatha. Lower-bounding of dynamic time warping distances for multivariate time series. Technical report, University of Massachusetts Amherst, 2002.
- [178] J. Remington and E. Wood. Formation of peripheral pulse contour in man. *Journal of Applied Physiology*, 9:433–442, 1956.
- [179] P. Renevey, R. Vetter, J. Krauss, P. Celka, and Y. Depeursinge. Wrist-located pulse detection using ir signals, activity and nonlinear artifact cancellation. In *23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 3, pages 3030–3033, 2001.
- [180] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26:43–49, 1978.
- [181] S. Salvador and P. Chan. FastDTW: Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11:561–580, 2007.
- [182] M. Santo-Tomes, F. Lopez-Jimenez, H. Machado, H. Aldrich, G. Lamas, and E. Lieberman. Effect of cigar smoking on endothelium-dependent brachial artery dilation in healthy young adults. *American Heart Journal*, 143:83–86, 2002.
- [183] M. Sareen, Abhinav, P. Prakash, and S. Anand. Wavelet decomposition and feature extraction from pulse signals of the radial artery. In *International Conference on Advanced Computer Theory and Engineering*, pages 551–555, 2008.
- [184] T. Sato, M. Nishinaga, A. Kawamoto, T. Ozawa, and H. Takatsuji. Accuracy of a continuous blood pressure monitor based on arterial tonometry. *Hypertension*, 21:866–874, 1993.
- [185] N. Scafetta, L. Griffin, and B. West. Hölder exponent spectra for human gait. *Physica A*, 328:561–583, 2003.
- [186] P. Segers and P. Verdonck. Role of tapering in aortic wave reflection: Hydraulic and mathematical model study. *Journal of Biomechanics*, 33:299–306, 2000.

- [187] Sensym. <http://www.sensym.com/>. Last viewed on January 20, 2005.
- [188] W. Sequeira. Yoga in treatment of carpal-tunnel syndrome. *Lancet*, 353:689–690, 1999.
- [189] S. Seydnejad and R. Kitney. Real-time heart rate variability extraction using the kaiser window. *IEEE Transactions on Biomedical Engineering*, 44:990–1005, 1997.
- [190] M. Sherebrin and R. Sherebrin. Frequency analysis of the peripheral pulse wave detected in the finger with a photoplethysmograph. *IEEE Transactions on Biomedical Engineering*, 37:313–317, 1990.
- [191] J. Shu and Y. Sun. Developing classification indices for Chinese pulse diagnosis. *Complementary Therapies in Medicine*, 15:190–198, 2007.
- [192] R. Singh. [www.gdpau.com/ayurvedaframe.html](http://www.gdpau.com/ayurvedaframe.html). Last viewed on June 10, 2008.
- [193] D. Skramsted. [http://www.phoenix.tcieee.org/004\\_piezo\\_film\\_blood\\_flow\\_sensor/phoenix\\_piezopulse.htm](http://www.phoenix.tcieee.org/004_piezo_film_blood_flow_sensor/phoenix_piezopulse.htm). Last viewed on December 1, 2007.
- [194] J. Smith and J. Kampine. *Circulatory Physiology-the Essentials*. Williams and Wilkins, 1984.
- [195] S. Soderstrom, G. Nyberg, M. O'Rourke, J. Sellgren, and J. Ponten. Can a clinically useful aortic pressure wave be derived from a radial pressure wave? *British Journal of Anaesthesia*, 88:481–488, 2002.
- [196] H. Solari and R. Gilmore. Relative rotation rates for driven dynamical systems. *Physical Review A*, 37:3096–3109, 1988.
- [197] S. Song, Y-Y. Qi, and J-F. Qiao. Research on de-noising of pulse signal based on fuzzy threshold in wavelet packet domain. In *International Conference on Wavelet Analysis and Pattern Recognition*, pages 103–106, 2007.
- [198] C. Stefanadis, E. Tsiamis, C. Vlachopoulos, C. Stratos, K. Toutouzas, C. Pitsavos, S. Marakas, H. Boudoulas, and P. Toutouzas. Unfavorable effect of smoking on the elastic properties of the human aorta. *Circulation*, 95:31–38, 1997.

- [199] C. Stefanadis, C. Vlachopoulos, E. Tsiamis, L. Diamantopoulos, K. Toutouzas, N. Giatrakos, S. Vaina, D. Tsekoura, and P. Toutouzas. Unfavorable effects of passive smoking on aortic function in men. *Annals of Internal Medicine*, 128:426–434, 1998.
- [200] G. Strang and T. Nguyen. *Wavelets and filter banks*. Wellesley-Cambridge Press, 1996.
- [201] J. Strus. *Sphygmicae artis jam mille ducentos annos peritae et desideratae, libri v a josephi struthio posnaniense, medico recens conscripti, basel*. (As transcribed by A. Noordergraaf, University of Pennsylvania, Philadelphia, PA), 1555.
- [202] Z. Struzik. Removing divergences in the negative moments of the multi-fractal partition function with the wavelet transform. Technical Report INS-R9803, CWI, 1998.
- [203] Z. Struzik. Direct multifractal spectrum calculation from the wavelet transform. Technical Report INS-R9914, Centre for Mathematics and Computer Science, Amsterdam, 1999.
- [204] Z. Struzik. Determining local singularity strengths and their spectra with the wavelet transform. *Fractals*, 8(2), 2000.
- [205] Z. Struzik. Revealing local variability properties of human heartbeat intervals with the local effective Hölder exponent. *Fractals*, 9(1):77–93, 2001.
- [206] Z. Struzik and A. Siebes. Outlier detection and localisation with wavelet based multifractal formalism. Technical Report INS-R0008, CWI, 2000.
- [207] L. Suganthi and M. M.Manivannan. Effect of upper arm cuff pressure on pulse morphology using photoplethysmography. In *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1792–1795, 2009.
- [208] S. Sur and S. Ghatak. An inexpensive arterial pressure wave sensor and its application in different physiological condition, Indian Institute of Technology

- Kharagpur, [arxiv.org/abs/physics/0512071](https://arxiv.org/abs/physics/0512071), 2005.
- [209] A. Taddei, A. Biagini, G. Distanti, M. Emdin, M. Mazzei, P. Pisani, N. Roggero, M. Varanini, R. Mark, G. Moody, L. Braaksma, C. Zeelenberg, and C. Marchesi. The European ST-T database: Development, distribution, and use. *Computers in Cardiology*, pages 177–180, 1990.
- [210] F. Takahashi and S. Abe. Decision-tree-based multiclass support vector machines. In *9th International Conference on Neural Information Processing*, volume 3, pages 1418–1422, 2002.
- [211] F Takens. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, 898:366–381, 1981.
- [212] K. Tan, K. Chan, and K. Choi. Detection of the QRS complex, P wave and T wave in electrocardiogram. In *1st International Conference on Advances in Medical Signal and Information Processing*, pages 41–47, 2000.
- [213] Y. Tang, Y-Q. Zhang, N. Chawla, and S. Krasser. SVMs modeling for highly imbalanced classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39:281–288, 2009.
- [214] Task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Circulation*, 93:1043-1065, 1996.
- [215] M. Taylor. An approach to an analysis of the arterial pulse wave I. Oscillations in an attenuating line. *Physics in Medicine and Biology*, 1:258–269, 1957.
- [216] M. Taylor. An approach to an analysis of the arterial pulse wave II. Fluid oscillations in an elastic pipe. *Physics in Medicine and Biology*, 1:321–329, 1957.
- [217] M. Thiel, M.C. Romano, J. Kurths, and P. Read. Estimation of dynamical invariants without embedding by recurrence plots. *Chaos*, 14:234–243, 2004.

- [218] E. Treo, M. Herrera, and M. Valentinuzzi. Algorithm for identifying and separating beats from arterial pulse records. *BioMedical Engineering OnLine*, 4, 2005.
- [219] T. Tsankov and R. Gilmore. Topological aspects of the structure of chaotic attractors in  $\mathbb{R}^3$ . *Physical Review E*, 69:056206–1–056206–11, 2004.
- [220] T. Tsankov, A. Nishtala, and R. Gilmore. Embeddings of a strange attractor into  $\mathbb{R}^3$ . *Physical Review E*, 69:1–9, 2004.
- [221] P. Tsui, L. Lin, C. Chang, J. Hwang, J. Lin, C. Chu, C. Chen, K. Chang, and C. Chang. Arterial pulse waveform analysis by the probability distribution of amplitude. *Physiological Measurement*, 28:803–812, 2007.
- [222] N. Tufillaro, H. Solari, and R. Gilmore. Relative rotation rates: Fingerprints for strange attractors. *Physical Review A*, 41:5717–5720, 1990.
- [223] V. Tuzcu and S. Nas. Dynamic time warping as a novel tool in pattern recognition of ECG changes in heart rhythm disturbances. In *IEEE International Conference on Systems, Man and Cybernetics*, pages 182–186, 2005.
- [224] H. Uguz, A. Arslan, and I. Turkoglu. A biomedical system based on hidden Markov model for diagnosis of the heart valve diseases. *Pattern Recognition Letters*, 28:395–404, 2007.
- [225] S. Upadhyay. *Nadi vijnana (Ancient pulse science)*. Chaukhambha Sanskrit Pratishthan, Delhi, 1986.
- [226] P. Vaitkevicius, J. Fleg, and J. Engel. Effects of age and aerobic capacity on arterial stiffness in adults. *Circulation*, 88:1456–1462, 1993.
- [227] V. Vapnik. *The nature of statistical learning theory*. Springer-Verlag, New York, 1995.
- [228] K. Veropoulos, C. Campbell, and N. Cristianini. Controlling the sensitivity of support vector machines. In *International Joint Conference on Artificial Intelligence*, pages 55–60, 1999.

- [229] C. Vlachopoulos, N. Alexopoulos, D. Panagiotakos, M. O'Rourke, and C. Stefanadis. Cigar smoking has an acute detrimental effect on arterial stiffness. *American Journal of Hypertension*, 17:299–303, 2004.
- [230] N. Volz. [http://www.ayurvedacollege.com/ayurvedicpulsediagnosis\\_000.htm](http://www.ayurvedacollege.com/ayurvedicpulsediagnosis_000.htm). Last viewed on January 20, 2008.
- [231] C. Wagner and P. Persson. Chaos in the cardiovascular system: An update. *Cardiovascular Research*, 40:257–264, 1998.
- [232] B. Wang, Y. Yang, and J. Xiang. Experiment design and power-spectral characteristics of pulse signals. *Chinese Journal of Physical Medicine*, 3, 1998.
- [233] H. Wang and Y. Cheng. A quantitative system for pulse diagnosis in traditional Chinese medicine. *27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 7:5676–5679, 2005.
- [234] J. Wang, X. Ning, and Y. Chen. Multifractal analysis of electronic cardiogram taken from healthy and unhealthy adult subjects. *Physica A*, 323:561–568, 2003.
- [235] K. Wang, L. Wang, D. Wang, and L. Xu. SVM classification for discriminating cardiovascular disease patients from non-cardiovascular disease controls using pulse waveform variability analysis. *Lecture Notes in Computer Science*, 3339:109–119, 2005.
- [236] K. Wang, L. Xu, D. Zhang, and C. Shi. TCPD based pulse monitoring and analyzing. *IEEE International Conference on Machine Learning and Computing*, 3:1366–1370, 2002.
- [237] L. Wang, K. Wang, and L. Xu. Lempel-Ziv decomposition based arrhythmic pulses recognition. *27th Annual International Conference of the Engineering in Medicine and Biology Society*, pages 4606–4609, 2005.
- [238] L. Wang, K. Wang, and L. Xx. Recognizing wrist pulse waveforms with improved dynamic time warping algorithm. In *3rd International Conference on Machine Learning and Cybernetics*, pages 3644–3649, 2004.

- [239] Y. Wang, T. Bai, and J. Kang. Acquisition of pulse characteristics based on wavelet modulus maximum principle. *Space Medicine and Medical Engineering*, 19:41–46, 2006.
- [240] H. Warner, H. Swan, D. Connolly, R. Tompkins, and E. Wood. Quantitation of beat-to-beat changes in stroke volume from the aortic pulse contour in man. *Journal of Applied Physiology*, 5:495–507, 1953.
- [241] C. Webber, M. Schmidt, and J. Walsh. Influence of isometric loading on biceps EMG dynamics as assessed by linear and nonlinear tools. *Journal of Applied Physiology*, 78:814–822, 1995.
- [242] C. Webber and J. Zbilut. Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, 76:965–973, 1994.
- [243] T. Weber, J. Auer, M. O’Rourke, E. Kvas, E. Lassnig, R. Berent, and B. Eber. Arterial stiffness, wave reflections, and the risk of coronary artery disease. *Circulation*, 109:184–189, 2004.
- [244] N. Wessel, C. Brueckner, H. Malberg, A. Schumann, F. Reimperger, K. Osterziel, A. Schirdewan, and A. Voss. Long-term symbolic dynamics for heart rate variability analysis in patients with dilated cardiomyopathy. *Computers in Cardiology*, 26:253–256, 1999.
- [245] N. Wessel, N. Marwan, U. Meyerfeldt, A. Schirdewan, and J. Kurths. Recurrence quantification analysis to characterise the heart rate variability before the onset of ventricular tachycardia. *Lecture Notes In Computer Science*, 2199:295–301, 2001.
- [246] N. Wessel, U. Schwarz, P. Sapanin, and J. Kurths. Symbolic dynamics for medical data analysis, 2001.
- [247] K. Wesseling, J. Jansen, J. Settels, and J. Schreuder. Computation of aortic flow from pressure in humans using a nonlinear, three-element model. *Journal of Applied Physiology*, 74:2566–2573, 1993.



- [248] J. Wijbenga, A. Balk, S. Meij, M. Simoons, and M. Malik. Heart rate variability index in congestive heart failure: Relation to clinical variables and prognosis. *European heart journal*, 19:1719–1724, 1998.
- [249] Wikipedia contributors. <http://en.wikipedia.org/wiki/sphygmomanometer>. Last viewed on September 1, 2006.
- [250] Wikipedia contributors. [http://en.wikipedia.org/w/index.php?title=Human\\_heart&oldid=353225735](http://en.wikipedia.org/w/index.php?title=Human_heart&oldid=353225735). Last viewed on June 9, 2009.
- [251] K. Wilkie. Human blood flow measurement and modelling. Technical Report 98181130, 2003.
- [252] I. Wilkinson, H. MacCallum, L. Flint, J. Cockcroft, D. Newby, and D. Webb. The influence of heart rate on augmentation index and central arterial pressure in humans. *The Journal of Physiology*, 525.1:263–270, 2000.
- [253] I. Wilkinson, C. McEniery, and J. Cockcroft. Atenolol and cardiovascular risk: An issue close to the heart. *The Lancet*, 367:627–629, 2006.
- [254] A. Wittmann. Will wireless win the war?: Making local wireless work for you. *Network Computing*, 5:58–70, 1994.
- [255] C. Xia, Y. Li, J. Yan, Y. Wang, H. Yan, R. Guo, and F. Li. Wrist pulse waveform feature extraction and dimension reduction with feature variability analysis. In *2nd International Conference on Bioinformatics and Biomedical Engineering*, pages 2048–2051, 2008.
- [256] G. Xiang, Y.-H. Liu, D. Ding, and Y.-T. Shen. An internet based pulse palpation system for Chinese medicine. In *IEEE International Conference on Intelligent Robots and System*, volume 2, pages 1481–1486, 2002.
- [257] L. Xu, M. Meng, and K. Wang. Pulse image recognition using fuzzy neural network. In *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3148–3151, 2007.

- [258] L. Xu, K. Wang, and L. Wang. Pulse waveforms classification based on wavelet network. *7th Annual International Conference of the Engineering in Medicine and Biology Society*, pages 4596–4599, 2005.
- [259] L. Xu, K. Wang, L. Wang, and N. Li. Pulse contour variability before and after exercise. *19th IEEE Symposium on Computer-Based Medical Systems*, pages 237–240, 2006.
- [260] L. Xu, K. Wang, and D. Zhang. Modern researches on pulse waveform of TCPD. In *IEEE International Conference on Communications, Circuits and Systems and West Sino Expositions*, volume 2, pages 1073–1077, 2002.
- [261] L. Xu, K. Wang, D. Zhang, and C. Shi. Adaptive baseline wander removal in the pulse waveform. In *IEEE Symposium on Computer-Based Medical Systems*, pages 143–148, 2002.
- [262] L. Xu, D. Zhang, K. Wang, , and L. Wang. Arrhythmic pulses detection using Lempel-Ziv complexity analysis. *EURASIP Journal on Applied Signal Processing*, 2006:1–12, 2006.
- [263] L. Xu, D. Zhang, K. Wang, N. Li, and X. Wang. Baseline wander correction in pulse waveforms using wavelet-based cascaded adaptive filter. *Computers in Biology and Medicine*, 37:716–731, 2007.
- [264] S. Yacin, M. Manivannan, and V. Chakravarthy. Pulse rate variability and gastric electric power in fasting and postprandial conditions. In *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2639–2642, 2009.
- [265] J. Yan, C. Xia, H. Wang, Y. Wang, R. Guo, F. Li, and H. Yan. Nonlinear dynamic analysis of wrist pulse with Lyapunov exponents. In *2nd International Conference on Bioinformatics and Biomedical Engineering*, pages 2177–2180, 2008.
- [266] Y. Yoon, M. Lee, and K. Soh. Pulse type classification by varying contact pressure. *IEEE Engineering in Medicine and Biology Magazine*, 19:106–110., 2000.

- [267] A. Zambanini, S. Cunningham, K. Parker, A. Khir, S. McG. Thom, and A. Hughes. Wave-energy patterns in carotid, brachial, and radial arteries: A noninvasive approach using wave-intensity analysis. *American journal of physiology*, 289:H270–H276, 2005.
- [268] J. Zbilut, Z. Hu, A. Giuliani, and Webber C. Singularities of the heart beat as demonstrated by recurrence quantification analysis. In *22nd Annual International Conference of the Engineering in Medicine and Biology Society*, volume 4, pages 2406–2409, 2000.
- [269] J. Zbilut, N. Thomasson, and Webber C. Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals. *Medical Engineering and Physics*, 24:53–60, 2002.
- [270] J. Zbilut and C. Webber. Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A*, 171:199–203, 1992.
- [271] D. Zhang, L. Zhang, D. Zhang, and Y. Zheng. Wavelet based analysis of doppler ultrasonic wrist-pulse signals. In *International Conference on BioMedical Engineering and Informatics*, volume 2, pages 539–543, 2008.
- [272] P.-Y. Zhang and H.-Y. Wang. A framework for automatic time-domain characteristic parameters extraction of human pulse signals. *EURASIP Journal on Advances in Signal Processing*, 2008:1–9, 2008.

---

# Publications

---

## Publications related to the thesis

1. Joshi, A.J., Chandran, S., Jayaraman, V.K. and Kulkarni, B.D.: **Nadi Tarangini: A pulse based diagnostic system.** *In Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2207–2210 (2007).* Also in Microsoft Techvista Doctoral Presentations 2007 (Third Prize).
2. Joshi, A.J., Chandran, S., Jayaraman, V.K. and Kulkarni, B.D.: **Arterial pulse system: Modern methods for traditional Indian medicine.** *In Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 608–611 (2007).*
3. Joshi, A.J., Chandran, S., Jayaraman, V.K. and Kulkarni, B.D.: **Arterial pulse rate variability analysis for diagnoses.** *In Proceedings of the International Conference on Pattern Recognition (2008).*
4. Joshi, A.J., Chandran, S., Jayaraman, V.K. and Kulkarni, B.D.: **Multifractality in arterial pulse.** *In Proceedings of the International Conference on Pattern Recognition (2008).*
5. Joshi, A.J., Jog, A., Chandran, S., Jayaraman, V.K. and Kulkarni, B.D.: **Topological**

**invariants of the arterial pulse.** In *PAKDD Workshop on Data Mining for Healthcare Management* (2010). To appear.

## Manuscripts

6. Joshi, A.J., Chandran, S., Jayaraman, V.K. and Kulkarni, B.D.: **A novel PSWT-DTW approach for analyzing pseudo-periodic signals.** *Under review in Computers in Biology and Medicine.*
7. Joshi, A.J., Ashwin, K.P., Chandran, S., Jayaraman, V.K. and Kulkarni, B.D.: **Symbolic dynamics based analysis of the arterial pulse.** Under preparation.

## Other related publications (not directly appearing in the thesis)

8. Joshi, A.J., Chandran, S., Jayaraman, V.K. and Kulkarni, B.D.: **Hybrid SVM for multiclass arrhythmia classification.** *International Journal of Functional Informatics and Personalised Medicine (IJFIPM)* (2010). Also appears in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine* (2009).
9. Joshi, A.J., Rajshekhar, Chandran, S., Phadke, S., Jayaraman, V.K. and Kulkarni, B.D.: **Arrhythmia classification using local Hölder exponents and support vector machine.** *Lecture Notes in Computer Science* **3776** 242–247 (2005). Also appears in *Proceedings of the 1st International Conference on Pattern Recognition and Machine Intelligence (PReMI)* (2009).
10. Jog, A., Joshi, A., Chandran, S. and Madabhushi, A.: **Classifying Ayurvedic pulse signals via consensus locally linear embedding.** In *Proceedings of Biosignals* (2009).

## Patents

- Dr. Ashok Bhat, Aniruddha Joshi, Anand Kulkarni, B. D. Kulkarni, V. K. Jayaraman, Sharat Chandran. **A non-invasive device Nadi Tarangini useful for quantitative detection of arterial nadi pulse waveform.** Submitted for PCT (international) application in August 2008.

## Acknowledgements

The journey throughout my PhD years has been very enriching and fulfilling. As I present this thesis, first and foremost I would like to thank my father for introducing me to the field of *Ayurveda*, and encouraging me to take up research in this field, and my mother who has always been very supporting with a pleasant smile :)

I would like to express my gratitude to my supervisor Professor Sharat Chandran for the opportunity of doing this work and for his supervision and advice during the complete course. He made sure that I had no administrative problems during my stay in IIT.

I thank my co-supervisors Dr. B. D. Kulkarni and Dr. V. K. Jayaraman for guiding me throughout my research, for the discussions and the insights I gained from them, without which this thesis would not have been possible. Their enthusiasm and passion about research motivated me a lot in my journey.

I am also grateful to Prof. V. M. Gadre and Prof. S. Mukherji for their valuable feedback and comments that helped me keep focus and progress in my work. They proved to be a great combination as they could guide me in both signal analysis and biomedical being their respective fields of expertise.

I am grateful to Vaidyaraj Ashok Bhat, all his students, and Dr. Jategaonkar who helped me understand not just the domain knowledge of *Ayurveda* and pulse-based methodology, but also how traditionally people must have transformed observations into the set of rules. All such GOD people like Vaidyaraj have made India proud with their heavy contributions in many fields including traditional medicine. I feel thankful that I got a chance to “actually meet” one of them and learn a few lessons of life.

I thank all the patients for allowing me to record their pulse and for providing me all the required information. My study has not been successful without their non-interfered involvement.

I am thankful to Anand Kulkarni who helped me during the initial stages of the project (actually it is the other way around) and Mr. Sharma for his valuable help in the hardware building.

I give gracious acknowledgment to my co-authors – Anand, Amod, Rajshekhar, Dr. Phadake and Ashwin – both for their technical contributions to the thesis and for being utterly enjoyable colleagues.

I am thankful to my lab friends in ViGIL and NCL for all the tea-sessions and discussions we had together. I really enjoyed the parties together, especially our novel *welcoming* parties. I learnt dedication and patience for the subject from senior members Avinash, Abhijit and Prakash.

The bond of friendship I developed with my Chai-group and Marathi Mandal friends will always be cherished in my heart. Nitin, Umesh, Salil, Smriti and me still have an e-tea twice a day. A day cannot be completed without T1 & T2. A special mention to Marathi mandal, where I met my lovely wife Meghana.

Appu, Biswarup, Manish, Vipul and me joined together for the intergrated-PhD program and finished many many things all together. The IIT life was just awesome because of all the togetherness. I cherish the fun nights at the hostel.

A special thanks to my friend Vinay, who has been an excellent colleague and a walking encyclopedia, providing advice on areas both technical and non. A heavy contribution in my publications is his constant encouragements and sometimes scolding.

I thank the KRESIT and CSE office staff in the Computer Science Department for providing advice and needed resources, and making the workplace environment most pleasant.

I would also like to thank CSIR for funding me during the course of my PhD.

Most of all, I would like to thank my family and especially my parents (again), for their absolute confidence in me, love, support and guidance.

Last but not the least, I would like to thank my beloved wife Meghana for her unconditional love, tasty dishes, and endless patience & understanding provided during these years. Her great support during the difficult moments, especially in the final year, helped me proceed towards completion. Also, many thanks for reading, proofing, and commenting on this dissertation, which is greatly improved by her efforts.

!!! Thank you GOD for making me do this !!!

Date: \_\_\_\_\_

Aniruddha J. Joshi



Data is available at <http://www.naditarangini.com>