

Temporal Pattern Recognition in Multiparameter ICU

Data

by

Mohammed Saeed

Bachelor of Science in Electrical Engineering
Purdue University, 1995

Master of Science in Electrical Engineering and Computer Science
Massachusetts Institute of Technology, 1997

Submitted to the Department of Electrical Engineering and
Computer Science in partial fulfillment of the requirements for the
degree of Doctor of Philosophy
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2007

© Mohammed Saeed, 2007. All rights reserved

The author hereby grants to MIT permission to reproduce and distribute
publicly paper and electronic copies of this thesis document in whole or in
part, and to grant others the right to do so.

Author: _____

Department of Electrical Engineering and Computer Science
May 29, 2007

Certified by: _____

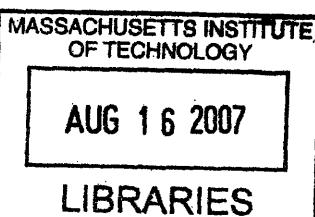
Roger G. Mark

Distinguished Professor of Health Sciences and Technology
Thesis Supervisor

Accepted by: _____

Arthur C. Smith

Chairman, Department Committee on Graduate Students



Temporal Pattern Recognition in ICU Data

by

Mohammed Saeed

Submitted to the Department of Electrical Engineering and Computer
Science in partial fulfillment of the requirements for the degree of Doctor of
Philosophy

Abstract

Intensive Care Unit (ICU) patients are physiologically fragile and require vigilant monitoring and support. The myriad of data gathered from biosensors and clinical information systems has created a challenge for clinicians to assimilate and interpret such large volumes of data. Physiologic measurements in the ICU are inherently noisy, multidimensional, and can readily fluctuate in response to therapeutic interventions as well as evolving pathophysiologic states. ICU patient monitoring systems may potentially improve the efficiency, accuracy and timeliness of clinical decision-making in intensive care. However, the aforementioned characteristics of ICU data can pose a significant signal processing and pattern recognition challenge---often leading to false and clinically irrelevant alarms.

We have developed a temporal database of several thousand ICU patient records to facilitate research in advanced monitoring systems. The MIMIC-II database includes high-resolution physiologic waveforms such as ECG, blood pressures waveforms, vital sign trends, laboratory data, fluid balance, therapy profiles, and clinical progress notes over each patient's ICU stay. We quantitatively and qualitatively characterize the MIMIC-II database and include examples of clinical studies that can be supported by its unique attributes. We also introduce a novel algorithm for identifying "similar" temporal patterns that may illuminate hidden information in physiologic time series.

The discovery of multi-parameter temporal patterns that are predictive of physiologic instability may aid clinicians in optimizing care. In this thesis, we introduce a novel temporal similarity metric based on a transformation of time series data into an intuitive symbolic representation. The symbolic transform is based on a wavelet decomposition to characterize time series dynamics at multiple time scales. The symbolic transformation allows us to utilize classical information retrieval algorithms based on a vector-space model. Our algorithm is capable of assessing the similarity between multi-dimensional time series and is computationally efficient. We utilized our algorithm to identify similar physiologic patterns in hemodynamic time series from ICU patients. The results of this thesis demonstrate that statistical similarities between different patient time series may have meaningful physiologic interpretations in the detection of impending hemodynamic deterioration. Thus, our framework may be of potential use in clinical decision-support

systems. As a generalized time series similarity metric, the algorithms that are described have applications in several other domains as well.

Thesis Supervisor: Roger G. Mark

Title: Distinguished Professor of Health Sciences and Technology

Professor of Electrical Engineering and Computer Science

Acknowledgments

When I first came to MIT, some people told me that it would be a challenging and daunting environment. There have certainly been occasions when times were difficult---but in truth---this experience has been an enjoyable, life-altering, and blissful chapter in my life. As I am writing these acknowledgements, I fully realize that I am a wealthy man. I am wealthy because I have a circle of friends and family that have showered me with their generosity, love, and happiness.

Professor Roger Mark has been a mentor, friend, and role model. He taught me a great deal about physiology, clinical medicine, and physiologic signal processing. I feel truly blessed for having met such a wonderful human being. We could just as easily converse about medicine, engineering, religion, politics, and family. His level of compassion for others and passion for educating and mentoring students is unrivaled. I wish him and his family all the joy and blessings that they so richly deserve.

Professor Peter Szolovits has been a wonderful research mentor and great teacher. He has injected insightful ideas when I needed them, and gently prodded me forward with his great sense of humor when I may have tended to stagnate. **Dr. Bill Long** has been a valuable member of my doctoral committee, and has provided me with a great deal of practical advice as I explored different aspects of my thesis.

Dr. Danny Talmor and Dr. Atul Malhotra are great colleagues to work with at BIDMC. Their insightful comments were particularly helpful in the clinical studies we developed from the MIMIC-II database. I look forward to many years of collaboration with them.

The people at LCP have made my studies at MIT extremely enjoyable. There are few words that can describe my admiration and feelings for **Rama Mukkamala**. I feel extremely fortunate to know somebody who is so hard-working, creative, passionate, loyal, and sincere. He really brought the word “fun” into LCP. When I start a ten minute phone call with him, it usually ends two hours later because we have so much to discuss. I treasure our close friendship and I am grateful for his support over the years.

Wei Zong is a constant professional in all senses of the word. He is unfailing in always providing his colleagues a helping hand. It is rewarding to know somebody who is so talented---yet so humble. I very much enjoyed every minute spent with Wei at MIT, Philips Medical Systems, and the swimming pool!

Thomas Heldt has been a great colleague over the years. As my officemate, he was an intelligent, principled, and hard-working peer. Outside of the lab, he has been a loyal and supportive friend. I could always count on him to make sure that I didn’t take myself too seriously with his witty zingers.

Omar Abdala is a great friend and colleague. I am very proud to know somebody who can combine his level of intellect with a passion for making our greater community a better place.

Matt Oefinger was great to have in the lab as a friend and colleague. His great sense of humor, positive outlook, and good nature were an integral part of LCP. Admittedly, I have yet to get over my loss to him in the LCP pizza-eating contest.

George Moody has provided me with so much help over the years. As we developed the MIMIC-II database, his experience was invaluable in moving us in the right direction. His passion for facilitating research has helped to sustain LCP over the years.

Andrew Reisner's wit, insight, and great sense of humor over the years were a welcome addition to LCP. I am grateful for the many times he taught me some very practical clinical medicine and pearls of wisdom about the dynamics of the medical profession.

Sherman is not only a virtuoso of the violin, but also a talented student and kind friend who is great to work with.

Professor George Verghese has made extremely valuable contributions to the MIMIC-II project. It is a pleasure to work with such a friendly and insightful individual.

Dr. Joseph Frassica has been a valuable mentor for many years. His understanding of ICU medicine and the patient monitoring industry is amazing.

Gari Clifford is an English gentleman in every respect. He is helpful to people in the lab and has a great sense of humor.

Mauro, Zaid, Li-wei, Isaac, Christine Leu, Anton, Zaid, Dewang, Brian Janz, and Benjamin Moody were all great contributors to LCP and wonderful people to work with.

Within the HST community, I have met many wonderful people.

Ali Shoeb has been a great friend over the years. His exuberance for learning and positive attitude with respect to every activity he engages in serves as a great example of the model student. His energetic personality makes every conversation with him enjoyable and truly memorable.

I first knew **Lucia Madariaga** as my team member in HST anatomy, and now I know her as my friend. Through the many enjoyable walks between Harvard and MIT, I came to appreciate and respect her spirituality, intelligence, and genuine kindness.

I have known **Nick Houstis** since high school in Indiana. His great sense of humor, loyalty to his friends, and absolute brilliance make him one of the most amazing people I have ever met.

Ray “L” Chan has been a great friend over the years. His unforgettable laughter and good nature always brightened my days. His positive outlook on life is an example I strive for.

Quite possibly, **Ying Zhang** takes the title as the world’s most considerate person. I remember when we had offices next to one another at Philips, she would bring me a glass of water when she heard the slightest cough out of me. You would think that somebody as kind as her is too good to be true---but she is for real!

I have had the pleasure of knowing some truly gifted and wonderful people in EECS over the years. It has been an honor to know **Professor Paul Gray** at MIT. I enjoyed being his TA in circuit class. He was always available to render his wealth of invaluable advice to me when I needed it. It was also a pleasure serve as a TA under **Professors Art Smith, Jacob White, and Jeff Lang**.

In the many environments I have worked in, there were always capable people making sure I didn’t stray too far off the path. **Marilyn Pierce** of EECS at MIT was always there for any questions I had about registration, graduation, and institute policies. **Patty** and **Linda** have been so helpful over the years at the HST office. At LCP, **Ken** has been extremely helpful anytime I needed his expertise in document processing or filling out obscure forms. **Karen** at Philips was always available to help me in every thing I did when all hope seemed lost! **Professor Rick Mitchell** at HST has always been a great leader and advisor for me and countless other students.

The first month I came to MIT as a naïve twenty one year-old kid from Indiana, I felt out of place and longed for the comforts of home and my family. While my dorm was next to a nuclear reactor, by the grace of God, my next door neighbor happened to be **Gassan Al-Kibsi**. Since I have known him, he has been my truest friend and brother. He has been an unfailing source of support, comfort, inspiration, and optimism. Whenever I feel challenged by a situation professionally or socially, I think about how Gassan would handle it, and the answer usually presents itself. We may now be physically separated by 8000 miles, but I feel his presence in my heart everyday.

Ayman Shabra is too humble of a person for even allowing me to praise him. He is the epitome of gentleness, sincerity, and thoughtfulness. Whenever I have needed his help, he has been there. For the kindness that he has shown me and his wisdom that has enlightened me, I will be in his debt for the rest of my life.

Babak Ayazifar has been in every respect an older brother that I can count on for moral support, professional and spiritual guidance, and his witty sense of humor. He has shaken me when I needed shaking, and calmed me when I needed calming. The pilgrimage we made to Mecca together will always be special because he was there with me. He and his wife **Sara** will always be dear friends.

Belal Helal has a unique ability to brighten the room with his smile. He has been a wonderful example of a man who combines intellect, loyalty to family and friends, and

glowing spirituality that shines on all those who know him. The many family parties we have had over the years with Belal and his wife, **Sarah**, and their wonderful children have made our lives at MIT more enjoyable.

Jalal Khan has always been a great role-model as a man who excels professionally while having an untiring commitment to his family and friends. I feel fortunate that such a good person is part of my life every time he calls me on the phone just to say “hi!” Every time I finish a long swim, I remember the countless hours Jalal spent in the children’s pool teaching this overgrown kid how to swim with the care and patience that only he possesses. Jalal and his wife, **Elie**, have been pillars among our community of friends.

Kashif Khan has impacted me in innumerable ways. His style of questioning and searching for the truth ---in science or religion--- should be made into a course for every truth seeker to enroll in. His remarkable gift of always seeing the good in people while forgiving their shortcomings enriched our circle of friends. As my close friend, I feel his positive impact every day of my life. Kashif and his wife, **Marilene**, enriched all our lives.

Milton (and Nausheen), Naim (and Rani), Asif (and Zaregul), Ali Merchant (and Soraya) are all dear friends that made life in Boston so enjoyable for me and my family over the years. I am humbled by how generous they have always been to us. Their friendship is a gift from the Almighty.

I am grateful to **Reem Al-Hashimy** for the kindness she has shown to my wife and me, and the valuable insight she has given me about the importance of showing compassion to patients and their families. I will never forget the great friends I made when I was studying at Purdue University. Hasan and Haydar Syed, and Gautham Nayak have always been great friends to me.

Larry Nielsen has been a great mentor for me at Philips. His has been a constant advocate for my research at Philips, and he always encouraged me to take risks and pursue my goals to the fullest extent. I have learnt a great deal from him about the practical challenges in patient monitoring as well as benefited from his wisdom in dealing with people.

Greg Raber put in so much of his time to write the software that has enabled us to create the MIMIC-II database. Simply put, this work would not have been possible without him. It is a pleasure to work with such a talented engineer who is fun-loving, passionate about his work and a real maverick who gets things done.

Jesus Santana and **Mike Bortonne** helped me every step of the way as I tried to validate the MIMIC-II system at Philips before I could use it in the hospital.

Carol Russell and **Steve Weisner** were quite generous in supporting my research activities and academic pursuits of the many years at Philips.

There are many other people that I have learnt a great deal from over the years at Philips and I thank them all: **Andres Aguirre, Walid Ali, Larry Eshelman, Joanne Foster, Scott Kresge, John Wang, and Dale Wiggins.**

Ten years ago I married Wejdan---the love of my life. She has supported me in every endeavor I have set forth on. She is my wife, the mother of my children, and my best friend. Every day as I arrive back from work, she has filled our home with joy and happiness. She has given up so much so that I could complete my MD-PhD over the past ten years. It will take me the next 100 years to recompense her unselfishness and compassion. My daughter, **Fatima**, and son, **Hasan**, are the joys of my life. One hug from them and any worry that I have quickly vanishes.

My brother **Ali** is the greatest brother anybody could dream for. He is gentle, unselfish, loyal, and supportive to the greatest extremes. He has been like a second father for my children and the backbone of our extended family. He is my friend, confidant, and ally. I cannot say anything else that would do him justice. The last year that he has been here at MIT has made this the most enjoyable year I could imagine.

My sisters, **Huda, Ziena, and Zainab** have always been a source of support and love even as I studied far from home. I feel so blessed that my children will grow up having such wonderful aunts.

MIT provided me with an education---but more importantly---I was blessed with a dear brother in **Ihsan Djomehri**. The purity and sincerity of Ihsan's heart provides me with great comfort. He has all the qualities I could ever ask for in a brother for me, husband for my sister, and uncle for my children. He has been so unselfish in showering me with his love and kindness. In short, he has been an angel.

My brother-in-law **Ridha** has always been an older brother I could count on for support, wisdom, and love. My brother-in-law **Issa** has always been kind to me. I am particularly grateful for the great journey to Mecca we took together.

The love and support my mother, **Khaola Saeed**, has always provided me is something I can never measure because it simply has no bounds. Her prayers and supplications have sustained me. She had devoted her entire life to comforting her children. Everything that I have ever accomplished is due to her struggles to care for me whenever I have needed her.

My father, **Mahdi Saeed** is my hero. He left a war-torn country and obtained a doctoral degree and helped to raise five children at the same time. He was deeply involved in our education from childhood through college. He provided me with guidance whenever I sought it out, and he comforted me when I was at my weakest. When I was sick with the chicken pox as a young child, he slept in the same small cramped room to comfort me for a week. When I had a major surgery, he flew to Boston for several weeks to be my cook, cleaner, and comforter.

Finally, I thank God for the many blessings He has given me and my family. I submit myself to Him, and seek His mercy. I ask that He protects me and my loved ones and increases my knowledge so that I can better serve mankind.

Contents

1.	Introduction	21
1.1.	Overview	21
1.2.	Thesis organization	25
2.	Background	27
2.1.	Summary of monitoring system capabilities	27
2.2.	Advanced patient monitoring algorithms	38
2.3.	ICU patient database to support advanced algorithm research	43
3.	Design and Characterization of a Large Scale Temporal ICU Database: MIMIC-II	45
3.1.	Introduction	45
3.2.	Overview of ICU databases.....	47
3.3.	Creation of the MIMIC-II Database.....	51
3.3.1.	Physiologic data acquisition from ICU bedside monitors.....	52
3.3.2.	Specifications of physiologic data.....	54
3.3.3.	Physiologic parameter trends	54
3.3.4.	ICU monitor alarms in MIMIC-II	57
3.3.5.	Specifications of clinical data	60
3.3.6.	Merger of physiologic and clinical data.....	64
3.3.7.	Annotation of merged ICU patient records	65
3.4.	Quantitative and qualitative characterization of the MIMIC-II database	65
3.4.1.	Comparison of Clinical and Physiologic Data	73
4.	Clinical Studies with MIMIC-II.....	76
4.1.	Analysis of agreement between noninvasive and invasive blood pressure measurements in ICU patients.....	77
4.1.1.	Review of oscillometric blood pressure measurement.....	78
4.1.2.	Methodology: Patient ABP and NIBP data and noise processing	81
4.1.3.	Methodology: Physiologic factors analysis.....	84

4.1.4. Methodology: Comparison between NIBP and ABP to detect hemodynamic instability in ICU patients.....	84
4.1.5. Results: Relative error analysis.....	86
4.1.6. Results: NIBP-ABP overall analysis of agreement.....	87
4.1.7. Results: Analysis of factors influencing NIBP-ABP agreement.....	91
4.1.8. Results: Hemodynamic instability detector performance	94
4.1.9. Results: NIBP-ABP trends prior to renal failure.....	96
4.1.10. Discussion: NIBP-ABP measurement of agreement.....	102
4.2. Diurnal variation in hemodynamic interventions in the ICU.....	105
4.2.1. Introduction	105
4.2.2. Methodology	107
4.2.3. Results: Hourly variations of blood pressure and heart rate	109
4.2.4. Results: Hourly variations of clinical data charting	111
4.2.5. Results: Diurnal variation in onset of hemodynamically significant interventions	112
4.2.6. Results: Temporal analysis of hypotensive events.....	115
4.2.7. Discussion	116
5. Novel Time Series Similarity Metric	121
5.1.1. Review of time series similarity algorithms.....	124
5.1.2. Introduction of novel wavelet-based similarity metric	128
5.1.3. Review of classical information retrieval model.....	129
5.1.4. Detection of similar time series as an information retrieval problem	134
5.1.5. Multiscale dynamics characterization of time series: wavelet-based approach	135
5.1.6. Review of wavelet convolutions	137
5.1.7. Methodology: Wavelet symbol generation	143
5.1.8. Methodology: Histogram smoothing using Parzen density estimators...	147
5.1.9. Results:	152
5.1.10. Results: Example queries from Library A.....	159
5.1.11. Results: Query performance with a noise stress test.....	161

5.1.12.	Results: Retrieval of similar computational model-generated ICU time series of Library B.....	162
5.1.13.	Results: Evaluation of retrieval algorithm with a benchmark time series database (SYNDATA)	163
6.	Prediction of Hemodynamic Deterioration in ICU Patients.....	166
6.1.	Problem formulation	169
6.1.1.	Selection and automated annotations of physiologic time series from ICU records	171
6.1.2.	Therapeutic profiles of hemodynamic stability and instability	172
6.1.3.	Limitations of hemodynamic instability criteria	175
6.2.	Methodology	176
6.2.1.	Physiologic trend pre-processing	177
6.2.2.	Derivation of estimated cardiac output	179
6.2.3.	Hemodynamic event library creation	180
6.3.	Results	182
6.3.1.	Performance of estimated cardiac output technique	183
6.3.2.	Evaluation of hemodynamic deterioration prediction alarm.....	185
6.3.3.	Optimal feature selection and analysis.....	189
6.3.4.	Comparison of physiologic similarity algorithm with conventional time domain feature descriptors	190
7.	Conclusions	200
7.1.	Major Thesis Contributions.....	200
7.1.1.	Development of a major new ICU database: MIMIC-II	200
7.1.2.	Clinical studies with the MIMIC-II database.....	201
7.1.3.	Development of a time series similarity algorithm using wavelet-based symbolic representations	202
7.1.4.	Novel algorithm to predict hemodynamic instability in ICU patients	202
7.2.	Future Work	203
7.2.1.	Next-Generation clinical database development.....	203
7.2.2.	Future clinical studies with the MIMIC-II database.	205
7.2.3.	Time series similarity metric.....	207

7.2.4. Hemodynamic instability prediction	208
---	-----

List of Figures

Figure 2-1: Physiologic signal processing flow diagram	32
Figure 2-2: Physiologic trends with noise.....	35
Figure 2-3: Example of ICU clinical data trends	37
Figure 3-1: MIMIC-II physiologic data archiving architecture	53
Figure 3-2: Examples of physiologic waves	57
Figure 3-3: Length of stay distributions for MIMIC-II ICU patients	68
Figure 3-4: Stratified ICU patient demographics and mortality rates.....	69
Figure 3-5: SAPS-I distribution & mortality rates.....	70
Figure 3-6: Distribution of ICD9 problems in MIMIC-II patients.....	71
Figure 4-1: Oscillometric method of blood pressure measurement	80
Figure 4-2: ABP noise rejection algorithm	83
Figure 4-3: MIMIC CaseID 4167 ABP & NIBP trend prior to pressor onset	85
Figure 4-4: Bland-Altman analysis of (systolic) NIBP-ABP agreement	89
Figure 4-5: Bland-Altman analysis of (mean) NIBP-ABP agreement.....	90
Figure 4-6: Bland-Altman analysis of (diastolic) NIBP-ABP agreement.....	91
Figure 4-7: NIBP-ABP (systolic) Error vs Age	92
Figure 4-8: Error (Mean NIBP-Mean ABP) vs Weight	93
Figure 4-9: Error (Sys NIBP - Sys ABP) as a function of heart-rate.....	94
Figure 4-10: MIMIC 9618 NIBP and ABP (systolic) blood pressure trends prior to creatinine increase	98
Figure 4-11: MIMIC 9618 NIBP and ABP (mean) blood pressure trends prior to creatinine increase	99
Figure 4-12: Bland-Altman analysis of NIBP-ABP (mean) error prior to creatinine increases	100
Figure 4-13: Bland-Altman analysis of NIBP-ABP (systolic) error prior to creatinine increases	101

Figure 4-14: Mean blood pressure (population-averaged) variations throughout 24-hour cycle	110
Figure 4-15: Mean heart rate (population-averaged) as a function of time of day	111
Figure 4-16: Rate of charting of hemodynamic variables (heart rate and blood pressure) as a function of time of day	112
Figure 4-17: Frequency of vaso-active medication onsets as function of time of day....	113
Figure 4-18: Frequency of IV fluid bolus onsets as function of time of day	114
Figure 4-19: Frequency of hypnotic/sedative medication onsets as function of time of day	114
Figure 4-20: Temporal variation of acute hypotensive events.....	116
Figure 5-1: Multiparameter hemodynamic trends from ICU patients prior to instability and reference time series during stability.....	123
Figure 5-2: Example of point-to-point error between two “similar” time series for the Euclidean distance metric.....	126
Figure 5-3: Example of FFT-based transformation of two different ramp time series ...	127
Figure 5-4: Example of document-to-vector conversion	130
Figure 5-5: Example of identifying "similar" sentences	132
Figure 5-6: Sentence similarity matrix displayed using color mapping of correlation coefficients	134
Figure 5-7: Example of Haar wavelet basis functions	137
Figure 5-8: Discrete wavelet transform utilizing Haar basis function to calculate wavelet detail coefficients	138
Figure 5-9: Discrete wavelet transform utilizing Haar function to calculate approximation coefficients	139
Figure 5-10: Filter bank implementation of DWT to calculate wavelet detail coefficients	140
Figure 5-11: Filter bank implementation to calculate wavelet approximation coefficients	141
Figure 5-12: Example of multiscale decomposition of a signal with Haar wavelets	142
Figure 5-13: Scheme for generation of quantized wavelet coefficients.....	145
Figure 5-14: Wavelet symbols created by quantization	146

Figure 5-15: Example of Parzen smoothing of a wavelet symbol histogram	149
Figure 5-16: Collection of wavelet symbol vectors from different time series	150
Figure 5-17: Examples of synthetic signals for Library A.....	153
Figure 5-18: Example from Library B: simulation of hemorrhage.....	155
Figure 5-19: Example from Library B: simulation of sepsis	156
Figure 5-20: Example from Library B: simulation of acute myocardial infarction	157
Figure 5-21: Examples of six classes from SYNTDATA library of time series	158
Figure 5-22: Example of increasing trend (from Library A) as query time series, and top 6 retrieved time series.	159
Figure 5-23: Library A sinusoidal signal as query (Q), and retrieved time series (1-5) .	160
Figure 5-24: Random walk time series from Library A.....	160
Figure 5-25: Time series retrieval performance in the presence of additive Gaussian noise	162
Figure 5-26: Retrieval and classification accuracy with SYNTDATA Library	165
Figure 6-1: Multiparameter hemodynamic trends from ICU patients.....	168
Figure 6-2: ICU record retrieval formulated as problem in information retrieval	170
Figure 6-3: Example of hemodynamic instability.....	173
Figure 6-4: Noise processing of physiologic trends.....	178
Figure 6-5: Flow diagram for hemodynamic instability using “similar” segments retrieved from reference ICU multiparameter segments	182
Figure 6-6: Bland-Altman analysis of agreement between estimated cardiac output (ECO) and thermodilution cardiac output (TCO).	183
Figure 6-7: Example of predictive hemodynamic deterioration alarm	187
Figure 6-8: ROC for predictive algorithm two hours to vasopressor onset	188
Figure 6-9: Effects of ECO on ROC	190
Figure 6-10: Example of line regression fitting.	191
Figure 6-11: ROC of time-domain regression and wavelet-symbolic representations for the prediction of hemodynamic instability.....	194
Figure 6-12: Example of similar multiscale dynamics in query and retrieved data segments	195
Figure 6-13: Best-fit lines representative of two different segments	196

Figure 6-14: Wavelet symbol class separation.....	197
Figure 6-15: Hemodynamic instability index with progressive deterioration over four hours	199

List of Tables

Table 2-1 Summary of major capabilities of ICU bedside patient monitors	34
Table 2-2 Summary of major clinical data types available from clinical information systems	36
Table 3-1: Comparison of Major ICU Database	50
Table 3-2: Physiologic trends available in MIMIC-II.....	55
Table 3-3: Summary of Bedside Monitor-Generated Alarms in MIMIC-II Database.....	59
Table 3-4: Description of clinical data within CareVue tables	61
Table 3-5: Description of supplementary clinical data tables	63
Table 3-6: Summary of MIMIC-II database demographics and outcomes.....	67
Table 3-7: Major cardiovasucalar interventions in the MIMIC-II database	72
Table 3-8:Availability of major physiologic signals in MIMIC-II patient records.....	73
Table 3-9: Distribution and comparison of physiologic parameters from monitors and nurse-validated charting data	75
Table 4-1: Absolute and Relative Error Analysis	87
Table 4-2: Analysis of factors influencing agreement between NIBP and ABP	92
Table 4-3: Sensitivity of hemondynamic instability detector	96
Table 4-4: False positive rate of detector during hemodynamic stability	96
Table 4-5: Result on NIBP-ABP error prior to acute renal failure	101
Table 4-6: List of Vaso-active Medications.....	108
Table 4-7: List of hypnotics and analgesics.....	109
Table 5-1: Feature vectors used to create feature matrix of individual sentences (or records).....	133
Table 5-2: Classifier accuracy with synthetic physiologic time series from Library B..	163
Table 5-3: Comparison of Retrieval Accuracy of Different Algorithms on SYNTDATA Library.....	165
Table 6-1: Hemodynamic therapeutic interventions	174
Table 6-2: ICU Record segment stability/instability criteria	175
Table 6-3: Summary of hemodynamic event library	181
Table 6-4: Informativeness of different trend parameters.....	189

Table 6-5: Time Domain Parameter Characterization	192
Table 6-6: Class separation of static(approximation) wavelet symbols.....	198
Table 6-7: Class separation of dynamic wavelet symbol.....	198

1. Introduction

Similarly to other physical systems such as manufacturing plants or large-scale computer networks, the function and status of the human body must, at times, be vigilantly monitored to ensure dangerous states are avoided when possible, and appropriate interventions are applied in a timely manner when needed. Individuals with life-threatening conditions (trauma victims, acutely-ill intensive care unit (ICU) patients) are physiologically fragile and unstable, and require close monitoring and rapid therapeutic interventions. A major role of medical personnel in clinical settings is to make frequent observations, and to obtain multiple quantitative measures of organ system function. Physiologic and clinical data can vary widely based upon a particular patient's pathophysiologic state, and the inherent noise adds another level of complexity to their analysis and interpretation. The resultant data must be assimilated and interpreted to develop pathophysiological hypotheses that motivate therapy. The need for real-time assimilation and interpretation of medical data presents a motivation for the development of automated physiologic monitoring systems with capabilities to rapidly assess a patient's physiologic state. Furthermore, computerized analysis of clinical data can potentially illuminate "hidden" information that can be used to optimize clinical care. The development of pattern recognition and machine learning techniques suited for processing such signals is a major challenge in the advancement to the next generation of monitoring and clinical decision-support systems. This thesis is focused on utilizing data-driven pattern recognition to extract meaningful information from ICU data sources for retrospective as well as real-time applications capable of identifying deterioration in ICU patients.

1.1. Overview

In many sectors of our society, advances in computer networking, microprocessor speeds, and information storage technologies have resulted in an explosion in the volumes of data that are generated. In the modern Intensive Care Unit (ICU), these

advances have been coupled with a significant evolution in biomedical sensors and devices capable of obtaining an impressive array of invasive and non-invasive physiologic measurements. Providing life support in the ICU is becoming an increasingly complex task, however, because of the growing volume of relevant data from clinical observations, bedside monitors, mechanical ventilators and a wide variety of laboratory tests and imaging studies. The enormous amounts of ICU data and its poor organization make its integration and interpretation time-consuming and inefficient. Furthermore, the projected increase in the population of elderly patients requiring close ICU monitoring and predicted shortages of nurses and intensivists may exacerbate the predicament facing the healthcare system of tomorrow. The “information overload” that results may actually hinder the diagnostic process, and may even lead to neglect of relevant data, resulting in errors and suboptimal ICU care [22]. On the other hand, the richness and detail of the collected data make it possible for a new generation of “intelligent” monitoring systems to track the physiologic state of the patient, employing the power of modern signal processing, pattern recognition, computational modeling, and knowledge-based clinical reasoning. With the advent of wearable biosensors, automated physiologic monitoring is expandable to arenas outside of traditional medical settings, such as military combat and home healthcare.

The Artificial Intelligence (AI) community was hopeful that AI would have a great impact on healthcare. However, a brief assessment of the current state-of-the-art in ICU patient monitoring suggests that AI contributions have been minimal. While there have been significant advances in the storage, networking, and acquisition of signals that can be simultaneously monitored, AI applications within patient monitoring systems have been primarily limited to arrhythmia analysis.

There have been several impediments to integrating “expert systems” and other AI technologies into medical settings in general, and ICU environments in particular. Ideally, an advanced system should have access to all the data generated for a single patient: physiologic waveforms (e.g. ECG, blood pressures), vital sign measurements (heart rate, temperature, cardiac output), lab results (blood gasses, microbiology, blood chemistry), medication profiles (IV drips, bolus medications), and clinical notes (nursing notes, patient history). Traditionally, different types of data have resided on separate

systems in an ICU. However, industry has made several recent advancements that will likely solve the challenge of integrating and accessing data from a single system [32].

Once the data are accessible to an advanced monitoring system, the challenges of processing and interpreting large medical data sets will be formidable. The physiologic and non-physiologic “noise” in medical data poses a significant challenge in developing monitoring systems capable of detecting important physiological events while maintaining an acceptable specificity. Moreover, there is a wide variation in features within medical data sets for a large and diverse patient population. A short-coming of previous AI approaches to patient monitoring is that they have relied upon a “black-box” approach to alarm generation. Alarms are generated with little or no supporting evidence or rationale based upon the available clinical data. For example, a pattern recognition system may warn that a patient is likely to have an adverse cardiovascular event in the near future. However, such an alarm may not highlight the features in the clinical data that are supportive of one diagnosis versus another. The “expert-system” approach to physiologic monitoring relies on the development of a set of rules used to interpret data sets. These rules are derived from knowledge bases of human physiology and medicine. As the number of physiologic signals available for analysis grows and the class of disease processes varies more, the task of codifying knowledge bases into rules becomes more complex. There will also be inherent differences between two or more experts on the optimal interpretation of physiologic data. Finally, in the domain of high dimensional physiologic patterns, there may be subtle patterns of physiologic significance that are not characterized by expert rules due to limitations in human knowledge.

An alternative approach to “expert systems” is to utilize data-driven techniques based upon machine learning and statistical pattern recognition. Statistical clustering, principal component analysis (PCA), support vector machines (SVM), and Bayesian graphical networks are among the most popular machine learning techniques [13]. These statistical techniques can be adapted to utilize “training” or “labeled” data to learn classification rules, or attempt to classify data in an unsupervised mode without the use of training data. The advantage of the former technique is that the human labels and annotations would ideally guide the pattern recognition system to produce similar classification decisions to human experts. However, the process of generating a “good”

training set that is representative of the overall population of data can be quite challenging. There are domains such as computational biology and bioinformatics where the sheer amount of data are so voluminous and relatively new, that no training sets even exist or the training sets are quite limited in extent [14].

Unsupervised statistical pattern recognition that does not utilize annotated data for learning can be quite useful in such domains as computational biology, document and web page retrieval, and financial data analysis [13]. Such frameworks attempt to derive novel rules from data by identifying statistically significant patterns and trends in data that can be separated into different classes. To implement such techniques for high dimensional physiologic data, the high-dimensional space is extremely difficult to interpret, and thus identify meaningful “clusters”. This general rule has been referred to as the “Curse of Dimensionality” within the AI community [2]. In practice, high-dimensional pattern recognition problems require far more training samples so that statistically meaningful clusters can be identified. The use of physiologic knowledge bases can still be incorporated into the implementation of unsupervised machine learning-based methods. For example, the data must be pre-processed to remove noise. Physiologic knowledge can be utilized for assessing the fidelity of the data as well as attenuating the noise when possible. Machine learning and pattern recognition techniques are usually presented with a feature vector that is ideally a “good” representation of the data. For example, in developing a machine learning algorithm to classify ECG rhythms, a feature vector may consist of R-R intervals, QRS widths, P-R intervals, QT intervals, T wave amplitude, and ST elevation/depression. These features were chosen based on cardiac electrophysiology knowledge as a set of data representative of the underlying physiology. Similarly, in multiparameter patient monitoring, the feature vectors can be constructed by leveraging human physiology knowledge. Finally, once significant high dimensional patterns are identified by processing large data sets, it is vital that such patterns have plausible physiologic interpretations so that optimal care and therapeutic interventions can be guided.

The automated analysis and processing of ICU data need not be limited to real-time applications. Retrospective analysis of ICU clinical and physiologic data can also be of significant importance in improving the provision of the healthcare in the ICU [66].

For example, retrospective analysis of therapeutic strategies that lead to better outcomes can optimize future patient care. The Joint Commission for the Accreditation of Healthcare Organizations (JCAHO) established quantitative criteria based on automatically acquired ICU clinical data to indicate the quality of care in an ICU. Pattern recognition and data mining of ICU clinical data sets may identify other quantitative metrics of ICU care provision. Retrospective analysis of physiologic data may also provide insight into the fidelity of ICU physiologic measurements. For example, the ever-increasing utilization of non-invasive technologies, such as non-invasive blood pressure (NIBP) modules, may sacrifice measurement accuracy, and lead to suboptimal care based on the interpretation of erroneous physiologic measurements.

The present thesis addresses several challenges in developing advanced physiologic monitoring algorithms. In particular, the major focus of this thesis is towards the development of data-driven frameworks for physiologic pattern recognition of hemodynamic deterioration in ICU patients. We describe the design, acquisition, and management of a major ICU patient database of unprecedented size and scope that will serve as an ideal resource for refining and evaluating our algorithms. We also present novel algorithms for data pre-processing, feature extraction, and pattern recognition that can be applied to multiparameter physiologic and clinical data. In particular, we motivate the use of feature extraction techniques, such as wavelets, that can characterize the dynamics of physiologic data at several scales. The major contribution of this thesis is the development of a framework to “learn” models from massive physiologic datasets by fusing robust statistical pattern recognition techniques with simple physiologic knowledge bases. By leveraging the availability of clinically rich data with modern signal processing, machine learning, and medical knowledge bases, we hope to illustrate the potential utility of our proposed framework for the development of next-generation advanced physiologic monitors.

1.2. Thesis organization

We now present the major organization of the thesis. In Chapter 2, a review of the relevant literature is included and motivations for the approaches we have adopted in

addressing some of the pertinent challenges in the development of next-generation patient monitoring systems. In Chapter 3, we describe a new multiparameter ICU patient database (MIMIC-II) that we have acquired and developed, as well as the software infrastructure necessary for its management. We provide quantitative characterizations of the physiologic and clinical data in our database. In Chapter 4, we include examples of retrospective clinical studies that process ICU physiologic measurements and quantitative therapeutic data from MIMIC-II and discuss their implications for improving ICU care. In Chapter 5, we introduce the signal-processing and pattern recognition framework we have developed for multiparameter physiologic data streams. We motivate the use of wavelet analysis and feature extraction for compactly representing physiologic signals for real-time monitoring algorithms as well as retrospective temporal databases. We demonstrate that this technique can serve as a robust method of assessing the similarity of multiparameter trends from massive time series databases. In Chapter 6, we apply our physiologic signal processing and pattern recognition framework with simple cardiovascular models to identify early sign of hemodynamic deterioration in ICU patients. We assess whether the statistical similarity between ICU events is correlated to their physiologic similarity. We develop applications for clinical-decision support and database mining tasks as well as near-real time monitoring. In Chapter 7, we summarize the major contributions of this thesis and discuss future extensions of our research within the domain of ICU patient monitoring as well as other arenas.

2. Background

ICU patients tend to be physiologically fragile, and require close observation and constant interventions to maintain homeostasis. Clinicians in the ICU are challenged with integrating and assimilating ever-increasing volumes of clinical data into the decision-making process while caring for patients. Optimal ICU care provision will become even more challenging with predicted shortages in nurses and intensivists [32]. Furthermore, shifts in patient demographics are resulting in a growing elderly patient population with chronic ailments and increased susceptibility to disease, thus requiring more vigilant monitoring. Advanced patient monitoring and clinical information systems may alleviate the “information overload” problem plaguing ICU clinicians by providing timely alarms, clinical decision-support tools, and facilitating the rapid integration and assimilation of ICU data. However, a brief survey of commercially available ICU patient monitoring and clinical information systems reveals that only a relatively small proportion of the acquired data are ever manipulated beyond simple tasks such as storage [see Table 2-1]. In this chapter, an overview of the many challenges in developing ICU patient monitoring systems is presented. This chapter is organized in the following manner. In the next section, we summarize the major data types that are typically acquired and processed with ICU patient monitoring and clinical information systems. We also provide a brief synopsis of the major data processing tasks that are integrated into current state-of-the-art commercial ICU systems. Then, we describe some of the major research initiatives in advanced patient monitoring algorithms. Finally, we motivate the need for developing a more comprehensive database as a resource for ICU monitoring algorithm development in general, and the research presented in this thesis in particular.

2.1. ***Summary of monitoring system capabilities***

The modern ICU has undergone a significant evolution with the rapid advances of hardware and software technologies. Biomedical sensors have been developed to monitor continuously several physiologic measurements including: ECG, blood gasses, cardiac output, arterial, venous, and pulmonary pressures, and blood chemistry, such as serum

glucose. Another important advance in sensors is the introduction of minimally invasive or non-invasive technologies to replace or augment more invasive measurement systems [2]. Imaging technologies such as ultrasound systems, Computer Aided Tomography, and MRI are rapidly emerging as significant tools to diagnose disease. Advances in silicon chip technologies and bio-assays have now made it possible to obtain rapid lab results from small blood and urine samples at the patient's bedside [70]. With the dawn of the era of genomics, it may become possible to acquire genomic or proteomic profiles of ICU patients to guide optimal treatment pathways [6].

Along with the increased capability to acquire numerous and complex physiological measurements, technology has also evolved to improve device connectivity as well as information storage. Relational database software is being leveraged for storing clinical laboratory data in an efficient and accessible manner. Disk storage capacities have grown significantly over the years such that a single personal computer (PC) can store a terabyte of data. Standards such as HL7 for device communication are also receiving more attention for their role in facilitating seamless communication between different devices in the ICU [52]. Finally, networking technologies have advanced to allow devices to send data at speeds beyond 1 gigabit/second across wired and wireless networks.

These technological advances in the acquisition, storage, and transmission of electronic data have resulted in both a challenge and an opportunity in advanced ICU patient monitoring. The challenge for clinicians is in interpreting and assimilating the plethora of available data into hypothesis for guiding optimal therapies provided to patients. However, as the volume of data continues to grow, clinicians are being overwhelmed in completing all their required tasks. The aforementioned technological advancements have also created an opportunity for developing advanced patient monitors and clinical information systems. Such advanced systems would ideally be capable of processing physiologic and clinical data and providing clinically relevant interpretations, alarms, and displays to support clinicians in deciding on appropriate care of their patients.

It is useful to define terms used frequently in this thesis. Terms such as “physiologic data,” will generally refer to higher resolution measurements that are acquired automatically at the bedside in a continuous or near-continuous fashion by a patient monitor. For example, physiologic trends and numerics include high-resolution data that are computed on a minute-to-minute (sampled at 1/60 Hz) basis such as heart rate, systolic/mean/diastolic blood pressures, oxygen saturation, respiration rate, and cardiac output (see Table 2-1 for an extensive listing of physiologic trends in ICU patient monitors). Typically, physiologic trends are derived after processing physiologic waveforms sampled at even higher rates (125 Hz) such as ECG and intra-arterial blood pressure (ABP). Physiologic waveforms are processed at the beat-to-beat level. The features extracted at the beat-to-beat level are then saved and averaged over one-minute intervals to compute a value for each feature that is used to create the minute-to-minute physiologic trend data (see Figure 2-1).

“Clinical data” will generally refer to sparsely and irregularly sampled diagnostic measurements as well as therapeutic parameters. Diagnostic clinical data includes analyses of manually collected specimens of ICU patients such as blood gasses, complete blood count (CBC), and blood chemistry. Diagnostic clinical data are often grouped into different categories related to body systems (or physiologic functions) such as hematology, fluid-balance, cardiovascular and pulmonary functions. Clinical data also includes therapeutic intervention parameters such as the type and dosage of a medication or the settings of a ventilator as well as free-form nursing progress notes. Finally, clinically data also encompasses patient vital statistics such as age, gender, and weight. Clinical information is typically stored in a clinical information system that is used by clinicians for charting. Table 2-2 includes a more extensive listing of the different clinical data that are pertinent to ICU patient care and explored in this thesis.

Figure 2-1 is an illustration of the major signal processing and pattern recognition modules that are found in state-of-the-art bedside patient monitors. Typically, several physiologic waveforms (e.g. ECG, intra-arterial blood pressure, respiration signal) are recorded from ICU patients. Once the signals have been discretely sampled, low-pass

digital filters are utilized to remove noise that is outside the frequency bands (0-40 Hz) of physiologic waveforms. Furthermore, specific filters optimized for the unique characteristics of the different physiologic waveforms are also utilized. For example, a high-pass filter may be utilized to remove low-frequency noise such as baseline wander from an ECG. Waveform signal processing is beyond the scope of this thesis and the reader is referred to [16] for a more in-depth review of physiologic waveform analysis.

After the physiologic waveforms are pre-processed, several different feature extraction and pattern recognition algorithms are applied to the respective waveforms to determine the parameters that characterize a patient's physiological state. In the case of ECG analysis, pattern recognition may be used to discern features in a signal that are indicative of physiologic processes instead of noise. For example, algorithms are used to identify different ECG features representing the electrical activation of the human heart such as the R-wave (ventricular depolarization) and T-wave (ventricular repolarization). The R-R time intervals (period of time between R-waves in successive beats) can then be computed to derive an instantaneous (beat-to-beat) heart rate. The beat-to-beat heart rates are then averaged over running windows of approximately 10 seconds to arrive at an updated heart rate to display on the patient monitor. Another level of averaging over a window size of 1 minute in duration is applied to generate minute-to-minute heart rate numerics (or trends). The specific averaging processes will be described in the next chapter (Chapter 3) with greater detail as related to the physiologic trends utilized in this thesis.

In addition to ECG, other signals such as arterial and pulmonary pressure waveforms are analyzed by bedside monitor algorithms to identify characteristic hemodynamic features. For example, a patient's beat-to-beat systolic, mean, and diastolic blood pressure are estimated from the intra-arterial blood pressure (ABP) waveform features [see Figure 2-1]. Then, beat-to-beat values are averaged over different window sizes to generate physiologic trends.

Many of the minute-to-minute as well as beat-to-beat numerics have threshold-based alarms associated with them. If a numeric exceeds the bounds of the pre-determined thresholds, an alarm is triggered and announced to the clinical staff. In addition to numeric-based alarming, ECG arrhythmia algorithms trigger alarms based on pattern recognition of high-resolution waveform data to identify waveform morphologies indicative of disturbances to the heart's electrical conduction system.

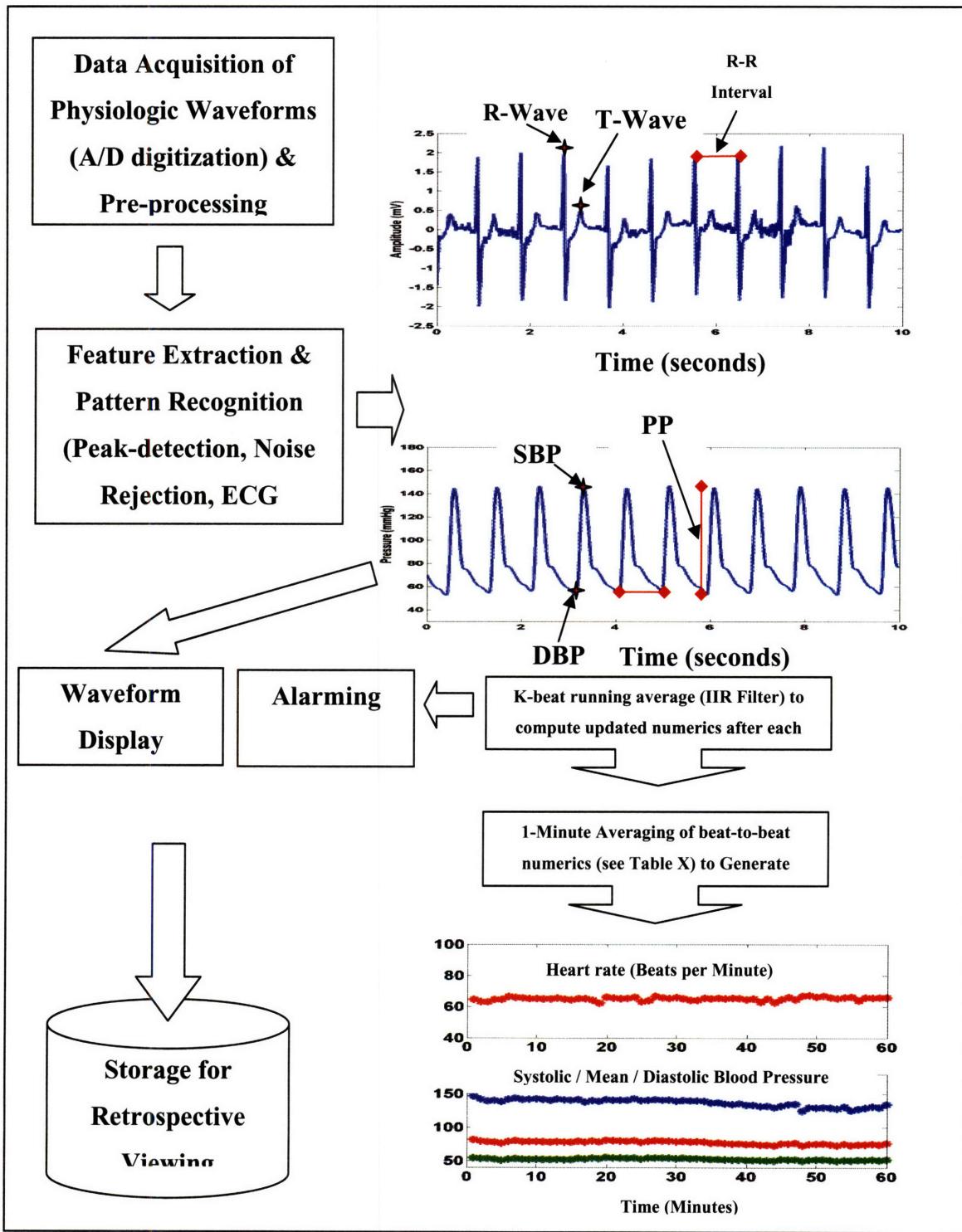


Figure 2-1: Physiologic signal processing flow diagram

A major focus of this thesis centers on developing algorithms that characterize physiologic trends of ICU patients that may be suggestive of hemodynamic instability. Physiologic trends generated by patient monitoring devices are corrupted frequently by noise. Noise is a major cause of false alarms and presents a challenge in developing more advanced monitoring algorithms. While there are several resources available in the literature that provide more extensive reviews of noise [16], a brief overview is included here to introduce the reader to various sources of noise in ICU patient monitoring data.

Physiologic waveforms are the highest resolution data and may be corrupted by noise processes. A noise source can be defined as any process that alters the value of a measured phenomenon (such as the radial artery blood pressure). Abrupt noise with rapid changes to a physiologic signal's value is frequently due to patient motion, and is merely one example of the type of noise that may appear in blood pressure trends. A clot formation in a pressure catheter over a time-course of minutes to hours may cause a gradual dampening of the acquired blood pressure waveform. A damped waveform will result in erroneous systolic and diastolic blood pressures. In pressures that are typically lower than systemic arterial blood pressure such as the pulmonary artery pressure and central venous pressure, shifting a patient's position relative to the height of a pressure transducer will cause a "step" change in the measured pressure. Ideally, advanced patient monitoring systems should identify changes in physiologic measurements that are due to underlying physiologic changes in a patient while suppressing noise.

Table 2-1 Summary of major capabilities of ICU bedside patient monitors

Acquired Signal	Recorded Parameters	Alarms Present
ECG	HR	Arrhythmia, HR, ST
Invasive Arterial Blood Pressure	HR (derived from blood pressure), Systolic/Mean/Diastolic BP	High/Low Threshold-based BP alarms
Pulmonary Artery Pressure	Systolic/Mean/Diastolic PAP	High/Low Threshold-based
Central Venous Pressure	CVP	High/Low Threshold-based
Non-invasive Blood Pressure (NIBP)	Systolic/Mean/Diastolic BP	High/Low Threshold-based
Respiration	Respiration Rate	High/Low Threshold-based
EEG	Bi-spectral Index (BIS)	High/Low Threshold-based
Cardiac Output (CO)	CO	High/Low Threshold-based
Mixed Venous O ₂ Saturation (SvO ₂)	SvO ₂	High/Low Threshold-based
Temperature	Temp	High/Low Threshold-based
Capnography	Expired CO ₂	High/Low Threshold-based
Ventilation Parameters	FiO ₂ , tidal volumes, etc	High/Low Threshold-based
Anesthesia Gases	Anesthesia partial pressures	High/Low Threshold-based
Trans-cutaneous Gases	O ₂ , CO ₂	High/Low Threshold-based

Figure 2-2 includes an example of approximately 140 hours of high resolution numerical trends of invasive arterial blood pressure that are generated continuously at one sample per minute by processing the acquired blood pressure waveform. In this example, the trends are corrupted by noise. In particular, the segment of data after 120 hours in the record is heavily corrupted by noise.

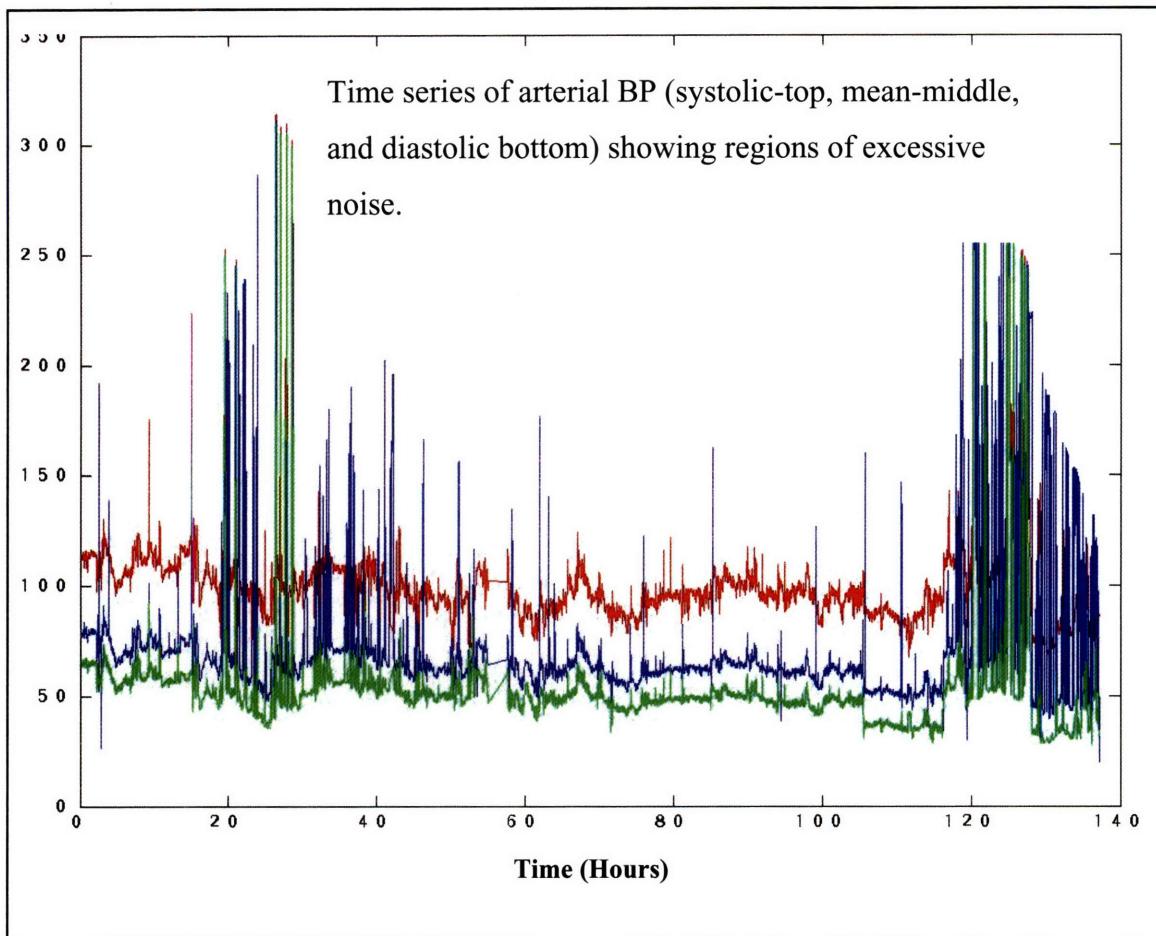


Figure 2-2: Physiologic trends with noise

Table 2-2 is a brief description of the major classes of clinical data stored in a clinical information system. Clinical information systems generally store sparser data than bedside patient monitors. Clinical data are also irregularly sampled and subject to erroneous values due to measurement error. However, while clinical information data are generally sparser, clinical data are far more diverse and may include hundreds of different types of laboratory measurements and therapeutic interventions. Clinical information systems have hitherto served as an electronic nurse charting system with little focus on advanced algorithm development capable of assessing a patient's physiology. For example, Figure 2-3 is a sample of clinical data that includes the nurse-validated vital signs and arterial blood gasses and pH over an interval of 100 hours. While there are physiologically significant changes to the acid-base status of this particular patient, clinical information systems have provided little or no interpretation of the stored clinical

data. Researchers have introduced the concept of clinical advisories that alert the clinical staff when a laboratory value is beyond physiologically acceptable bounds [65]. However, several pathophysiologic states (requiring different treatments) can result in a common physiologic measurement deviating from a normal value. For example, both cardiogenic shock and hypovolemic shock may result in low blood pressure alarms and thus, require different therapies. Aggressive intravenous fluid administration would be an appropriate therapy in a state of hypovolemic shock but may be detrimental if given to a patient in cardiogenic shock.

Table 2-2 Summary of major clinical data types available from clinical information systems

	Example	Typical Charting Interval
Nurse-validated Vital Signs	HR, BP, RR	Hourly
Medication Interventions	IV Medication Drip Rate	Hourly
Ventilator Settings	PEEP, PIP, FiO ₂ , Ventilation Mode, Tidal Volume	Every 8 Hours or more frequently with changes to settings
Clinical Laboratory Measurements	Serum Sodium, Complete Blood Count (CBC)	Every 12 to 24 Hours
Fluid Balance I/O	Hourly IV Fluid Input, Hourly Urine Output	Hourly
Nursing Progress Notes	Free-form text notes summarizing patient state	Every 8-12 Hours

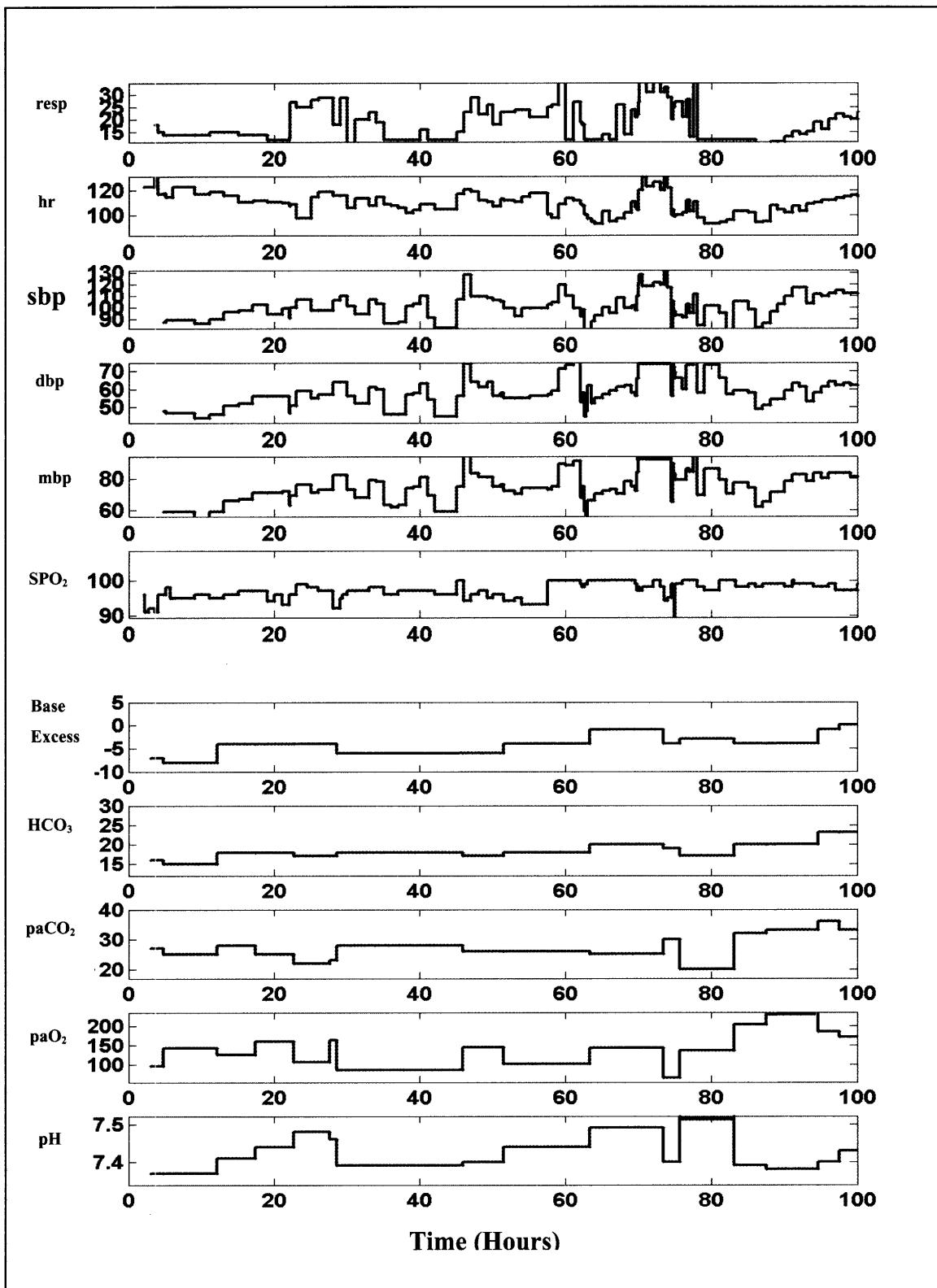


Figure 2-3: Example of ICU clinical data trends

2.2. Advanced patient monitoring algorithms

An intriguing and important challenge is to explore the extent to which the wealth of available (and machine-accessible) patient data can be used to formulate automatically dynamic pathophysiologic models of the patient's changing clinical status. Such models or hypotheses would provide rational structures around which to present the data to clinicians, would provide the basis for more sophisticated and sensitive alarms, and would play a pivotal role in developing decision support paradigms to guide therapy. Over the past fifteen years, research in academia and industry has begun to address the general area of "intelligent patient monitoring" [52].

The only acquired physiologic signal that is processed with sophisticated algorithms to provide clinically significant interpretation is the ECG. Most bedside monitoring systems are capable of generating arrhythmia alarms. However, the majority of physiologic signals and parameters are never interpreted beyond simple single-parameter comparisons to high and low thresholds. For example, multiparameter algorithms capable of identifying a gradual increase in heart rate concomitant with a decrease in arterial blood pressure may be more specific for shock, but have not been utilized in monitoring algorithms. The development and deployment of such algorithms in an FDA-regulated industry are considerable. Such advanced algorithms that operate on multiparameter physiologic data require large sets of "real" data for algorithm refinement, performance characterization, and FDA validation studies. In the following discussion about prior research in the area of intelligent patient monitoring, the availability of large sets of realistic ICU measurements across diverse patient populations emerges as one of the primary reasons that such algorithms have not been deployed in clinical settings for real-time monitoring applications.

Intelligent patient monitoring (IPM) refers to patient monitoring systems that are capable of one or more of the following tasks related to the interpretation and display of

physiologic patient data: multiparameter processing and interpretation, temporal reasoning and trend prediction, physiologic modeling and hypothesis generation, and sophisticated display and visualization of physiologic data.

A major challenge to a system performing any of the aforementioned tasks in intelligent patient monitoring is the inherent noise present in physiologic and clinical data. In this thesis, “noise” refers to the corruption of quantitative measurements of physiologic function due to non-physiologic sources. For example, noise in an ECG is commonly due to electrode movement or poor contact with the skin. Such noise corruption often leads to false arrhythmia alarms [16] or improper heart rate calculations. Noise may be additive to the underlying signal, such as a mild ECG baseline wander due to breathing, or it can completely replace an underlying signal, such as a sensor saturation or signal dropout.

Signal processing for monitoring has many challenges. Physiologic waveforms such as ECG and invasive arterial blood pressure (ABP) can be corrupted by non-physiologic and physiologic sources, and derived parameters such as systolic blood pressure (SBP) or heart rate (HR) may be corrupted by algorithmic or instrumentation errors. Furthermore, there are intervals in which such data may be missing, as when a patient is being transported between units. A consequence of noise includes the generation of false monitoring alarms.

Traditional monitoring alarms are rules that have been defined to trigger the enunciation of a message to warn the clinical staff of a dangerous physiologic state such as an arrhythmia, low blood pressure, apnea, or low oxygen saturation. The rules are triggered when the monitored physiologic signal or an abstraction of the signal is compared to a decision boundary that delineates an alarm from a “normal” state.

Several studies have suggested that over 80% of alarms generated by the current generation of monitors can be classified as “false-positive” or “clinically irrelevant.” [75, [76]. The high number of false alarms leads not only to sleep deprivation for patients and stress for patients and staff, but also to wasted time, resources, and neglect of truly

dangerous events. For example, one study recorded thirty-three separate audio sounds in the ICU and asked staff to listen to the tones individually in a quiet room; they correctly identified only 50% of the critical alarms [9]. Multiple monitors, each capable of producing multiple false alarms, are the norm in typical ICUs; in some units, it is possible to count more than forty alarm sources for a single bed – the mechanical ventilator, telemetry, pulse oxymetry, infusion pumps, etc. -- essentially none of which are integrated to provide more useful data.

Advanced monitoring systems must deal robustly with noise, artifact, and data loss such as illustrated by Figure 2-2. To address the challenge of reducing false alarms, there has been a significant amount of research in the area of noise detection and noise suppression in physiologic data [74]. Zong and colleagues developed a fuzzy-logic based algorithm to detect noise in ABP waveforms by fusing information from the ABP and ECG waveforms. Zong's algorithm significantly improved the specificity of the blood pressure alarms of ICU patients while preserving the alarm sensitivity [84]. Tsien et al. developed trend-based noise detection algorithms using several different machine-learning techniques such as neural networks and support vector machines [74]. Multiparameter data streams were projected onto predefined templates, and machine-learning algorithms were trained on annotated ICU data such as HR, ABP, and O₂ saturation. These techniques were successful in significantly improving the specificity of vital sign alarms. The data sets were limited in size, however (~300 hours total), and did not include all the measurements that are captured in a ICU. The prospective nature of Tsien's study resulted in the acquisition of an insignificant number of physiologic events with hemodynamic instability. Thus, it is difficult to evaluate the clinical efficacy of such algorithms. There is clearly a need for an extensive collection of real-world data from ICU patients to support further development and evaluation of advanced monitoring technology.

Once data have been pre-processed to reduce artifacts and noise, the next signal-processing task involves extracting features from waveforms, vital signs, and clinical measurements. The sampling frequency and regularity of different measurements may

vary widely; for example, heart rate can be calculated every second, but there may be several hours between consecutive arterial blood gas measurements. Information from different sources can also be inconsistent; for example, an invasive ABP value may drop dramatically during periods of cuff inflation, while the HR measurement from the ECG remains unchanged. Thoraval et al. developed a ventricular rhythm tracking algorithm that fused analysis from the ECG and the invasive ABP waveforms in short, but noisy datasets using simple binary logic [73]. However, to properly evaluate the performance of such an algorithm under different stresses, richer physiologic data sets are needed that include different pathophysiologic states, effects of therapies, different classes and severities of noise, and different patient populations.

Several recent studies describe examples of prognostically or therapeutically useful physiologic signals and parameters. One landmark study showed improved outcome in early goal-directed therapy in sepsis using a protocol that followed continuous ScvO₂ as well as urine output, MAP, and CVP [61]. Identification of additional useful parameters that may lead to earlier warning for potential decompensation, more timely treatment, and more accurate determination of prognosis requires large amounts of data. Larger data sets not only increase the statistical power of such studies, but can also better characterize the performance of monitoring algorithms across different patient populations, pathophysiologic states, and treatment protocols.

Another challenge that has received limited attention among researchers centers on developing algorithms with predictive capabilities. Arrhythmia algorithms are quite sensitive to the onset of ventricular fibrillation or tachycardia. However, little progress has been made in developing algorithms that can predict the likelihood of a future episode of ventricular fibrillation based on present and past data. Such an alarm may help focus the clinical staff's attention to a problem when it is more manageable and requires less aggressive interventions. Since true ventricular fibrillation is such an "infrequent" event in the ICU, collection of a database with a sufficient number of true ventricular fibrillation episodes and false alarms would greatly aid researchers engaged in the development of such an algorithm.

Acuity scores such as APACHE and SAPS I-II, were developed to predict patient mortality in the ICU and to categorize the severity of illness of patients for clinical studies and unit performance metrics. Typically, a set of observed clinical variables and vital signs are assigned a certain weighting based upon how far their values are beyond normal ranges. The composite score is the summation of these weighed observations and is mapped to a likelihood of ICU mortality [42]. However, these acuity measures were not optimized to guide therapy or alert the staff of impending problems. Rather, such acuity metrics are utilized as measures of ICU performance in terms outcomes or length of stay. The development of real-time acuity metrics that are optimized to outcomes other than mortality may prove to be useful in guiding more timely therapy. For example, one can imagine optimizing a formula based on a set of dynamic physiologic and clinical variables with respect to an outcome such as the development of hypotension or shock.

Several researchers have proposed model-based frameworks for interpreting physiologic and clinical data. Zhao et al. developed an expert system for interpretation of hemodynamic data of ICU patients using a lumped-parameter cardiovascular simulation and a knowledge base of the static cardiovascular system. Although the decision process did not incorporate the valuable information in past values of physiologic trends and assumed the data to be noise-free, this expert system performed comparably to human experts over several simulated data sets [82]. The SIMON project was another effort toward the development of a model-based monitoring system [10]. This prototype system estimated the physiologic state of patients and predicted the temporal evolution of monitored variables, and also ranked the importance of monitored data streams based upon the context.

The reader is referred to [26] for a more in-depth review of ICU alarms research. The aforementioned studies exemplify only a limited set of the challenges confronting researchers seeking to develop and deploy ICU monitoring systems. A major commonality between all such projects centers on the limited access to sufficiently rich, “real” clinical data support ongoing research and validation of novel algorithms.

2.3. ICU patient database to support advanced algorithm research

To develop robust automated patient monitoring algorithms, large amounts of well-characterized clinical test data are needed. The complexity of such data renders hopeless efforts to develop analytic techniques by intuition or reasoning alone. It is essential to develop and test algorithms with real data, and to be able to perform such tests repeatedly and reproducibly as algorithm refinements evolve. The availability of large training sets becomes even more paramount when dealing with algorithms that attempt to process and classify several parameters and measurements simultaneously.

The process of creating a comprehensive ICU database requires development of a high-throughput data archiving system capable of storing several physiologic and clinical data streams, detailed profiles of therapeutic interventions, relevant diagnoses and problem lists, and contextual understanding of the patient ICU stay from nursing and/or physician progress notes. The introduction of clinical information systems [15] has greatly facilitated much of this data collection. Access to high-resolution physiologic waveforms requires understanding and decoding of proprietary vendor data communication and storage formats.

Along with the “raw” data that is acquired, clinician-reviewed annotations may be developed by retrospectively reviewing the archived ICU records. Accurate and detailed annotations are not only “gold standards” to evaluate monitor performance, but would also facilitate research in data mining and knowledge-discovery using rich clinical data sets. Advances in computer networking, disk storage, and processing power now make it possible to acquire and organize an ICU patient database of unprecedented size and scope. An annotated database that is rich in physiologic and clinical information will be a

tremendous asset to the research community at large, and may help realize the development of next-generation intelligent patient monitoring systems.

In the next chapter, we describe the new database that we have developed and quantitatively characterize its relevant attributes. We also provide a review of other ICU databases that have been developed and compare their features and specifications to our database.

3. Design and Characterization of a Large Scale Temporal ICU Database: MIMIC-II

3.1. *Introduction*

Chapter 2 provided an overview of ICU patient monitoring. ICU patient monitoring systems are now able to acquire an impressive array of detailed measurements from each patient. In the long term, monitoring systems are needed that not only report these measurements to human users but also form pathophysiological hypotheses that best explain the rich and complex volume of relevant data from clinical observations, bedside monitors, mechanical ventilators and a wide variety of laboratory tests and imaging studies. Such systems should reduce the ever-growing problem of information overload, and provide much more clinically relevant and timely alarms than today's disparate limit-based alarms.

The development of such advanced monitoring and clinical decision support systems requires large amounts of well-characterized clinical test data. Digital recordings are ideal test data for these purposes, as has been well demonstrated in research on cardiac arrhythmia and monitoring of transient ischemia [19]. The need for large-scale physiologic databases becomes even greater when the algorithms under development operate on several physiologic signals. Current ICU bedside patient monitors primarily deploy algorithms designed to process a single class of physiologic signals such as analyzing the human electrocardiogram for the presence of an arrhythmia. However, newer algorithms may fuse several continuous physiologic signals with intermittent

laboratory measurements to arrive at a more integrated approach to patient monitoring and clinical decision support. In general, as the dimensionality of the input data increases, an algorithm requires a much larger training set of data to optimize its performance [2]. Thus, in developing a new ICU patient database, particular emphasis should be placed on developing a high throughput archiving system that would enable the acquisition of a sufficiently large number of ICU patient data sets.

Large scale physiologic databases can also be utilized to support data mining research to identify and characterize novel multidimensional dynamic patterns that would be challenging for the un-aided human to recognize. For example, identifying correlations between the time of day, ICU staffing levels, and simultaneous occurrence of hypotension, tachycardia, and oliguria (low urine output) may provide useful quantitative feedback regarding the impact of ICU staffing shortages on patient care. Such data mining algorithms may also uncover unique physiologic signatures that may be predictive of future hemodynamic instability. Shoemaker et al developed a database of trauma patients to study the development of novel acuity scoring algorithms based on noninvasive data recorded from trauma patients [67]. Retrospective analysis of ICU physiologic databases can also be utilized to identify iatrogenic causes of patient morbidity and mortality. For example, Schoenfeld et al [68] utilized ICU records to study different measures of efficacy in clinical trials of treatments for acute respiratory distress.

The major goal of this chapter is to introduce the MIMIC-II database as a new resource for developing ICU patient monitoring algorithms. The database will be described and characterized in terms of its attributes so that a researcher can define the scope of questions that can be investigated with the MIMIC-II database. This chapter is organized in the following manner. The next section provides an overview of previous ICU databases. Section 3.3 describes the new database (MIMIC-II) that we have developed. The methodology for acquiring and integrating ICU patient records is briefly reviewed. Section 3.4 provides a quantitative and qualitative analysis of MIMIC-II to characterize the patient demographics and morbidities as well as the physiologic and clinical data in the database.

3.2. Overview of ICU databases

Several efforts have been made in developing ICU databases aimed at improving various aspects of patient care. ICU databases can be classified along several different attributes. Perhaps the most significant attributes include: number of records, patient populations (e.g. medical, trauma, coronary, neonatal), sparseness and completeness of ICU records (e.g. sparse clinical data, clinical progress notes, higher resolution physiologic, therapeutic intervention profile, waveforms), inclusion of clinician annotations, and the availability of the database to other researchers (e.g. commercial, academic, freely downloadable). Table 3-1 includes several of the more widely-published ICU databases that have been developed by researchers. A brief comparison of the qualities of these different databases is included.

Within the critical care community, well-known databases including APACHE [83] and Project IMPACT [8] have resulted in the acquisition of hundreds of thousands of ICU patient cases from dozens of hospitals throughout the United States of America. Researchers have utilized these databases to support clinical studies. Among the main contributions of PROJECT IMPACTICU and APACHE are supporting the development of several classes of acuity scoring systems such as APACHE I-IV, SAPS, SOFA, and MPM [42]. These databases have also supported researchers in developing guidelines for various ICU practices such as goal-directed therapy for septic patients [61]. However, these aforementioned databases failed to capture high resolution information such as waveforms, minute-to-minute vital signs, and hourly charts of medications, labs, and fluid balance. Higher resolution data are necessary for ensuring that transient but clinically significant physiologic events are not overlooked. Also, clinician-validated annotations to serve as “gold-standards” are not available. As demonstrated in previously published databases, annotations can be an invaluable resource for improving patient monitoring algorithms [19].

Among the first multiparameter ICU databases, the Massachusetts General Hospital / Marquette Foundation (MGH/MF) Waveform Database was collected in 1992 in Massachusetts General Hospital critical-care units, during surgery, cardiac catheterization, or other electrophysiology studies [35]. The MGH database was primarily designed to support research in waveform analysis and includes several channels of hemodynamic and ECG waves. However, the database does not contain bedside monitoring alarms, numerics, and relevant clinical and laboratory data.

The SIMON (Signal Interpretation and MONitoring) project began at Vanderbilt in 1998 and is among the largest multiparameter trauma ICU databases [10]. SIMON was developed to support research into the development of intelligent patient monitoring algorithms that fuse physiologic and clinical data. SIMON is an ongoing project and has thus far resulted in the collection of over 3000 ICU patient records. The SIMON database consists of several streams of second-to-second vital signs (heart rate, mean blood pressure, respiration, etc) acquired from bedside monitors as well as laboratory measurements. The vast majority of SIMON patient records do not contain actual physiologic waveforms nor do they include records of therapeutic interventions and relevant free-text clinical data such as nursing progress notes.

The IMPROVE database was collected in Finland and resulted in the acquisition of 59 ICU records [36] with significant cardiovascular and pulmonary disorders. However, each record was approximately 24 hours in length. The recordings were annotated in real-time by physicians at the bedside to denote significant physiologic state changes in a patient. IMPROVE records include relevant clinical and laboratory data for each patient record. Following the IMPROVE database, the same research group developed the IBIS database and included similar types of ICU data from 200 patients with significant neurological disorders.

The MIMIC (Multiparameter Intelligent Monitoring in Intensive Care) database (referred to as MIMIC-I in this thesis) [51] was also developed to support research in intelligent patient monitoring in the ICU. MIMIC-I was dependent on cumbersome data acquisition technologies that required a great deal of human overview during the recording period.

Researchers attempted to record 24 to 48 hours of continuous hemodynamic, respiratory, and ECG waveforms from bedside monitors of patients that were deemed to have an increased likelihood of hemodynamic instability during the recording interval. However, because it was not feasible to record an entire ICU stay for a patient, a recording period of 24 hours may not have been sufficient to encompass a period of hemodynamic instability. The relevant clinical data were manually extracted from paper records by research nurses. The low-throughput data collection methodology resulted in the assimilation of approximately 100 patient records with high resolution waveforms (500Hz), physiologic parameters, bedside alarms, and a relational database to store the associated clinical and laboratory data. The major clinical problems are available for each record, but significant hemodynamic events were not annotated. Towards the end of the MIMIC-I project, newer data acquisition technologies were available that allowed for the development of MIMIC-II.

Table 3-1: Comparison of Major ICU Database

	MGH	SIMON	IMPROVE	IBIS	APACHE & IMPACT-ICU	MIMIC-I	MIMIC-II
Category of Database	Data from surgery, cath. or electro-physiologic studies	Trauma ICU	ICU with (hypovolemia, heart failure, sepsis, or respiratory problems)	ICU and OR patients needing neuro-monitors	ICU (data gathered for large-scale outcomes and protocol studies)	SICU MICU CCU Patients with hemodynamic instability	SICU/MICU/CCU U/CSRU data from patients with hemodynamic instability and/or other disorders
Number of Records	250	~3000	59	200	> 100,000	100	(~2100 Complete records to date with physiologic and clinical data) ~14000 records with clinical data only
Record Length	12-86min 1hr avg	Entire ICU stay (3.2 day mean stay)	24 hrs	3 hrs	Entire ICU stay	24-48 hrs	Entire ICU stay (7.2 day mean stay)
Physiologic Waveforms	3 ECG/ABP/ PAP/CVP LAP/CO2	None	2 Leads of ECG, ABP, PAP, CVP, RESP, AWF, AWP, O2, CO2, AA, Evoked Potentials	2 Leads of ECG, 2 Channels of EEG, ABP, PAP, CVP, RESP, AWF, AWP, O2, CO2, AA, Evoked Potentials	NA	All Monitored Waveforms (Multi-lead ECG,ABP, PAP,CVP, RESP)	Up to 4 Simultaneously monitored waveforms Typically: (Two Leads of ECG,ABP,PAP)
Vital Signs & Numerics	NA	1 Value every second for all monitored	1 Value every 2 Minutes for all monitored and derived numerics	1 Value every 2 Minutes for all monitored	NA	1 Value every second for all monitored	1 Value every 1 Minute for all monitored and derived numerics

		and derived numerics		and derived numerics		and derived numerics	
Bedside Alarms	Available	NA	NA	NA	NA	Available	Available
Laboratory/ Clinical Data	NA	Partial Data	Available	Available	Available	Available	Available
Clinician Notes	NA	NA	NA	NA	NA	Available	Available
Therapy Profiles	NA	NA	Nursing Actions Only	NA	NA	Available	Available
ICD9 Codes / Problem Lists	NA	NA	Available	Available	Available	Available	Available
Annotation	NA	NA	Patient State & Nursing Actions	Patient State & Physician Free Text	NA	NA	Ongoing
Commercial or Academic	Academic	Academic	Academic	Academic	Commercial (Owned by Cerner Corporation)	Academic	Academic
Availability	WEB	Private	Fee-based availability	N/A	Fee-based availability	WEB	Only physiologic data available via web

3.3. Creation of the MIMIC-II Database

There were several formidable challenges in creating the MIMIC-II database. The acquisition and integration of several disparate streams of physiologic and clinical data needed detailed understanding of proprietary data formats and database schema. Furthermore, advanced monitoring algorithms that process very high dimensional data require an unprecedented large set of representative training examples. Thus, the data acquisition technology needed capabilities of simultaneously recording from several ICU beds over an entire ICU stay. Once the data are acquired, an open-source, user-friendly, secure, and scalable database was developed to allow researchers to analyze and annotate ICU patient records [63]. In order to meet these challenges, a data acquisition system was

designed in close collaboration with the aid of the patient monitoring system vendor (Philips Medical Systems-Andover, MA).

We will describe the technology that was used to prospectively acquire physiologic signals and clinical data generated from ICU bedside monitors and clinical information systems. Then, we describe the process by which MIMIC-II ICU patient records are synthesized and integrated into cohesive patient records that can be annotated by clinicians. We also describe the major issues in developing a generalized framework for physiologic and clinical database development that can be extended across different clinical sites.

3.3.1. Physiologic data acquisition from ICU bedside monitors

In a typical ICU, each patient's physiologic state is monitored in real-time by the placement of several electrodes and catheters to acquire physiologic waveforms such as ECG and arterial blood pressure. The signals are analyzed and displayed by a bedside patient monitor that may alarm if a monitored physiologic variable deviates from an acceptable range. An ICU may also have a networked monitoring architecture so that all the bedside monitors can transmit their acquired data to a central computer. This centralized terminal (referred to as a "central station") allows a clinician (nurse or physician) to observe the status of the monitored variables of all ICU patients in a unit at a central location. The networked central computer (database server) is capable of saving the monitoring data of each patient for up to 96 hours so that a patient's data may be analyzed retrospectively. The acquired data are purged from the database server once a patient is discharged from the unit. Through collaboration with the monitoring system vendor (Philips Medical Systems, Andover, MA), an archiving system was developed to create permanent archives of the physiologic data residing in the central database server. The monitoring data were recorded in the vendor's proprietary data formats [38] and was subsequently translated to the MIT WFDB format [19].

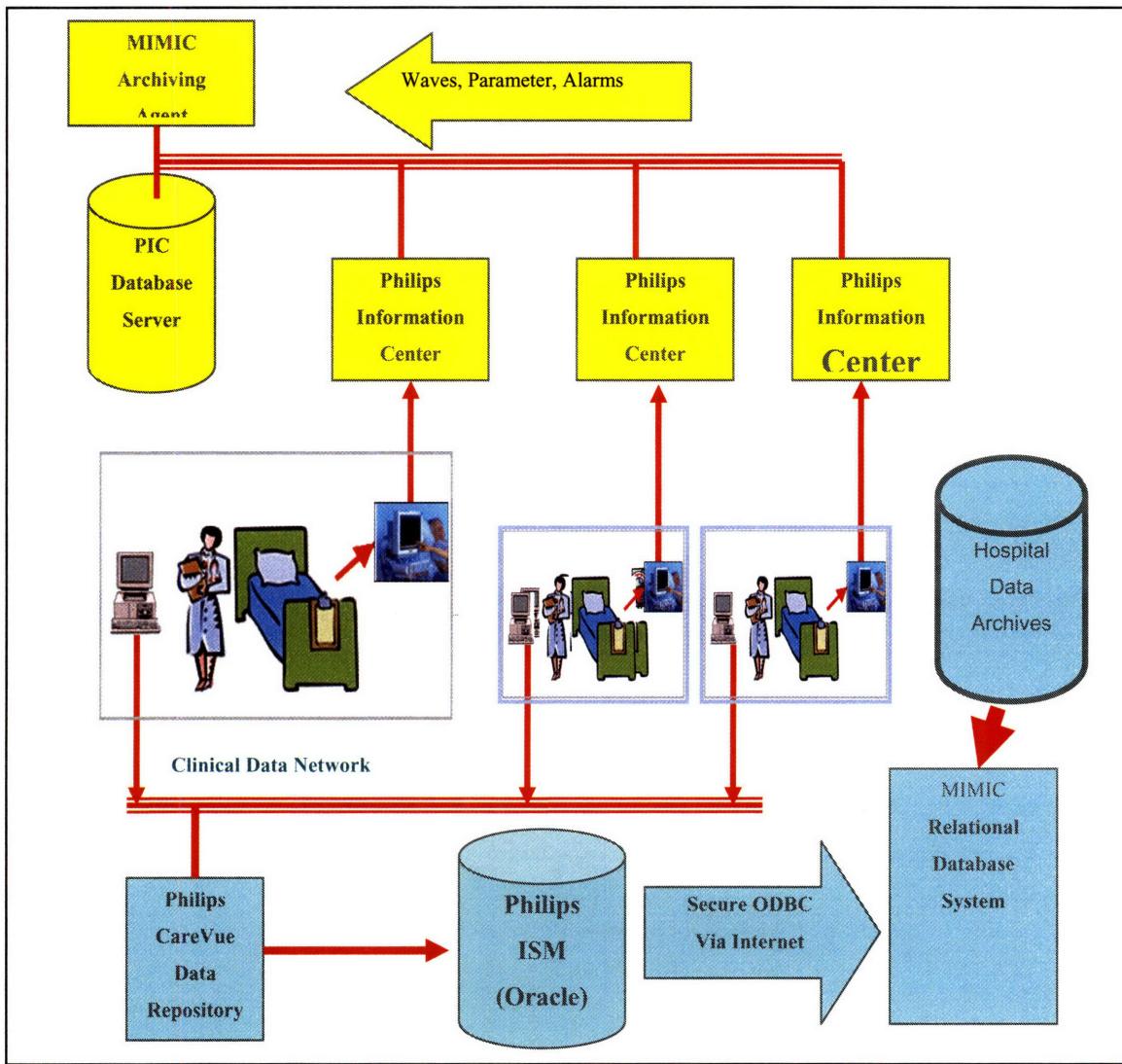


Figure 3-1: MIMIC-II physiologic data archiving architecture

We monitored up to 48 simultaneous ICU beds from Medical Intensive Care Unit (MICU), Coronary Care Unit (CCU), Cardiac Surgical Recovery Unit (CSRUs), Surgical Intensive Care Unit (SICU), and Trauma SICU (T-SICU). Each patient record commenced with a patient admission and ended with a final discharge from the ICU. Each record consisted of four continuously monitored waveforms (2 Leads of ECG, Arterial Blood Pressure, and Pulmonary Artery Pressure) sampled at 125 Hz, 30 1-minute parameters (HR, BP, SpO₂, Cardiac Output), and monitor-generated alarms and in-ops. “In-ops” refer to monitor-generated alarms that indicate states such as electrodes being taken off a patient, and thus, precluding the monitoring of the intended physiologic

signals. The waveforms and parameters were originally sourced from Philips CMS bedside patient monitors (Philips Medical Systems, Andover, MA). The bedside monitoring data were then transmitted to a Philips Information Center Database Server (PICDBS). With the assistance and cooperation of the manufacturer, a customized archiving agent was developed to query the PICDBS. The archiving agent, equipped with a 200 Gigabyte hard drive, continuously retrieved and stored the waveform, parameter, and alarm data from all the monitored ICU beds. At approximately two-week intervals, all completed records were downloaded from the archiving agent and subsequently purged from the archiving agent. The physiologic data were retrieved from the archiving agent onto hot-swappable USB 2.0 or IEEE1394 (“Firewire”) hard-drives.

3.3.2. Specifications of physiologic data

Physiologic Waveforms

As mentioned previously, the physiologic waveforms can include up to 4 simultaneously monitored signals. The archiving system was configured to archive 2 leads of ECG, 1 arterial blood pressure (ABP) waveform, and 1 pulmonary artery pressure (PAP) waveform. Figure 3-2 is an example of a set of physiologic waveforms included in a MIMIC-II patient record. The archiving technology did not allow for prioritization or substituting pressure waveforms. Thus, if an ABP or PAP was not available, the total number of recorded waveforms for that interval would decrease. The ECG waves are saved with 8-bit quantization at a sampling rate of 125 Hz using a peak-picking algorithm from an original 500 Hz ECG signal [28]. The pressure waveforms are saved with 8-bit quantization at a resolution of 125 Hz using conventional decimation from the original 500 Hz waveform signals.

3.3.3. Physiologic parameter trends

Physiologic trends and numerics include high-resolution data that are computed on a minute-to-minute (sampled at 1/60 Hz) basis such as heart rate, systolic/mean/diastolic arterial blood pressures, oxygen saturation, respiration rate, and cardiac output (see Table 3-2 for a listing of physiologic trends available in the MIMIC-II database).

Table 3-2: Physiologic trends available in MIMIC-II

Trend Classes (1 sample/minute)
Heart Rate
Systolic (SBP), mean (MBP), diastolic (DBP) arterial blood pressures
Systolic (SPAP), mean (MPAP), diastolic (DPAP) pulmonary arterial blood pressures
Oxygen Saturation (SpO_2)
Respiratory Rate (Resp)
Cardiac Output (Thermodilution Method:TCO) {intermittent measurements}
Pulmonary Artery Wedge Pressure (PAWP) {intermittent measurements}

Typically, physiologic trends are derived after processing physiologic waveforms such as ECG and ABP. Physiologic waveforms such as ECG and ABP are processed by monitoring algorithms at a beat-to-beat level to produce parameters for each beat. The respective parameters of each beat may include, for example, an instantaneous heart-rate from an ECG signal or a beat's systolic, mean, and diastolic blood pressure in an ABP waveform. The beat-to-beat parameters are then averaged over a pre-defined window size (where the window size is based upon the number of beats within a window). Equation 1 is an infinite impulse response (IIR) filter that is used to compute the running average ($P_{avg}[i]$) for pressure parameters that is updated on a beat-to-beat basis. The number of beats in a window is dependent on the signal being analyzed. After this first stage of averaging the beat-to-beat numerics, another level of averaging is utilized on a minute-to-minute basis. Thus, the final physiologic parameters represent a one-minute average of the IIR filter output during each non-overlapping minute.

Equation 1

$$P_{avg}[i] = P_{avg}[i-1] + \frac{(P[i] - P_{avg}[i-1])}{k[i]}$$

$P_{avg}[i]$ = averaged pressure value numeric

$P[i]$ = beat-specific pressure (systolic, diastolic, mean) for beat i

k = weighting factor [2-30], for venous and pulmonary pressure

$k = 8$, for arterial and intracranial pressure

$k = 2$ for all i greater than 0, else 1 (at start up)

The aforementioned description for the derivation of physiologic parameters (trends) is applicable to those parameters that are continuously acquired on a beat-to-beat basis such as heart-rate, ABP, and SpO₂. Other continuously acquired parameters such as respiration are derived using logic over longer time scales. The reader is referred to [50] for a more in-depth overview of the processing of respiration signals. Finally, aperiodic (intermittent) parameters such as cardiac output (thermodilution-derived) and pulmonary artery wedge pressure are included in the minute-to-minute physiologic trends (when measured) with several hours between successive valid measurements.



Figure 3-2: Examples of physiologic waves

3.3.4. ICU monitor alarms in MIMIC-II

The monitor-generated alarms are also archived with each MIMIC-II record. Table 3-3 includes a listing of all the alarm classes available in the MIMIC-II database. The alarms cover a broad range of physiologic signals such as ECG (arrhythmia alarms), blood pressure, oxygen saturation, and respiration. A severity is associated with each alarm, and is denoted by a red, yellow, or green label. In general, red alarms are considered “life-threatening” alarms that are indicative of rapid patient deterioration. For example, alarms indicative of asystole, ventricular fibrillation, and apnea are intended to alert the clinical staff to a patient that may need immediate life-saving interventions. Yellow alarms are triggered when a condition has been detected that may be classified as clinically

significant but not immediately life-threatening. For example, a high rate of premature ventricular contraction (PVC) detected from an ECG signal would trigger a yellow alarm but may not signify rapid patient deterioration. Green alarms are generally utilized to indicate monitoring system problems such as disconnected electrodes resulting in the loss of ECG monitoring. A time stamp indicating when an alarm condition has been triggered is also included.

Table 3-3: Summary of Bedside Monitor-Generated Alarms in MIMIC-II Database

Red ECG Alarms
Asystole
VFib/VTach
VTach
Extreme Brady
Extreme Tachy
Red Pressure Alarms
Pressure Disconnects
Red Respiration Alarms
Apnea
Vent Disconnect
Vent Failure
Red SpO2 Alarm
Extreme Desat
Yellow Alarms
Non Sustain VTach
Run PVCs
Pair PVCs
Pacer Not Capture
Pacer Not Pace
Missed Beat
High Pressure (Arterial, Pulmonary)
Low Pressure (Arterial, Pulmonary)
All Inops (Green)
ECG Inops
Pressure Inops
Respiration Inops
SpO2 Inops
Other Inops
Alarms Off {Suspend, Standby & All Alarms Off}

3.3.5. Specifications of clinical data

Along with monitor-generated physiologic data, there is a wealth of clinical data that can be used to characterize an ICU patient's physiologic state. The clinical data are available in electronic clinical information systems. To capture the relevant clinical information for each monitored patient, we leveraged the use of the Philips Information Support Mart (ISM) that was interfaced with the unit's clinical information system, CareVue (Philips Medical Systems). The ISM is an Oracle relational database that warehouses clinical information such as lab results, nurses' text notes, medications, fluid balance, and patient demographics. Customized scripts were written in SQL to query the ISM database and retrieve the clinical information for each monitored patient. The retrieved ISM records were uploaded into an open-source Postgres relational database. The data schema for the clinical information in MIMIC II was based on the schema reported in the Philips ISM User's Guide (Philips Medical Systems, Andover, MA). The most relevant tables that contained data salient to the present research are summarized briefly in Table 3-4.

Table 3-4: Description of clinical data within CareVue tables

Table Name	Clinical Data	Purpose
CensusEvents	Admit Timestamp, Discharge Timestamp	Defining dates and length of ICU stay. The information regarding discharge status (e.g. outcome, discharge destination) was incomplete.
ChartEvents	Clinical Laboratory Results (e.g. CBC, Blood Chemistry), Ventilator Settings, Nurse Validated Vital Signs	Clinical data used to assess body system function, document therapeutic parameters (ventilator settings, intra-aortic balloon pump, pacemakers, catheter insertion sites), and near-hourly charting of validated vital signs (e.g. heart rate, blood pressure, Oxygen saturation).
DPatients	Patient Names, Dates of Birth, Medical Record Number	Identifying patient information is used to match clinical data with monitor-generated physiologic data. Table DPatients is subsequently de-identified to preserve patient anonymity and comply with HIPAA standards.
MedEvents	Continuous (IV drip), Medications	MedEvents represents a time series profile of medication hourly rates (dosages) that are charted. Bolus medications are not included in MedEvents.
IOEvents	Hourly Input (IV, PO)	This table includes a comprehensive listing of the various fluids administered to a patient on an hourly basis are charted here (e.g. crystalloid solutions, blood products, total parenteral nutrition (TPN)). Also, output (e.g. urine, chest tubes, stools) is charted with a varying frequency (from one to several hours between successive measurements)
TotalBalEvents	Net hourly and 24 hour fluid balance calculation	Clinical information system generated summation and integration of net fluid balance on a 1- and 24-hour basis.

Utilizing the clinical information system as a high-throughput method of accessing a patient's clinical data did not encompass all the clinical data generated for each ICU patient. In order to facilitate a more complete characterization of a patient's ICU stay, clinical and demographic data that were unavailable in CareVue were obtained from several different hospital information systems. Table 3-5 includes a listing of the supplemental data that were gathered from other hospital information systems. The clinical and laboratory measurements of ICU patients admitted from the Emergency

Department (ED) were included in the MIMIC-II database when available. Clinical data that are not normally recorded in CareVue (such as urinalysis) was also obtained for each ICU patient. Patient demographics and length-of-stay were obtained so that retrospective outcomes studies and other related questions could be investigated using MIMIC-II. The discharge summary for each ICU patient was archived so that a physician's synopsis of a patient's stay could be included in MIMIC-II. The Computerized Provider Order Entry (CPOE) data for each MIMIC-II patient were obtained so that the actual and contemplated clinician's therapeutic interventions could be recorded for each patient. The CPOE data partially addresses limitations in the CareVue medication profile of each patient. Clinicians chose not to include IV bolus and oral medications in CareVue even if their orders were included in CPOE. However, it should be noted that CPOE only provides information about an intervention being ordered and does not confirm the actual administration of a therapeutic intervention. ICD-9 codes were obtained for each patient to serve as a standardized list of patient problems and interventions for each MIMIC-II patient. Lastly, radiology and ECG interpretations were also retrieved for patients when available. While these additional data streams do not form a complete electronic medical record, they substantially improve the richness of MIMIC-II so that a wide range of scientific studies can be supported in ICU patient monitoring and clinical decision support research.

Table 3-5: Description of supplementary clinical data tables

Clinical Data Class	Description
Emergency Department Data	Vital signs and clinical laboratory measurements measured from an ED patient prior to ICU admission.
ICD-9	Codes used for billing that describe the problems assigned to a patient
Length of Stay / Demographics	Length of stay in an ICU bed, and status {stable, deceased, to nursing home, to rehabilitation services} upon discharge
Clinical Laboratory Measurements	Laboratory measurements not archived in CareVue clinical information system (e.g. urine chemistries)
Microbiology	Results of cultures sent from patient specimens
12 Lead ECG Reports	Interpretation of 12-lead ECG
Echo Interpretations	Interpretation of Echo
Radiology	Radiology impression of diagnostic imaging (x-ray, Computerized Tomography, Ultrasound)
Discharge Summary	Physician summary of patient ICU course upon discharge
Computerized Provider Order Entry	Listing of provider orders (medications, procedures)

3.3.6. Merger of physiologic and clinical data

The first step in merging disparate monitor-generated physiologic data and clinical data from the hospital clinical information system includes matching data records corresponding to the same patient. The monitor-generated data included a unique identifier (referred to as a CASEID) with a patient name (first and last name) and medical record number (MRN). The name and MRN fields were manually entered by a nurse into the networked central station when a patient was admitted. Often (approximately 30% of cases), one or more identifier fields were not filled for an admitted patient. The CareVue clinical information system also included a unique patient identifier (referred to as a PID) for each ICU stay of a patient. The CIS also includes identifying information such as a patient's name and MRN which is automatically input through the hospital-wide information system when a patient is admitted to a unit.

If the patient's identifying information (name, and MRN) was available with a physiologic data record (indexed by a CASEID), the respective physiologic data record was matched to the corresponding clinical information record from CareVue. There were two stages to the merger process. The first stage included matching names and medical record numbers (when available) from the monitor-generated data records and the clinical data records from CareVue. The second stage included comparing the similarity of the physiologic trends from the higher resolution monitoring data (approximately 1 sample per minute) with the nurse-validated vital sign trends in the clinical information system sampled on an hourly basis. Briefly, determination of trend similarity included assessing the overlap (intersection) in the available trends that were present in the physiologic and clinical data. For example, if the nurse-validated clinical data included hourly measurements of central venous pressure and pulmonary artery pressure, then the monitor-generated physiologic trends of a correctly matched record should also contain non-zero values for these signals during the same time interval. If the set of vital sign trends from physiologic and clinical records are disjoint, it is sufficient to detect a mismatched pair of records. However, a matched set of trends is not sufficient to guarantee a correct pairing of physiologic and clinical data records. Thus, the physiologic

trend values were sampled at the same times that nurse-validated clinical data were available for the respective trend, and the actual time-varying trends from both sources were correlated with one another. A high level of correlation (greater than 0.8) was deemed sufficient to establish that a clinical data record was matched to the correct physiologic data record.

3.3.7. Annotation of merged ICU patient records

A major goal in developing the MIMIC-II database as a research resource is to include physician annotations of the major physiologic events occurring over a patient's ICU stay. An "annotation station" was developed to allow physicians to view and annotate a rich class of physiologic and clinical data. The annotation process is not a major focus of this thesis, and the reader is referred to [1] for a more extensive review of both the annotation station as well as the overall annotation process.

3.4. Quantitative and qualitative characterization of the MIMIC-II database

In the preceding section, the process and technology used to develop the MIMIC-II database was described. In this section, several qualities of the MIMIC-II database are statistically and clinically analyzed. In particular, the overall patient population in the MIMIC-II database is characterized along several demographic, diagnostic, and therapeutic criteria.

Table 3-6 includes an analysis of the patient demographics and outcomes of the MIMIC-II patient population. As the MIMIC-II database is undergoing constant development and growth, a patient data set of 2094 records were chosen for analysis and to be representative of the overall database. The average length of stay (and hence, length of a patient record) is 7.2 days. However, there is a wide variation due to outliers with prolonged stays even though approximately half the patients have stays of less than or equal to 3 days.

The overall mortality rate was 14.1% within the MIMIC-II patient population. The mortality rate can also be analyzed by stratifying across different ICU units as well as patient acuities. As shown in Figure 3-4, the mortality rate is significantly different from unit to unit. The highest mortality rates (approximately 20%) are associated with SICU and MICU patients, and the lowest mortality rate (~7%) occurred in the CSRU. The difference in mortality rates between the different units reflects a difference in patient demographics between these units as well as common medical problems and acuities associated with the respective units. For example, the age of patients in different units is significantly different. Figure 3-4 also illustrates the difference in mortality rates stratified across different age groups.

Table 3-6: Summary of MIMIC-II database demographics and outcomes

Number of patients	2094 (with complete physiologic and clinical data) ~14900 (with clinical data only and no physiologic waveforms)
Length of Stay (days)	
Mean	7.2
Median	3 days
Mortality	14.1%
Age (years)	
Mean +/- Standard Deviation	65.4 +/- 16.7
Median	67.7
Patient Unit Distributions	
CCU	560
SICU	125
CSRU	643
MICU	760
Gender Distribution	
Female	886 (42.4%)
Male	1206 (57.7%)
Acuity Characterization (SAPS-I Score)	
(Mean +/- St. Dev)	13.5 +/- 5.6

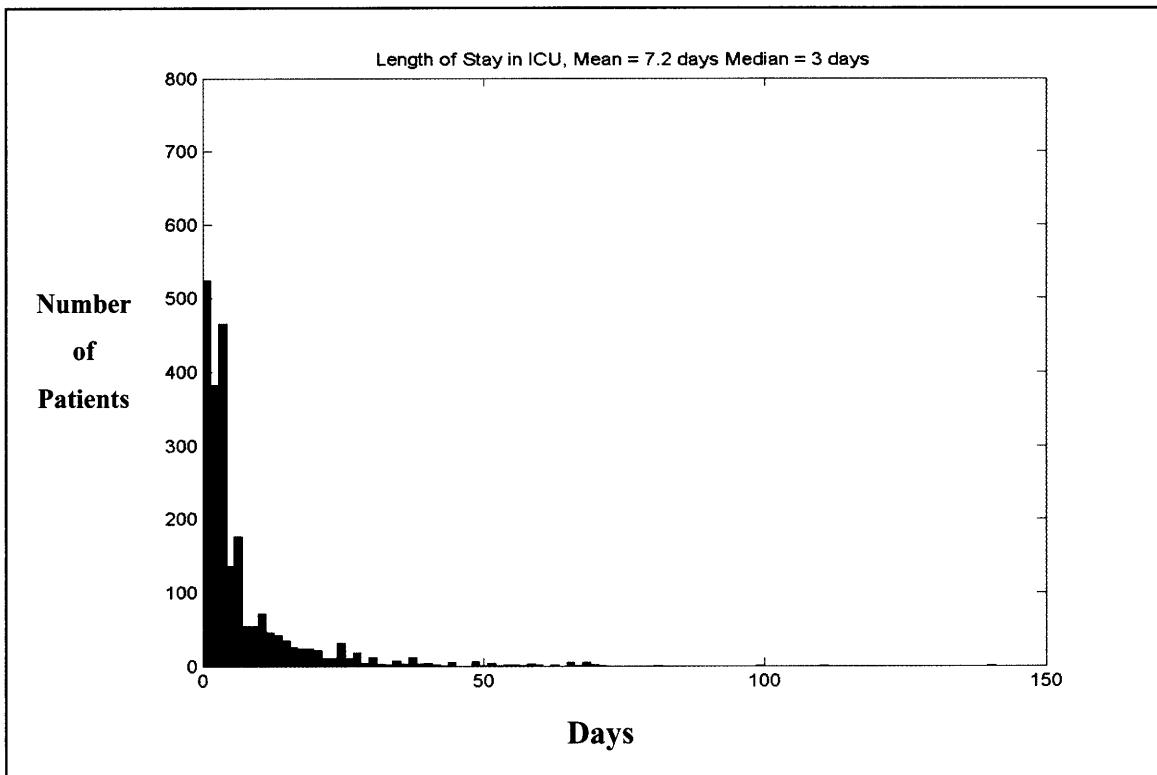


Figure 3-3: Length of stay distributions for MIMIC-II ICU patients

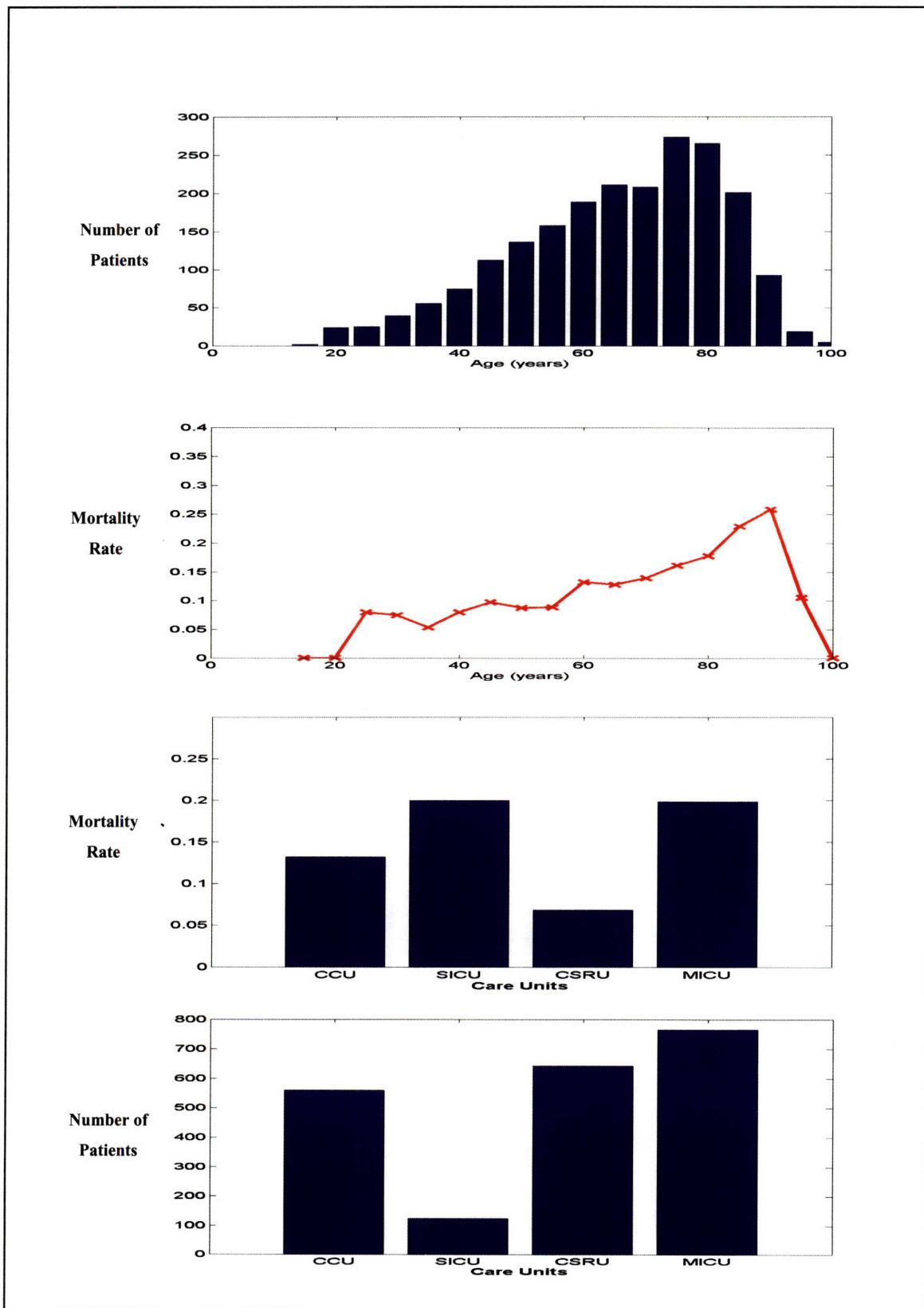


Figure 3-4: Stratified ICU patient demographics and mortality rates

Another method of characterizing the MIMIC-II patient population includes an analysis of patient acuities or severity of illness. Acuities can be based on diagnostic as well as therapeutic criteria. For example, the Simplified Acute Physiology Score (SAPS) was introduced as a metric to predict ICU patient outcomes (mortality) and assess the critical care unit performance as a function of the severity of illness of patients. There have been several refinements to the original SAPS algorithm [41]. The SAPS-I formula was chosen for its simplicity and requiring only readily available clinical laboratory measurements and vital signs. The mortality rate in MIMIC-II patients is trended as a function of SAPS-I score in the cumulative distribution curve of Figure 3-5. As the SAPS-I score increases, the mortality rate significantly increases. The overall trend is consistent with previous studies [41]. The SAPS-I distribution and formula also provides a means for researchers to study patient groups from MIMIC-II with similar overall acuities.

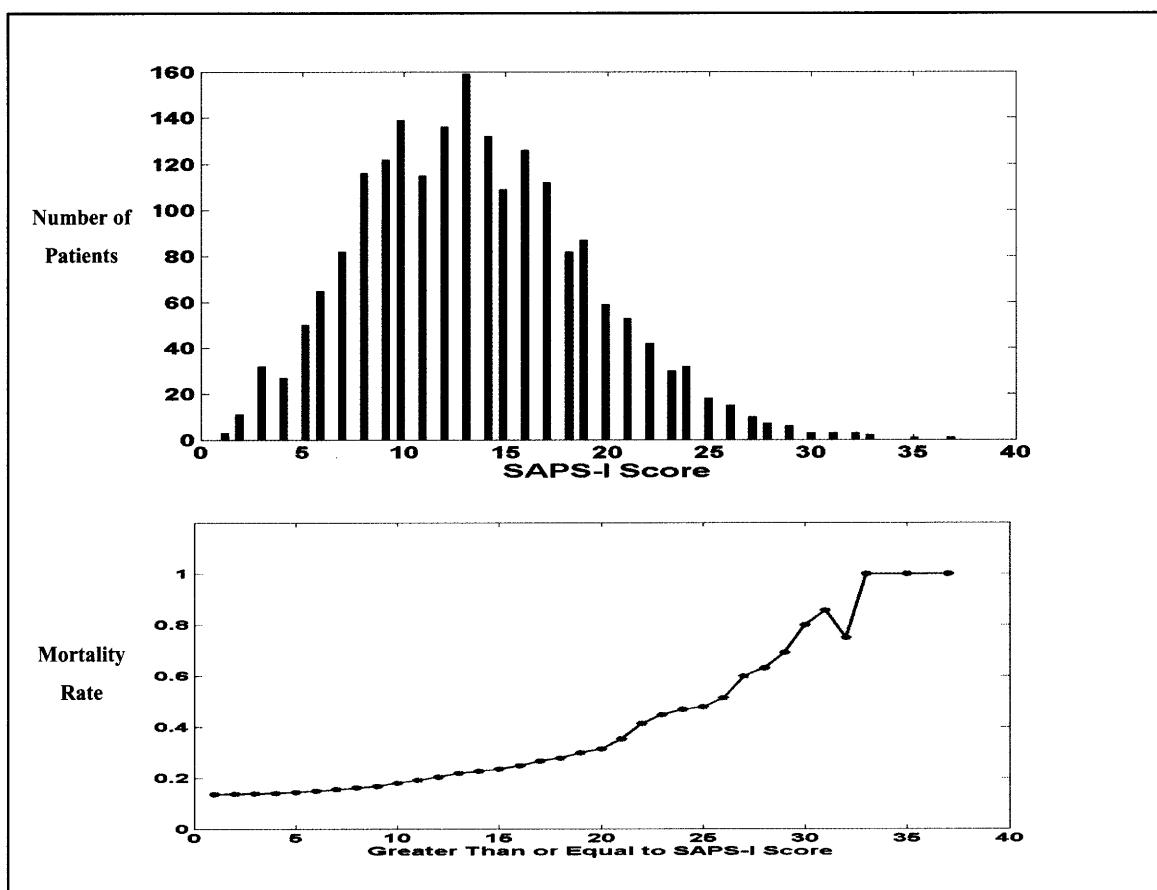


Figure 3-5: SAPS-I distribution & mortality rates

The ICD9 codes provided with each patient record can be utilized as a high-throughput method of identifying patients with specific problem lists that may interest a researcher. The ICD9 codes reflect what the clinical staff at the time identified as the major problems that were taken into consideration when treating the patient. Figure 3-6 is a graphical representation of the distribution of major problems within the MIMIC-II patient population. A patient generally has a number of problems as classified by the ICD9 codes. In subsequent chapters, the ICD9 codes will be utilized for identifying patients with similar hemodynamic instability patterns.

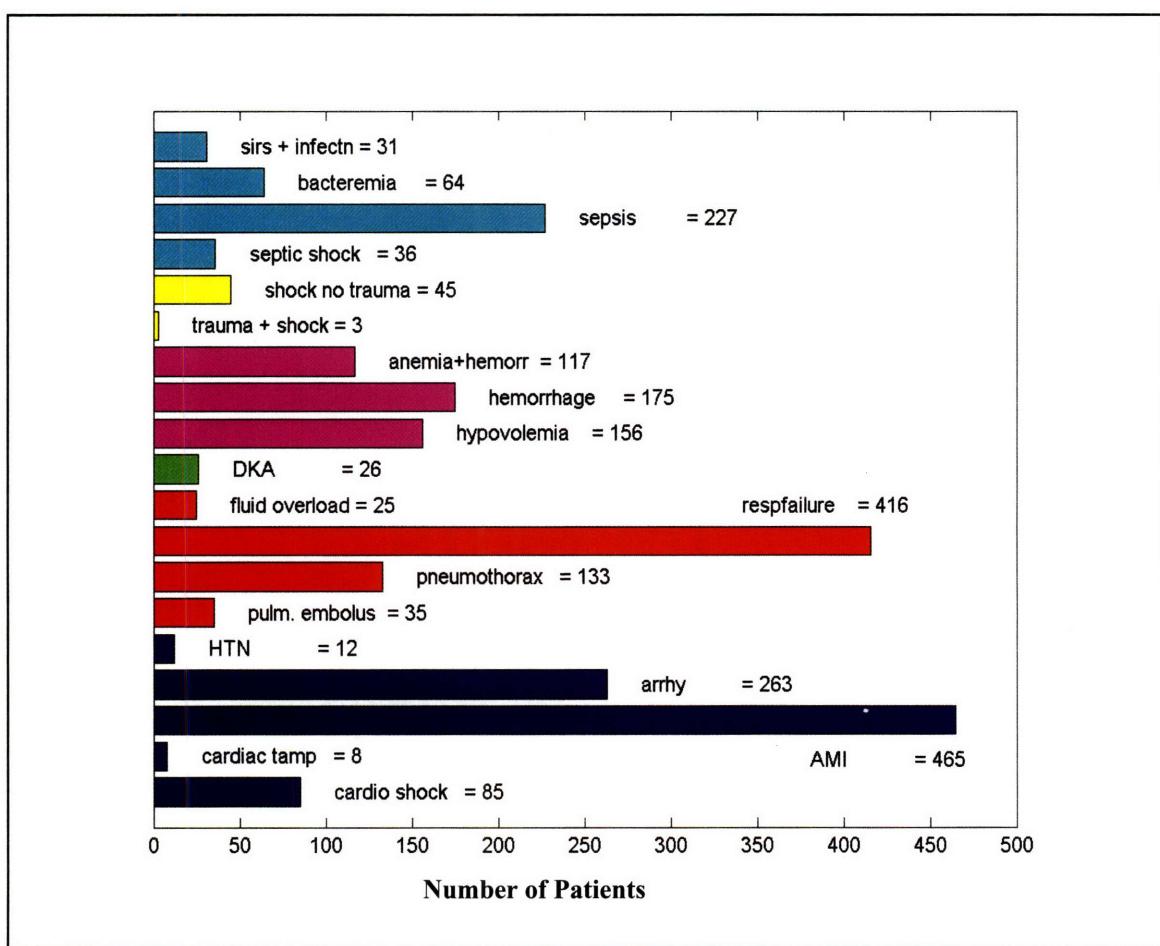


Figure 3-6: Distribution of ICD9 problems in MIMIC-II patients

Patients in the MIMIC-II database were also characterized based upon several hemodynamically significant interventions. The class of intervention may be indicative of the underlying pathophysiology. For example, an intra-aortic balloon pump (IABP) may be indicated for a patient with significant cardiogenic shock. Thus, selecting patients with IABP interventions for further analysis is likely to yield a patient set rich in cardiogenic shock events. As Table 3-7 indicates, there are different mortality rates associated with patients receiving different classes of interventions. The highest mortality rate (19.4%) was associated with patients on mechanical ventilation while the lowest mortality rate (5.5%) was associated with patients requiring pacing. The analysis does not establish a causal relationship between an intervention and a mortality rate. However, the patients that are sicker and hence, more likely to expire, may require more significant interventions to maintain homeostasis.

Table 3-7: Major cardiovasucular interventions in the MIMIC-II database

Major Cardiovascular Interventions	Number of Patients	Mortality Rate
Ventilator	1186	19.4%
Vaso-active Medications {Neosynephrine, Levophed, Dopamine, Dobutamine, Epinephrine}	1069	18.7%
Blood Products {whole blood, packed red blood cells, fresh frozen plasma, cryoprecipitates}	1015	17.0%
Intra-aortic Balloon Pump	129	16.4%
Pacemaker	475	05.5%

Another method that can be used to characterize an ICU patient includes analyzing the physiologic signals that were recorded from the patient. The availability of a certain physiologic signal may be indicative of the underlying problem that the ICU staff was attempting to monitor. As Table 3-8, demonstrates, certain physiologic signals such as ECG and Non-invasive blood pressure (NIBP) are measured on almost every patient while other signals are not as common (such as Pulmonary Artery Pressure). Thus, the availability of a pulmonary artery pressure signal may indicate that a patient may be hemodynamically unstable in comparison to a patient with only a non-invasive cuff-based blood pressure.

Table 3-8: Availability of major physiologic signals in MIMIC-II patient records

Measured Physiologic Signals	% of Total Patients (N=2094)
ECG	98%
Respiration	96%
Oxygen Saturation	97%
Intra-arterial Blood Pressure	60%
Non-invasive blood Pressure	92%
Central Venous Pressure	38%
Pulmonary Artery Pressure	29%
PAWP	5%
Cardiac Output	13%
Intra-cranial pressure	1.4%

3.4.1. Comparison of Clinical and Physiologic Data

The characterization of the MIMIC-II database is based upon analysis of both clinical and physiologic data. As defined previously, the physiologic data refers to bedside-monitor generated signals such as ECG and blood pressure waveforms as well as numerics such

as heart rate and oxygen saturation (SpO_2). Physiologic data are typically sampled at high resolution rates of 125 Hz for waveforms and 1 sample/minute for numerics. The clinical data refers to data extracted from a clinical information system that includes diverse laboratory measurements, nurse-validated vital signs, and interventions that are typically recorded on an hourly time scale. The bedside-monitor generated vital signs are subject to noise corruption. Thus, prior to including vital signs into a patient's hourly chart in a clinical information system, a nurse is charged with validating (or filtering) the monitor-generated data first. The nurse-validation step is intended to increase the fidelity of the vital signs from a clinical information system. However, the results as shown in Table 3-9demonstrate that there is a statistically insignificant difference in the overall vital sign numerics when the nurse-validated and monitor-generated data are compared. The results suggest that the higher-resolution monitor-generated physiologic data can be exploited for identifying salient hemodynamic events with automated algorithms. In subsequent chapters of this thesis, such algorithms will be described in greater detail and their performance assessed with the MIMIC-II database.

Table 3-9: Distribution and comparison of physiologic parameters from monitors and nurse-validated charting data.

	Nurse-Validated	Bedside Parameters
Heart Rate	84.3 +/- 13.5	82.3 +/- 14.2
Intra-arterial Blood Pressure		
Systolic	121.4 +/- 18.3	118.7 +/- 19.2
Diastolic	59.8 +/- 9.8	58.1 +/- 10.4
Mean	80.8 +/- 11.6	78.1 +/- 12.3
Non-invasive Blood Pressure		
Systolic	Not Specified	115.5 +/- 19.0
Diastolic		53.0 +/- 11.6
Mean		74.0 +/- 12.1
CVP	11.2 +/- 3.9	10.5 +/- 4.7
Mean PAP	30.1 +/- 6.6	25.5 +/- 7.6
PAWP	17.0 +/- 5.8	17.0 +/- 5.8
CO	5.4 +/- 2.1	5.4 +/- 2.1
Respiratory Rate	19.3 +/- 3.5	18.6 +/- 4.0
SpO₂	97.0 +/- 2.4	97.2 +/- 2.1

4.Clinical Studies with MIMIC-II

The MIMIC-II database can potentially support a broad array of clinical studies that may provide additional insight into ICU patient care, biomedical sensor performance, and validate novel physiologic signal processing and pattern recognition algorithms. In particular, the relatively high-resolution physiologic and clinical data that is available allows for the posing of novel questions in the field of ICU patient monitoring. The large size and high-resolution characteristics of the database also pose significant challenges to undertaking valid clinical studies. The inherent noise that often corrupts the invalidated data is of particular concern when one is attempting to test clinical hypotheses with the MIMIC-II database.

There are several possible studies for evaluating biomedical sensor modalities that are susceptible to noise in the ICU. As an example of one such study, the performance of automated noninvasive blood pressure modules was evaluated using the MIMIC-II database. An accurate method of monitoring of blood pressure is of fundamental importance in ICU patient care. The deterioration of blood pressure in an ICU patient requires a vigilant therapeutic response from the clinical staff to prevent end-organ damage.

The large scale of the MIMIC-II database and the availability of high-resolution physiologic and therapeutic data permit novel insight into the temporal vigilance of clinicians in response to physiologic patterns suggestive of hemodynamic deterioration. Patient records from the MIMIC-II database were utilized to study the diurnal variation in the incidence of hypotensive episodes (concomitant decreases in mean arterial pressure

below 60 mmHg and increases in heart rate) and the corresponding diurnal variation in the initiation of therapeutic interventions.

This chapter is organized in the following manner. The next section (4.1) describes a retrospective analysis of the degree of agreement between cuff-based non-invasive blood pressure measurements and invasive (intra-arterial) blood pressure measurements in ICU patients. Then section 4.2 includes a studies the diurnal pattern of hemodynamic deterioration as well as therapeutic responses of ICU clinicians.

4.1. Analysis of agreement between noninvasive and invasive blood pressure measurements in ICU patients.

ICU patients are in general physiologically fragile and require constant physiologic monitoring and support. An intra-arterial catheter is commonly accepted as the most accurate method of measuring arterial blood pressure [4]. However, noninvasive technologies that can provide accurate measurements of arterial blood pressure may be preferable for several reasons: noninvasive technologies have reduced risk of nosocomial infections, invasive catheters can be a cause of emboli, invasive monitoring may be relatively more restrictive of a patient's ambulation, and noninvasive monitoring may be psychologically less stressful to the patient and his or her family.

Arterial blood pressure is among the most routinely monitored signals in an ICU patient. As reported in the previous chapter, approximately 60% of MIMIC-II patients are monitored for a portion of their overall ICU stay with an invasive arterial catheter. Noninvasive methods of automatically measuring arterial blood pressure using an inflatable cuff and the oscillometric method have been previously described in the literature [5]. Several vendors have introduced automated noninvasive blood pressure (NIBP) modules for in-hospital patient monitoring. Several studies have evaluated NIBP performance in comparison to intra-arterial blood pressure measurements [21]. However, most such studies have only included a limited number of patients and did not monitor

the patients over extended ICU stays of several days or more. As the number of subjects and duration of monitoring increases, it may be possible to study the performance of NIBP with respect to different physiologic states such as hypotension or shock. Previous prospective studies were carefully designed to control for optimal NIBP setup parameters such as choosing the appropriate cuff sizes for a given arm circumference, constant calibrations, and arm positioning. In reality, the clinical staff in a busy ICU environment will often overlook these set-up parameters and procedures. As researchers attempt to utilize NIBP measurements for real-time patient monitoring algorithms, it is essential to understand the measurement accuracy of NIBP technologies in realistic ICU environments. ICU clinicians need to know whether NIBP measurements can be trusted in routine ICU settings.

The aim of the present study was to compare the blood pressure measurements obtained from automated NIBP with respect to intra-arterial blood pressure in a large ICU population over extended monitoring intervals.

4.1.1. Review of oscillometric blood pressure measurement

The oscillometric method for measuring blood pressure is the most common automated and non-invasive method used to measure arterial blood pressure in ICU patients [4]. An oscillometric blood pressure device determines blood pressure by detecting a sequence of oscillations in cuff pressure [18]. As demonstrated in Figure 4-1, the cuff around the upper arm is inflated so that the cuff pressure exceeds the systolic blood pressure. The high trans-luminal pressure occludes the underlying brachial artery so that the pressure downstream of the point of occlusion drops to 0 mmHg. The cuff pressure is then slowly deflated automatically so that the cuff pressure approaches the systolic blood pressure. The pulsatile arterial flow causes oscillations superimposed on the cuff pressure. As the cuff pressure continues to drop, the magnitude of the oscillations increases as the cuff pressure approaches the intra-arterial mean blood pressure. Then, the oscillation amplitudes decrease as the cuff pressure approaches the intra-arterial diastolic blood pressure.

Results from empirical [18] and modeling studies [77] have supported the general practice of choosing the cuff pressure during maximal oscillations as an accurate estimate of the intra-arterial mean blood pressure. However, the ad-hoc algorithms that different vendors have utilized for determining systolic and diastolic blood pressure have rarely been disclosed [18]. The accuracy of these algorithms has been reported in several studies when compared to invasive intra-arterial blood pressure (ABP), which is considered the most accurate (“gold standard”) method of measuring arterial blood pressure. Many of these studies have shown clinically significant discrepancies between the non-invasive oscillometric blood pressure and ABP [46]. In general, the NIBP systolic and diastolic blood pressures were reported to be the least accurate measurements, whereas the NIBP mean blood pressure was deemed to be the most reliable measurement when using the oscillometric method. In the next section, the description of the retrospective analysis of agreement between NIBP and ABP within the MIMIC-II database will be provided.

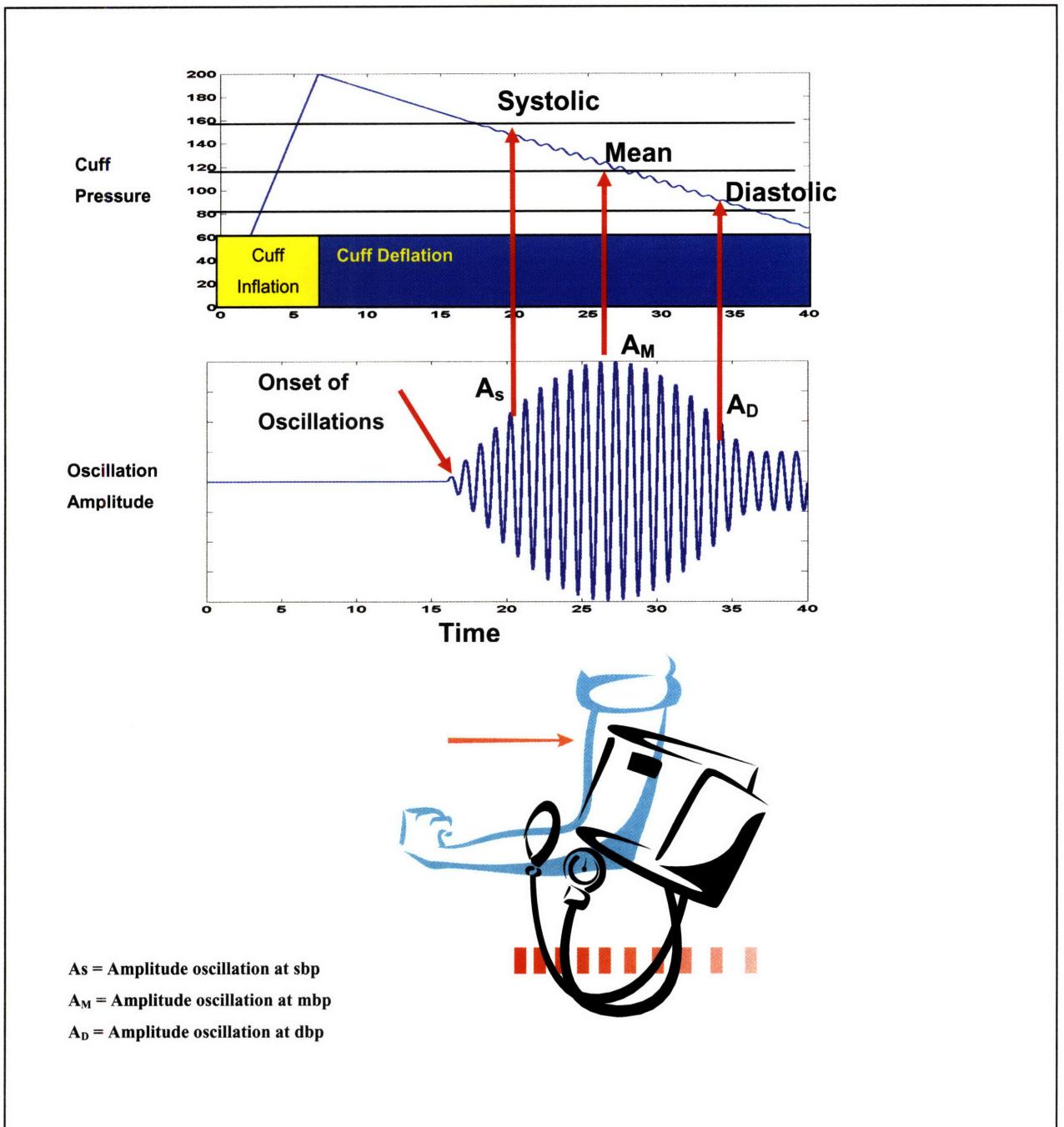


Figure 4-1: Oscillometric method of blood pressure measurement

4.1.2. Methodology: Patient ABP and NIBP data and noise processing

The available 1-minute average intra-arterial and NIBP blood pressure trends were analyzed from all the records in the MIMIC-II database. The M1008A NBP module (Philips Medical Systems, Andover, MA) is the standard NIBP module utilized in the ICUs that were the source of data for MIMIC-II. From among those records with simultaneous NIBP and intra-arterial blood pressure measurements, a subset of patient records were included in the present study that met the following criteria: 1) at least 20 or more individual NIBP measurements spanning a minimum of 8 hours were available with simultaneous valid intra-arterial blood pressure measurements, 2) the difference between the mean intra-arterial blood pressure and mean NIBP was less than 50 mmHg (an error greater than 50 mmHg is likely due to a systematic error that does not reflect the actual measurement technologies).

The methodology of this study includes an automated algorithm for removing measurement pairs in which the intra-arterial blood pressure was deemed to be noisy. Dampening of the arterial line can be a frequent source of error in invasive blood pressure measurements [46]. Typically, the observed arterial waveform morphology changes over time such that the pulse pressure tends to decay towards 0 mmHg without a significant change in the mean blood pressure. Thus, a comparison of a simultaneous non-invasive blood pressure measurement would typically result in the systolic NIBP being significantly greater than the observed intra-arterial systolic blood pressure. The difference between the NIBP mean blood pressure and intra-arterial mean blood pressure would however be significantly less. Thus, in order to remove measurement pairs in which the ABP measurement is unreliable, a simple set of rules were developed. The schematic in Figure 4-2 details the rules used in identifying noisy ABP measurements. Essential to the rule process is the utilization of an “estimated” mean blood pressure (EMBP) based upon the well-known formula (see Equation 2) for estimating mean blood pressure from the systolic and diastolic blood pressure [59].

Equation 2

$$EMBP = \frac{(2 * DBP + SBP)}{3}$$

If a significant difference between the measured mean arterial blood pressure (monitor-derived) and the estimated mean arterial blood pressure is found, then the pressure measurements (diastolic, mean, and systolic) are rejected.

Based upon the aforementioned exclusion criteria, 302 patient records were included in a comparative analysis of intra-arterial (invasive) blood pressure (ABP) and non-invasive blood pressure (NIBP) in ICU patients.

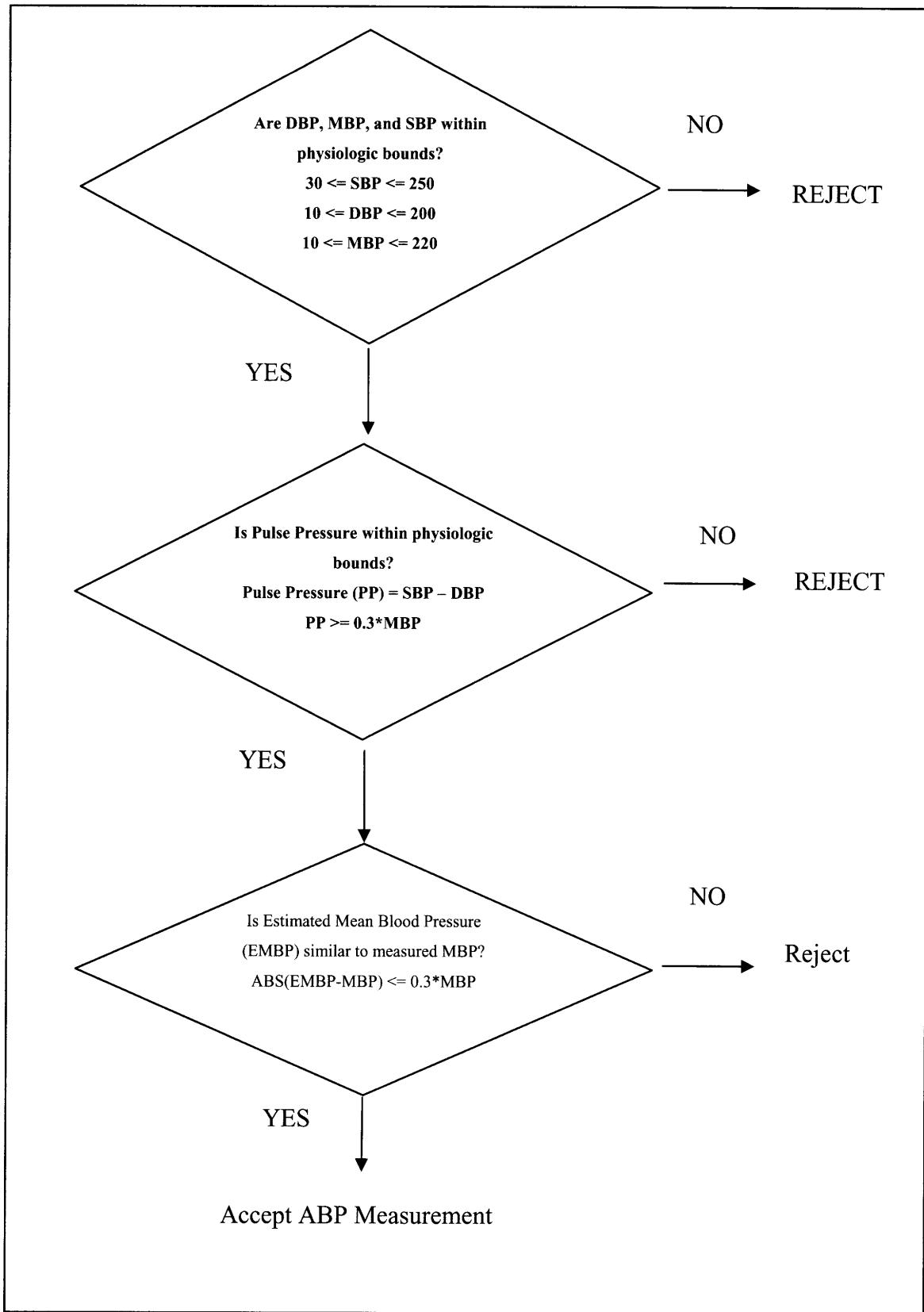


Figure 4-2: ABP noise rejection algorithm

4.1.3. Methodology: Physiologic factors analysis

Several investigators have reported that discrepancies between NIBP and ABP measurements may be correlated to several physiologic factors [4]. For example, the appropriate cuff size for a particular patient is generally a function of a patient's weight. Inappropriate cuff sizes will lead to inaccurate NIBP measurements. The MIMIC-II database affords one the opportunity to analyze the error between ABP and NIBP as a function of the patient weight. Arterial stiffness is another contributing factor to NIBP inaccuracies that has been mentioned in the literature [77]. Arterial stiffness can be affected by pathological processes as well as age-related stresses and various vaso-active medications. The error between NIBP and ABP was analyzed as a function of such factors as age as well as the presence or absence of vaso-active medications when NIBP-ABP paired measurements were available.

4.1.4. Methodology: Comparison between NIBP and ABP to detect hemodynamic instability in ICU patients

A primary reason for monitoring a patient's blood pressure is to assess end-organ perfusion. Ideally, clinicians should respond to deterioration in blood pressure by initiating appropriate therapy to preclude end-organ damage. Vaso-active medications ("pressors") are one of the most commonly used interventions in the ICU. Vaso-active medications' pharmacological effects include the increase of total peripheral resistance as a means of increasing blood pressure. Levophed and neosynephrine were chosen as examples of the most potent pressors that are indicative of hemodynamic instability. The MIMIC-II database was utilized to assess the possibility of developing a detector of hemodynamic instability using NIBP or ABP. In particular, the pressure measurements (NIBP and ABP) during the period of time preceding the onset of a vaso-active medication were utilized to develop a simple threshold-based detector. An example of NIBP and ABP trends prior to the onset of a vaso-active medication is provided Figure 4-3. The period of time for the inclusion of measurements spanned from 5 minutes to 90 minutes prior to the onset of therapy (vaso-active medication). This interval is utilized

because of the assumption that a deterioration of systemic blood pressure (acute hypotension) was noticed by the clinical staff and prompted the initiation of therapy.

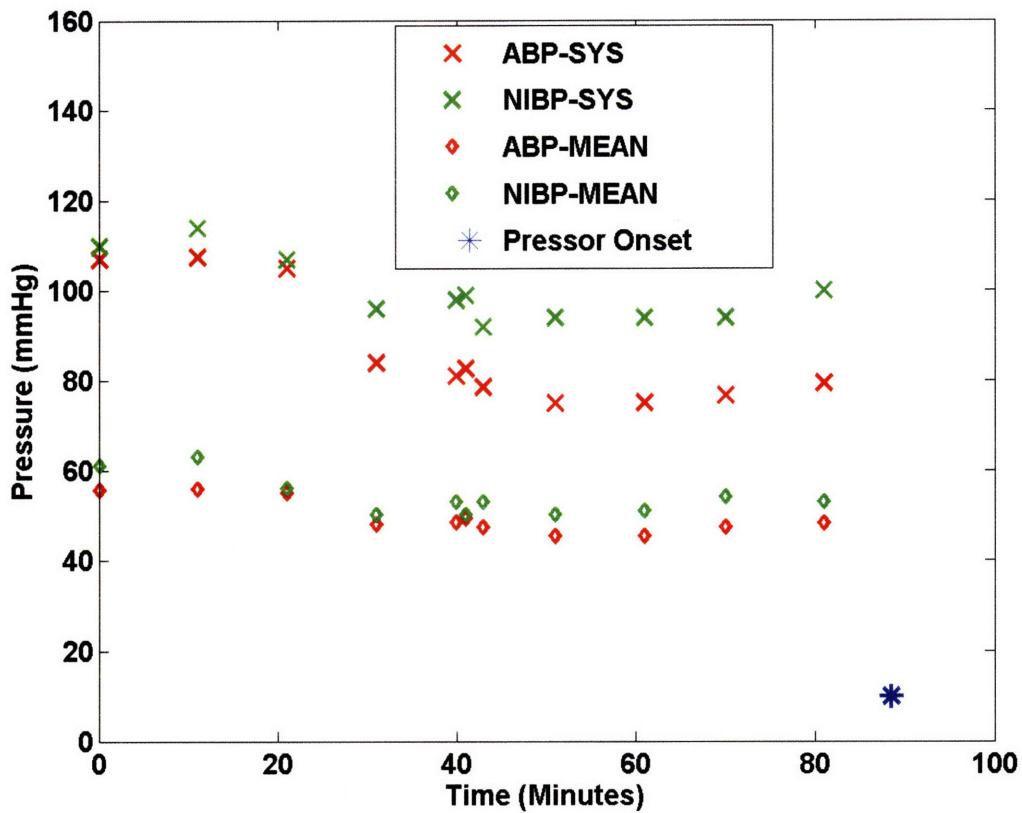


Figure 4-3: MIMIC CaseID 4167 ABP & NIBP trend prior to pressor onset

The utilization of vaso-active medications can be considered a therapeutic criterion for a fiducial point indicating hemodynamic instability. However, it may also be possible to leverage other diagnostic data (apart from blood pressure) as being indicative of hemodynamic instability. When the cardiovascular system fails to perfuse the kidneys adequately due to a drop in systemic blood pressure, a form of pre-renal failure may develop. The hourly urine output may drop to levels below 20 ml/hr. The fluid balance information in the MIMIC-II database was utilized to identify intervals of transient and sustained oliguria. The NIBP and ABP measurements during an interval prior to the onset of oliguria were analyzed. Oliguria is in general a more slowly evolving process that may

take several hours to notice. In comparison, acute hypotension may develop over a few minutes. Thus, the interval of time included up to 12 hours prior to the onset of oliguria.

Along with oliguria, other diagnostic variables that are utilized to assess renal perfusion and the glomerular filtration rate (GFR) include serum creatinine and blood urea nitrogen (BUN). A significant increase in the values of these variables from their baselines is often indicative of acute renal failure. The creatinine and BUN values were processed to identify significant transient increases between consecutive measurements occurring within an interval of 36 hours. The NIBP and ABP measurements during an interval prior to the onset of oliguria were analyzed. The time window of pressure measurements included all measurements within 18 hours prior to the detected increase in BUN or creatinine.

4.1.5. Results: Relative error analysis

Data are expressed as mean with standard deviation where applicable. Categorical (partitioned) data were expressed with percentages. The difference between measurement methods was assessed with Bland-Altman analysis [3] and the paired Student's t-test.

Tracking relative changes in blood pressure over time may be as important as the overall absolute value of blood pressure in determining the physiologic state of a patient. Thus, the pair-wise NIBP and ABP measurements were analyzed for agreement with respect to changes in pressure. For each systolic, mean, and diastolic pressure time series of a MIMIC-II patient, the overall average value of the time series was subtracted from each individual value, and the result was then divided by the respective average. The resultant time series reflected the percent change in an individual measurement from the overall mean of the time series. The error between the paired values of percent changes from NIBP and ABP is then analyzed by pooling the time series from all patients. The normalization procedure that is described will result in the removal of bias between the

two measurement techniques. Thus, the relative agreement between the methods can be summarized by the standard deviation of the error and is included in Table 4-1.

4.1.6. Results: NIBP-ABP overall analysis of agreement

Table 4-1: Absolute and relative error analysis

<i>25,345 pair-wise measurements analyzed from 302 patients</i>	NIBP	ABP	Absolute Mean Error (NIBP-ABP)	Absolute Standard Deviation of Error (NIBP-ABP)	Correlation Coefficient to Average of NIBP and ABP	Relative Pair-wise Standard Deviation of Error (NIBP-ABP)
Systolic	114.8+/- 23.8	115.3+/- 30.0	-.5	17.5	-.37	11.1%
Mean	77.5+/- 18.7	72.8+/- 15.5	-4.7	10.7	-.32	12.2%
Diastolic	51.5+/- 14.9	58.3+/- 14.7	-6.8	10.5	.02	17.2%

The overall agreement between NIBP and ABP is summarized in Table 4-1. The systolic blood pressure had the lowest (absolute) overall bias (-0.5 mmHg) between NIBP and ABP. However, the standard deviation of error for systolic blood pressure was the highest (17.5 mmHg). The diastolic blood pressure had the highest (absolute) bias (-6.8 mmHg) and a standard deviation of error of 10.5 mmHg. The bias and precision error for diastolic blood pressure are the higher than either systolic or diastolic after normalizing the error by the population mean diastolic pressure (51.5 mmHg). The precision error between NIBP and ABP was lowest using mean blood pressure after the same normalization. The bias error (-4.7 mmHg) between NIBP mean and ABP mean is consistent with previous findings [4].

As Figure 4-4 and Figure 4-5 demonstrate, the error between NIBP and ABP for systolic as well as mean had a statistically significant correlation to the underlying pressure ($p < .005$). In general, at lower pressures (systolic blood pressure less than 90 mmHg) the NIBP systolic pressure tends to over-estimate the ABP systolic pressure. At higher pressures (systolic blood pressure greater than 150 mmHg) the NIBP systolic blood pressure under-estimated the ABP systolic pressure. The NIBP mean pressure had a significant bias error under both normotensive and hypertensive pressures (mean pressure greater than 70 mmHg). The error in diastolic blood pressure did not exhibit a significant correlation to the underlying pressure.

Another measure of agreement between NIBP and ABP can include the assessment of relative changes with respect to the mean of a trend. Thus, by replacing each measurement by its percent change from the mean value of the respective trend, and dividing by the mean of the trend, a percent change from the mean can be calculated. The difference in the percent change from the mean of NIBP and ABP can then be compared. Table 4-1 includes a summary of the error between NIBP and ABP relative changes for systolic, mean, and diastolic parameters. The diastolic NIBP and ABP had the poorest level of agreement (17.2% error). The relative errors between NIBP and ABP for systolic and mean pressure were 11.1% and 12.2%, respectively.

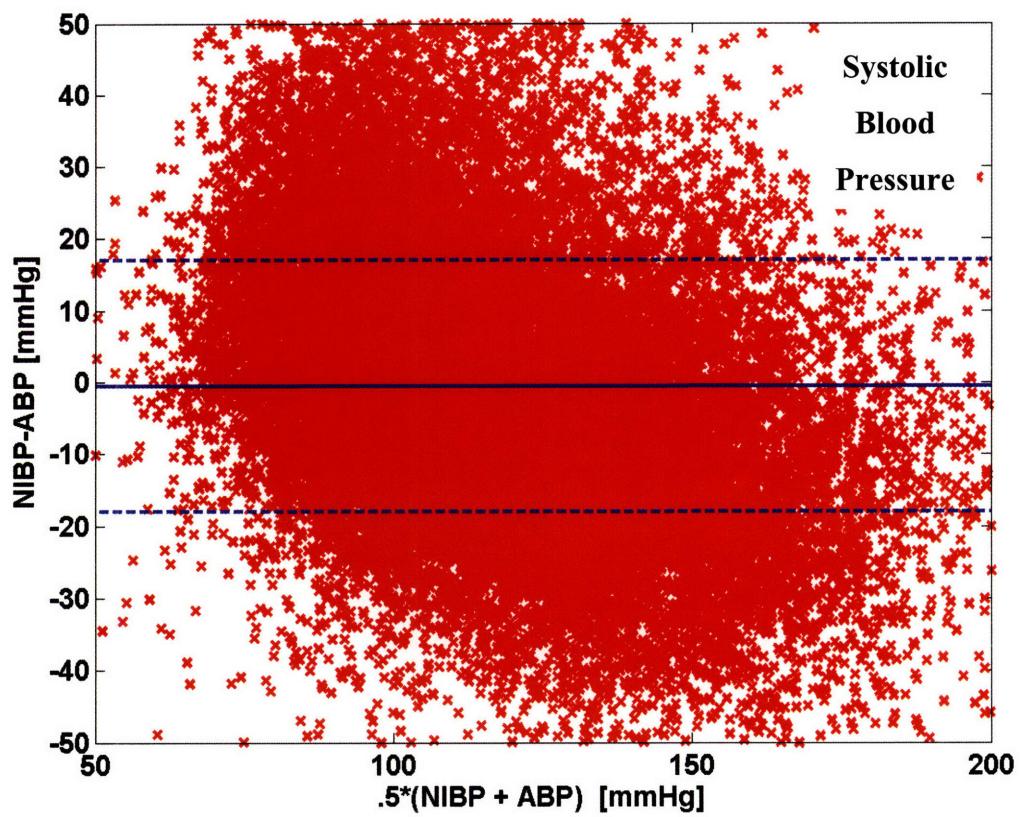


Figure 4-4: Bland-Altman analysis of (systolic) NIBP-ABP agreement

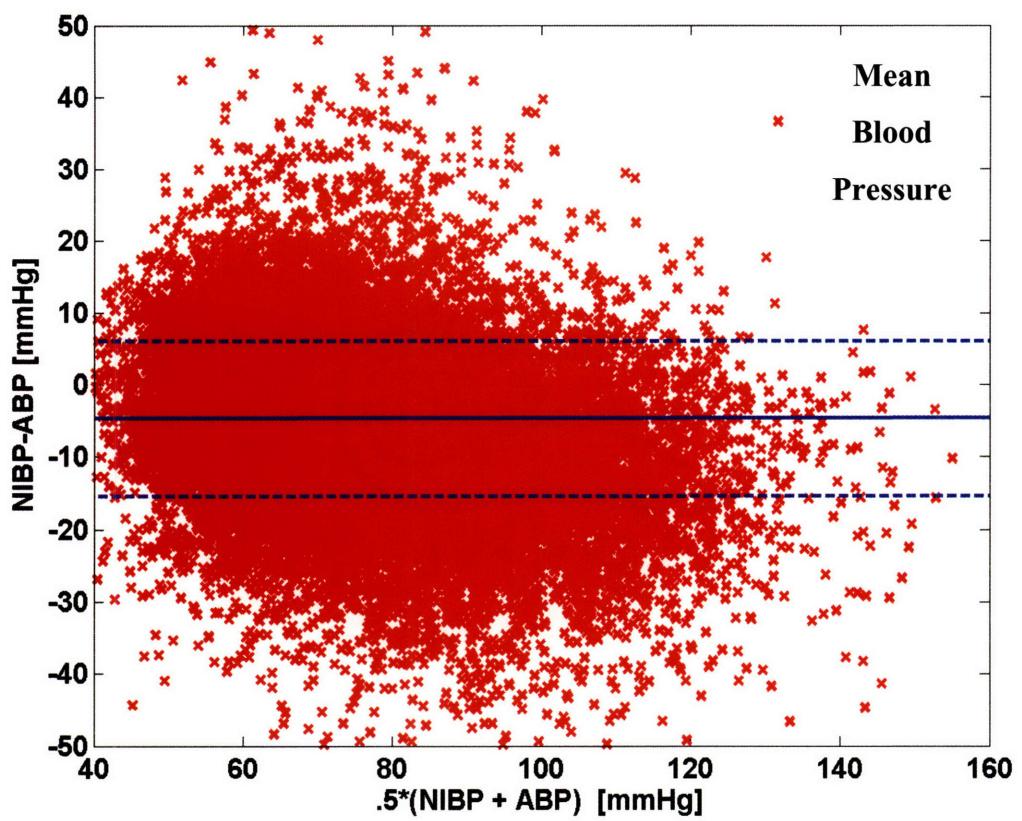


Figure 4-5: Bland-Altman analysis of (mean) NIBP-ABP agreement

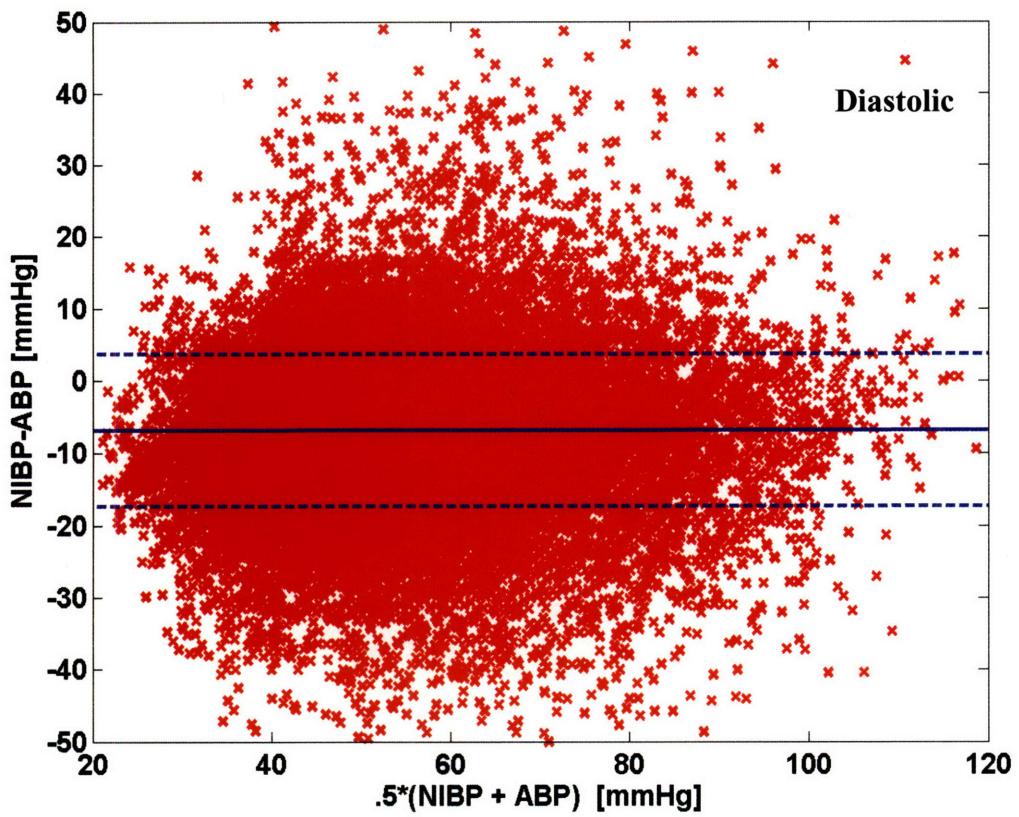


Figure 4-6: Bland-Altman analysis of (diastolic) NIBP-ABP agreement

4.1.7. Results: Analysis of factors influencing NIBP-ABP agreement

The error (NIBP-ABP) between paired pressure measurements was correlated to various factors and the results summarized in Table 4-2. Among the factors that were analyzed, the error between NIBP and ABP measurements was found to be most correlated to the measured mean arterial blood pressure. The error correlated poorly with other factors such as heart rate, age, and patient weight.

Table 4-2: Analysis of factors influencing agreement between NIBP and ABP

<i>Correlation Coefficient between NIBP-ABP error and various physiologic factors</i>	Mean Blood Pressure	Heart Rate	Age	Weight	Presence of Vaso-active Medication
Systolic	-0.46	.16	-0.09	0.02	0.08
Mean	-0.56	.10	-0.04	0.00	0.09
Diastolic	-0.28	.02	0.03	-0.06	0.03

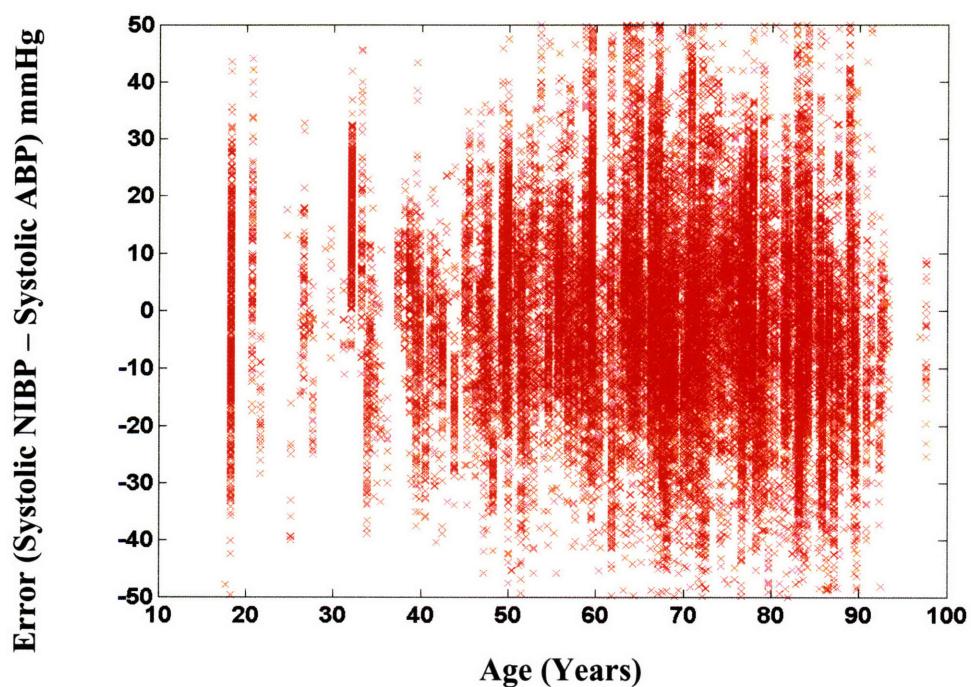


Figure 4-7: NIBP-ABP (systolic) Error vs Age

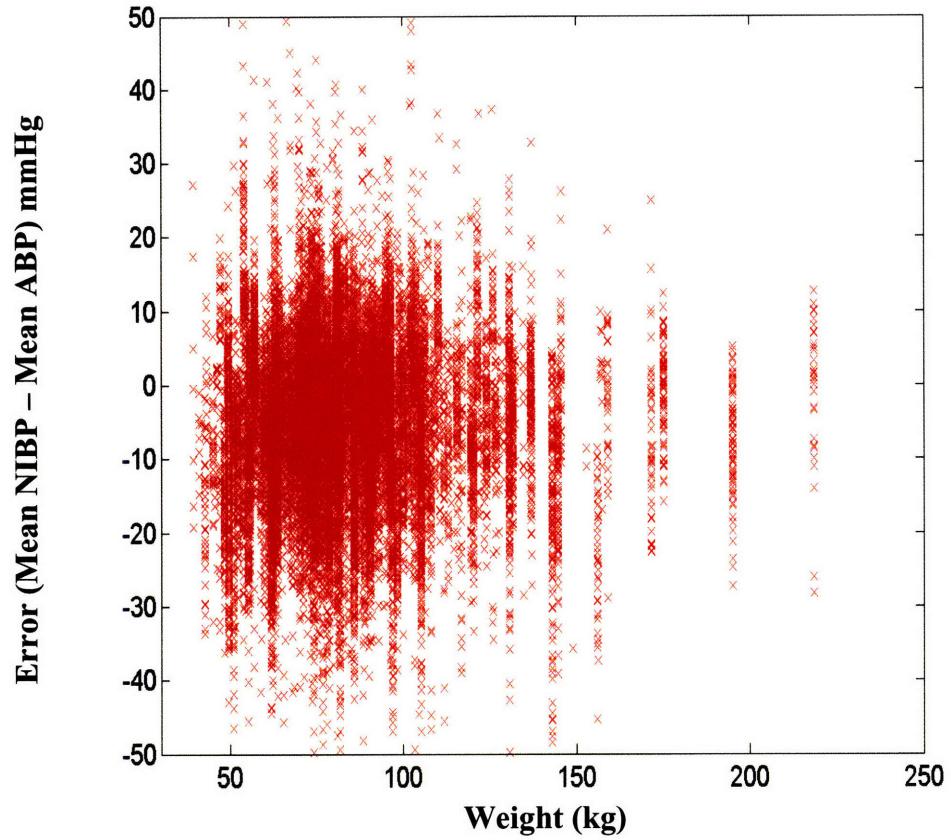


Figure 4-8: Error (Mean NIBP-Mean ABP) vs Weight

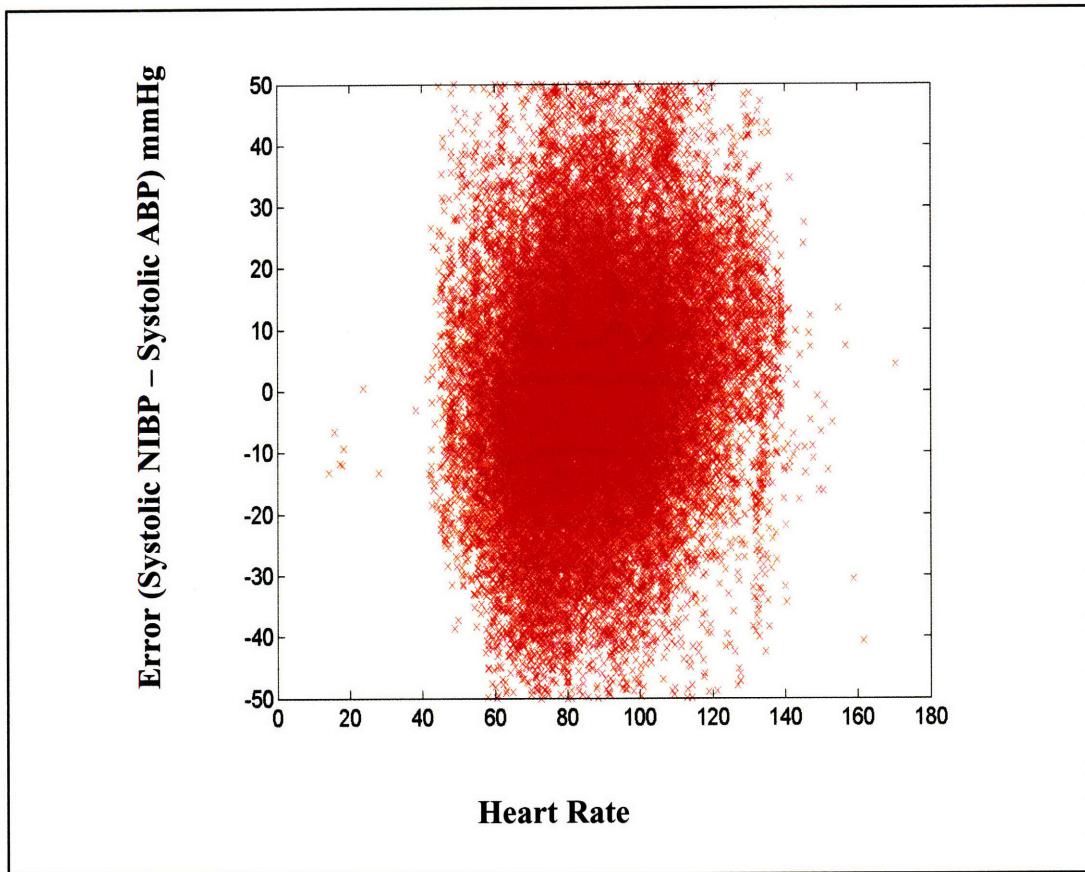


Figure 4-9: Error (Sys NIBP - Sys ABP) as a function of heart-rate

4.1.8. Results: Hemodynamic instability detector performance

NIBP and ABP measurements are evaluated as features for a simple threshold-based hemodynamic instability detector. The motivation for this study is to provide insight into the potential overall clinical utility of such measurements in hemodynamic monitoring. The NIBP and ABP features were evaluated using a simple clinical rule for the detection of hypotension. The respective NIBP and ABP data distributions from periods prior to instability were analyzed using a threshold-based detector. A systolic blood pressure below 80 mmHg or a mean blood pressure below 60 mmHg was chosen as the threshold for the detection of hypotension. These aforementioned thresholds using the intervals of data during hemodynamic instability quantitatively characterize the sensitivity of a detector for clinically significant hypotension. Sensitivity is defined as the percentage of

measurements below the hypotension threshold in the monitoring interval prior to the onset of a vaso-active medication. The false positive rate of such a detector can be studied using data from intervals of relative stability (no pressors such as neosynephrine or levophed given within 12 hours before or after a pressure measurement is taken). Ideally, a detector would not alarm during a period of stability, and thus the percentage (defined as false positive rate) of measurements below the respective hypotension thresholds should be minimal. The results of the detector performance are summarized in Table 4-3 and Table 4-4.

This study is limited by the assumption that the ICU staff is constantly vigilant to transient hypotension and thus, initiates vaso-active medications when it is appropriate to do so, and does not use such medications when they are unnecessary. Thus, we are assuming that the error between NIBP and ABP is not correlated to the level of vigilance of the ICU staff.

During periods of hemodynamic instability (90 minutes prior to an onset of a pressor), the mean blood pressure (MBP) using invasive ABP was the most sensitive detector as indicated by 51% of measurements being below a threshold of 60 mmHg (see Table 4-4). The MBP as measured with NIBP had a comparable sensitivity of 47%. However, when using the systolic blood pressure (SBP) from NIBP, the sensitivity of an NIBP-based detector is significantly less (21% of points below threshold during instability).

The false positive rate (as measured by the percentage of points below the hypotension threshold during a period of stability) was also assessed and summarized in Table 4-4. The SBP-based threshold using NIBP had the lowest false positive rate (2%). The MBP-based threshold using ABP had a higher false positive rate at 14%. A limitation of the study of false-positives is that there may have been a transient episode of hypotension without a subsequent intervention on the part of the clinical staff.

Table 4-3: Sensitivity of hemodynamic instability detector

<i>531 paired measurements (pooled) from 76 patients</i>	SBP < 80 mmHg	MBP < 60 mmHg
NIBP	21%	47%
ABP	41%	51%

Table 4-4: False positive rate of detector during hemodynamic stability

<i>8603 paired measurements from 282 patients</i>	SBP < 80 mmHg	MBP < 60 mmHg
NIBP	2%	17%
ABP	6%	14%

4.1.9. Results: NIBP-ABP trends prior to renal failure

The preceding analysis focused on sensitivity analysis of NIBP and ABP as predictors of hemodynamic instability. The definition of hemodynamic instability was based on a therapeutic criteria set including the use of vaso-active medications. However, the use of a medication may have been influenced by the blood pressure values prior to the onset of the medication. In this section, we define purely diagnostic criteria to evaluate the agreement between NIBP and ABP during an unstable physiologic state. The systemic blood pressure is measured as a surrogate for end-organ perfusion pressures. Low perfusion pressures can lead to organ failure. For example, renal failure may ensue in the presence of sustained hypotension or shock. An example of acute renal failure preceded by a drop in blood pressure is given by Figure 4-10 (NIBP and ABP systolic blood pressures). The same patient's NIBP and ABP mean blood pressures are plotted in Figure 4-11. In this ICU example, the NIBP systolic blood pressure did not decrease over time to the same extent as did the NIBP mean blood pressure. However, the systolic (invasive) ABP did decrease in time. This example may suggest that the systolic NIBP may be a poor indicator of pre-renal failure. The MIMIC-II database allows for studying this relationship in a large ICU patient populations.

Table 4-5 includes the results for blood pressure measurements using ABP and NIBP prior to renal failure in ICU patients. Acute renal failure was identified as a serum creatinine increase greater than 30% within 36 hours [46]. Those patients with simultaneous NIBP and ABP measurements within 36 hours prior to renal failure were included for analysis.

As Table 4-5 demonstrates, there was a significant difference between systolic NIBP and ABP prior to renal failure episodes (as defined by increases in creatinine). Systolic NIBP had a positive bias of 4.15 mmHg and standard deviation of error of 18.6 mmHg in comparison to systolic ABP. The error between NIBP and ABP in measuring mean systemic pressure included a smaller bias (-2.4 mmHg) and standard deviation (10.6 mmHg). Bland-Altman analysis of error for NIBP-ABP is included in Figure 4-12 (for mean pressures) and Figure 4-13 (for systolic pressures).

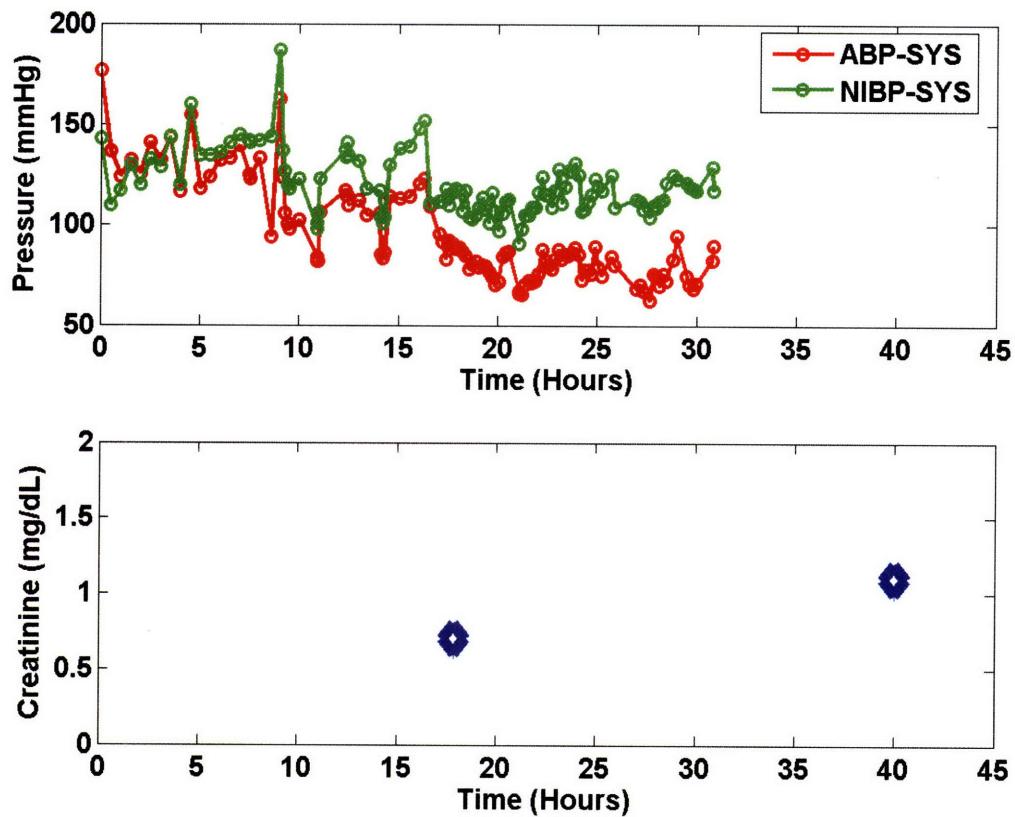


Figure 4-10: MIMIC 9618 NIBP and ABP (systolic) blood pressure trends prior to creatinine increase

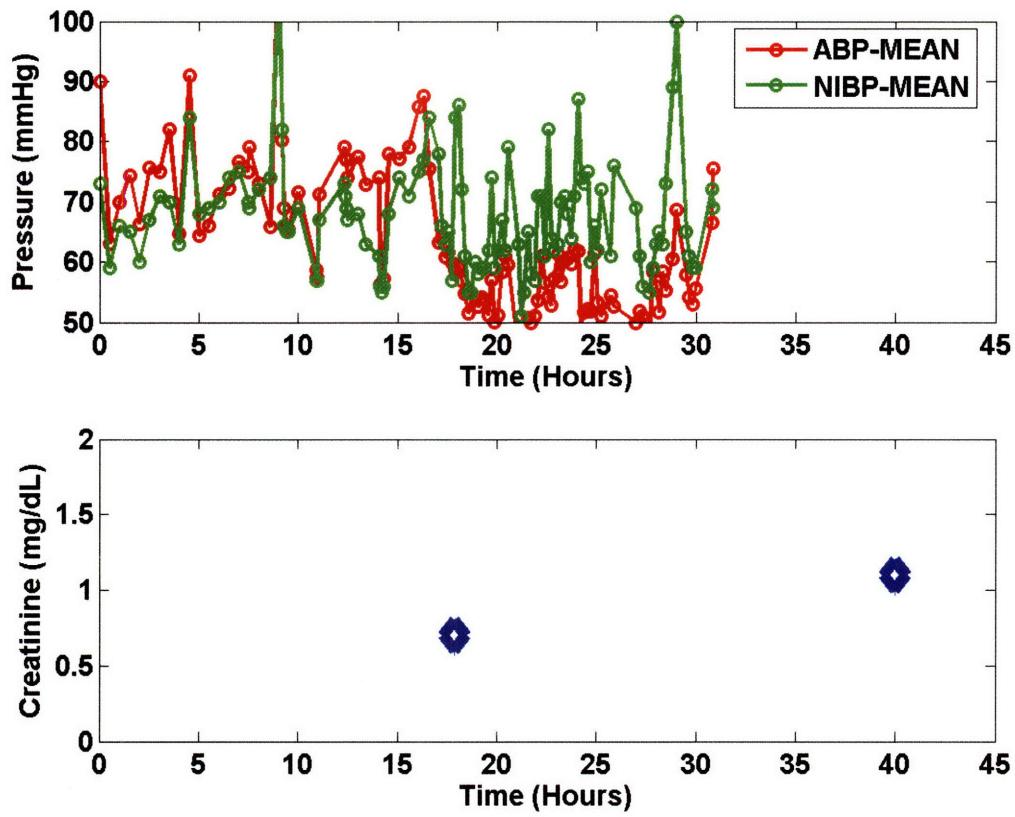


Figure 4-11: MIMIC 9618 NIBP and ABP (mean) blood pressure trends prior to creatinine increase

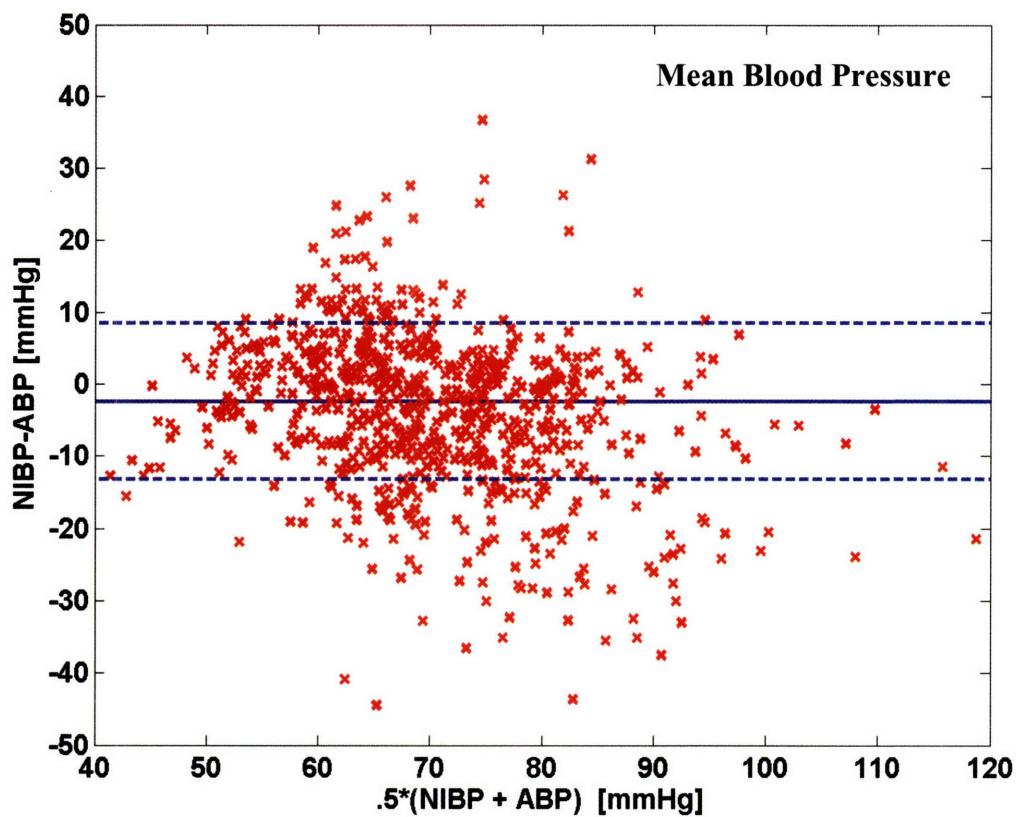


Figure 4-12: Bland-Altman analysis of NIBP-ABP (mean) error prior to creatinine increases

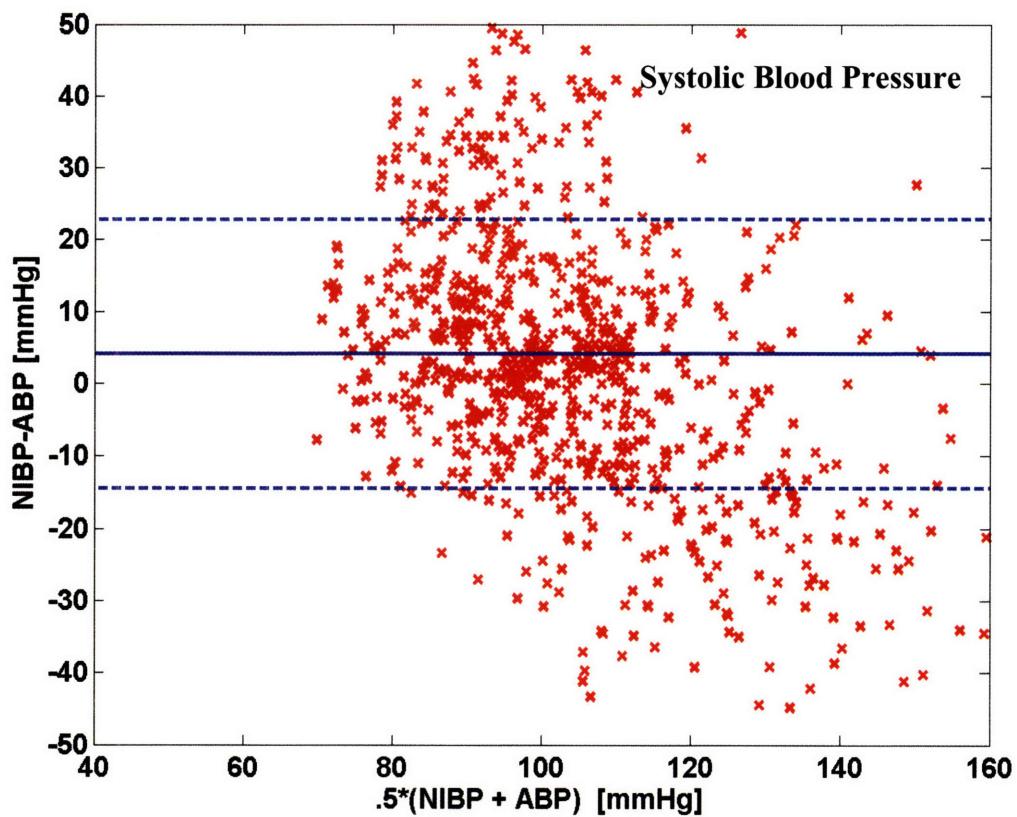


Figure 4-13: Bland-Altman analysis of NIBP-ABP (systolic) error prior to creatinine increases

Table 4-5: Result on NIBP-ABP error prior to acute renal failure

<i>Analysis of NIBP-ABP error within 18 hours prior to creatinine increase (989 paired measurements from 63 patients)</i>	BIAS +/- Precision (mmHg)	Correlation Coefficient
Systolic	4.15 +/- 18.6	-0.45
Mean	-2.4 +/- 10.9	-0.25
Diastolic	-6.0 +/- 10.2	0.12

4.1.10. Discussion: NIBP-ABP measurement of agreement

The utilization of the MIMIC-II database to analyze retrospectively the agreement between NIBP and ABP measurements in an ICU patient population revealed novel and clinically significant discrepancies. Previous studies comparing NIBP and ABP measurements were prospective in nature and limited to a smaller patient population. The unique aspects of the current study include the size of the patient population, various pathophysiologic states (such as renal failure) under which patients are monitored, and “realistic” non-ideal conditions in which pressure measurements are performed.

Previous studies have regarded the invasive ABP measurements as the “gold-standard” measurement by which non-invasive blood pressure measurement accuracies are assessed. However, the retrospective and large-scale nature of this study necessitated that an algorithm be utilized to reject “noisy” ABP measurements (see Figure 4-2). Furthermore, the Bland-Altman analysis which utilized the average of NIBP and ABP measurements was not reliant on the ABP measurement as serving as the “gold standard.” Finally, NIBP and ABP measurements were independently analyzed for serving as sensitive and specific indicators and/or predictors of various physiologic states.

The overall analysis revealed that a clinically significant bias and precision error between NIBP and ABP measurements was most pronounced when patients were hypotensive as defined by the average of NIBP and ABP systolic measurements (or NIBP and ABP mean arterial pressures). The agreement between NIBP and ABP was significantly worse when using systolic blood pressures for comparison during hypotension ($SBP < 80$ mmHg). The NIBP and ABP mean blood pressures had a higher level agreement in terms of bias and precision during hypotension ($MBP < 60$) than the systolic pressures.

The analysis of agreement was also extended across different physiologic states. The NIBP systolic blood pressure was found to have poor sensitivity in indicating when a

patient had a hypotensive episode as defined by the initiation of vasopressor therapy. However, the NIBP mean blood pressure was found to be nearly equally sensitive as the ABP mean and systolic blood pressures. Another physiologic state that was assessed included renal failure. In evaluating the blood pressure measurements prior to renal failure, the systolic NIBP was shown to have a clinically significant bias and precision error in comparison to systolic AB. The bias and precision error was most pronounced during hypotension prior to renal failure and greater than the error demonstrated when analyzing the overall population. On the other hand, the mean NIBP measurements had less bias and precision error when compared to the mean ABP measurements.

This finding is particularly salient when one considers a physiologically optimal definition of hypotension. The mean pressure is the true driving pressure for peripheral blood flow [46]. However, many clinicians utilize the systolic blood pressure for defining hypotension and shock [references]. For example, several ICU patient acuity formulae include the systolic blood pressure as a core measurement while excluding the mean blood pressure. A simple indicator of shock is a formula called the “Shock Index” [58]. The shock index is defined as the ratio of the heart rate to the systolic blood pressure. However, the findings of this study would suggest that the conventional shock index would lack sensitivity if reliant upon NIBP systolic blood pressure.

The preference for systolic blood pressure as the de-facto parameter indicative of hypotension is perhaps based on the limitations of the most popular non-invasive measurement technology---the auscultatory method [18]. Prior to the introduction of automated NIBP measurements, clinicians would assess blood pressure with a simple cuff and use a stethoscope to find the Korotkoff sounds that are indicative of the systolic and diastolic blood pressures. Alternatively, clinicians could use a cuff and simply palpate the wrist for the radial pulse. Such simple techniques would readily provide a systolic blood pressure measurement, but not the mean pressure. Thus, clinical practice was influenced to utilize the measurements that were available at the time. The findings of the present study may support the re-evaluation of a more clinically optimal definition of hypotension and shock.

A principal motivation for the current study is to evaluate the performance of NIBP in a realistic ICU setting and provide guidance to enhancing the performance of NIBP algorithms. The retrospective analysis of agreement between NIBP and ABP highlighted clinically significant discrepancies between the two methods in measuring the systolic blood pressure. A common method to identify the systolic and diastolic blood pressure using the oscillometric technique is to assume that the ratios of oscillation amplitudes (at the systolic and diastolic blood pressures) to maximal amplitudes (corresponding to the mean blood pressure) are constant (see Figure 4-1). Geddes et al found that the ratio between the oscillation amplitude occurring at the systolic blood pressure (A_S) and maximal oscillation (A_M) amplitude (at the mean blood pressure) was equal to 0.62 (average of all subjects) in dogs with simultaneously measured invasive and cuff pressures [18]. However, across the different animal subjects, there was clinically significant variability in the optimal ratios used to identify systolic and diastolic blood pressures. Similar studies have been conducted in humans by measuring the oscillation amplitudes with respect to the Korotkoff sounds obtained from the auscultatory method. Ursino et al described a physical model to characterize factors affecting the agreement between oscillometric and invasive blood pressure measurements [77]. Their model predicted that aforementioned ratios (systolic/mean, and diastolic/mean) of the cuff pressure oscillation amplitudes are dependent on the underlying mean blood pressure, pulse pressure, heart rate, and blood vessel wall stiffness. Many of these parameters can widely vary across an ICU patient population, and within an individual ICU patient's stay. Thus, experimental and model-based studies suggest that the characteristic ratios that relate the systolic and diastolic oscillations should not be considered constant values across an ICU population. The retrospective study in this thesis also supports a re-evaluation of the ad-hoc algorithms used to identify SBP and DBP. In particular, the significant bias (overestimation of ABP) in the comparison between NIBP and ABP systolic pressures during hypotension suggests that the ratios relating the oscillation amplitudes at systolic blood pressure to the maximal oscillations (that signify the mean blood pressure) should be decreased when the mean blood pressure is found to be low. Thus, the results of this study suggest a pressure-dependent ratio for systolic blood

pressure identification may significantly improve the performance of NIBP during episodes of hypotension. The pressure-dependent functional used to optimally tune the ratio of amplitudes need not be linear, and can be optimized through regression of the NIBP/ABP dataset in MIMIC-II.

4.2. Diurnal variation in hemodynamic interventions in the ICU

4.2.1. Introduction

There has been recent interest in studying the diurnal pattern of medical interventions in the ICU [17]. Ideally, the collective vigilance of ICU clinicians to transient changes in a patient's physiology should be consistent throughout a 24-hour cycle. The large scale of the MIMIC-II database and the availability of high-resolution physiologic and therapeutic data allows for novel insight into the temporal vigilance of clinicians in response to physiologic patterns suggestive of hemodynamic deterioration.

Failure to respond promptly to a physiologic pattern that is suggestive of deterioration can arguably be considered a medical error of omission. The reduction of medical errors and iatrogenic injury has emerged as a major focus in the efforts to improve the delivery of care in the ICU environment [62]. Medical errors are estimated to cause between 44,000 and 98,000 deaths each year [27]. Indeed, one of the motivations for the development of clinical decision support systems has been the belief that such technologies can significantly reduce medical errors and adverse events within hospital environments. Several researchers have utilized epidemiologic methods to study the incidence of adverse events and medical errors in intensive care. Medical errors have been defined as a “failure of a planned action to be completed as intended or the use of a wrong plan to achieve an aim.” An adverse event is defined as “any injury due to medical management, rather than the underlying disease” [39]. The major studies on the occurrence rate of in-hospital medical errors have focused primarily on errors of commission rather than errors of omission. However, several researchers have suggested that errors of omission are far more common than errors of commission [Brigham study,

others]. A main challenge in studying errors of omission in the ICU is identifying physiologic events that may meet certain treatment criteria but are otherwise ignored or not acted upon promptly. Furthermore, it is important to identify factors that may be linked to errors of omission in order to promote better clinical practice in the ICU.

A particular aspect of clinicians' vigilance that has been receiving increased attention is diurnal variation in vigilance. For example, the use of medical emergency teams (METs) to respond to ICU patient crises was analyzed as a function of the time of day [17]. METs differ from traditional "code teams" that respond to cardiac arrests in that the rationale for the deployment of METs is to offer earlier interventions to prevent a patient from deteriorating. Thus, METs are ideally used to prevent a cardiac arrest instead of responding to one. The results of two major studies on the activation of METs concluded that there is significant time-dependent variability in the hospital ability to consistently detect conditions that meet MET activation criteria [17, [31]]. However, both of these studies were limited by the availability of high-resolution physiologic data and were dependent on sparser paper-based nursing charts. While cardiac arrest represents one class of hemodynamic events, the MIMIC-II database allows for the study of a broader set of physiologic states that require therapeutic interventions.

Among the principal roles of ICU clinicians is to monitor and maintain the hemodynamic stability of ICU patients. Signs of hemodynamic instability may include significant drops in systemic blood pressure and concomitant changes in heart rate. During periods of hemodynamic instability, hypotension may result in poor end-tissue perfusion and, ultimately, end-organ failure. The maintenance of adequate blood pressures in critically ill patients is most frequently achieved through the administration of intravenous fluids, vaso-active medications, and inotropic agents [46].

The MIMIC-II database was utilized to study the diurnal variation in the incidence of hypotensive episodes. The diurnal variation in the initiation of hemodynamically significant vaso-active therapeutic interventions was also studied. Factors that may influence the need for therapeutic interventions, such as the hemodynamic patterns of an

ICU patient, were also characterized as a function of time of day. In the next section, the methodologies for identifying hypotensive events as well as the therapeutic interventions that are suggestive of hemodynamic instability are described. Then, the major results are summarized with particular focus on the variability in the incidence of hemodynamic events and interventions as a function of the time of day. Finally, a discussion of the major findings and the potential ramifications for ICU practice are provided.

4.2.2. Methodology

A basic level of analysis of hemodynamics included computing the summary statistics (mean and standard deviation) of specific vital sign variables (nurse-validated) as a function of time of day. For example, all the heart rates stored in the CareVue database were first sorted by the hour of the day (with rounding to the nearest hour) in which they were recorded. Then, all the heart rate values for each hour are grouped together and the respective means and standard deviations are calculated. This temporal analysis is repeated for other variables such as mean blood pressure, respiration rate, and oxygen saturation.

The MIMIC-II database presented unique challenges in developing definitions for hypotension based upon quantitative and objective criteria. In particular, the high resolution and volume of physiologic and clinical data coupled with inherent noise in the data required the use of automated algorithms to identify hypotensive episodes. These automated algorithms utilized different quantitative definitions for hypotension that were dependent upon a combination of therapeutic and diagnostic data.

The first definition of a hypotensive event relied strictly on the availability of diagnostic vital signs data. In particular, definitions were developed that identify events suggestive of acute changes in a patient's hemodynamic stability. The nursing chart data from CareVue was utilized as the source of vital signs data for this aspect of the study. A hypotensive event was defined any point in time in which the following three criteria have all been met: 1) the nurse-verified mean blood pressure (MBP) was less than or equal to 60 mmHg, 2) the MBP has dropped by more than 10 mmHg within two hours,

and 3) there is a concomitant increase in heart rate by more than 10 beats/minute within two hours. This aforementioned multiparameter definition was utilized to identify those events that are likely to reflect actual acute changes to physiology rather than noise in the diagnostic measurements.

A shortcoming of relying upon purely diagnostic (monitoring) data in defining hypotension is that such data are subject to noise. Noisy data may lead to the false detection of hypotensive events. Thus, another definition of hypotension was developed that was dependent upon therapeutic criteria. Two of the most common therapeutic interventions utilized in the ICU in response to acute hypotension include intravenous fluid infusions and vaso-active medications. An intravenous fluid bolus was defined by satisfying one of the following two criteria: 1) total intravenous fluid input exceeding 800 ml within one hour, or 2) total intravenous input exceeding 1200 ml within two hours. A vaso-active medication is typically given to a hypotensive patient in order to increase total peripheral resistance and thereby raise the systemic blood pressure. A list of vaso-active medications is provided in Table 4-6.

Analgesics and hypnotics are common medications utilized in ICU practice to maintain a patient's comfort. However, the administration of certain analgesics (e.g. morphine) and hypnotics (propofol) has been linked to iatrogenic acute hypotension in ICU patients [references]. Thus, analysis of the hourly variation in medications that may cause hypotension may provide further insight into the temporal hemodynamic variations in ICU patients.

Table 4-6: List of Vaso-active Medications

dopamine
epinephrine
Levophed (Norepinephrine)
Neosynephrine

Table 4-7: List of hypnotics and analgesics

Medication Name
Propofol
Morphine
Ativan
Fentanyl

The fiducial point that was used to demarcate a hemodynamic event was determined by the time of onset of a vaso-active medication. The onset of administration was chosen because of the high likelihood that the initiation of the treatment was in response to the clinical staff's detection of hemodynamic deterioration.

Finally, those hemodynamic events defined by an onset of a vaso-active medication were further analyzed to characterize the blood pressure values prior to therapy initiation. The motivation for this method of analysis is based upon a desire to characterize the hemodynamic patterns that lead to the initiation of therapy. Furthermore, one can study if the patterns are statistically similar throughout the 24-hour care cycle. Thus, prior to each medication onset, those hemodynamic events that had available minute-to-minute resolution mean blood pressure measurements were identified. Then, the blood pressure measurements up to one hour prior to the medication initiation were processed to identify the hourly mean value as well as the minimum pressure value.

Another class of medications that is commonly used in the ICU includes hypnotics and analgesics. However, while such medications are intended to sedate a patient and maintain adequate patient comfort, certain hypnotics and analgesics have also been linked to causing hypotension in a patient [Propofol reference]. The incidence of propofol onsets was analyzed as a function of time of day in the MIMIC-II patient population in a manner similar to the vaso-active medication onsets.

4.2.3. Results: Hourly variations of blood pressure and heart rate

The population-averaged heart rates and mean blood pressures are plotted as a function of time of day as shown in Figure 4-14 and Figure 4-15, respectively. The unpaired t-test and Ranksum test were used to determine if there were statistically significant fluctuations as a function of time of day. The temporal fluctuations in both blood pressure and heart rate were minor when normalized by the standard deviations, but statistically significant ($p < 0.005$). The population-averaged heart rate was lowest at 2 AM (85.4 ± 17.4 bpm) and highest at 7 PM (88.2 ± 17.6 bpm). The population-averaged mean blood pressure was lowest at 2 AM (79.0 ± 16.6 mmHg) and highest at 8 AM (81.9 ± 17.6 mmHg).

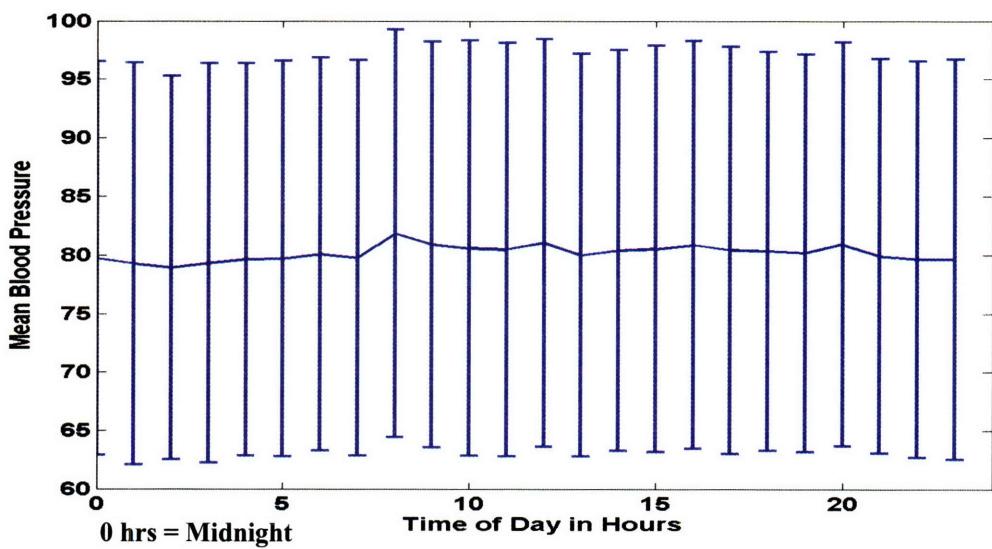


Figure 4-14: Mean blood pressure (population-averaged) variations throughout 24-hour cycle

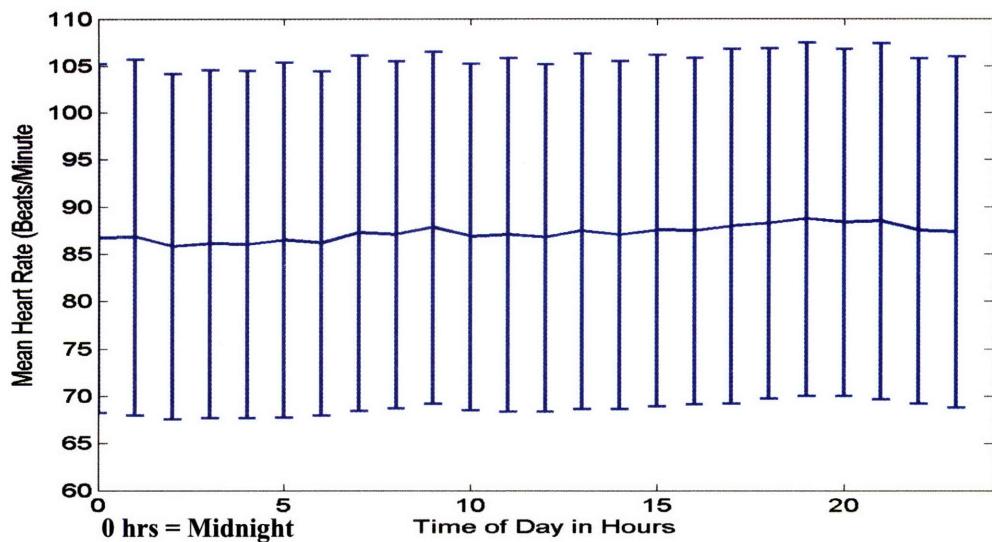


Figure 4-15: Mean heart rate (population-averaged) as a function of time of day

4.2.4. Results: Hourly variations of clinical data charting

Nursecharted vital signs (e.g. blood pressure, heart rate, respiration, oxygen saturation) are typically entered every hour in the clinical information system. The rate of charting of vital signs was trended over the 24-hour cycle. As demonstrated in Figure 4-16, the charting rate was uniform throughout the 24-hour cycle except for two noticeable decreases (approximately 20% decrease) in frequency at 7 AM and 7 PM. Nurse shift changes also occur at these two hours.

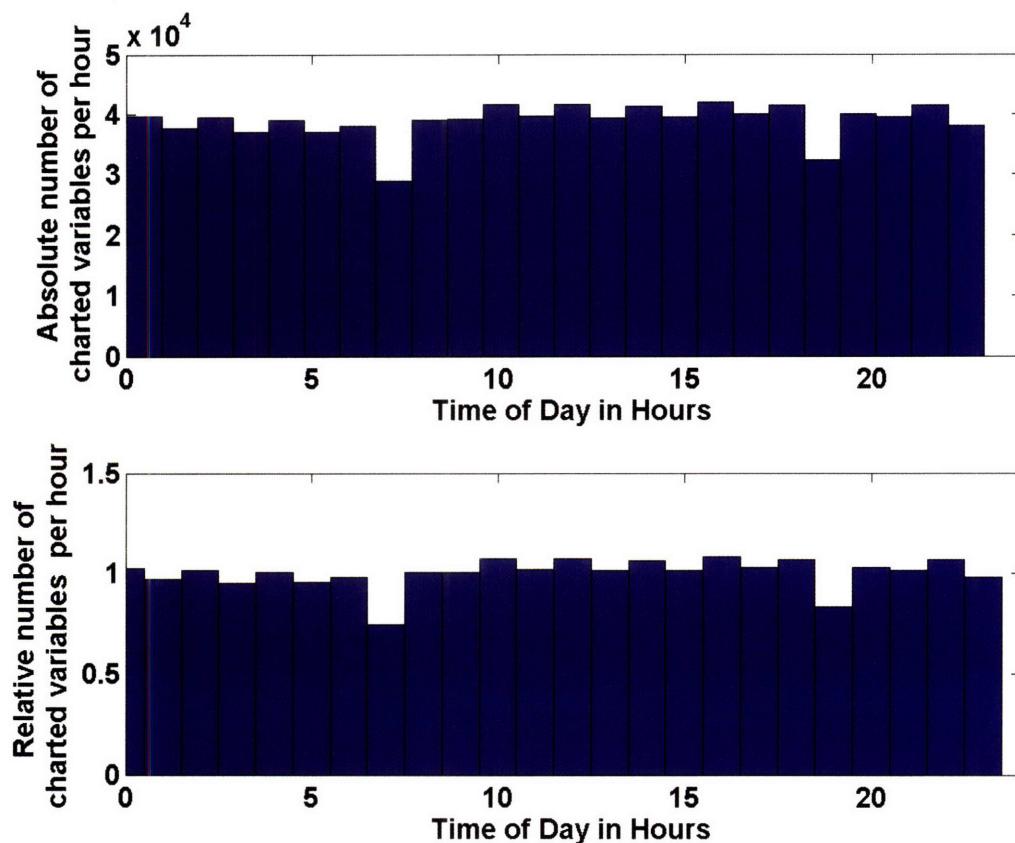


Figure 4-16: Rate of charting of hemodynamic variables (heart rate and blood pressure) as a function of time of day

4.2.5. Results: Diurnal variation in onset of hemodynamically significant interventions

The frequency of onset of hemodynamically significant therapeutic interventions was charted as a function of time of day (in hours). In Figure 4-17, the total number of vaso-active medication onsets (initiation of a medication) is trended as a function of time of day. The MIMIC-II patients with vaso-active medication onsets were randomly sorted into 5 different groups and the mean and standard deviations of the hourly medication onsets are shown in Figure 4-17(b). There was a statistically significant decrease in the number of vaso-active medication onsets at 7 AM (lowest hour of total onsets) in comparison to 8 PM (peak hour of total onsets) using the Student's t-test at ($p < 0.05$).

Another method to treat hypotension is the administration of a bolus of intravenous fluid. In a manner similar to the analysis of hourly variations in vaso-active medication onsets, hourly variations in the total fluid bolus administrations in patients are shown in Figure 4-18. Patient randomization into five groups allowed for the calculation of hourly means and standard deviations. The charted hourly statistics demonstrated significant differences as a function of time of day. The lowest number of fluid bolus administrations was at 7 AM, while the peak hour was at 2 PM ($p < 0.05$).

The variation in the total hourly initiations of an analgesic or hypnotic medications is graphed in Figure 4-19. Patient randomization into five groups allowed for the calculation of hourly means and standard deviations. The charted hourly statistics demonstrated significant differences as a function of time of day. The lowest number of total onsets of hypnotics/analgesic medication administrations was at 7 AM, while the peak hour was at 11 AM ($p < 0.05$).

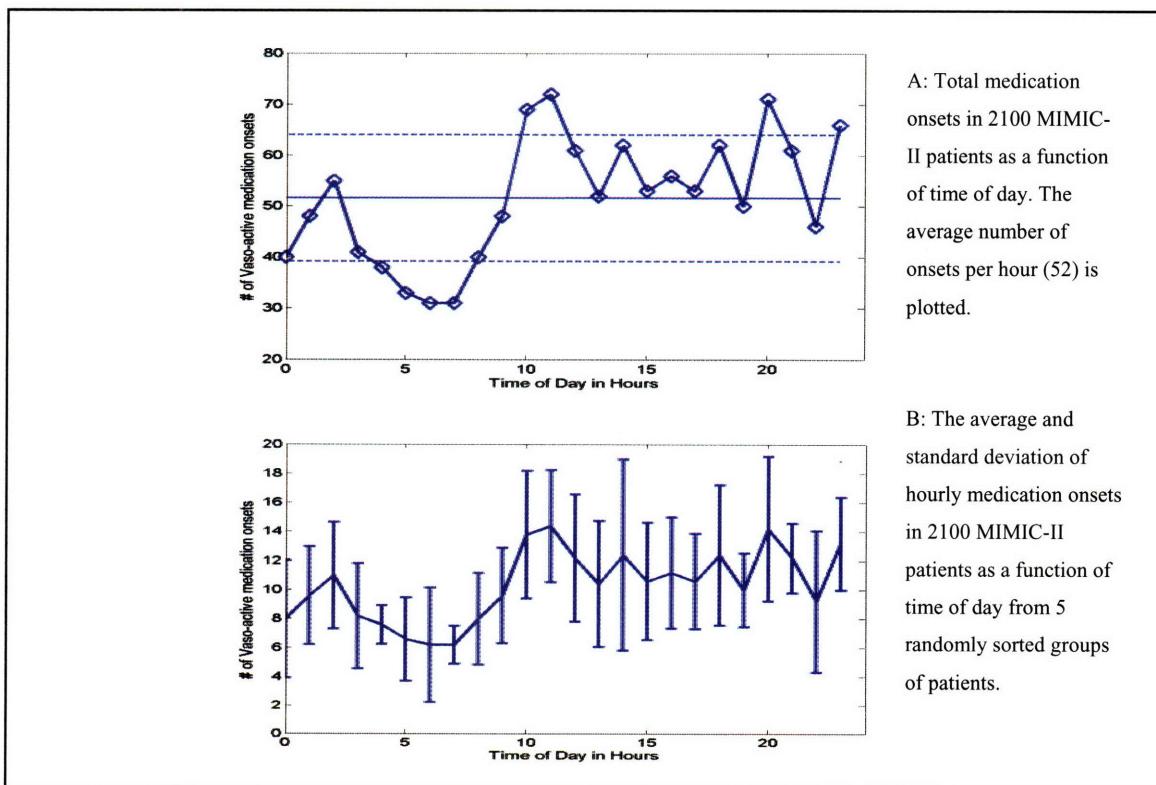


Figure 4-17: Frequency of vaso-active medication onsets as function of time of day

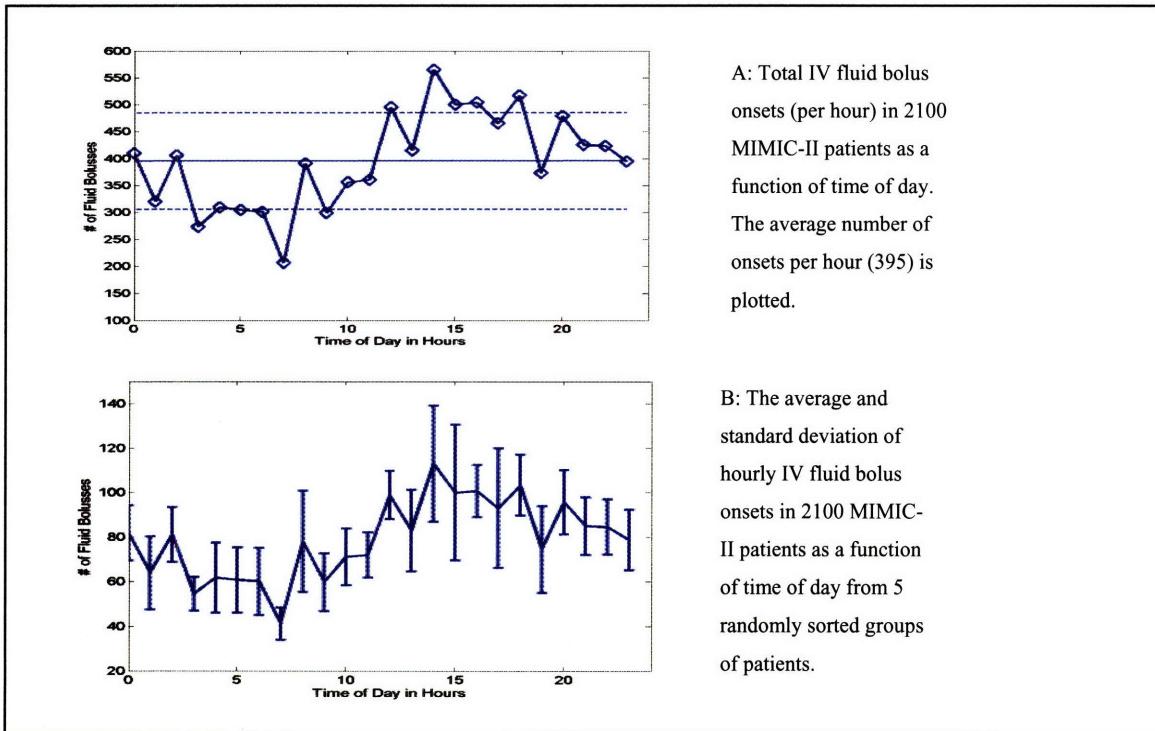


Figure 4-18: Frequency of IV fluid bolus onsets as function of time of day

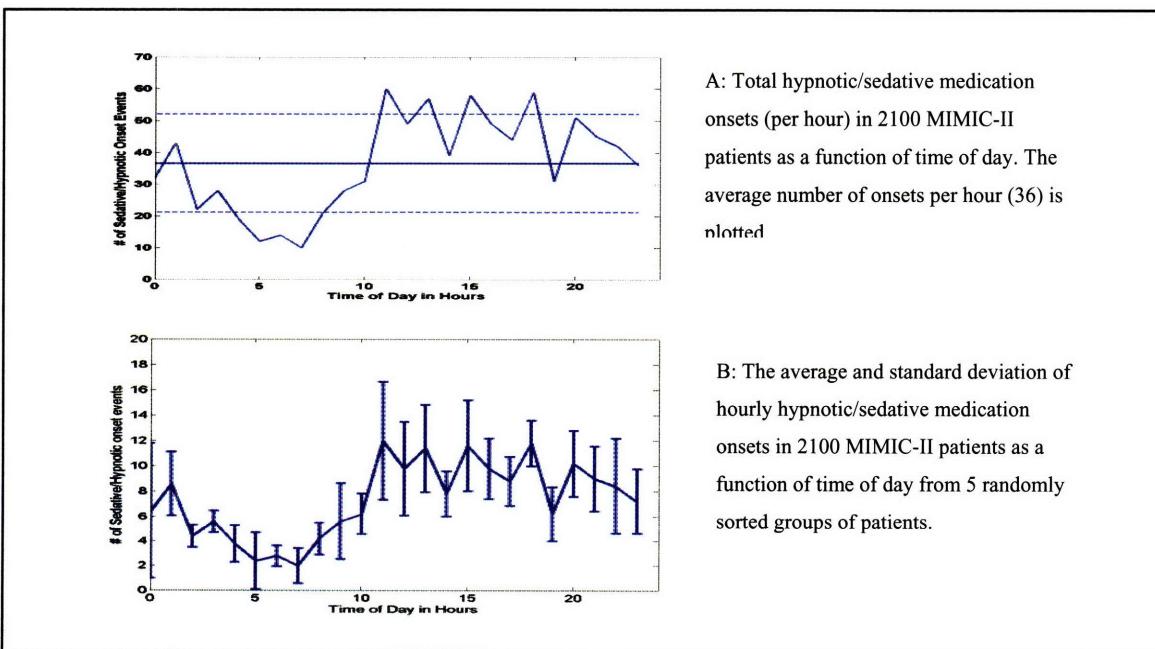


Figure 4-19: Frequency of hypnotic/sedative medication onsets as function of time of day

4.2.6. Results: Temporal analysis of hypotensive events

The number of acute hypotensive events (see methodology for definition of hypotensive event) was charted as a function of time of day as shown in Figure 4-20(a). The hypotensive events were identified by analysis of the nurse-validated heart rates and blood pressures that were charted in the clinical information system. There was significant hourly variation in the incidence of hypotensive events. The peak incidence of acute hypotensive events occurred at 5 PM while the minimal number of events occurred at 9 AM. However, there was not as distinct of a diurnal pattern in the incidence of hemodynamic events as there was in the initiation of hemodynamic interventions.

In Figure 4-20(b-c), analysis of blood pressure values prior to the initiation of vaso-active therapy is shown. In particular, the high-resolution minute-to-minute monitor-generated arterial blood pressures (mean) were analyzed during the 60 minute window prior to the initiation of therapy. The average value of the MBP measurements during the 60-minute window (Figure 4-20b) and the minimum value (Figure 4-20c) are shown. The hourly window representing the time between 7AM-8AM had the lowest average and minimum MBP. The window representing the time between 4PM-5PM had the highest average and minimum MBP. Statistical analysis using the two-sample t-test found significantly different average and minimal MBP values prior to vaso-active medication onsets as a function of the time of day ($p < 0.05$).

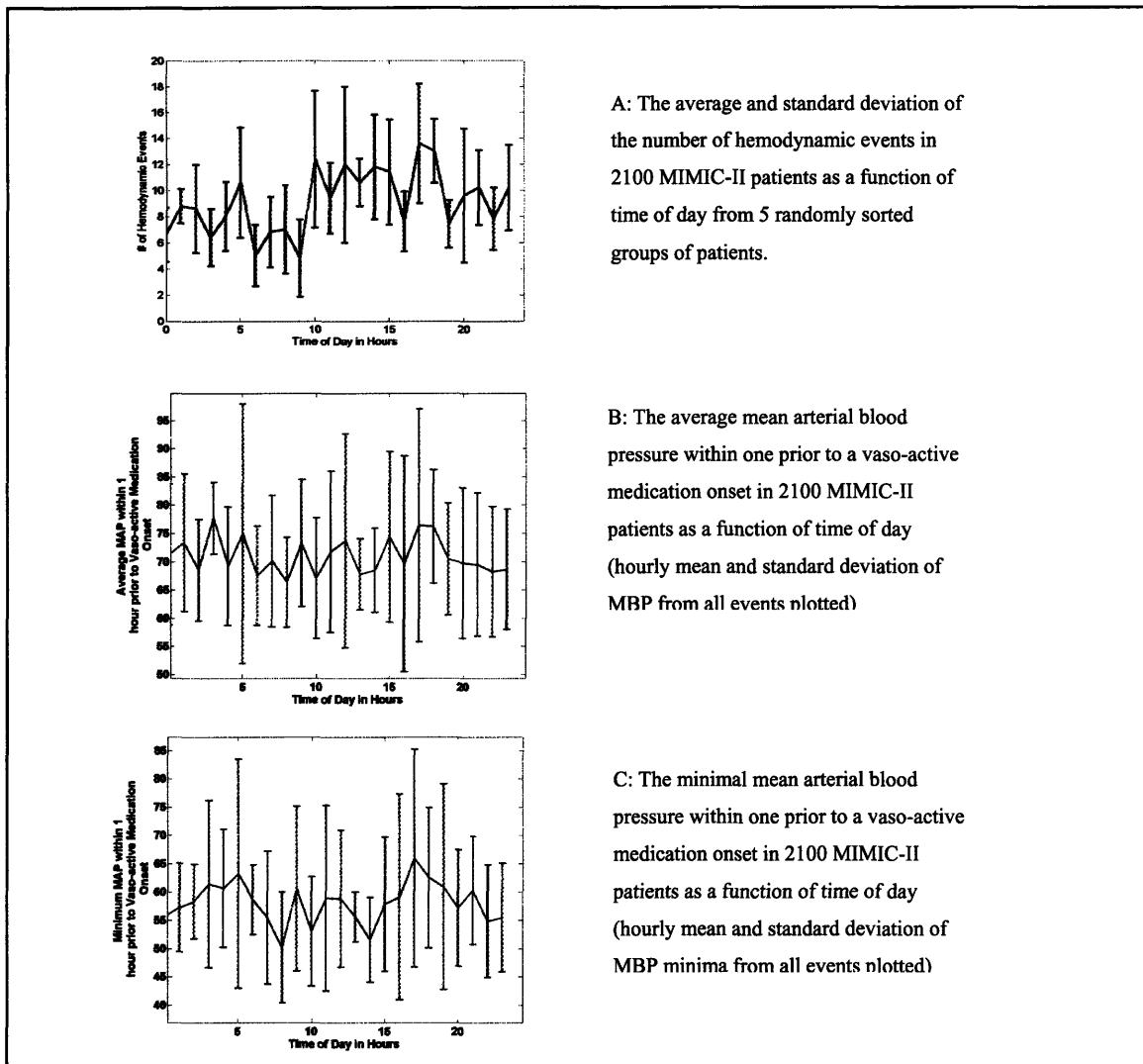


Figure 4-20: Temporal variation of acute hypotensive events

4.2.7. Discussion

The MIMIC-II database was utilized to study the temporal variations in hemodynamic of ICU patients. The major contribution of this study involved the analysis of high-resolution therapeutic and diagnostic (monitoring and nurse-validated) vital signs measurements over the 24-hour daily cycle in a large cohort of ICU patients.

The baseline analysis of vital signs (mean blood pressure and heart rate) demonstrated minor but statistically significant variations over the 24-hour cycle of care. The heart rate

and blood pressure changes were less than the diurnal variations observed in normal healthy subjects [57]. The autonomic control system exerts a powerful influence on the blood pressure and heart rate. The changes in heart rates and blood pressures in healthy subjects over a 24-hour cycle are linked strongly with the sleep-wake cycle [57]. However, ICU patients are subject to various therapeutic interventions that modulate both the autonomic control system (via vaso-active medications) and the sleep-wake cycle (via hypnotics/analgesic medications). The timing of interventions may be independent from patient to patient. The temporal analysis of the onset of new vaso-active medications and hypnotics suggests that there are intervals where medications are more likely to be started. This time-dependent variation may result in the small but statistically significant variation of hourly averaged heart-rates and blood pressure time series in the ICU patient population as a whole. The massive sample size (greater than 300,000 measurements) used in this study from the MIMIC-II database is most likely the principal reasons that such hourly differences in heart rate and blood pressure are statistically significant.

The heart-rate and blood pressure trends were also analyzed to identify episodes of acute hypotension (see Figure 4-20). While there were significant hourly variations in the incidence of acute hypotensive as a function time, there was less of a distinct transition in the rate of hypotensive events over the 24-hour cycle (day vs night). The peak incidence of hypotensive events (based only on blood pressure and heart rate time series) occurred at 5 PM. The peak hour also fell close to the hour of peak administration of vaso-active medications and (IV bolus) fluids which occurs at 7 PM. However, there were several hours with significant hemodynamic events that occurred after midnight, thus night time was not as quiescent as the therapeutic profiles (histograms) would suggest.

The nurse-validated vital sign data were utilized for certain aspects of the analysis of temporal hemodynamic variations. Nurse-validated vital sign data are typically charted at hourly intervals in the clinical information system. While the overall charting frequency as determined by the analysis of the entire ICU patient population was mostly uniform, there were two significant decreases in charting frequency at 7 AM and 7PM. The decreases in charting frequency occurred at the hours of nursing-shift changes. Clinician

shift changes in the ICU should ideally result in minimal changes to the collective vigilance and completion routine tasks on the part of the staff. Hourly charting of vital signs by nurses is important so that clinically significant trends suggestive of physiologic status changes are detected in a timely fashion. Failure to chart clinical data during shift changes may represent a transient decrease in workflow effectiveness. The time-dependent automated histogram analysis of hemodynamic variables can perhaps be investigated as a novel method of measuring workflow effectiveness during shift changes.

The initiation of hemodynamically significant therapeutic interventions was studied as a function of time of day. Vaso-active medications such as Norepinephrine or Neosynephrine are presumably started in response to a patient's acute hemodynamic deterioration. There are no studies that we are aware of that would suggest that an ICU patient is more likely to deteriorate at specific times of the day due to underlying disease processes. In particular, a recent study by Galhotra et al demonstrated that the incidence of MET activation events and cardiac arrests did not have a diurnal variation in the ICU [17]. However in the MIMIC-II database, the initiation of therapeutic interventions that may be appropriate to treat hemodynamic deterioration showed a strong diurnal variation (see Figure 4-17 and Figure 4-18). Furthermore, the initiation of analgesic and hypnotic medications that may be indicated when a patient complains of pain and discomfort also demonstrated a similar diurnal variation (see Figure 4-19). A possible explanation of these results is that hypnotics like propofol have known hypotensive side effects, and thus, their initiation causes an increased likelihood of hemodynamic deterioration. However, a similar temporal histogram analysis of hemodynamically significant interventions (pressors and fluid boluses) in patients that were not given propofol or morphine still demonstrated similar diurnal variations in intervention onsets.

Another level of temporal analysis fused physiologic data from bedside monitors with therapeutic information such as the administration of vaso-active medications. As mentioned previously, a therapeutic intervention such as the initiation of a vaso-active medication may be preceded by a transient drop in blood pressure. Thus, the blood

pressure can be considered the independent variable. The blood pressure was characterized during the one-hour period prior to the initiation of vaso-active medications. Figure 4-20 shows that both the average and minimal value of the MBP over the window vary significantly over the 24-hour cycle. Furthermore, the hourly differences are far greater in magnitude than the hourly differences in baseline blood pressure (see Figure 4-14). Typical ICU protocols call for the initiation of vaso-active medications when the mean (or systolic) blood pressure drops below a predefined threshold. Ideally, the ICU staff's vigilance to acute hypotension should remain constant over the 24-hour cycle. The minimum mean blood pressure prior to the initiation of vaso-active medications was smallest in the 7AM-8AM window (52 mmHg) and highest in the 4PM-5PM window (65 mmHg). These findings suggest that, on average, the clinical staff allows the blood pressure to drop significantly lower between 7AM-8AM before initiating appropriate therapy.

The increasing focus on the use of ICU guidelines and standards of care [66] is--in part--reflective of an effort to reduce medical errors. Developing an objective and quantitative metric for identifying medical errors of omissions is a challenge in the ICU environment. For example, allowing a patient's blood pressure to drop precipitously low can be considered an error of omission. Whether such an error would lead to a poorer outcome or longer ICU stay is unknown. However, a prolonged deficit in end-organ perfusion can compromise a patient's physiologic stability [46].

In the preceding analysis of diurnal variation in therapeutic interventions and vital signs, significant hourly differences were found. The 24-hour care cycle includes periods of time in which patients are more likely to be subject to different procedures and diagnostic studies. Furthermore, the workflow of the clinical staff includes several time-dependent administrative tasks and staffing changes. However, vigilance to a fragile ICU patient's physiologic status changes should remain optimal over the entire 24-hour cycle if possible. The novel methodology utilized in this study may be considered as a high throughput data analysis tool to characterize ICU consistency as a function of the time of day. The resultant analysis may focus a clinician's attention to windows of time where

the delivered care is inconsistent with typical unit performance or clinically accepted standards of care. The identified inconsistencies may warrant further analysis. If a lack of vigilance is identified as the primary cause of inconsistency, changes to clinical practice may be considered as corrective measures.

5.Novel Time Series Similarity Metric

In the previous chapters, we introduced the MIMIC-II database as a new resource for supporting research in patient monitoring and clinical decision support algorithms. The availability of multiple streams of high-resolution physiologic signals as well as salient therapeutic data allows for the development of novel pattern recognition algorithms. In Chapter 2, we motivated the need for developing algorithms capable of fusing information from several different parameter time series (such as heart rate, blood pressure, and respiration rate) to support clinicians in identifying clinically significant physiologic conditions as they evolve. We summarized several different approaches such as rules-based expert systems as well as different types of pattern recognition algorithms (neural networks, fuzzy logic, and multidimensional clustering). A particular area of pattern recognition and data mining that has received significant research in the past involves the development of algorithms that identify “similarities” between different time series [33]. The definition of “similarity” among time series is a vague concept. The metrics that are used to identify similar signals or records are typically based on quantitative functions. However, the quantitative similarity between two or more signals does not necessarily translate into a meaningful practical similarity. Thus, a major challenge in developing physiologic similarity metrics with applications to ICU patient records is identifying statistical (quantitative) metrics that also demonstrate physiologic or clinical similarity.

Time series similarity metrics are not restricted to medical domains. The ongoing advances in computer processing power, networking, and data storage have enabled modern computers with capabilities of generating, processing and storing terabytes of data. Often the data are time-varying, such as data from biomedical sensors attached to patients in a hospital intensive care unit (ICU) or seismic recordings from geological studies. Massive volumes of data can be readily archived in a digital data warehouse to

support research in automated data mining and knowledge discovery. Ongoing research in time series datmining includes developing algorithms that identify “similar” temporal patterns in a collection of time series [1]. For example, one may consider the following queries:

Q1: Identify corporations with similar stock price fluctuations with respect to a change in oil price.

Q2: Identify a group of ICU patients with similar changes in their heart rate and blood pressure trends prior to an episode of severe hypotension.

A time series in general and an ICU patient record in particular, can be of varying lengths --- from a few hours to several hundreds of hours. Multiparameter records can consist of several different data streams that convey unique and important information. In meteorology, weather patterns may be characterized by humidity, temperature, and wind speeds. In MIMIC-II, patient records include high resolution (125 Hz) physiologic waveforms of ECG that monitor the heart’s electrical activity, and vital parameter recordings acquired at a resolution of 1 sample per minute to follow a patient’s blood pressure, heart rate, and blood oxygen saturation. Also, there are fluid balance data and medication drip rates charted on a near-hourly basis. Finally, there are clinical laboratory values that are sparsely recorded at a rate of approximately 1 to 3 samples a day.

The aforementioned characteristics of the MIMIC-II database pose several challenges for developing suitable temporal similarity metrics. Ideally, a similarity metric should be capable of comparing two records that are of different lengths. Furthermore, a temporal similarity metric must be capable of characterizing an ICU record using more than one data stream. For example, a slowly increasing heart rate trend over several hours accompanied by a concomitant decrease in blood pressure (or cardiac output) may be indicative of an internal bleed in a patient. Thus, a similarity metric that fails to capture dynamical relationships between two or more parameters would be of limited use in identifying internal bleeds in a large-scale ICU patient database. A database consisting of terabytes of time series data may have unique patterns that have hitherto been

undiscovered. Thus, for data mining and “knowledge-discovery” applications, an algorithm cannot be heavily dependent on user-tunable parameters or inputs. All of these aforementioned challenges are compounded by the inherent noise in ICU data.

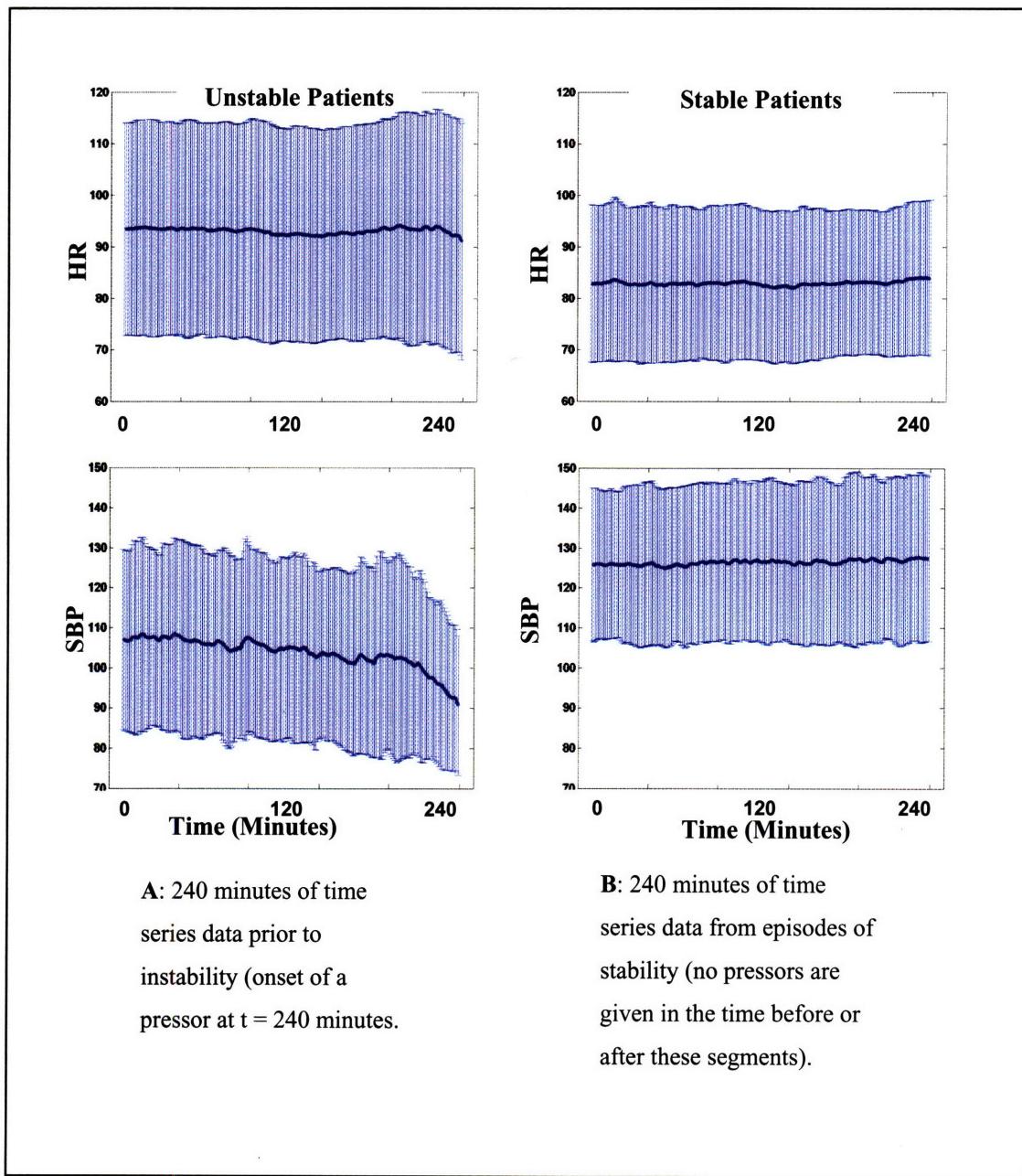


Figure 5-1: Multiparameter hemodynamic trends from ICU patients prior to instability and reference time series during stability.

Figure 5-1 includes examples of two channels of physiologic time series (heart rate, systolic blood pressure (SBP)) from 88 ICU patient episodes prior to hemodynamic deterioration and reference time series during intervals of stability. The trends are population averaged (with means and standard errors) and demonstrate that there is significant variability from patient to patient. However, data mining may reveal that there are “signatures” in multidimensional physiologic space that may be predictive of hemodynamic deterioration. The identification of similar signatures shared by patients with similar pathophysiologies is a major goal in our ICU clinical decision support research.

In this chapter, we introduce a new temporal similarity metric based on transformation of time series data into an intuitive symbolic representation. This symbolic transform allows us to model a temporal record in a manner similar to popular information-retrieval (IR) models of documents or web pages. The classical IR algorithms utilize a high-dimensional vector-space model in which each element of the vector represents the number of occurrences of a given word in the document [64]. Thus, the structure of the document is characterized by this “term frequency vector” (TFV). In order to transform time-series into a collection of “words” or symbols, a transform is needed that compactly represents the salient characteristics of single and multiparameter time series. We demonstrate that a wavelet-based representation of temporal records offers an intuitively appealing and computationally efficient solution.

The chapter is organized in the following manner: in the next section, we provide a brief overview of some popular methods used to assess the similarity between time series. Then, we describe the methodology for implementing our new trend similarity algorithm. We assess our algorithm’s performance using multidimensional time series from publicly available and standardized time series databases. We also assessed its performance in classifying physiologic time series generated from a computational model of the cardiovascular system.

5.1.1. Review of time series similarity algorithms

Keogh et al [33] provide an excellent survey of methods developed for the retrieval of similar time series. Previous algorithms can be grouped into time-domain methods and transform-based methods. The simplest time-domain algorithm for computing a similarity metric between time series is the Euclidean distance between two discrete time series $x[n]$ and $y[n]$ where the distance between the two series is defined as:

Equation 3

$$D(x, y) = \sqrt{\sum_{n=0}^M (x[n] - y[n])^2}$$

While the Euclidean distance metric is rather simple, its shortcomings exemplify the challenges in developing more robust time series similarity metrics. The Euclidean distance metric assumes that discrete time series in a database have the same length and are uniformly sampled from their original continuous time processes. Euclidean distance algorithms in particular, and most time-series similarity metrics in general assume that signals are aligned so that “similar” signals will have similar dynamics at the same points in time. For example, Figure 5-2 demonstrates two sinusoidal signals that have similar amplitudes with slightly different frequencies and phase shifts. The calculated point-to-point error is significant and would result in a large value for a distance metric between the two signals. To overcome these constraints, modifications to the Euclidean distance metric have been utilized based on the principle of time-warping where signals are “stretched” or “compressed” so that pertinent features are aligned in time[33]. However, such signal processing methods may significantly change the unique characteristics of a signal and require careful tuning parameters. Windowing and segmentation techniques have been developed to divide a signal into a set of sub-sequences which allows greater flexibility in matching time series by using shifting operations [25].

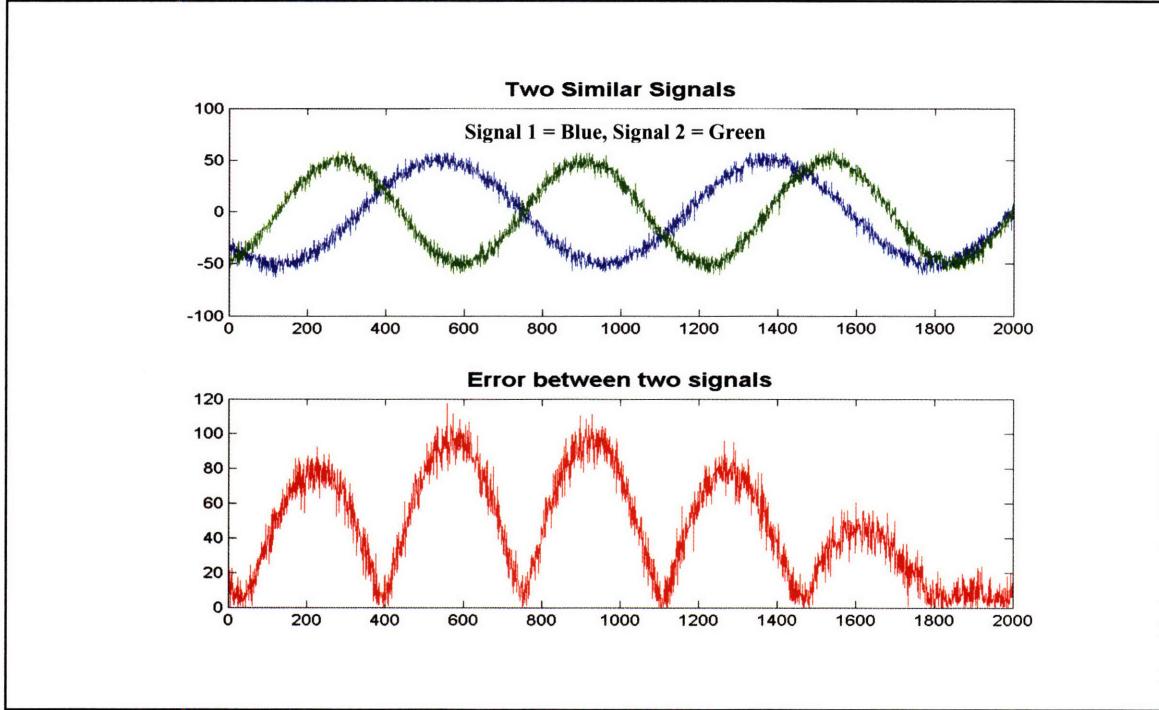


Figure 5-2: Example of point-to-point error between two “similar” time series for the Euclidean distance metric.

Transform-based techniques have been applied to the time series similarity metrics to overcome the limitations of Euclidean and time-warping methods that are restricted to the time-domain. The transform-based techniques project time series of interest onto a set of functions such as sinusoids or principal components [25]. The data transformation reduces the dimensionality of the original times series and facilitates the use of machine learning techniques in matching similar time series.

Fourier-based transform techniques are commonly utilized to represent a time series. However, transformations of signals from the time domain to the transform (Fourier) domain may at times be non-intuitive. For example, in Figure 5-3, two different “ramp” function time series are presented with their respective Fourier Transforms. Only the magnitude of the Fourier transforms is included (the Fourier transform has complex representations). Because the phase information is excluded, the two different ramp functions appear to have very similar magnitude values for their respective Fourier coefficients. Thus, this simple example demonstrates that phase information cannot be neglected in utilizing the Fourier transform to represent time series. An attractive feature

of a time series similarity algorithm is to identify why two or more signals are similar to one another upon the request of a user. However, the phase information of a signal is not an intuitive and simple concept to convey. Thus, the requirements for needing both magnitude and phase components of a Fourier transform would be an unappealing method for querying “similar” time series.

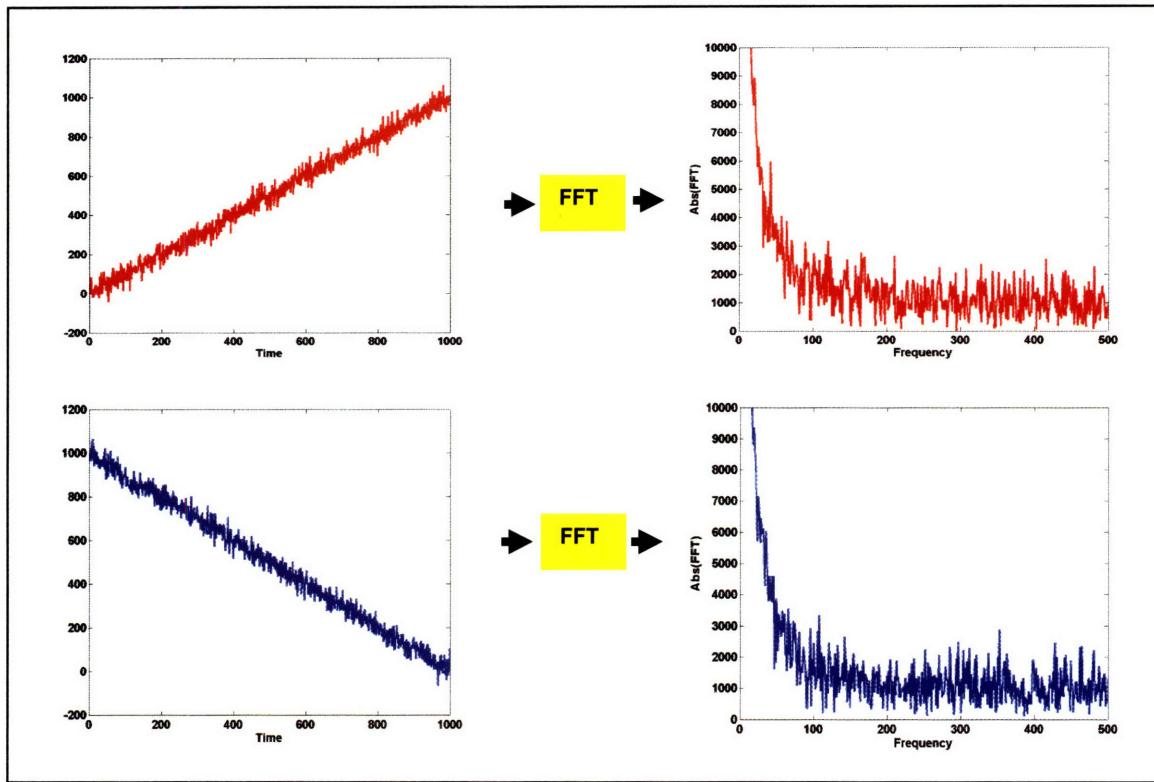


Figure 5-3: Example of FFT-based transformation of two different ramp time series

While an improvement over time-domain techniques, transform-based similarity metrics are still an active area of research. In particular, computationally efficient algorithms for real-time (online) applications are sought that can identify similar multidimensional time series from large databases.

In reviewing the literature, we will now focus on algorithms that are based upon “symbolic” representations of time series. The SAX (Symbolic Aggregate Approximation) method introduced by Lin and Keogh [44] is among the most popular symbolic methods utilized for time series similarity searches. The SAX method segments

a time series into equal length segments with piece-wise constant approximations, and then assigns a symbol (or letter) to each segment. The symbolic transformation of each segment is based upon a probabilistic mapping of numeric data to symbols. Thus, a time series is transformed into a string of letters (or symbols). Then, time series are matched by assessing the differences of the symbols from equal-length strings. An attractive feature of the SAX algorithm is that it allows for significant dimensionality reduction. SAX also provides for a lower bound that guarantees no false-dismissals of a time series within a certain Euclidean distance from the query signal. While there may be a mathematical elegance to providing guarantees of no false dismissals in a Euclidean sense, the previously described limitations of Euclidean distance metrics may limit the practical importance of such a guarantee.

Another recent method that has been published during the course of the present thesis work is the MVQ (Multiresolution Vector Quantized) approximation algorithm of Megalooikonomou et al [47]. The MVQ method segments a time series at several resolutions (or scales). Instead of fitting piecewise constant approximations to each segment, the MVQ technique finds optimal vectors using the traditional Generalized Lloyd Algorithm (GLA) and generates a codebook. The vector codes then become the symbolic representation of the time series. In [47], the authors have shown that the MVQ technique has superior performance in comparison to other state-of-the-art time series retrieval algorithms using standard time series test databases. One of the major limitations of the MVQ technique is that it requires training to generate the optimal vector codebook and has several parameters that need to be tuned.

5.1.2. Introduction of novel wavelet-based similarity metric

We introduce a new temporal similarity metric based on transformation of time series data into an intuitive symbolic representation. This symbolic transform allows us to model a temporal record in a manner similar to popular information-retrieval (IR) models of documents or web pages. The classical IR algorithms utilize a high-dimensional vector-space model in which each element of the vector represents the number of

occurrences of a given word in the document [60]. Thus, the structure of the document is characterized by this “term frequency vector” (TFV). In order to transform time-series into a collection of “words” or symbols, a transform is needed that compactly represents the salient characteristics of single and multiparameter time series. We demonstrate that a wavelet-based representation of temporal records offers an intuitively appealing and computationally efficient solution.

In the next section, we briefly review the vector space model that is commonly implemented in information retrieval systems such as document indexing and web-search engines [64].

5.1.3. Review of classical information retrieval model

In this section, we describe the classic information retrieval (IR) model and define terms that are frequently used in describing the novel extensions and applications of the IR model to time series signals in general, and physiologic time series in particular. In this thesis, a “record” is used to refer to a set of words or numerical values that collectively characterize a particular entity in a database. For example, a collection of words in a specific sequence can be members that uniquely characterize a particular record such as a document or a web-page. The collection of records can consist of a library of documents or the internet (collection of web-pages). A sequence of observations (samples) from a multidimensional time-varying process can form a record that represents the original time-varying process. A record can also consist of heterogeneous members such as numbers and words. For example, the MIMIC-II database can be seen as a collection of ICU patient records. Each record consists of several different physiologic time series and text-based entries such as nursing progress notes and discharge summaries.

In order to identify and retrieve “similar” records in a database, the original raw record must be converted into a representation that characterizes the salient features of that respective record. The classical text-based information retrieval (IR) systems represent records using the vector-space model. Using the method of Salton et al, unique words are

chosen from a document to represent the various dimensions of the vector [64]. Thus, the value of a specific element in the vector is equal to the frequency of occurrence of the word in the document. Figure 5-4 includes an example of how a simple document may be transformed into a vector. Another important observation concerning the Salton model is that the original ordering of the terms in a document is no longer preserved in the term-frequency vector.

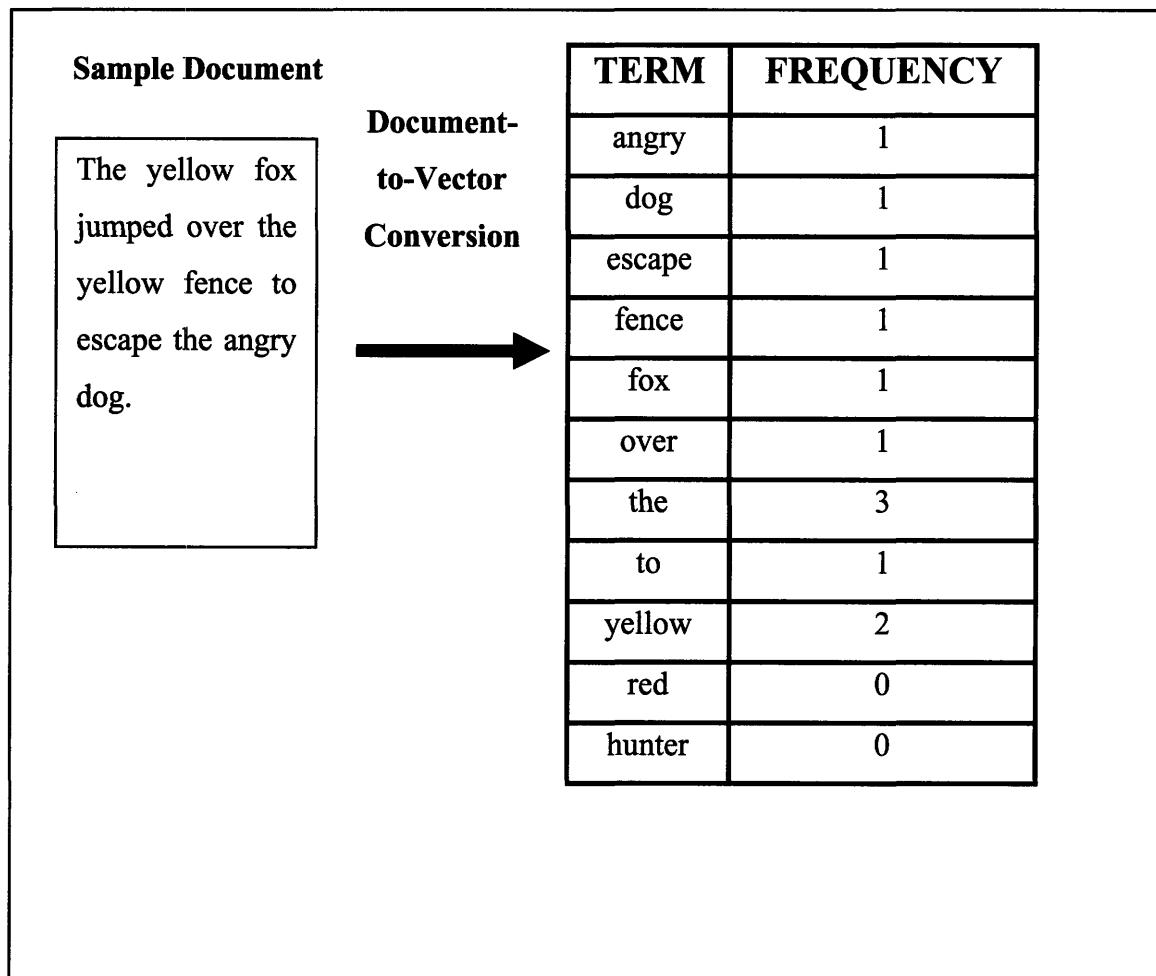


Figure 5-4: Example of document-to-vector conversion

As is demonstrated in Figure 5-4, the vector-space model transforms text-based records (documents or web-pages) into mathematical representations. This transformation allows for the utilization of classical pattern recognition approaches used in statistics for various detection problems. The various dimensions of the feature vector need not be weighted

equally. For example, words that may appear frequently in text-based records such as the words ‘the’ or ‘to’ are generally not helpful in characterizing the important properties of a record, and thus, receive less weight in the overall vector weighting scheme. In fact, uninformative words can be eliminated from an overall feature as a means of dimensionality reduction. A study of captioned words on British television concluded that approximately 250 words accounted for 68% of all the spoken words [29]. A popular term weighting method to account for the significant difference in the prior probability of a word being utilized in a given sentence is referred to as the term frequency, inverse document frequency method (TF*IDF) [60]. A common variant of the TF*IDF algorithm is described in Equation 4. The IDF weight of a term f_i , is defined as:

Equation 4

$$IDF(f_i) = \log \frac{N}{f_{i,n}}$$

where N is equal to the number of records in the database, and $f_{i,n}$ is equal to the number of records in the database that have at least one occurrence of the term f_i . Thus, the IDF weight tends to de-emphasize terms in a vector that occur frequently across many records in a database (where $f_{i,n}$ tends to be large). The final value, $W(f_i)$, of each element in the term frequency vector for a given record is defined as:

Equation 5

$$W(f_i) = TF(f_i) * IDF(f_i)$$

where $TF(f_i)$ is the number of occurrences of term f_i in a given record.

1. New Alligator exhibit at Boston Zoo to open this fall.
2. Alligator swallows Boston jogger near Charles River.
3. New gourmet restaurant on Charles to serve Alligator on menu.
4. Exotic menu to include Alligator at the Cajun Restaurant.
5. Stone Park Zoo will feature Alligator display

Figure 5-5: Example of identifying "similar" sentences

Table 5-1: Feature vectors used to create feature matrix of individual sentences (or records).

Term	Sentence 1 (Number of occurrences in sentence)	Sentence 2 (Number of occurrences in sentence)	Sentence 3 (Number of occurrences in sentence)	Sentence 4 (Number of occurrences in sentence)	Sentence 5 (Number of occurrences in sentence)
Alligator	1	1	1	1	1
Restaurant	0	0	1	1	0
Jogger	0	1	0	0	0
Swallows	0	1	0	0	0
Menu	0	0	1	1	0
Zoo	1	0	0	0	1
Gourmet	0	0	1	0	0
Boston	1	1	0	0	0
Charles	0	1	1	0	0
River	0	1	0	0	0
Cajun	0	0	0	1	0

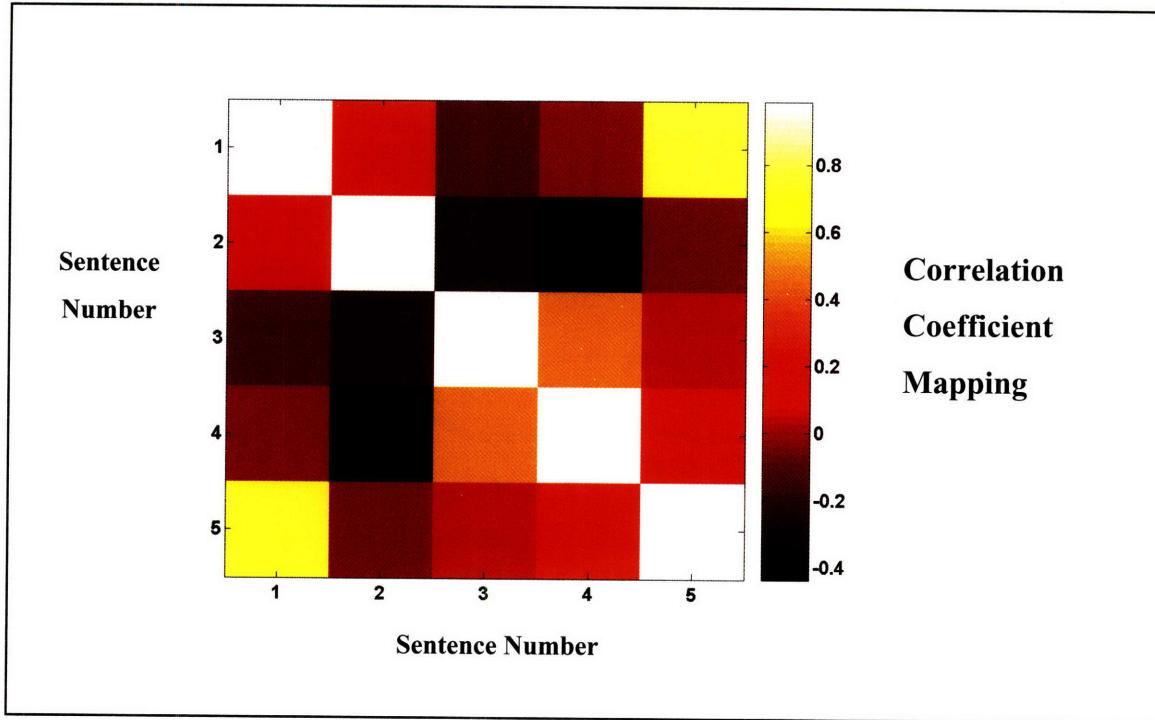


Figure 5-6: Sentence similarity matrix displayed using color mapping of correlation coefficients

The example presented in Figure 5-5, Table 5-1, Figure 5-6 for matching similar sentences also highlights several challenges that are nontrivial. For example, in the simple term frequency vector model, each dimension or “term” is modeled to be independent from other dimensions or terms. However, certain terms are related and convey information about similar concepts. For example, the words “gourmet” and “menu” both convey information that would suggest that the record may be referring to concepts related to food or restaurants. However, the term frequency vector would consider such elements to be independent in a probabilistic model. There are many other challenges to constructing optimal term feature vectors that include language modeling and are beyond the scope of this thesis. The interested reader is referred to [64] for more detailed reviews.

5.1.4. Detection of similar time series as an information retrieval problem

While IR-techniques are popular for indexing records consisting of collections of words (documents), very little research exists on applying the IR model to records of multidimensional time series. A principal challenge confronting researchers is developing an adequate mapping algorithm that can transform time-series into a set of “words” or symbols. The symbols that are derived should adequately describe the structure and “uniqueness” of a time-series. The use of wavelets to characterize the multi-resolution (multiscale) structure of a time-varying signal offers an intuitively appealing and computationally efficient algorithm for symbol generation. There is also a physiological basis to focus on multiscale transforms. Physiologic control systems include short-term (within seconds) and long-term (within hours or days) mechanisms for modulating the physiologic state of a patient. In the next section, we describe the methodology used to derive a new wavelet-based symbolic transform and similarity metric.

5.1.5. Multiscale dynamics characterization of time series: wavelet-based approach

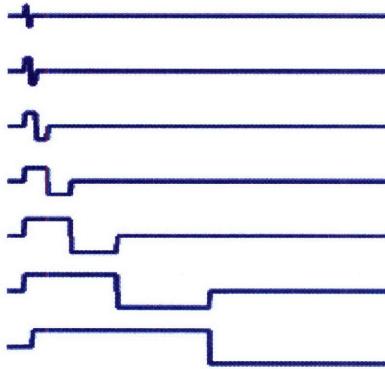
Wavelets have become increasingly important in areas of signal processing such as data compression, signal de-noising, and feature extraction in pattern recognition [45]. Wavelets are basis functions that can be used to decompose time-varying signals into terms of averages and differences at several different time scales. Several researchers have discussed many of the attractive properties of wavelets [71]. Wavelets have some properties comparable to other popular transform techniques like Fourier analysis. However, wavelets are localized in time, whereas Fourier coefficients represent signal energy components defined over a signal’s entire support (time-span). In the present research, we take advantage of the time localization property of wavelets to develop a novel time series similarity metric. Next, we briefly review the implementation scheme used for computing the wavelet transform of a signal.

The continuous wavelet transform (CWT) for a continuous signal $f(t)$ is defined as:

Equation 6

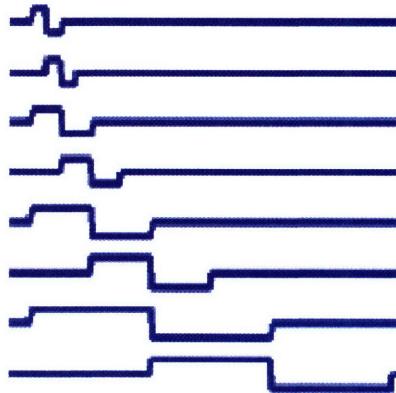
$$CWT(f, j, b) = j^{-1/2} \int_{-\infty}^{\infty} f(t) w(2^{-j}t - b) dt$$

where j is a scale parameter and b is a shifting parameter. In Equation 6, the wavelet basis function is characterized by $w(t)$. The CWT is calculated by continuously correlating shifted and scaled versions of $w(t)$ with $f(t)$. The CWT of $f(t)$ transforms a one-dimensional signal into two dimensions indexed by j and b . A wavelet coefficient characterizes the signal dynamics at a certain scale j and a certain location in time indicated by b . Coefficients at coarse scales (where j is large) characterize global dynamics of a signal, whereas finer scale (where j is small) characterize finer detail dynamics of a signal. The CWT results in highly redundant information in coefficients when b and j are varied in small steps. In practice, signals are discretely sampled and thus, a discrete version of the CWT is often utilized and is referred to as the Discrete Wavelet Transform (DWT). There are several families of wavelet basis functions that could be utilized in the CWT or DWT. Wavelet families are chosen based upon desired properties such as smoothness, computational complexity, and orthogonality. Numerous reviews in the literature provide excellent overviews of the properties of wavelets and can be consulted [53]. The Haar wavelet is perhaps the simplest wavelet to implement and can be represented by successive local-difference and local-averaging operations. As can be seen in Figure 5-7, the Haar wavelets at different dyadic scales are orthogonal to one another, and can be used to form a complete basis in discrete space. Thus, a discrete signal, $x[n]$, can be perfectly reconstructed by applying an “inverse” discrete wavelet transform on the wavelet coefficients. For example, a signal $x[n]$, that is defined over 16 samples in duration requires only 14 detail wavelet coefficients and 2 approximation coefficients to be perfectly reconstructed.



Harr Basis Functions:

By projecting a signal, $x[n]$, on each wavelet basis function, the resulting coefficient characterizes the signal dynamics at the respective wavelet scale. Thus, the coarser wavelet scales can be used to represent changes in a time-series occurring over longer time intervals of several hours, whereas the finer scales can be used to represent changes over a few minutes. Haar functions can be chosen based on their scale and shift properties to be orthogonal to one another.



Over-complete Wavelet Basis

Function: Wavelets are shifted so that there are two basis functions of the same support (where values are non-zero), but are non-orthogonal to one another. In practice, over-complete wavelet representations are implemented by over-sampling operations using a filter bank as illustrated in Figure 5-10.

Figure 5-7: Example of Haar wavelet basis functions

5.1.6. Review of wavelet convolutions

Discrete Haar wavelet basis functions, $w_j[n]$ and $w_{j+1}[n]$ (as illustrated in Figure 5-7) can be utilized in the DWT. As we change j , the wavelet basis functions can be dilated or contracted as illustrated in Figure 5-7. The signal dynamics at scale j can be characterized for any signal $f[n]$ by convolving $f[n]$ with $w_j[n]$ to yield an over-sampled set of wavelet coefficients, $h_{j*}[n]$:

Equation 7

$$h_{j*}[n] = w_j[n] * f[n]$$

However, $h_{j*}[n]$, still represents an over-sampled (redundant) version of the wavelet coefficients of scale j . In practice, $h_{j*}[n]$, where $w_j[n]$ has support of $2M$ points, is decimated by a factor of $2M$, which leads to the wavelet coefficients, $h_j[n]$, of scale j . In the over-complete discrete wavelet transform implemented in this research, the decimation was only by a factor of M or less.

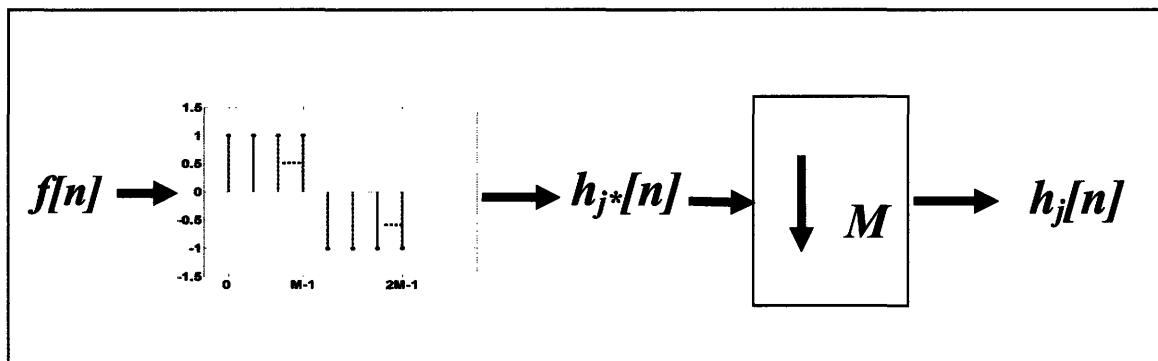


Figure 5-8: Discrete wavelet transform utilizing Haar basis function to calculate wavelet detail coefficients

Along with the Haar wavelet basis function $w_j[n]$, a smoothing operation on the signal is utilized to characterize the global trend of a signal by a moving average filter to yield a smoothed approximation, $a[n]$, of the original signal $f[n]$.

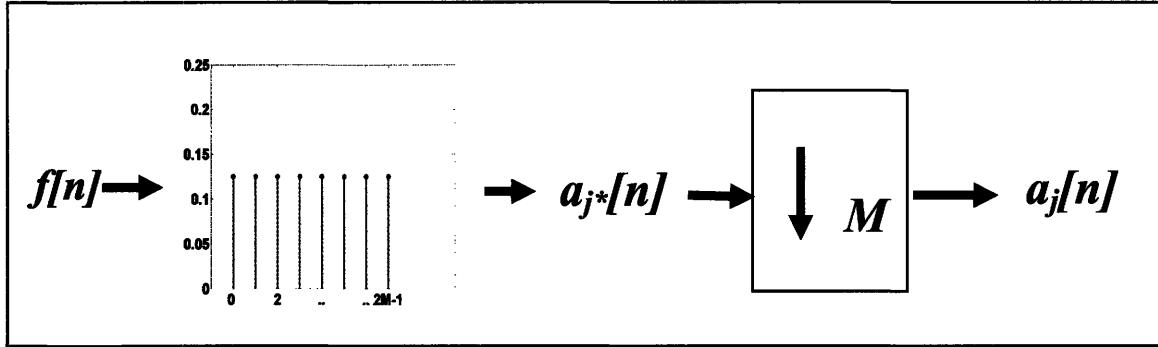


Figure 5-9: Discrete wavelet transform utilizing Haar function to calculate approximation coefficients

In order to obtain the multiscale collection of wavelet coefficients $w_j[n]$ and approximation coefficients, $a_j[n]$, investigators commonly utilize efficient multirate filter bank implementations [62]. In the implementation utilized in this research, a modified filter bank implementation is utilized and is described in Figure 5-10 and Figure 5-11.

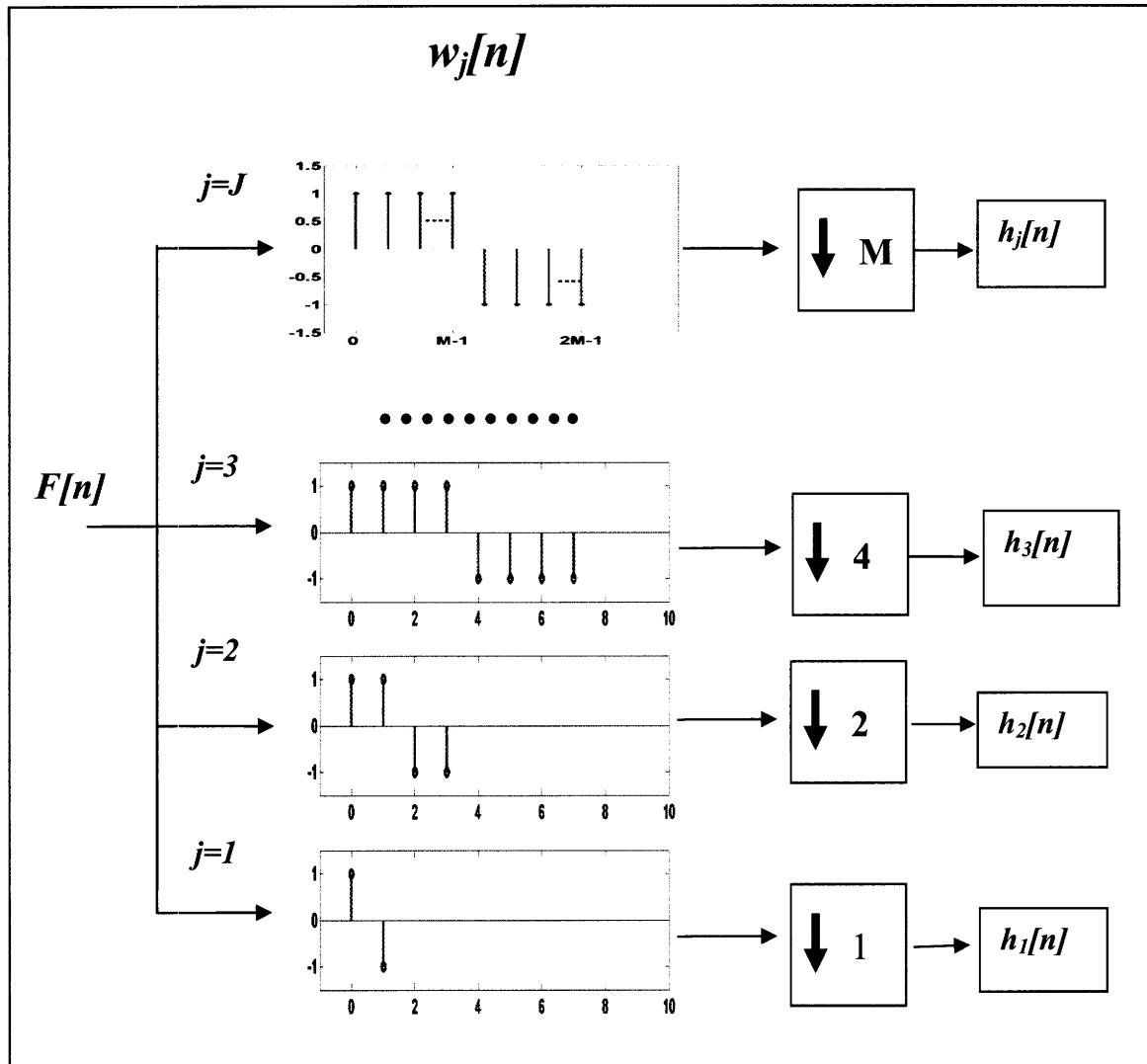


Figure 5-10: Filter bank implementation of DWT to calculate wavelet detail coefficients

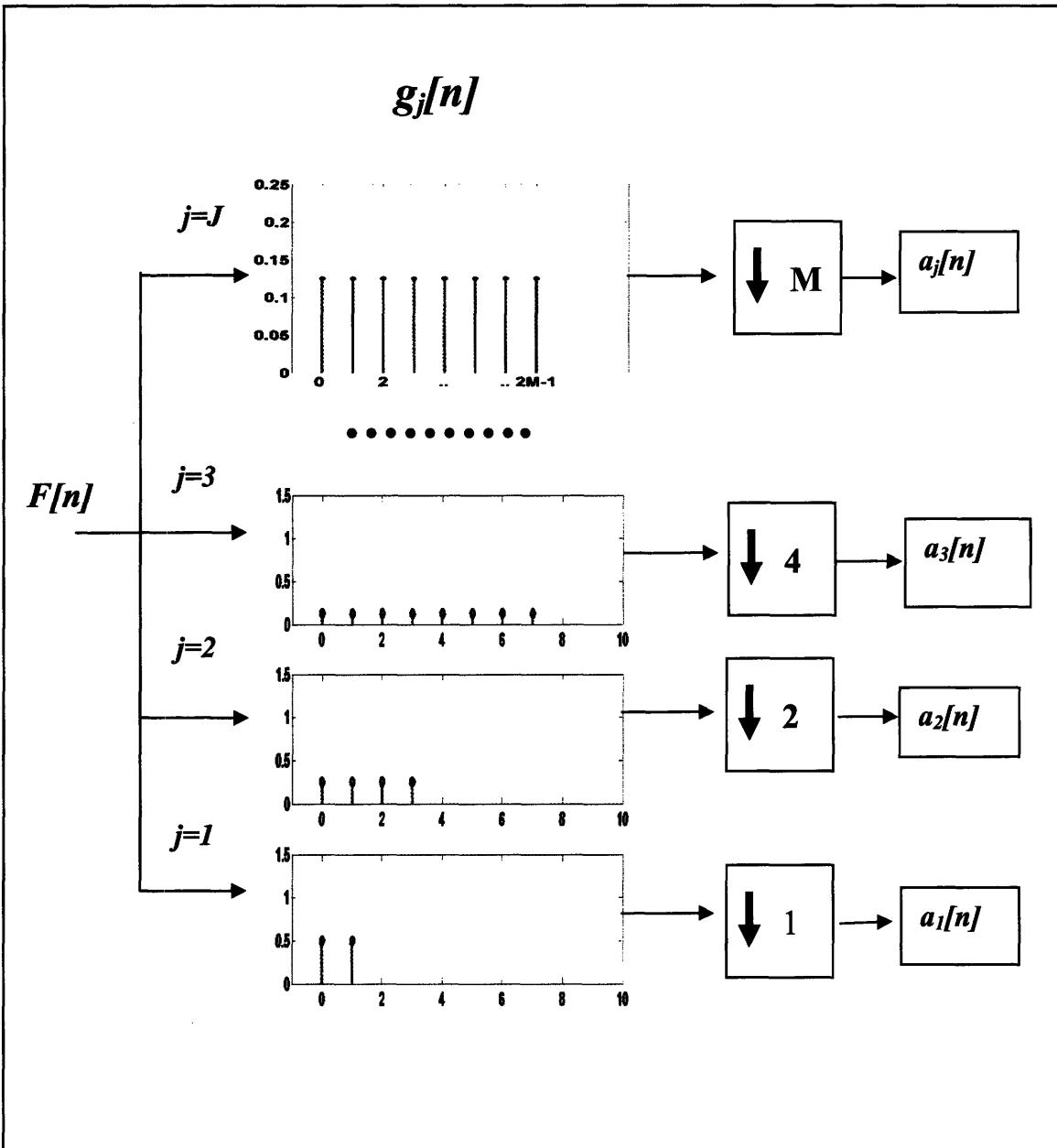


Figure 5-11: Filter bank implementation to calculate wavelet approximation coefficients

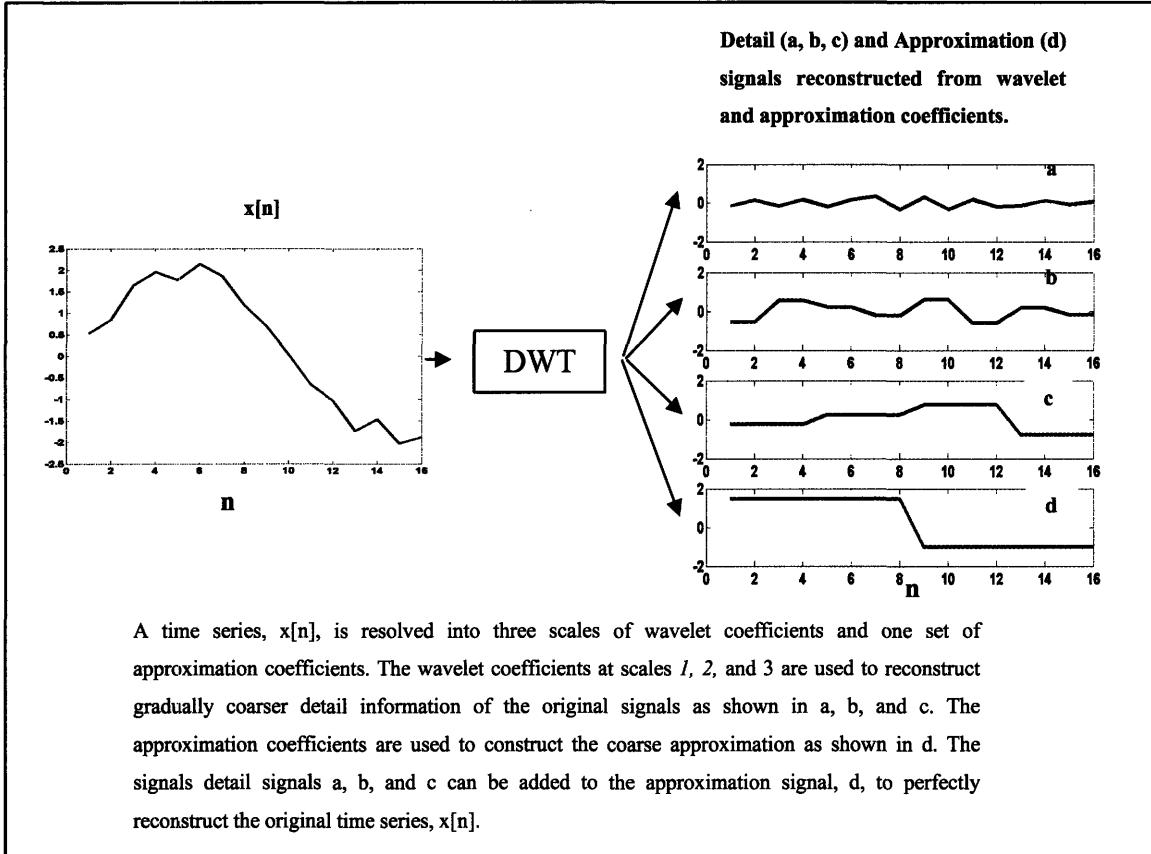


Figure 5-12: Example of multiscale decomposition of a signal with Haar wavelets

The simulated signal in Figure 5-12 illustrates how any signal can be resolved into multiscale components that describe the dynamics as well as the approximation (low-frequency) characteristics using wavelet analysis. For example, the three detail signals from the wavelet coefficients have been reconstructed at scales 1, 2, and 3. As can be seen, each scale has successively coarser detail information extracted from the original signal, $x[n]$. The approximation coefficients are used to reconstruct the crudest approximation of the signal without the wavelet coefficients. The three detail (wavelet) trends and one approximation trend can be added to perfectly reconstruct the original signal. The time domain signal, $x[n]$, consists of 16 samples. The wavelet information at scales 1, 2, and 3 consist of 8, 4, and 2 coefficients respectively. The approximation information consists of 2 coefficients. Thus, the total collection of wavelet and approximation coefficients consists of 16 points as well. The detail and approximation coefficients can be used to reconstruct the different detail and approximation information of the original signal, $x[n]$.

We have chosen to utilize an over-complete form of the Haar wavelet transform which does not have the orthogonality property and is a slight modification of the framework proposed in [55]. An over-complete expansion allows us to increase the expressive power of our feature vector while retaining the computational simplicity of the DWT. In practice, this is accomplished by simply over-sampling the wavelet coefficients. Thus, it will result in a greater number of coefficients than the conventional DWT. If there is energy in a signal which falls at the edge of a wavelet defined at scale j with compact support of length $2M$, the over-complete implementation would better characterize the temporal dynamics because the signal of interest would be projected onto the next wavelet which would have the same compact support $2M$ but be shifted by only $2M-1$ points. In the classical DWT, the next wavelet on the same scale would be shifted by $2M$ and thus potentially “miss” edges in a signal at a certain scale. In this thesis, an over-complete discrete wavelet transform has been utilized by reduced decimation factors (2^k), where $k < m$, for each scale j .

5.1.7. Methodology: Wavelet symbol generation

In order to transform a time series into features that can be utilized in classic information-retrieval vector space models, the wavelet coefficients at each scale are quantized into discrete symbols as illustrated in Figure 3. The quantization method that was chosen is the simple uniform quantizer. The number of symbols for each scale is a tunable parameter. We now define the methodology used to create the wavelet symbol vectors.

Time series in general and physiologic time series in particular, can be treated as continuous time varying processes. To simplify computation and handling of such signals, the amplitudes of time series are quantized to varying resolution. Quantization is advantageous because it restricts the possible values a signal can assume. Ideally, the quantization should not result in the loss of physiologically meaningful information. For

example, if a mean arterial blood pressure signal was quantized to either 0 mmHg or 100 mmHg, then significant physiologic information would be lost in the quantization process. Different physiologic signals require different quantization schemes because of the differences in their dynamic ranges. For example, arterial blood pH can only have physiologic ranges between 6.5 and 7.7, whereas a patient's heart rate may vary from 0 to 300 beats per minute. The quantization schemes for these two different signals must accommodate the difference in statistics between these signals. The aforementioned examples of quantization are referring to the actual time-domain value of the physiologic signals. However, the same arguments would apply to the respective wavelet coefficients of such signals in the wavelet domain. We now describe the quantization technique we utilize for generating quantized wavelet coefficients of time series.

A collection of time series $x_i[n]$ are resolved into their respective wavelet coefficients at several different scales. The time series $x_i[n]$, are assumed to be of a similar class. For example, $x_i[n]$ may represent a set of arterial blood pressure time series from several different patients. The assumption that time series are representative of similar dynamics is important within the domain of physiologic signal in particular because different signals may have vastly different characteristic dynamics as shown in the aforementioned example of arterial pH and arterial blood pressure.

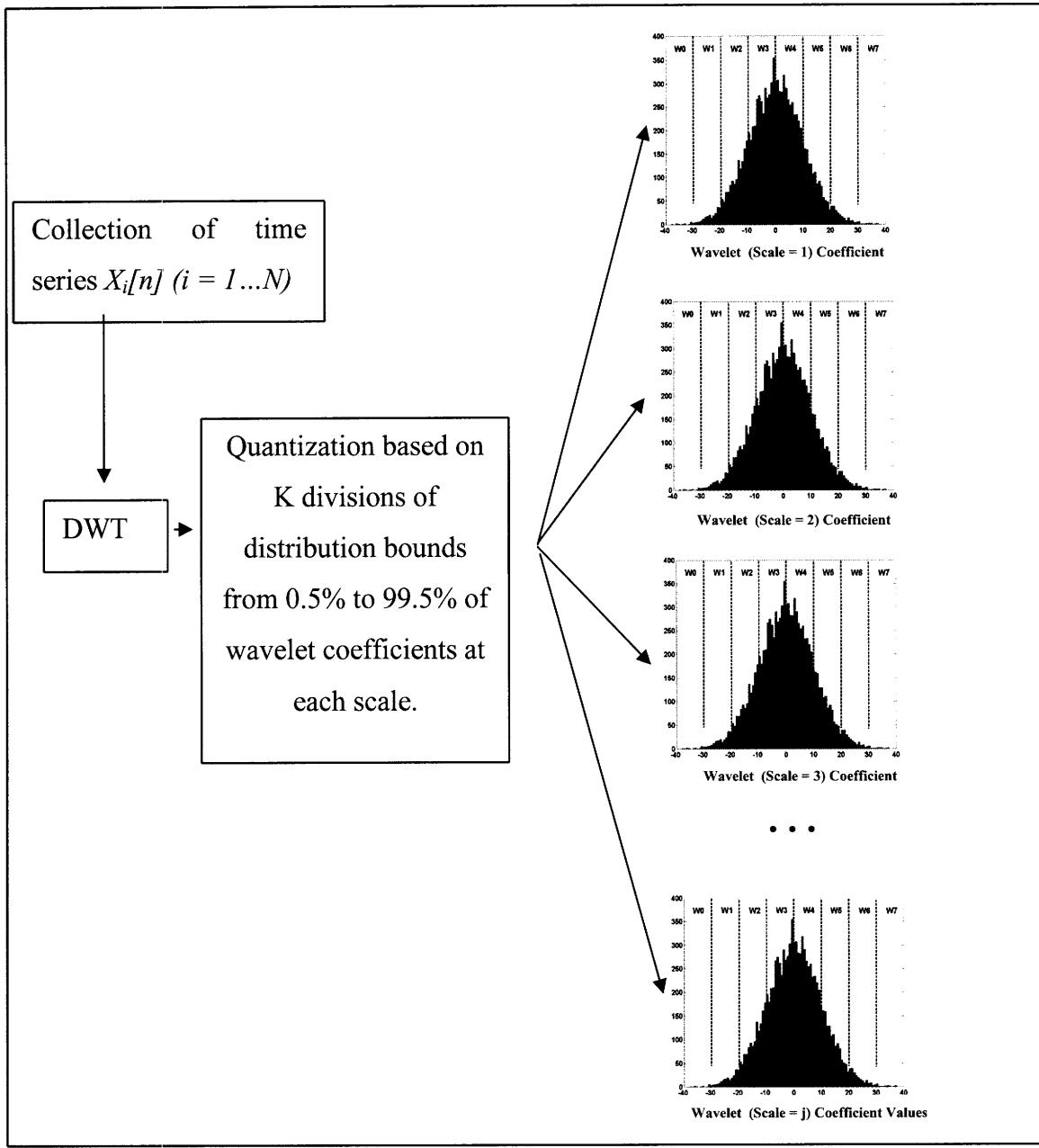


Figure 5-13: Scheme for generation of quantized wavelet coefficients

The schematic presented in Figure 5-13 demonstrates the overall wavelet coefficient quantization process. We model a wavelet coefficient probability distribution at each scale j and find the minimum and maximum wavelet coefficient values that bound the wavelet coefficient distribution from the lower 0.5% to the upper 99.5% values. The advantage of identifying these bounds in a wavelet coefficient distribution is that such a

process accommodates different signals that may have different dynamic ranges (physiologic bounds) while rejecting outliers within a class.

The range of values between these bounds is uniformly split in K divisions. K is chosen based upon a desired level of granularity of the wavelet coefficients and is a compromise between data compression and acceptable signal loss. Each division is then assigned a unique ‘symbol’, w_{jk} (where j is the scale of the wavelet and k is one of K divisions in the distribution). Thus, any wavelet coefficient that falls within one of these divisions is represented by its respective symbol. Thus, a time-domain signal will be represented by a collection of “symbols” or “terms.” This transformation allows us to apply the information retrieval (IR) models that have been previously described for text-based records to time series.

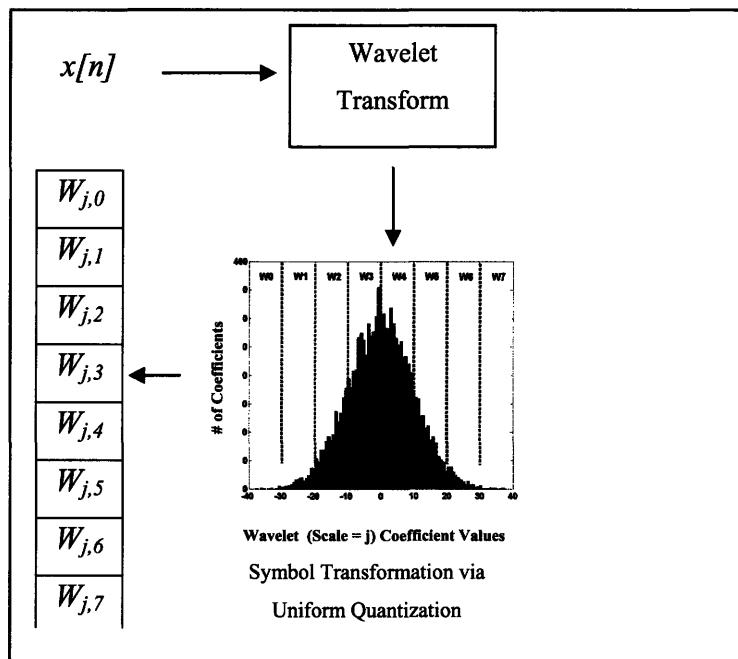


Figure 5-14: Wavelet symbols created by quantization

To generate the high-dimensional wavelet symbol vector for a given time series, the elements of the feature vector are assigned values equaling the number of wavelet coefficients (at the respective scale, j) that are quantized to the corresponding wavelet

symbol of those elements. Thus, the wavelet symbol vector is similar to a histogram of the frequency of wavelet symbols over all the scales. The wavelet symbol vectors of each scale, j , are concatenated to create one final high-dimensional vector.

5.1.8. Methodology: Histogram smoothing using Parzen density estimators

A histogram is used as a quantitative characterization of the distribution of values sampled from a set of measurements [13]. As shown in Equation 8, a histogram is simply created by finding the number of samples from the total collection of samples of coefficients $w_{j,i}(i=1..N)$ at a certain scale, j , that take on a certain value for all possible values y .

Equation 8

$$hist(y) = \sum_{i=1}^N \delta(w_{j,i} - y)$$

where $\delta(y) = 1$ when $y = 0$, else $\delta(y) = 0$ for all other y . If the samples are assumed to be from a random variable, then as the number of samples approaches infinity, the normalized histogram will approximate the probability density function (pdf), or probability mass function (pmf) in the discrete random variable case. However, in practice, the number of samples is limited and other techniques can be applied to estimate the pdf of a random variable.

The parzen density estimator is one of the most popular nonparametric techniques of estimating a pdf. The reader can consult [13] for a more in-depth review of nonparametric pdf estimation techniques. In the present application of wavelet-based symbolic representations, this feature of parzen density estimators is particularly useful. As shown in Figure 5-14, the wavelet symbol bins are organized so that symbols that are assigned to neighboring bins in the feature vector are more closely related to one another than are symbols that are mapped to distant bins. The attractive feature of the parzen

density estimator is that it can be used to effectively “smooth” a conventional histogram by weighting the values of neighboring bins of a histogram (see Figure 5-15). The similarity or dissimilarity between individual symbols is a significant difference between the time series similarity model we are developing and the traditional text-based information retrieval models. The text-based retrieval model presented earlier in the chapter assumed that the words in the limited feature vector were independent and their individual similarities were not modeled.

Equation 9

$$p_N(y; \sigma, j) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\sigma} K\left(\frac{y - w_{j,i}}{\sigma}\right)$$

Equation 10

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$$

We utilized a parzen density estimator with a Gaussian smoothing function as defined by Equation 9 and Equation 10. The variance, σ , of the Gaussian function is a tunable parameter. In practice we used a value of $\sigma = 0.5$. Relatively small changes (0.125 to 1.0) in the variance parameter did not result in significant changes to overall performance.

Since we are utilizing an over-complete wavelet transform technique (low decimation factor in the filter bank implementation), the wavelet coefficient values already have “smooth” histograms. Thus, the Parzen density estimators did not significantly change the results in the examples used in this thesis. However, if one would use a more highly decimated wavelet transform, the Parzen density estimators would make a greater difference.

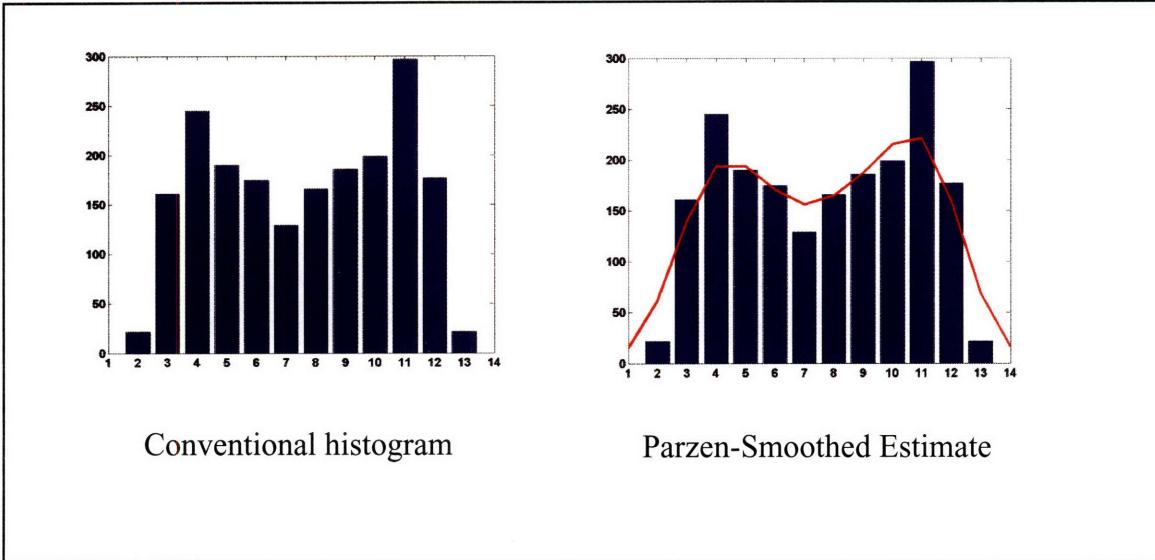


Figure 5-15: Example of Parzen smoothing of a wavelet symbol histogram

Loss of Timing Information of Wavelet Coefficients

As can be seen from Equation 6, a wavelet coefficient has three inherent values that characterize a signal in the transform domain: 1) the scale of the wavelet coefficient, 2) the time value (or shift with respect to the beginning of the time-domain signal), and 3) the actual value of the coefficient. Figure 5-16 includes a collection of N different wavelet symbol feature vectors derived from N different time series. The scales of the wavelet coefficients are still preserved in the symbol vector based on the element location. The value of each wavelet coefficient is subject to a lossy quantization process when the symbols are generated. However, the quantized value still provides for an “approximate” value of the original wavelet coefficient. The number of wavelet coefficients with a specific value and scale is represented by the value of the element, and is color-coded based on an intensity mapping for visualization purposes. However, the relative time or sequence of a wavelet coefficient is discarded in the overall wavelet symbol feature vector. The loss of the timing information in our wavelet symbol feature vector is consistent with the IR model used in text-based record retrieval. The sequence of occurrences of words in a sentence or document is also not utilized, but the retrieval performance was not significantly degraded. We hypothesize that loss of the wavelet

timing information will not compromise the performance of the time-series retrieval algorithm.

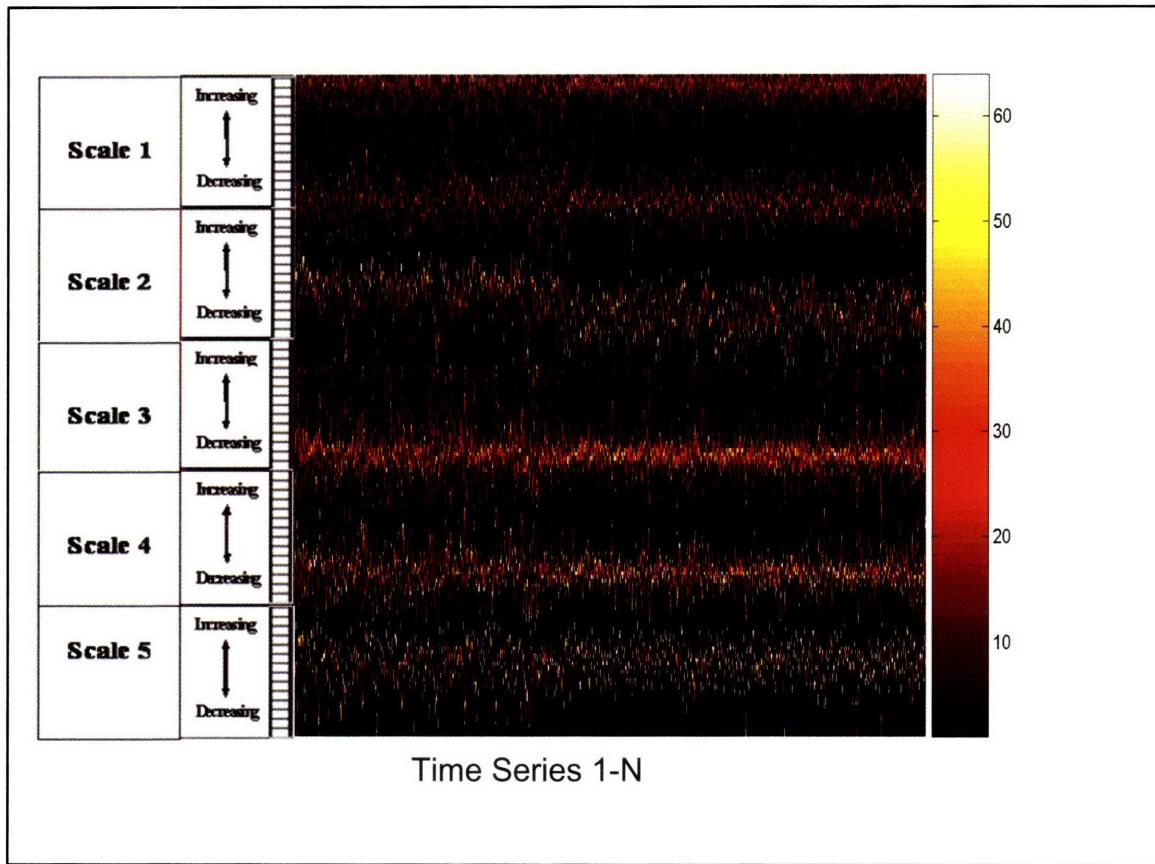


Figure 5-16: Collection of wavelet symbol vectors from different time series

The value of each element is then further weighted by use of a modification of the popular “Inverse Document Frequency” (IDF) weighting scheme [64]. The IDF weight of an element, $w_{j,i}$, is defined as:

Equation 11

$$IDF_{classic}(w_{j,i}) = \log \frac{N}{f(w_{j,i})}$$

where N is equal to the number of records in the time series database, and $f(w_{j,i})$ is equal to the number of records in the database that have at least one wavelet coefficient that is quantized to the symbol, $w_{j,i}$. For example, $w_{j,i}$ may represent the wavelet symbol indicating a systolic blood pressure time series, $x[n]$, has a segment of data where systolic blood pressure decreased by approximately 20 mmHg when averaged over a 4-hour time scale. An intuitively appealing aspect of the IDF weighting scheme is that the wavelet symbols that are less frequently observed in the database of time series are more heavily weighted. Thus, the data-driven IDF weighting scheme favors matching time series records that have similar “rare” temporal dynamics.

The IDF as defined by Equation 11 was developed for text-based record retrieval. It does not consider the intra-record frequency of a symbol. Thus, if a given term, $w_{j,1}$, appears at a rate of 100 times in every record in the database, and another term, $w_{j,2}$, appears 2 times in every record, they would still have the same IDF weight. Experimentally, we found that ignoring the intra-record frequency of symbol for time series retrieval applications led to poor performance. Thus, we developed a modified functional for a weighting scheme that does not ignore the intra-record frequency of a symbol.

Equation 12

$$IDF(w_{j,i}) = \log \frac{N}{f(w_{j,i}) * \sum_{k=1}^N f(w_{j,i,k})}$$

In Equation 12, $f(w_{j,i})$ represents the total number of records in the database that contain at least one occurrence of the $w_{j,i}$ symbol. The function, $f(w_{j,i,k})$ represents the frequency of occurrence of a given wavelet symbol $w_{j,i}$ in a specific record, k , in the database.

The final term frequency vector that includes all the wavelet symbols, $(w_{j,i}..w_{J,I})$, of a time series, $x[n]$, is defined as:

Equation 13

$$TFV(x) = [IDF(w_{j,i}) * w_{j,i} .. IDF(w_{J,I}) * w_{J,I}]$$

The distance between the term frequency vectors of two time series, $x[n]$ and $y[n]$ is calculated by computing the correlation coefficient of the two vectors:

Equation 14

$$D(x, y) = p \langle TFV(x), TFV(y) \rangle$$

5.1.9. Results:

Description of Synthetic Time Series Databases

The algorithms for assessing similarities in time series were evaluated with libraries of synthetic signals. The first library included several different classes of signals: sinusoids, random walks, increasing trends, and decreasing trends. The synthetic signals were created in MATLAB (Mathworks, Natick, MA). The following equations were the characteristic equations used to generate three different classes of signals:

Equation 15

$$y[n] = y[n - 1] + randn[n]$$

Equation 16

$$y[n] = A \sin[2\pi(f_i n + \Phi_i)]$$

Equation 17

$$y[n] = r_i * n$$

Equation 15 describes a random-walk time series. The `randn()` function in MATLAB was utilized to add a Gaussian noise variable to every consecutive sample. Equation 16 is a representative equation for synthetic sinusoidal signals. The frequency, f_i , and phase, Φ_i , of each signal was varied to create several different sinusoids. Equation 17 represents a simple ramp function that represents continuously increasing or decreasing time series based upon the chosen values for r_i . Other classes of signals included “saw-tooth”

patterns. Based upon the above equations, 100 synthetic time series were generated and are referred to as *Library A*. Example time series from Library A are shown in Figure 5-17.

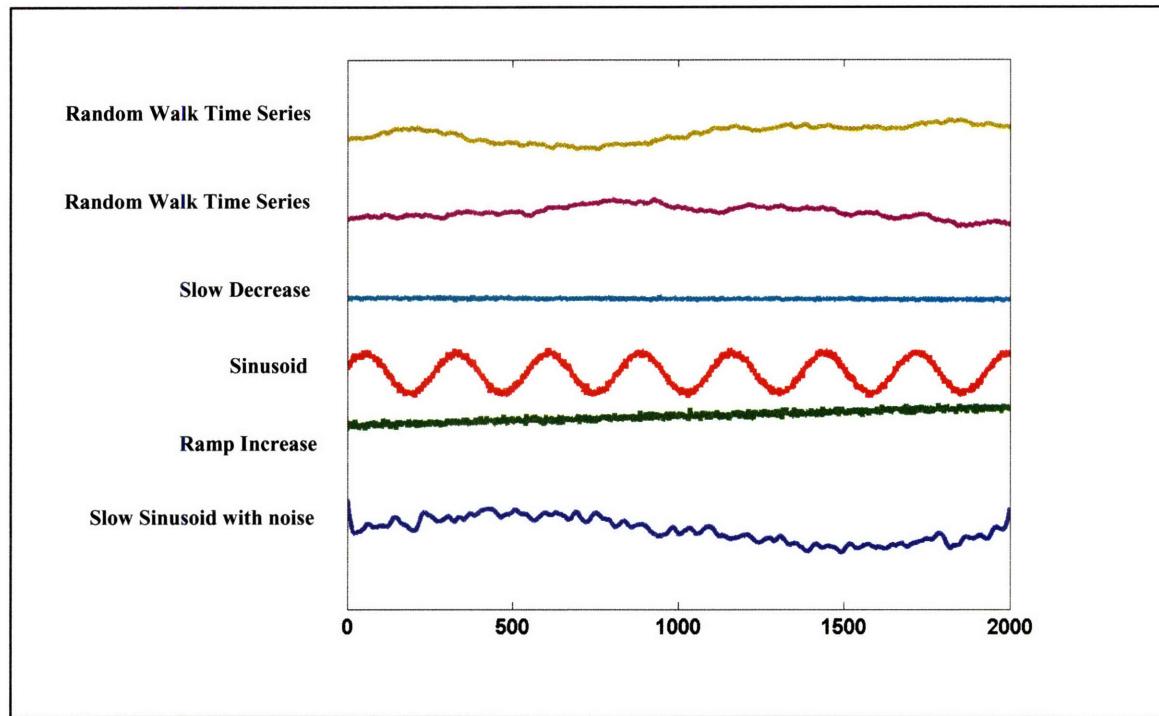


Figure 5-17: Examples of synthetic signals for Library A

The second library, *Library B*, was generated by utilizing the commercially available physiology simulator software, QCP Simulator-Research Edition (Biological Simulators, Jackson, MS). The QCP algorithm is based on detailed physiologic models of several organ systems and their respective feedback mechanisms. A scripting module allowed for the generation of three different physiologic classes including hemorrhage, left-heart failure, and sepsis. The different scenarios were simulated by manipulating various physiologic parameters of the cardiovascular system (see QCP User's Manual). Within each class, various severities of abnormalities were simulated. For example, the simulator was used to generate 12 different severities of hemorrhage by modifying the bleeding rate (see Figure 5-18). Other profiles were generated for acute myocardial infarction by varying the degree of left-heart contractility (see Figure 5-20). Sepsis was difficult to

simulate with the QCP program due to its complex pathophysiology. For example, the capillary permeability which is thought to change in human sepsis, was not accessible in the QCP program. Thus, only the total peripheral resistance (TPR) was decreased by various degrees to simulate the vasodilation that is known to accompany sepsis (see Figure 5-19). By examining the corresponding figures, we see that these different physiologic processes may produce distinct multidimensional hemodynamic patterns. In addition to the 36 simulated physiologic records, 12 additional records were generated using low-pass filtered noisy random time series using the `randn()` function in MATLAB and a low-pass filter. Thus, *Library B* consisted of a total of 80 records.

In order to simulate the inherent noise that is found in most physiologic signals, the simulated signals in *Libraries A* and *B* were corrupted with additive Gaussian noise of varying levels by utilizing the `randn()` function in the MATLAB environment. The variance of the Gaussian noise was modified in order to study the performance of the time series retrieval algorithm under different noise stress levels.

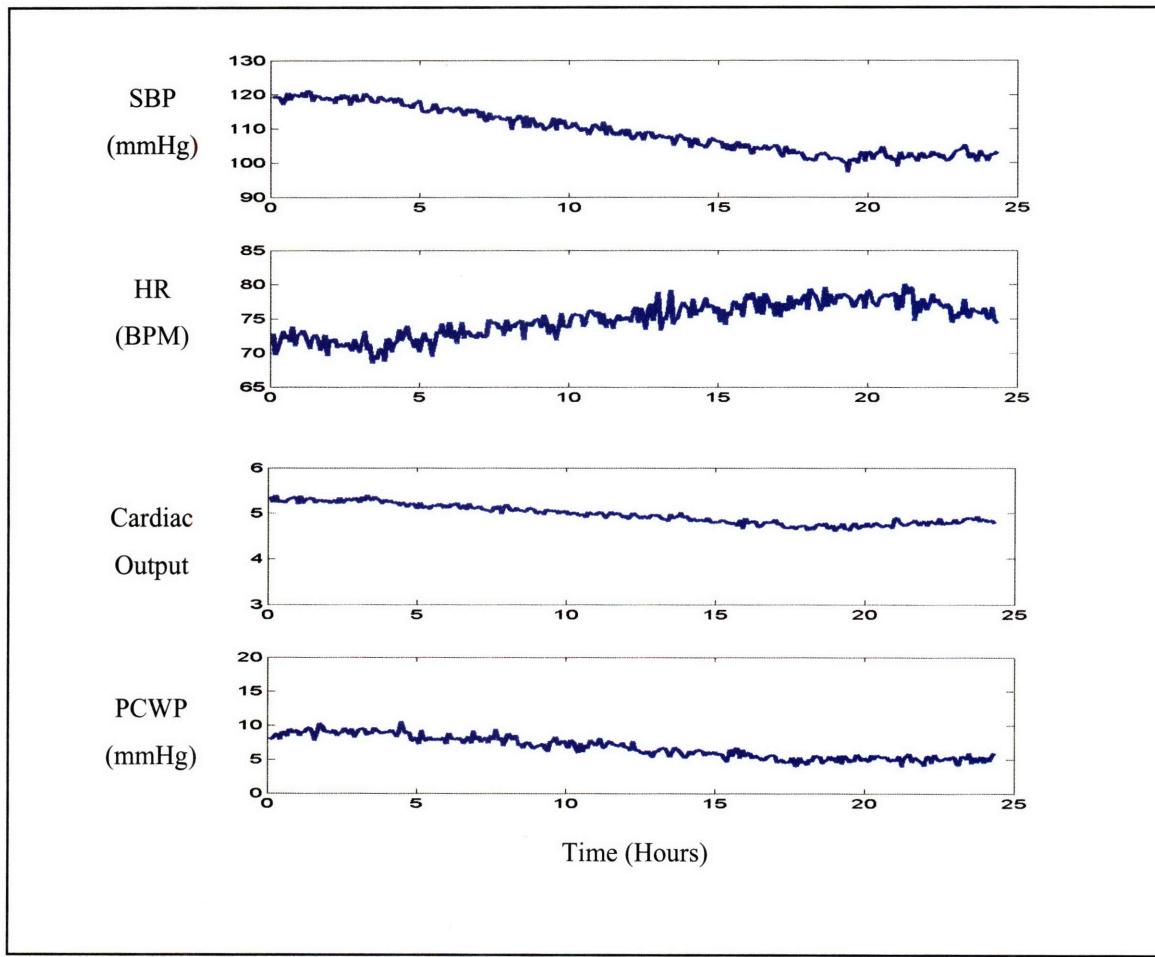


Figure 5-18: Example from Library B: simulation of hemorrhage

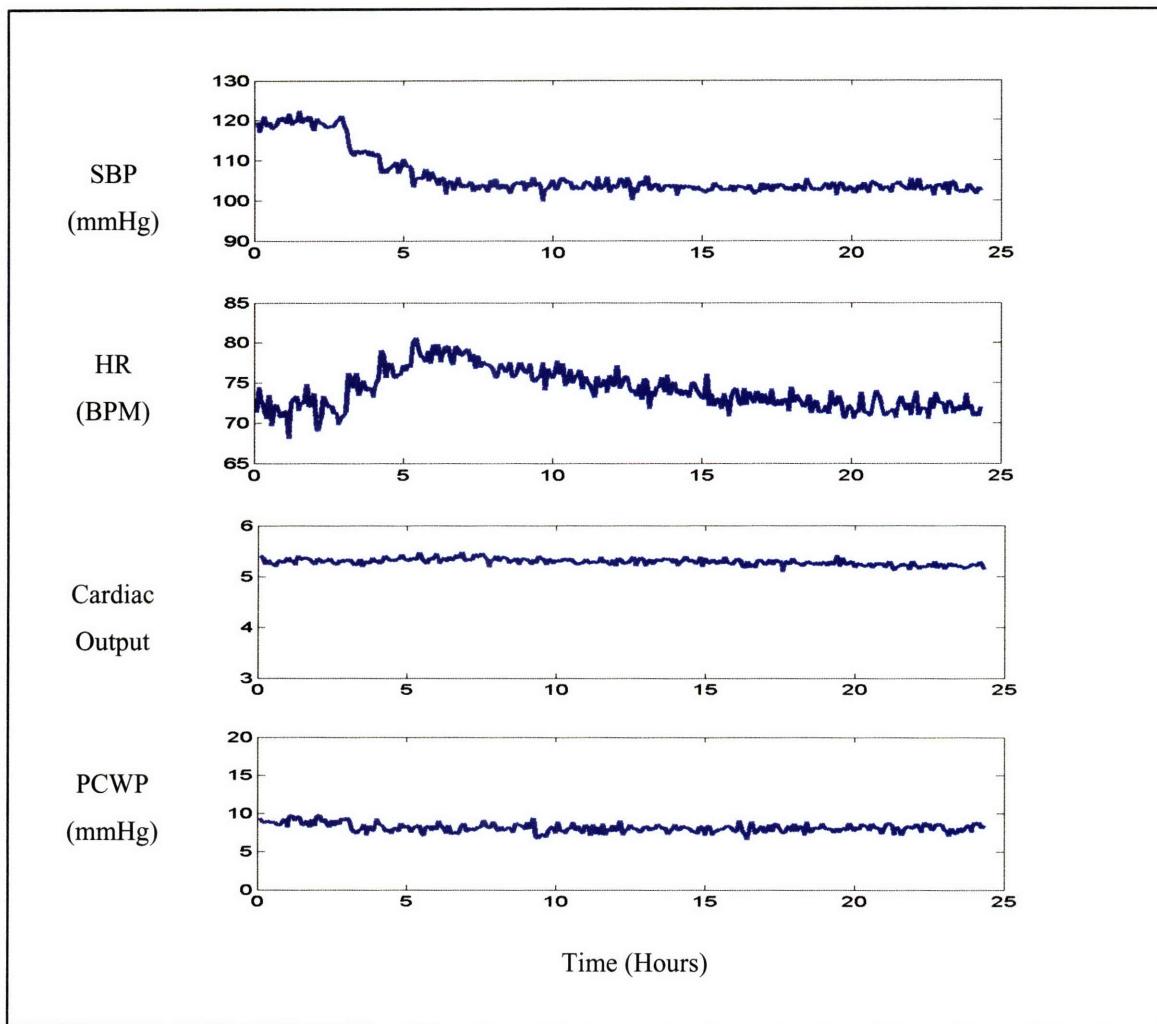


Figure 5-19: Example from Library B: simulation of sepsis

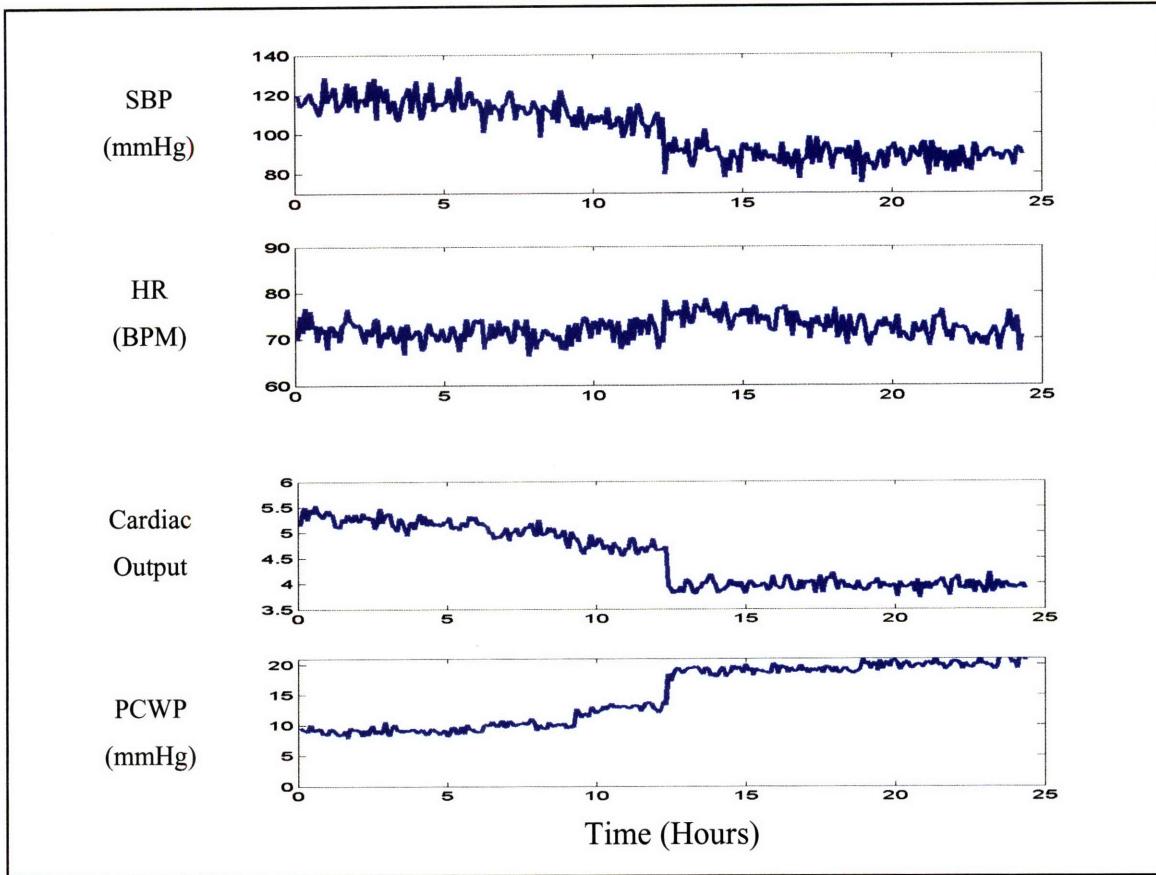


Figure 5-20: Example from Library B: simulation of acute myocardial infarction

In order to compare different time series indexing methods, it is essential to test different algorithms on standard datasets that all researchers have access to. The Control Chart synthetic dataset (SYNCDATA), for the UCI KDD archive [47] has been previously utilized by other researchers for benchmarking time series similarity algorithms and was utilized to test our algorithm as well. The SYNCDATA library contains 600 examples of synthetically-generated control charts (of 60 points). There are six classes (each with 100 examples) of time series data in the SYNCDATA library: Normal, Cyclic, Increasing trend, Decreasing trend, Upward shift, and Downward shift. Examples of the SYNCDATA library are shown in Figure 5-21.

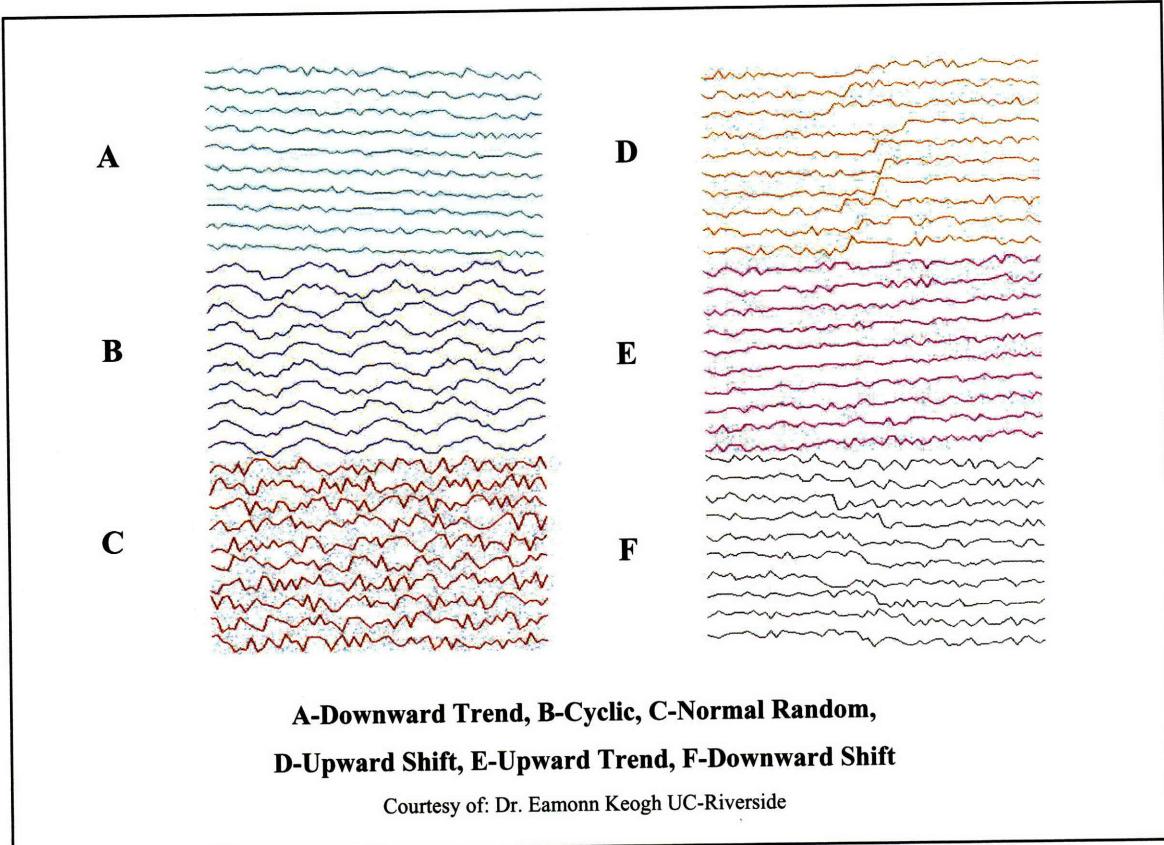


Figure 5-21: Examples of six classes from SYNTDATA library of time series

5.1.10. Results: Example queries from Library A

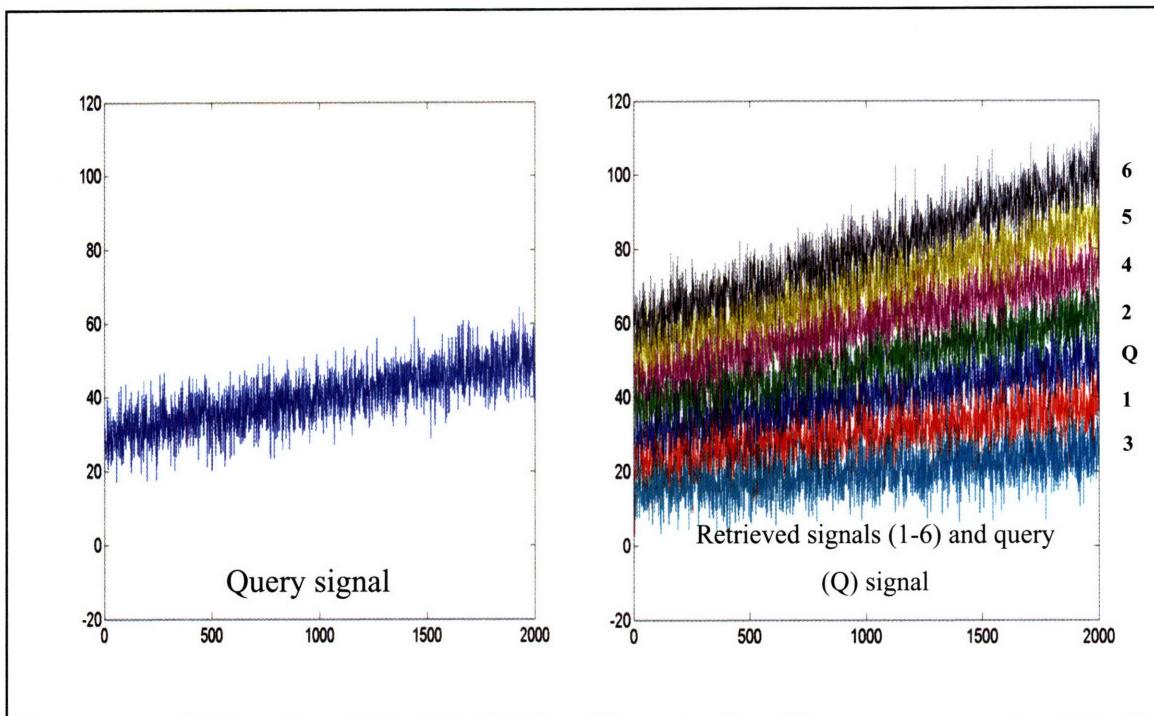


Figure 5-22: Example of increasing trend (from Library A) as query time series, and top 6 retrieved time series.

Several different classes of signals from Library A were utilized as query signals to test the performance of the time series similarity algorithm. Figure 5-22 demonstrates the performance of our retrieval algorithm when sample time series (increasing ramp function) is utilized as a query (Q), and the resultant time series (in ranked order of similarity) are returned. The retrieved signals are the most similar based upon the parameters of the ramp function that are used to generate them.

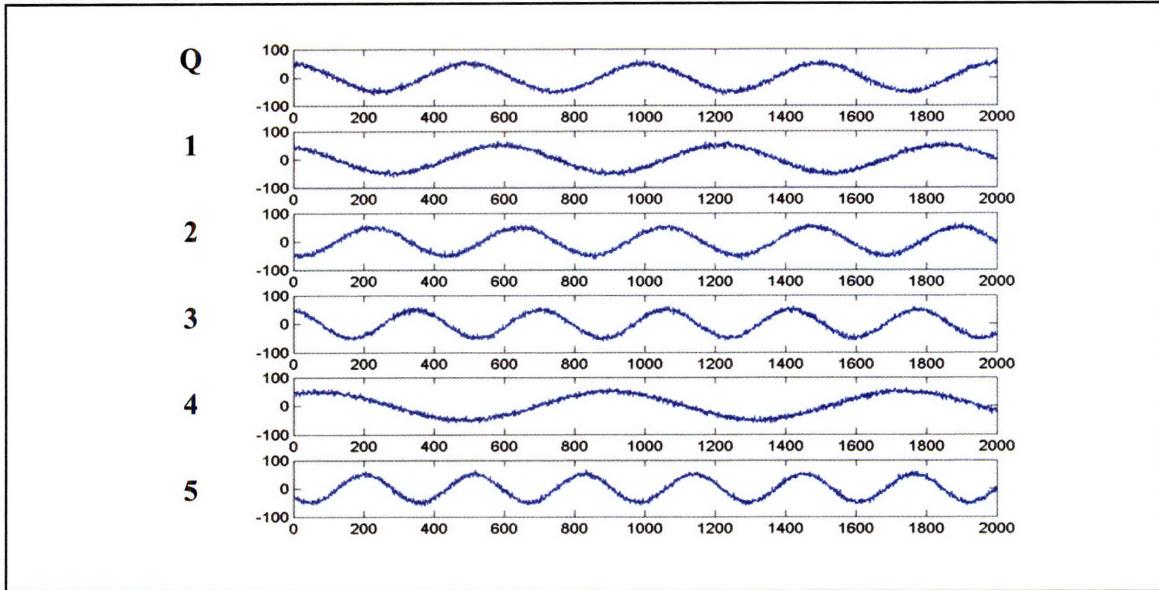


Figure 5-23: Library A sinusoidal signal as query (Q), and retrieved time series (1-5)

Results with different sinusoids from Library B demonstrated that the wavelet-based retrieval algorithm was able to retrieve the five most similar sinusoidal time series when a sinusoid (Q) was submitted as a query signal. In the query example of Figure 5-23, the amplitudes of the sinusoids were constant over all the time series, but the frequencies and phases were different.

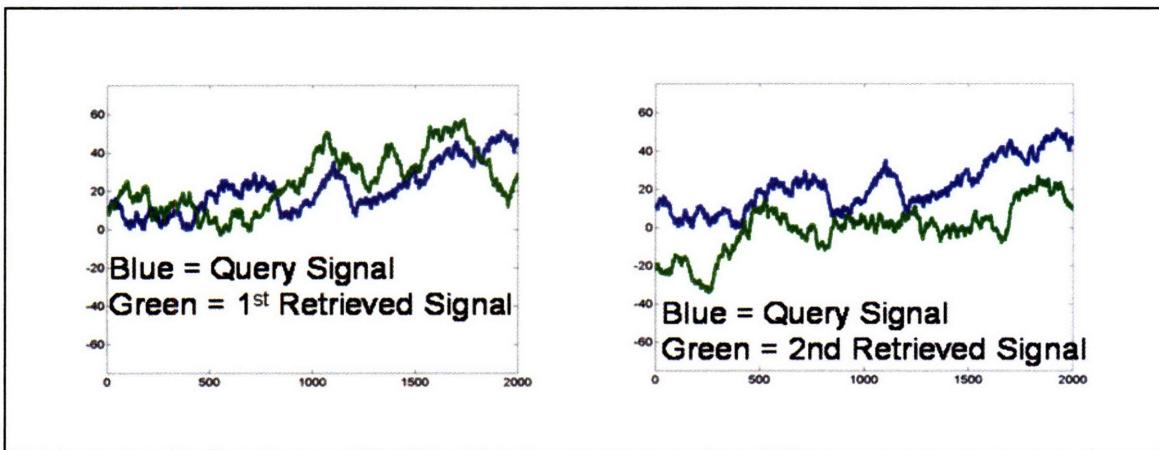


Figure 5-24: Random walk time series from Library A

Finally, random walk time series were used to test the retrieval algorithm performance on more complex signals. The objective performance on random-walk time series data were difficult to assess. As shown in Figure 5-24, the two time series that were retrieved had subjective characteristics that were similar to the original query.

5.1.11. Results: Query performance with a noise stress test

Real-world signals are frequently corrupted by noise. The time series retrieval algorithm was assessed for robustness to additive Gaussian noise. In Figure 5-25, the results are shown for two sample query simulations from Library A. The left panel represents a query sinusoidal signal and the nine most similar time series that were retrieved. The right panel shows the results after all the signals in Library A are corrupted by additive Gaussian noise with a standard deviation of 50. In the presence of noise, the most similar noisy time series that are retrieved correspond exactly to the same series in the noise-free test. The other retrieved signals also have sinusoidal structure. The simulations demonstrated the robustness of the wavelet-based symbolic feature vector in characterizing the most important characteristics of a time series even in the presence of noise.

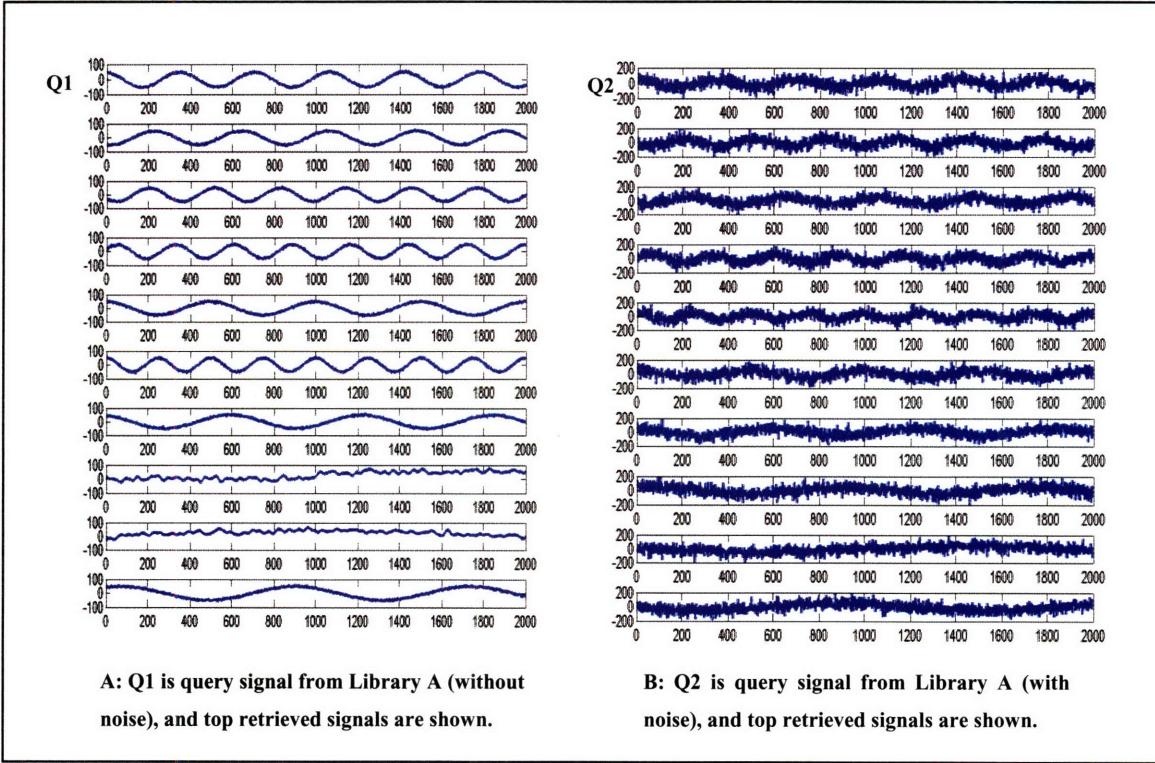


Figure 5-25: Time series retrieval performance in the presence of additive Gaussian noise

5.1.12. Results: Retrieval of similar computational model-generated ICU time series of Library B

In order to assess the performance in retrieving similar physiologic patterns in Library B, the mean blood pressure, cardiac output, and pulmonary capillary wedge pressure trends from the 36 cases were used as queries. For each of the 36 cases, the 5 most similar cases were retrieved from Library B. The similarity metric (based upon the correlation between wavelet symbol feature vectors) was used to weigh the likelihood of belonging to a class. The accuracy in classifying each of the three classes (hemorrhage, sepsis, acute myocardial infarction) is given in Table 5-2. The classifier achieved a high accuracy. The few cases that were misclassified were typically examples where the severity of the disease was very limited.

Table 5-2: Classifier accuracy with synthetic physiologic time series from Library B

	Classifier Accuracy for 12 cases of each class when retrieving 5 most similar cases	3 most similar cases
Hemorrhage	10/12 = 83.3%	10/12
Sepsis	9/12 = 75%	10/12
Acute Myocardial Infarction	12/12 = 100%	12/12

5.1.13. Results: Evaluation of retrieval algorithm with a benchmark time series database (SYNTDATA)

The SYNTDATA time series were also assessed to compare the accuracy of our time series retrieval algorithm with other algorithms that have been tested with the SYNTDATA library. To assess algorithm accuracy, we used the same methodology described in [47] and compared our results to the statistics they reported for other algorithms.

To test the retrieval accuracy of our algorithm on the SYNTDATA library, each one of the 600 time series was used a query signal. Retrieval accuracy, r , is defined by Equation 18 as the number of retrieved time series, q_m , that are from the same class as the query signal, divided by the total number of time series in the database that are from the same class as the query signal, C_N .

Equation 18

$$r = \frac{q_m}{C_N}$$

We determined the retrieval accuracy by selecting the top 99 similar signals that are retrieved by the algorithm, and then finding the subset of which belongs to the same class as the query signal. Thus, the retrieval accuracy can be assessed for each query signal.

We then averaged the retrieval accuracy over all 600 signals to arrive at the final retrieval accuracy for the database.

The performance statistics on the SYNTDATA are reported in Table 5-3 and Figure 5-26. Both our algorithm and the MVQ algorithm report the highest accuracy in comparison to other published methods as well as the simple Euclidean distance metric. The MVQ algorithm is a relatively new time series retrieval algorithm that has been reported to have superior performance over other techniques such as SAX, DFT, and DTW algorithms[47]. Our algorithm achieved an overall accuracy of 0.83 on the challenging SYNTDATA dataset which is equal to the accuracy of the MVQ algorithm. Many of the time series with subtle changes are quite difficult to classify even by a human. Thus, the similar accuracies of both algorithms on this dataset may suggest that an even higher accuracy will be difficult to achieve due to the nature of the dataset. A major advantage our algorithm enjoys over the MVQ algorithm is that the MVQ algorithm requires extensive training data from a subset of the SYNTDATA library in order to create an optimal codebook of features. Our algorithm is unsupervised and thus, requires no training data. Moreover, our algorithm has few tunable parameters. We modified 2 parameters (number of symbols per scale, and the width of the parzen kernel) and found that the results were not significantly altered. The MVQ algorithm requires tuning of 6 parameters in addition to training of a codebook. The Euclidean method had a very poor retrieval accuracy of 0.51.

We also assessed the classification accuracy of our algorithm by classifying a query signal with the class label of the most common class among the retrieved time series. Our retrieval algorithm achieved a classification accuracy of 0.95. Thus, 95% of the time series were assigned to the correct class from among the six classes in the database. These results motivate the use of our algorithm in classifying ICU physiologic time series to possible classes that may be suggestive of different states of physiologic stability.

Table 5-3: Comparison of Retrieval Accuracy of Different Algorithms on SYNTDATA Library

Algorithm	Retrieval Accuracy
Our Algorithm	0.83
MVQ Algorithm	0.83
Euclidean	0.51

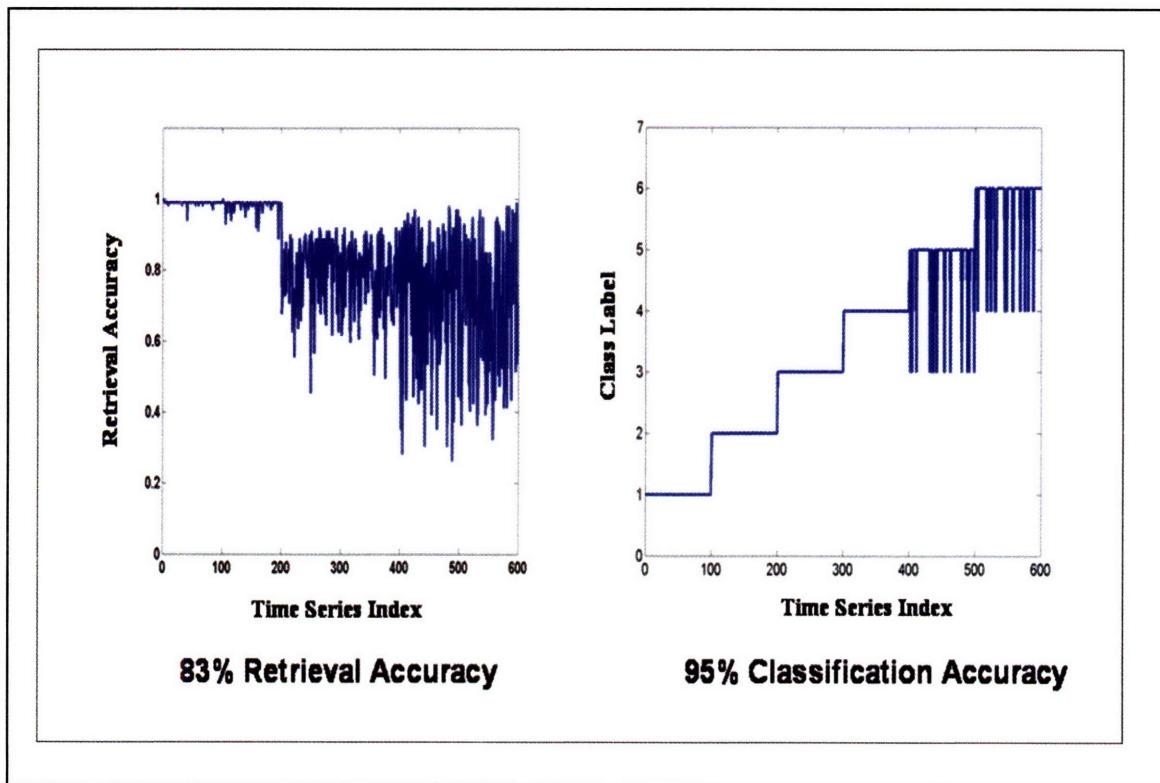


Figure 5-26: Retrieval and classification accuracy with SYNTDATA Library

6.Prediction of Hemodynamic Deterioration in ICU Patients

In Chapter 5, we introduced a new temporal similarity metric for the retrieval of similar time series. We evaluated the performance of our retrieval algorithm on synthetic time series generated from different classes of waveforms such as sinusoids, random walks, and ramps. In addition, a computational physiologic model was used to generate time series that simulate the hemodynamics of patients in different ICU disease states such as sepsis, hemorrhage, and heart failure. The algorithm's retrieval performance using the physiologic model-based time series was robust in the presence of additive noise. Furthermore, the retrieval accuracy of our algorithm on a publicly available time series database (SYNTDATA library) was shown to have superior overall performance in comparison to several of the most popular time series indexing algorithms. In this chapter we will apply the similarity metric to real physiologic time series from the MIMIC-II database.

A modern ICU bedside monitor typically measures several signals from an ICU patient. As previously noted, each signal is typically processed and classified independently of other signals. Several parameters (vital signs) are displayed by the monitor without an overall quantitative characterization of the state of the patient. A busy clinician is then challenged to track several simultaneous parameters from each of the patients that he or she may be caring for. One example of instability that ICU clinicians react to is in the domain of hemodynamic deterioration. The classic sequence of events in hemodynamic deterioration involve drops in a patient's blood pressure followed by therapeutic interventions intended to raise the systemic blood pressure to a range that is compatible with homeostasis. There are several mechanisms that may contribute to the deterioration

of blood pressure in a patient and include: heart failure, hypovolemia, septic and aseptic inflammatory syndromes [69]. A high level of vigilance is required to appropriately react to and treat hypotension. Recent reports [37] have found that the duration of hypotension prior to the initiation of therapy was the critical determinant of survival in human septic shock. Thus, the development of monitoring algorithms that can predict impending deterioration in ICU patients may improve clinical vigilance in a busy ICU.

Figure 6-1 includes examples of three channels of physiologic time series (heart rate, systolic blood pressure, and mean blood pressure) from 88 ICU patient episodes prior to hemodynamic deterioration. In these cases, the fiducial point at $t = 240$ minutes indicates the point in time when a vasopressor medication was initiated as a treatment for hypotension. The physiologic trends are population averaged (with means and standard errors on a minute-to-minute basis) and displayed from four hours prior to the point of intervention (intervention at $t = 0$ minutes). The composite trends with error bars demonstrate that there is significant variability from patient to patient. However, data mining may reveal that there are “signatures” in multidimensional physiologic space that may be predictive of hemodynamic deterioration. A number of clinical studies have shown that certain patterns of physiologic deterioration may precede cardiopulmonary arrests [17]. We hypothesize that analysis of hemodynamic time series with our wavelet-based symbolic similarity metric may illuminate patterns with significant physiologic similarity as well.

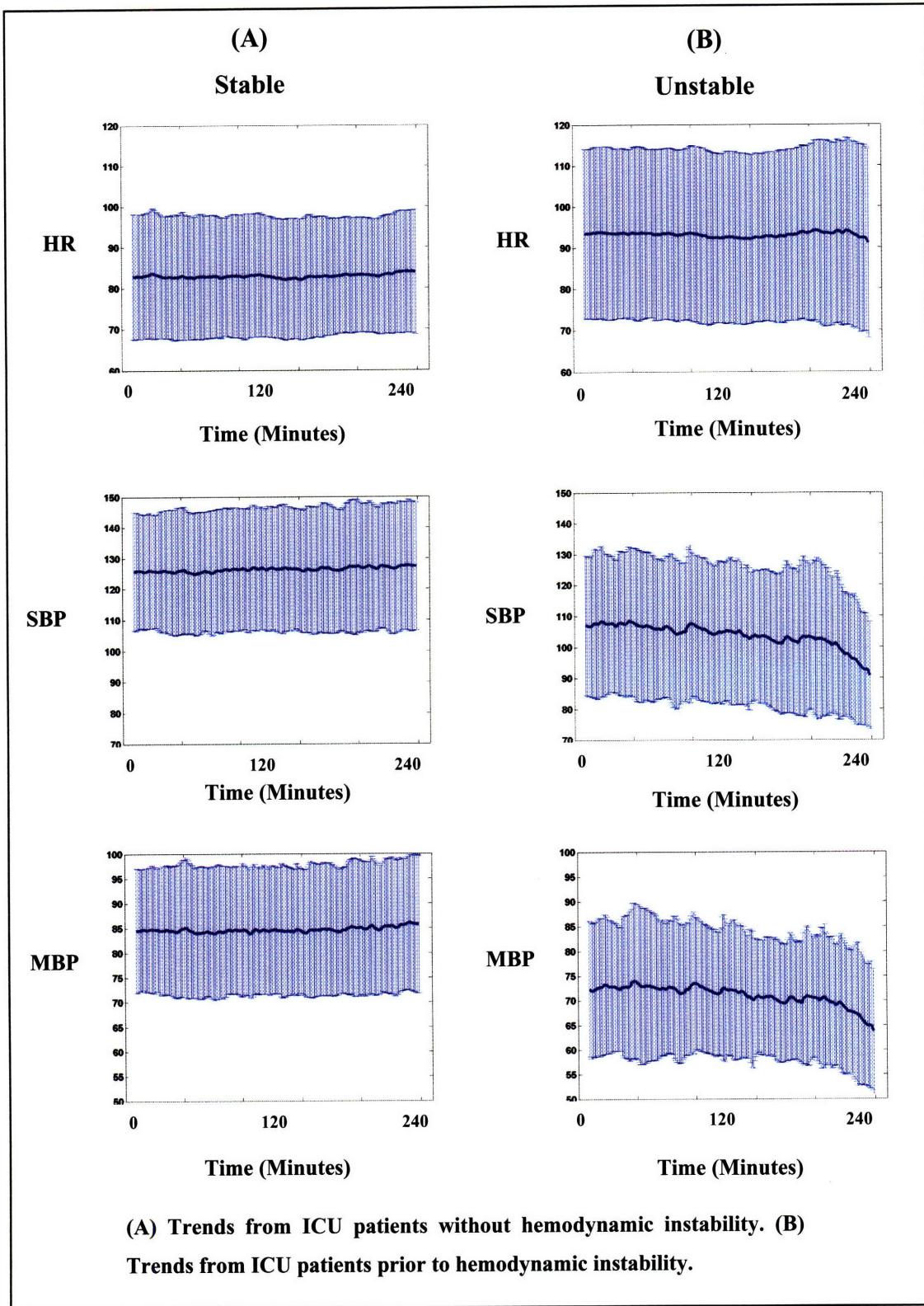


Figure 6-1: Multiparameter hemodynamic trends from ICU patients.

Chapter 6 is organized in the following manner. The next section provides a problem formulation for hemodynamic deterioration we wish to address using the MIMIC-II database. We then describe the subset of data segments in the MIMIC-II database that are ideal for developing and evaluating predictive algorithms for hemodynamic deterioration in ICU patients. Then, we describe the feature extraction process that is based on the use of the framework for wavelet symbolic representation of time series that was introduced in Chapter 5. Finally, we evaluate the performance of our algorithm for predicting hemodynamic deterioration in ICU patients.

6.1. Problem formulation

Hemodynamic monitoring is utilized for obtaining real-time assessments of the functioning of the cardiovascular system in fragile ICU patients. Several studies have shown that identifying ICU patients with physiologic deterioration such as multi-organ failure or sepsis at earlier stages in the disease process results in improved outcomes [37]. Shoemaker et al developed an algorithm that predicted outcomes in trauma patients based upon similarities in a patient's observed vital signs and diagnoses with a database of other patients with known outcomes [67]. Their algorithm's ultimate purpose is to provide an early warning of impending deterioration by fusing several monitored physiologic variables into one composite indicator of overall stability.

We motivate our current problem of hemodynamic instability prediction in a manner similar to how we motivated the overall time series retrieval challenge in Chapter 5. In particular, we will cast the problem of predicting hemodynamic deterioration as a problem in information retrieval such as is commonly done in document and web-page indexing. As shown in Figure 6-2, query search terms are commonly input into an internet search engine with access to a vast internet database of web pages. The search engine processes the query and returns a ranked list of relevant web pages. In a similar manner, it may be possible to submit an ICU record which consists of one or more physiologic time series as a query to an algorithm with access to a vast ICU database of

physiologic time series. The algorithm will process the query time series and identify “similar” physiologic time series in the database from other patients. The retrieved records may also have associated outcomes that may be utilized to predict the outcome in an ICU record utilized in the query. The retrieved physiologic time series are ranked based on their statistical similarity to the query time series. We hypothesize that statistical similarity may also imply physiologic similarity and thus, clinical inference based on known outcomes from retrieved time series may be suggestive of an outcome in the query ICU physiologic time series.

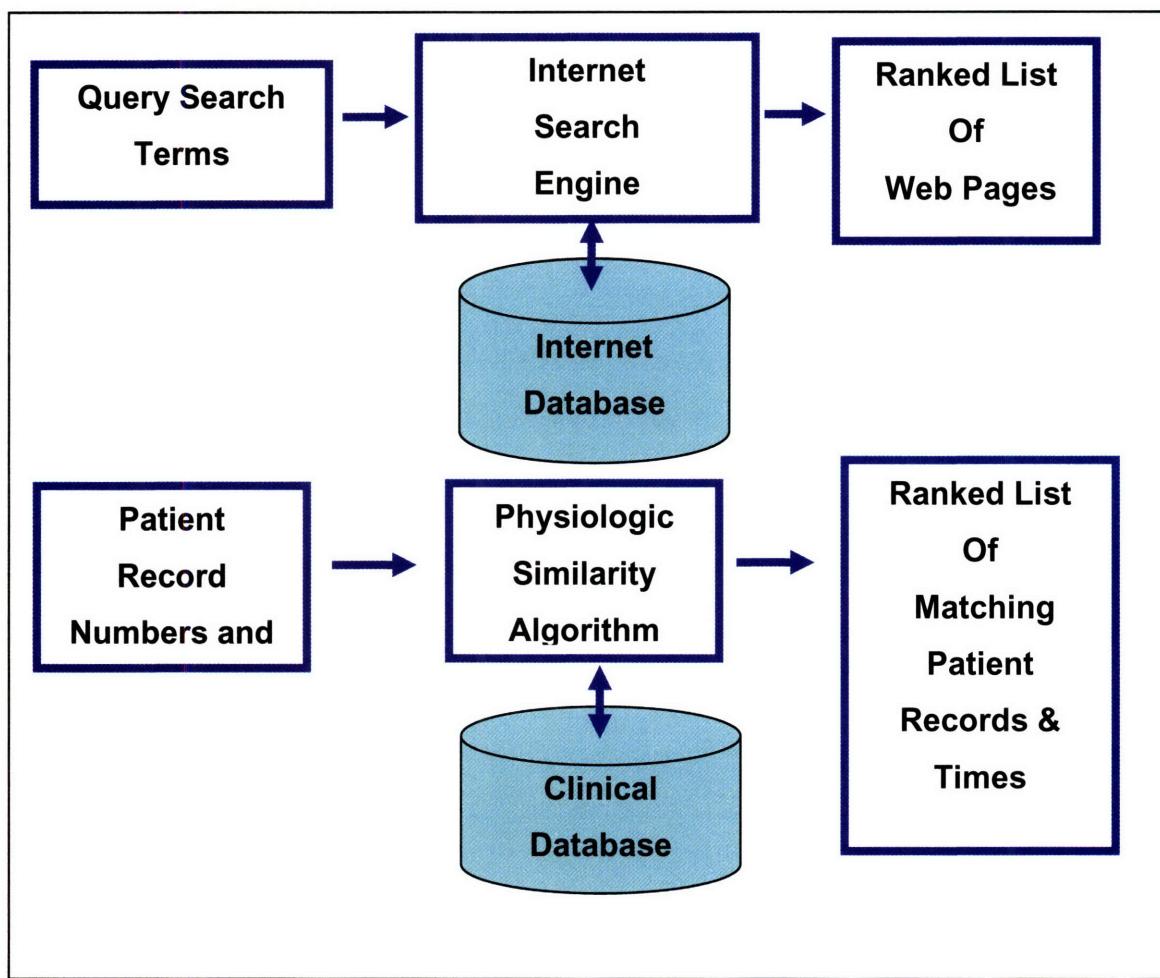


Figure 6-2: ICU record retrieval formulated as problem in information retrieval

6.1.1. Selection and automated annotations of physiologic time series from ICU records

In each MIMIC-II record there are several physiologic signals. Different physiologic signals convey unique information about different organ systems. For example, the ventilator parameters and arterial blood gasses are used to monitor the pulmonary system. We are particularly interested in hemodynamic deterioration and the cardiovascular system. The physiologic signals that are frequently monitored in hemodynamically unstable ICU patients and readily available in MIMIC-II include heart rates and invasive arterial blood pressures. The availability of different physiologic signals is often limited by their inherent invasiveness. For example, pulmonary artery pressure and cardiac output monitoring would be physiologically informative, but are infrequently measured due to the invasiveness that is required to make such measurements. Based on the readily available physiologic time series, we seek to classify the hemodynamic state of an ICU patient into one of two states (or classes): A) Impending Hemodynamic Instability (deterioration), and B) Hemodynamic Stability. While this classification scheme is simplistic, it is also based on conventional therapeutic decision making in the ICU. In particular, hemodynamic instability is defined as the state in which patient requires aggressive vasopressor support in order to maintain adequate blood pressure and vital organ perfusion. Hemodynamic stability is assumed when a patient is not treated with aggressive therapeutic support to maintain adequate systemic blood pressure.

To evaluate the performance of our predictive (classification) algorithm, it would be advantageous to have access to “gold-standard” labels from clinician annotations that label the ICU record segments into periods of hemodynamic stability and instability. However, developing and implementing an objective annotation process that can retrospectively label the ICU records is a challenging and time-consuming task. Thus, we will evaluate the performance of our predictive algorithm based upon the available therapeutic data time series of ICU patients.

6.1.2. Therapeutic profiles of hemodynamic stability and instability

An ICU patient typically receives aggressive therapeutic interventions when the clinical staff concludes that the patient is becoming hemodynamically unstable. In the modern ICU, the patient's blood pressure is closely controlled so that it does not deviate from a physiologically acceptable range. In Figure 6-3, physiologic time series from an ICU patient are shown. At approximately 16 hours into the record, the clinical staff reacted to the drop in blood pressure by starting Neosynephrine, a potent vasopressor. The hourly IV fluid was not sufficient to maintain hemodynamic stability. However, as can be seen, prior to the onset of the vasopressor, the trajectory of the blood pressure at 14 hours started to drop. An algorithm that could have provided an early warning two hours before the point of vasopressor intervention may serve as a useful clinical decision support tool.

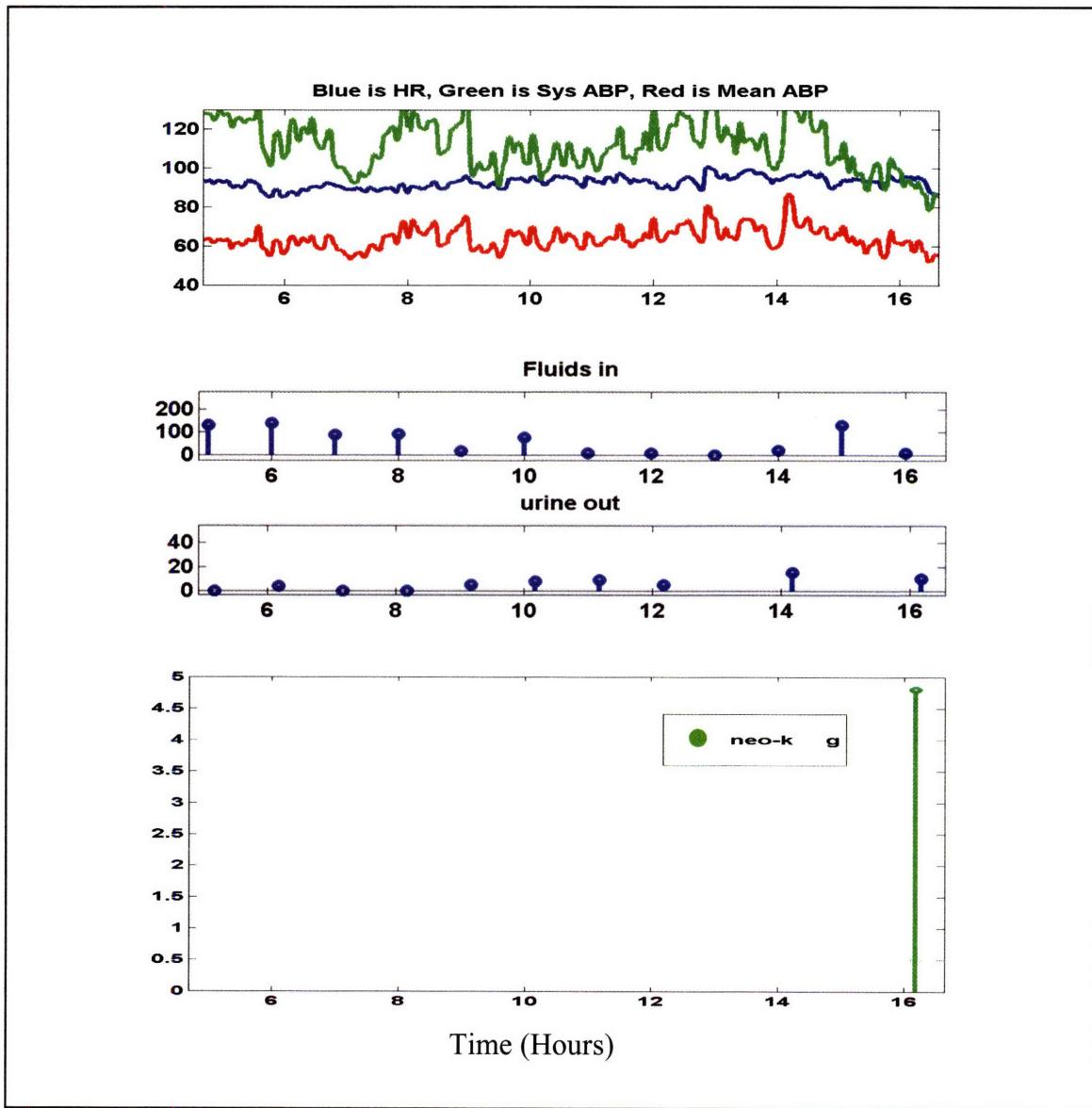


Figure 6-3: Example of hemodynamic instability

The means of controlling the blood pressure include the use of drugs that increase systemic resistance (vasopressors) or heart function, as well as intravenous fluid to increase intravascular volume. Table 6-1 is a listing of the classes of interventions and relevant examples of each intervention that are commonly found in the MIMIC-II database. The use of a vasopressor is typically reserved for the most hemodynamically unstable patients. The onset of a vasopressor intervention (such as Levophed) was interpreted to suggest that a patient is hemodynamically unstable. To simplify the

classification process as well as account for the variability in clinical decision making, the actual dosage of a vasopressor was not utilized.

Nearly all ICU patients (stable and unstable) receive certain levels of maintenance IV fluids to replace insensible losses and to maintain homeostasis. However, patients that may be showing signs of intravascular volume depletion may receive significant amounts of IV fluids within a short period of time. A stable patient state required that no vasopressors are utilized and only minimal IV fluid is being administered. The patient state was classified as an intermediate state (neither stable nor unstable) when a significant amount of fluid (see Table 6-1) was administered, and hence, was not used in the library.

Table 6-1: Hemodynamic therapeutic interventions

Class of Medication	Example of Medication	Thresholds (if applicable)	Conditions that negate stability
Vasopressors	Dopamine Neosynephrine Levophed	Any Dose	Any Dose
IV Fluids	IV Fluid or Blood Products	N/A	Greater than 1.2 L/hour Crystalloid or 2 units PRBCs/hr

Table 6-2 includes the criteria used for classifying data segments of ICU records as being indicative of stability or future instability. The selection of the particular therapeutic criteria allowed for a high-throughput and automated annotation technique of ICU record segments, while meeting clinically acceptable definitions of hemodynamic stability.

Table 6-2: ICU Record segment stability/instability criteria

	Hemodynamic Stability	Hemodynamic Instability
Minimum Window Size	2 hours	2 hours
Med Criteria	No onset of a pressor for at least 12 hours after the end of the segment of data (window)	Onset of a new pressor within N hours of the end of the segment of data (window)
Fluid Criteria	Less than N L/hr for all hours within window	N/A
Physiologic Data Availability	Required heart rate and invasive ABP 1-min time series numerics	Required heart rate and invasive ABP 1-min time series numerics
Other Exclusions	Intra-aortic balloon pumps Pacemakers	N/A

6.1.3. Limitations of hemodynamic instability criteria

In MIMIC-II, we utilized the continuous medication and fluid-balance data to automatically identify the segments of stability and instability. This framework implicitly assumes that the clinicians' judgment with respect to therapeutic support is correct. However, as we have previously stated, effective vigilance in the ICU is difficult to achieve at all times, and thus, there may be occasions when a clinician's decision making is suboptimal. Thus, there may be episodes of time in which a therapeutic response in the form of a vasopressor would have been appropriate, but was not given. Also, there may be episodes in which the use of a vasopressor was unnecessary. Development of an unbiased annotation and adjudication process to deal with such discrepancies is a challenging task, and was not utilized.

As previously mentioned, documentation of clinical data are not without errors. The time an intervention was made (such as increasing the level of an IV medication) may be off by up to 30 minutes or more. Furthermore, certain medications may not have been charted in the appropriate field and were, thus, not factored into the labeling of a window of data.

6.2. **Methodology**

In the preceding section, we motivated the clinical need for predicting hemodynamic deterioration in ICU patients. In this section we will describe the methodology that was utilized for developing a novel algorithm for predicting hemodynamic deterioration. MIMIC-II records are rich in several physiologic signals. We sought to use those signals that are commonly found in most hemodynamically unstable patients (patients who received vasopressor therapy). For example, most hemodynamically unstable patients will be monitored with an ECG and invasive intra-arterial catheter to track systemic blood pressure. The most hemodynamically unstable patient may also have Swan-Ganz catheters for tracking pulmonary artery and central venous pressures as well as cardiac output [72]. While such hemodynamic measurements are helpful in assessing the hemodynamic status of a patient and optimizing therapy [46], they are considered to be invasive and are utilized only in a minority of patients (those patients that may be *the* most hemodynamically compromised within an ICU).

In our signal selection process, we chose to include the heart rate and arterial blood pressure trends (sampled on a minute-to-minute interval) because they are readily available and informative. We chose to exclude the recorded invasive PAP, CVP, and cardiac output measurements since they were not readily available in all patients. We determined that their inclusion may bias our results since patients with PAP, CVP, or CO trends would be more prone to having episodes of hemodynamic instability. Others have factored the mere presence of such signals into severity of illness scores [42]. Furthermore, the PAP and CVP measurements are typically low pressures and subject to severe noise corruption which limits their use in practice. However, the cardiac output

measurement is not typically as noisy and provides valuable information about the mechanical work of the heart. It would be advantageous to include cardiac output trends in our signal analysis algorithms. However, there lack of availability in all patients was a limiting factor. To overcome this challenge, we sought to employ cardiac output estimation techniques using only the blood pressure and heart rate trends. We and others have demonstrated that estimates of cardiac output may be derived within an acceptable level of accuracy by processing the blood pressure and heart rate data [72]. Prior to deriving an estimated cardiac output time series, the blood pressure and heart rate trends require pre-processing to remove noise. In the next section, we describe a simplified knowledge-base framework for filtering physiologic time series.

6.2.1. Physiologic trend pre-processing

As previously noted, physiologic trends such as blood pressure and heart-rate are inherently corrupted by noise. We utilized a combination of physiologic-knowledge based noise detection and outlier removal to process the one-minute parameter heart rate and blood pressure data. The schematic in Figure 6-4, provides an overview of the artifact and noise removal algorithm.

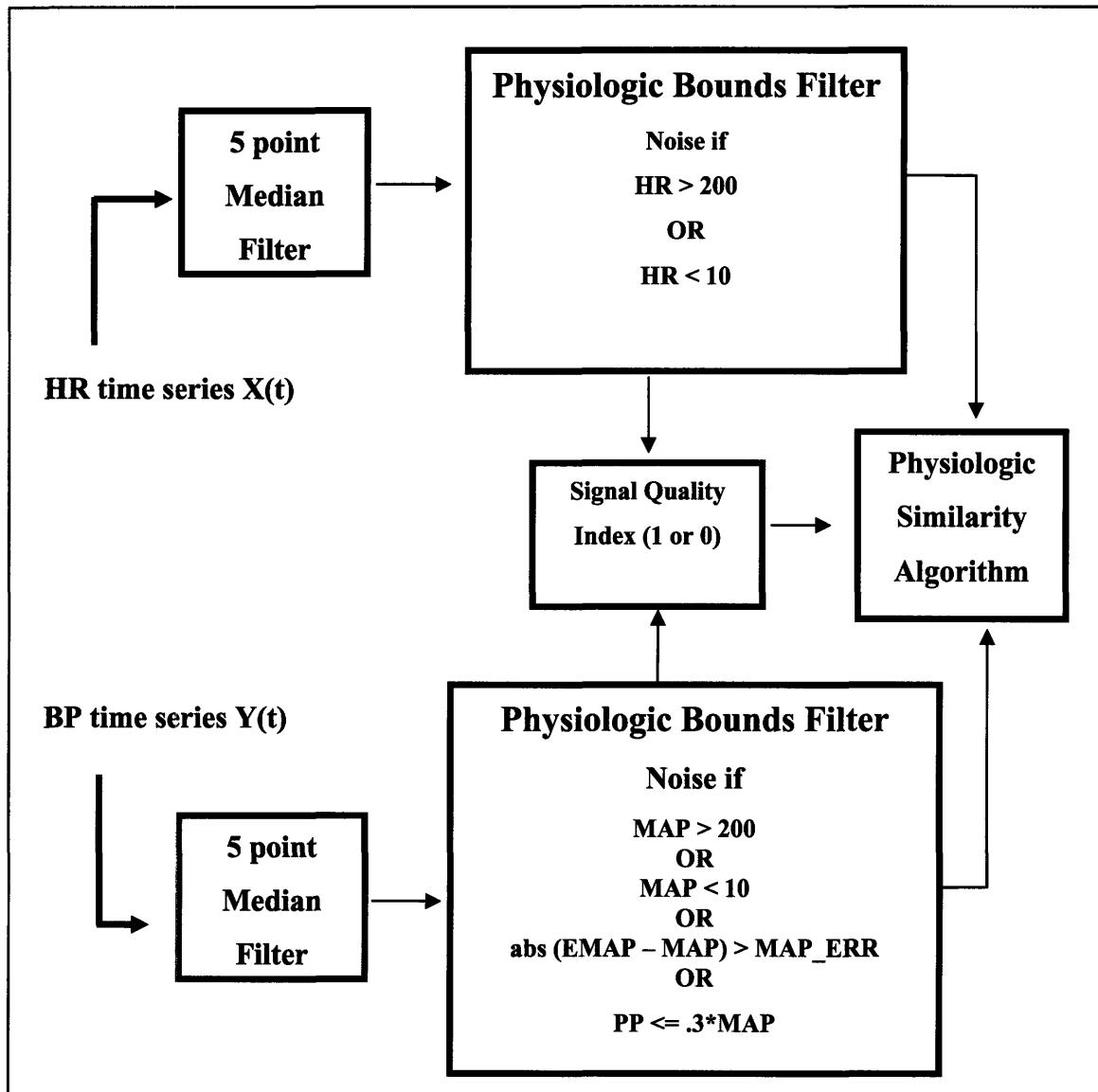


Figure 6-4: Noise processing of physiologic trends

As shown in Figure 6-4, the one-minute heart rate and blood pressure (systolic, mean, and diastolic) are first filtered with a 5 point median filter to remove outliers. Then, physiologic bounds are applied to the all of the physiologic time series as described in the schematic. The actual mean arterial pressure (MAP) is compared to absolute thresholds as well as to the estimated mean arterial pressure (EMAP). The formula to derive EMAP from the measures systolic and diastolic arterial blood pressures was explained in Chapter 4. The difference between **EMAP** and **MAP** is compared to a threshold (**MAP_ERR**) and is used to identify noisy data when the threshold is exceeded. In order to exclude blood

pressure trends that are damped, the pulse pressure (*PP*) must be greater than 30% of the *MAP*. The heart rate trend is deemed noisy if the heart rate value exceeds 200 beats per minute (bpm) or is less than 10 bpm. For those values of heart rate and blood pressure that were deemed to be noisy based on these physiologic bounds filters, they were replaced by linearly interpolating values from neighboring points. Thus, the process yields de-noised heart and blood pressure trends as well as a signal quality index. The signal quality index is binary value (where 1 equates to noise in any of the physiologic time series and 0 equates to noise-free data) for each minute (or sample). If more than 25% of the one minute samples are deemed to be noisy for a specific hour segment of data, then that segment is not included in the overall analysis of deterioration for a given patient. Thus, the patient state will not be assessed over an hour interval if there is excessive noise in the physiologic trends.

There are physiologic conditions, such as a transient arrhythmia, that may result in an abrupt change in a signal over a small time interval of one or two minutes. The median filtering may suppress such real physiologic changes. However, the embedded arrhythmia alarms within a bedside monitor should ideally detect such transient episodes. The focus of the present work is to predict longer-term changes in a patient state (over hours) that maybe signatures of deterioration. Thus, the framework presented here will not address physiologic processes that may occur over a few seconds to a couple of minutes.

6.2.2. Derivation of estimated cardiac output

Sun et al [72] utilized the MIMIC-II database and compared the performance of several previously published blood pressure-based cardiac output estimation techniques with the recorded thermodilution-based cardiac output. A modified algorithm based on the work of Liljestrand et al [72] demonstrated the best performance in comparison to other cardiac output estimation techniques.

Equation 19

$$ECO = k_{co} \frac{HR \times (SBP - DBP)}{MAP}$$

The estimated cardiac output we derived by using the de-noised heart rate and blood pressure trends. Using a slight modification of the method of Liljestrand, the estimated cardiac output (ECO) is defined in Equation 19. The pulse pressure (PP) is defined as the difference of the systolic blood pressure (SBP) and diastolic blood pressure (DBP). The constant of proportionality, k_{co} , is related to the arterial compliance. In practice, the arterial compliance is not a known parameter. Thus, the estimated cardiac output must either be calibrated to actual thermodilution cardiac output measurements, or the estimated cardiac output should be interpreted as a signal that is proportional to the actual cardiac output. Intermittent cardiac output is typically measured using a Swan-Ganz catheter with the thermodilution technique which requires an expert user at the bedside. Since few patients have thermodilution cardiac output measurements, ECO cannot be calibrated in most MIMIC-II patients. Thus, the actual value of ECO is difficult to interpret physiologically. We focus on changes to ECO over time as being more meaningful because they would ideally reflect the same percent change in actual cardiac output if we assume that k_{co} remains constant. In [72], the blood pressure waveforms were first processed to arrive at the blood pressure parameters and signal quality estimates. In this thesis, we estimate cardiac output using only the one-minute heart-rate and blood pressure trends to reduce the amount of processing that would have been required for 125 Hz waveform data. Thus, we rely on the bedside monitor estimates of blood pressure (systolic, mean, and diastolic) and heart rate.

6.2.3. Hemodynamic event library creation

To test our physiologic similarity metric as a predictor of hemodynamic deterioration, we define two classes (or physiologic states) that we will attempt to differentiate. The “Unstable” class will be ICU record segments consisting of physiologic trend data prior to an onset of a significant vasopressor. The minimal length of a segment of data was at least two hours of available trend data prior to a pressor onset. The “Stable” class includes ICU patient record segments consisting of physiologic trend data in which there are no significant pressor medications given 12 hours before the start of the window and 12 hours after the end of the window. The input features to our predictive algorithm

include 1-minute averaged heart rate and arterial blood pressure trends as well as estimated cardiac output (ECO).

The unstable class consisted of 170 segments (windows) of data from 119 different patients. The stable class consisted of 162 segments (windows) of data from 128 different patients. The Hemodynamic Event library is summarized in Table 6-3.

Table 6-3: Summary of hemodynamic event library

	Number of Segments	Number of Unique Patients	Total Monitoring Hours
Stable	162	119	1420 hours
Unstable	170	128	765 hours

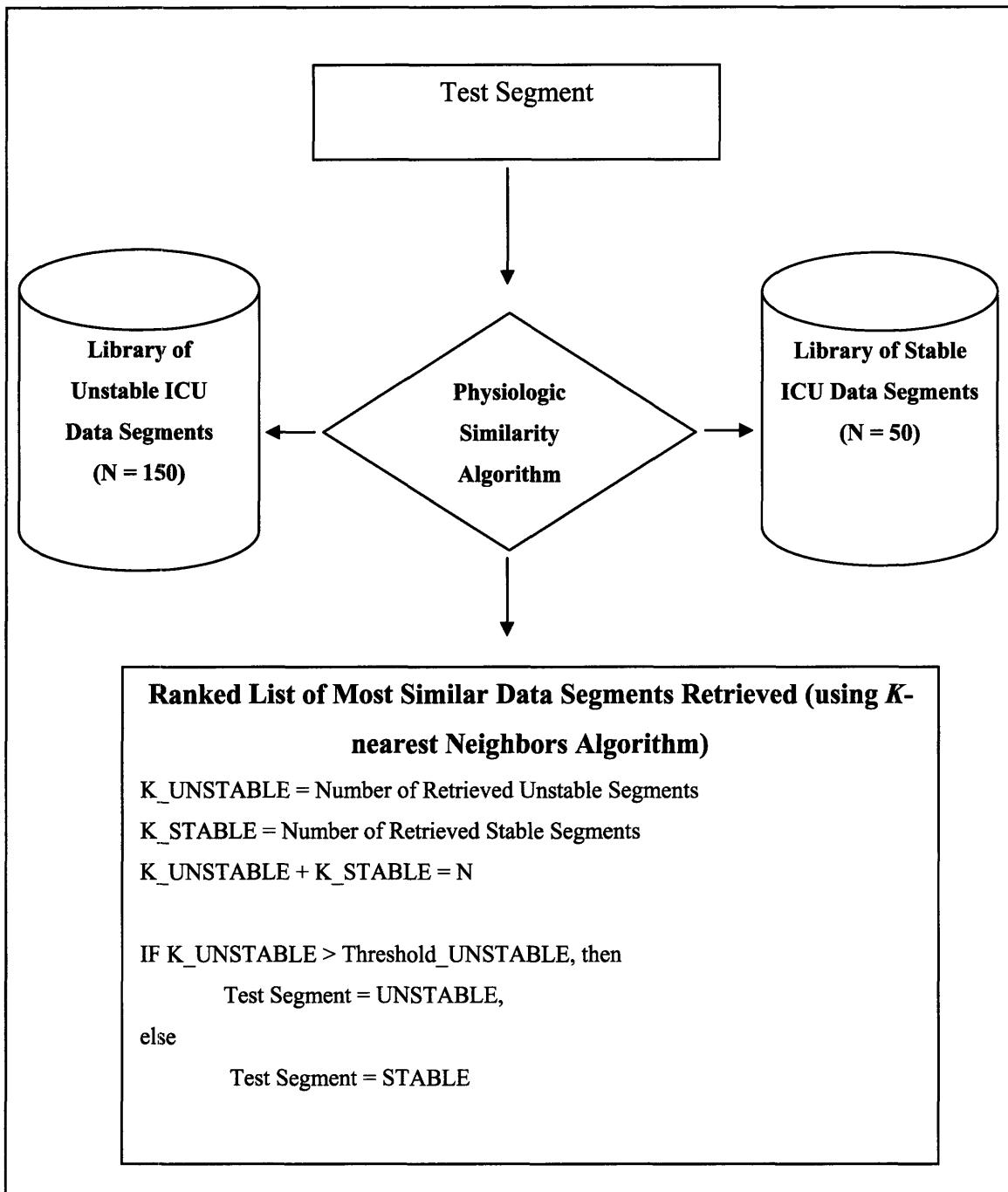


Figure 6-5: Flow diagram for hemodynamic instability using “similar” segments retrieved from reference ICU multiparameter segments

6.3. Results

In this section, we will characterize the performance of the hemodynamic deterioration prediction algorithm with real MIMIC-II patient data. We will also characterize the important physiologic factors that are important in affecting the sensitivity and specificity

of our predictive algorithm. We also present the output of the predictive algorithm with a physiologically meaningful interpretation in term of changes in physiologic variables. Before we present the results of the predictive algorithm, we briefly analyze the performance of the estimated cardiac output (ECO) technique in terms of its agreement with actual cardiac output measurements that are available in a subset of MIMIC-II patients. The ECO trend is utilized as one of the features in our wavelet-based symbolic feature vector, and thus, its agreement with actual CO is important to characterize.

6.3.1. Performance of estimated cardiac output technique

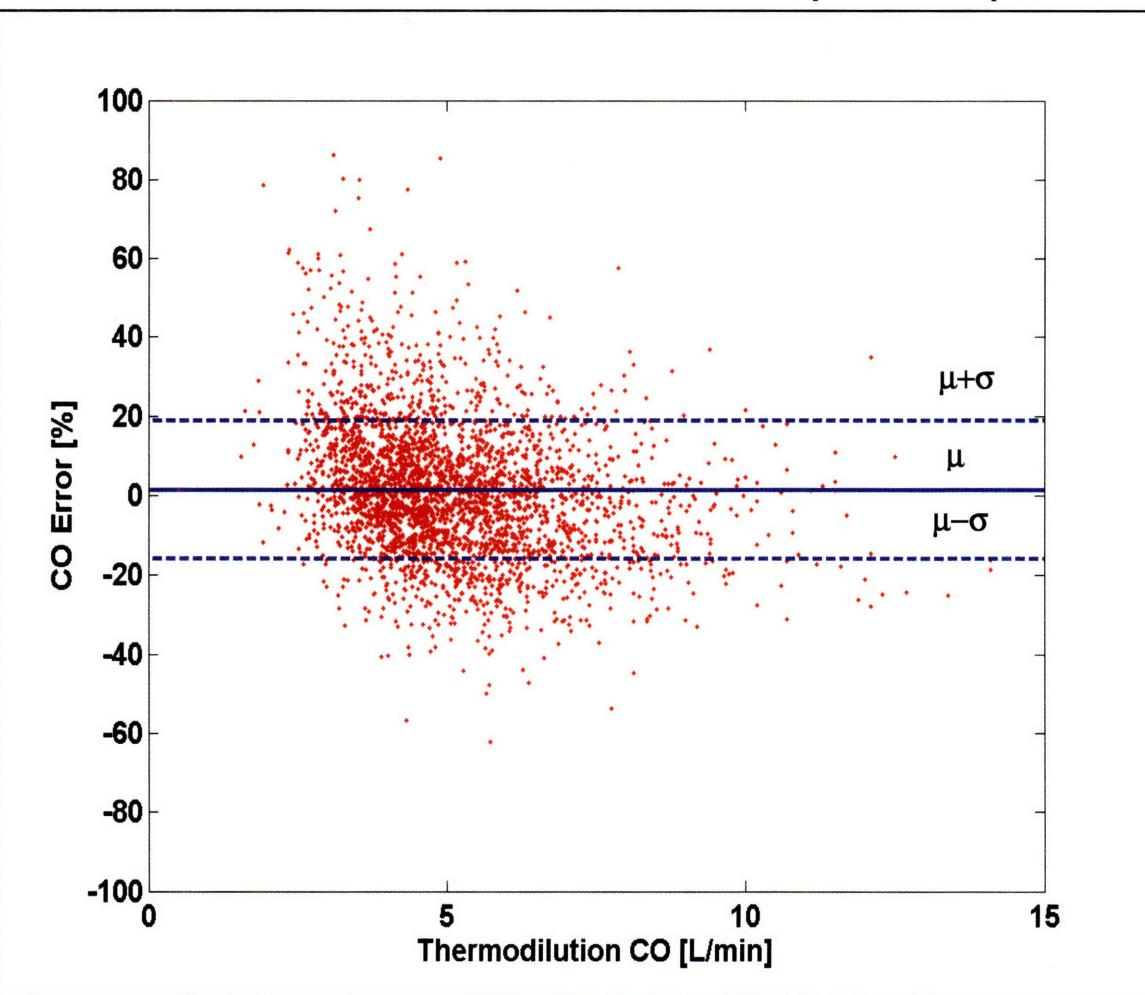


Figure 6-6: Bland-Altman analysis of agreement between estimated cardiac output (ECO) and thermodilution cardiac output (TCO).

The method described by Equation 19 to estimate cardiac output was compared against actual thermodilution cardiac output (TCO) measurements in a subset of patients. For each patient with available cardiac output measurements and invasive arterial blood pressure, the estimated cardiac output (ECO) was calibrated so that the mean ECO over the patient-stay was equal to the mean of the actual TCO measurements over that stay. The error for each measurement was calculated as the difference between the estimated CO and actual TCO normalized by the actual TCO. The Bland-Altman analysis of error is displayed in Figure 6-6 in terms of the percent error between ECO and TCO as a function of TCO. There are 2881 individual TCO measurements pooled from 290 patients. The bias (1.5%) is small since the calibration method described above resulted in the same mean for each patient ECO and TCO time series. The precision error (standard deviation of the error pooled from all patients) was 17.3%. Some patients had more TCO measurements during an ICU stay and thus contributed more points for the overall analysis. If the precision error is averaged for the 290 patients equally (without weighting by the number of TCO measurements), the average precision error is then calculated to be 15.6%. The error between ECO and TCO was uncorrelated with measured TCO ($p = -0.28$).

The results of the present study are significant because the size of the study in terms of the number of patients is --- to the best of our knowledge --- among the largest reported. Other studies have utilized analysis of the arterial blood pressure waveform to derive an estimated cardiac output with one of several different pulse contour analysis techniques. We previously published a comprehensive analysis of 11 different techniques to estimate cardiac output changes by analyzing the ABP waveforms [72]. The present results demonstrate that utilizing only the ABP trend and HR trend is sufficient to achieve a comparable level of agreement between ECO and TCO in comparison to several recently published papers using waveform-level analysis. Finally, this cardiac output study demonstrates that the MIMIC-II database is an ideal platform for measuring the accuracy of an ECO algorithm. MIMIC-II data are representative of the realistic dynamics with which cardiac output estimation algorithms may be challenged. The challenges include physiologic waveform dynamics across different hemodynamic states during an ICU stay

as well as waveforms that may be corrupted by noise and suboptimal maintenance (e.g. clot-formation in catheters or improper calibration of transducers).

Some of the ECO measurements were significantly different from the corresponding TCO measurements in certain patients. Such significant errors may be due to several factors: noise in the blood pressure or heart rate trend that was not detected with our simple signal quality estimator, inaccurate TCO measurements [54], and limitations of the models described in [72] to estimate CO. While these errors may preclude the ECO from being clinically accepted as a replacement for TCO, we hypothesize that the ECO may still be a useful derived measurement for detecting which patients may deteriorate and require aggressive hemodynamic interventions prior to a significant drop in blood pressure. In the next section we characterize the performance of our predictive algorithm for hemodynamic deterioration.

6.3.2. Evaluation of hemodynamic deterioration prediction alarm

We evaluated the performance of our predictive algorithm for hemodynamic instability. We chose M ($M = 150$) segments of trend data that were 120 minutes in length. The trends were first processed with our simple signal quality estimator as described in Figure 6-4. Segments which had greater than 10% of data deemed to be noisy were excluded from the analysis. Each test segment was utilized as a query to the physiologic trend similarity algorithm. The similarity between the test segment and a library of representative stable and unstable segments from other patients was assessed. The schematic in Figure 6-2 summarizes the classification algorithm. For each query segment, the top N most similar library segments are retrieved. In practice, we set $N = 40$ to yield the best performance. The number of retrieved unstable library segments ($K_{UNSTABLE}$) was divided by N and compared to a threshold ($THRESH_{UNSTABLE}$). If the threshold was exceeded, then the test segment was classified to be unstable. Otherwise, the test segment was classified to be stable. The

percentage of unstable library segments among the top N retrieved segments was defined as the “*Instability Index*.”

In Figure 6-7, we provide an example of a predictive alarm that is based on the physiologic similarity algorithm. In this example, the input is a two-hour sliding window of data. Every 15 minutes, the instability index is updated with newly available data. This particular example includes a normal systolic blood pressure that would generally not be indicative of imminent hemodynamic instability. As reported in Table 6-4, the systolic blood pressure is the parameter with the greatest class separation between stable and unstable patients. Four hours (at $t=120$ minutes) prior to the point of deterioration (vasopressor onset at $t=360$ minutes), the systolic blood pressure is far more similar to a stable patient than an unstable patient. However, there are fluctuations in blood pressure as well as heart rate that occur over 30 to 60 minute time scales that are prominent in this record at $t = 130$ minutes and $t = 220$ minutes. These fluctuations are captured in the multiscale wavelet representation. The instability index continues to increase because these fluctuations are more “similar” to fluctuations in the reference hemodynamically UNSTABLE library as compared to the STABLE library segments.

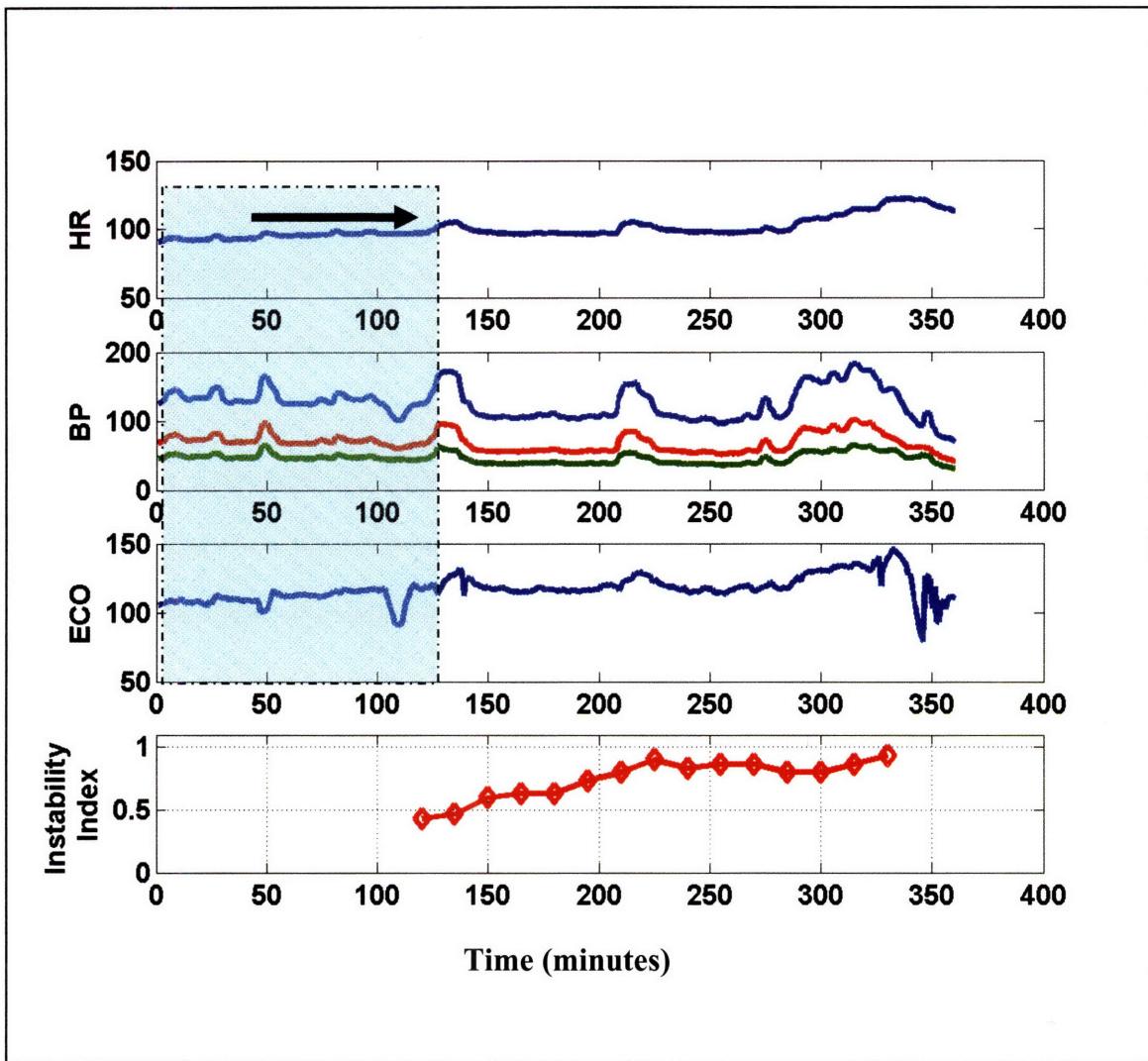


Figure 6-7: Example of predictive hemodynamic deterioration alarm

To assess the performance of the classifier, the classifier labels were compared to the “true” labels as determined by the corresponding therapeutic profile associated with each segment of 120 minutes of data. If there was a pressor onset within 4 hours of the end of the test segment, the true class was set to UNSTABLE. If there were no pressor onsets within 12 hours of the end of the segment, the true segment class was set to STABLE.

The performance of the classifier was characterized in terms of sensitivity and specificity. The sensitivity and specificity are defined as:

$$Sensitivity = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

where TN (true negative) equals the number of correctly labeled stable events, FN (false negative) equals the number of true unstable events that were labeled as stable, FP (false positive) equals the number of true stable segments that were labeled as unstable, TP (true positive) equals the number of correctly labeled unstable events.

We performed several cross-validation studies where the reference segments for the STABLE and UNSTABLE libraries were varied, as were the query segments. Furthermore, the threshold (THRESH_UNSTABLE) was also varied to determine the optimal value to yield the best performance in terms of sensitivity and specificity. A characteristic receiver operator curve (ROC) was constructed as is shown in Figure 6-8. A sensitivity of 0.76 and specificity of 0.82 was achieved. The area under the curve as a measure of overall accuracy equaled 0.83.

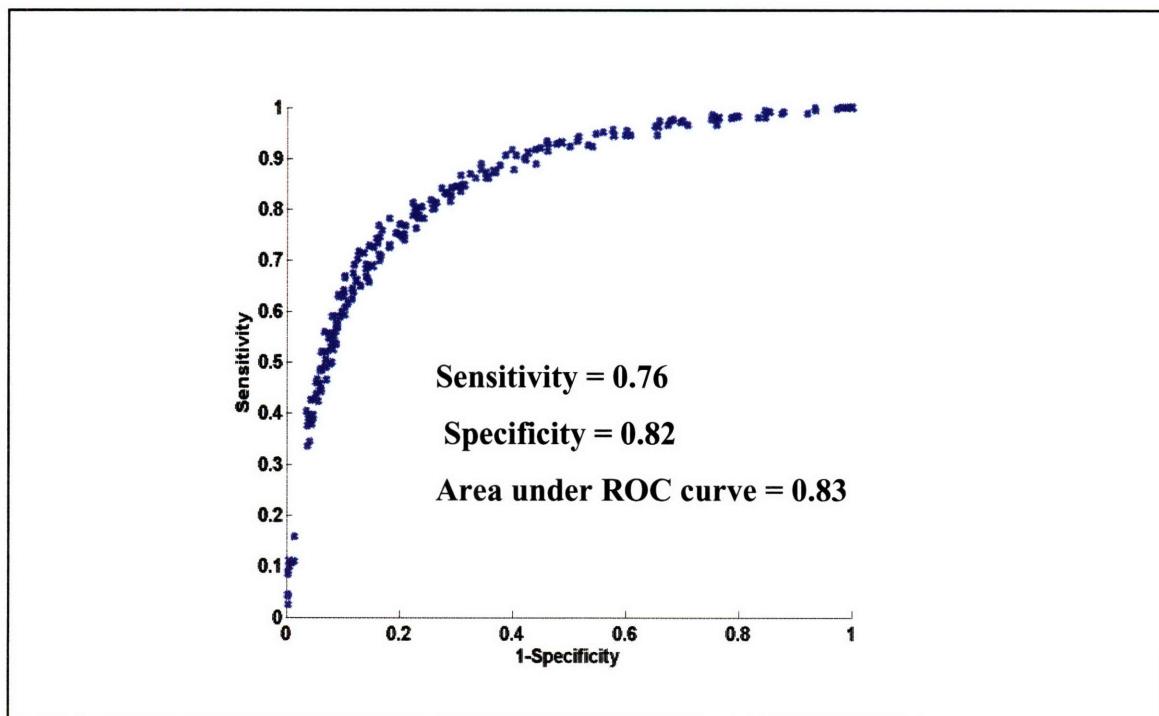


Figure 6-8: ROC for predictive algorithm two hours to vasopressor onset

6.3.3. Optimal feature selection and analysis

The four physiologic parameters that were studied include heart rate (HR), systolic blood pressure (SBP), mean blood pressure (MBP), and estimated cardiac output (ECO). The present study allows for an analysis into the informativeness of each trend towards the prediction of future instability. To determine the level of informativeness of each trend, different trend combinations were utilized in the sample queries to be submitted to the physiologic similarity algorithm. The informativeness of each trend combination was evaluated in terms of the overall specificity and sensitivity of the predictive algorithm to identify hemodynamic instability. As shown in Table 6-4, the systolic blood pressure was the most informative single trend parameter for predicting deterioration. The mean blood pressure as a single parameter had more specificity than systolic blood pressure, but less sensitivity. The best performance was achieved by combining heart rate, systolic blood pressure, and estimated cardiac output. The addition of mean blood pressure did not improve the ROC. As can be seen in the ROC included in Figure 6-9, estimated cardiac output (ECO) improves the overall classifier performance when combined with other parameters.

Table 6-4: Informativeness of different trend parameters

Feature	Sensitivity	Specificity	Area under ROC
SBP	69	79	0.74
MBP	63	82	0.72
HR	63	62	0.61
ECO	52	73	0.62
SBP + HR	77	73	0.79
SBP + HR + MBP	73	79	0.79
SBP + HR + ECO	76	82	0.83
SBP + HR + MBP + ECO	77	81	0.83

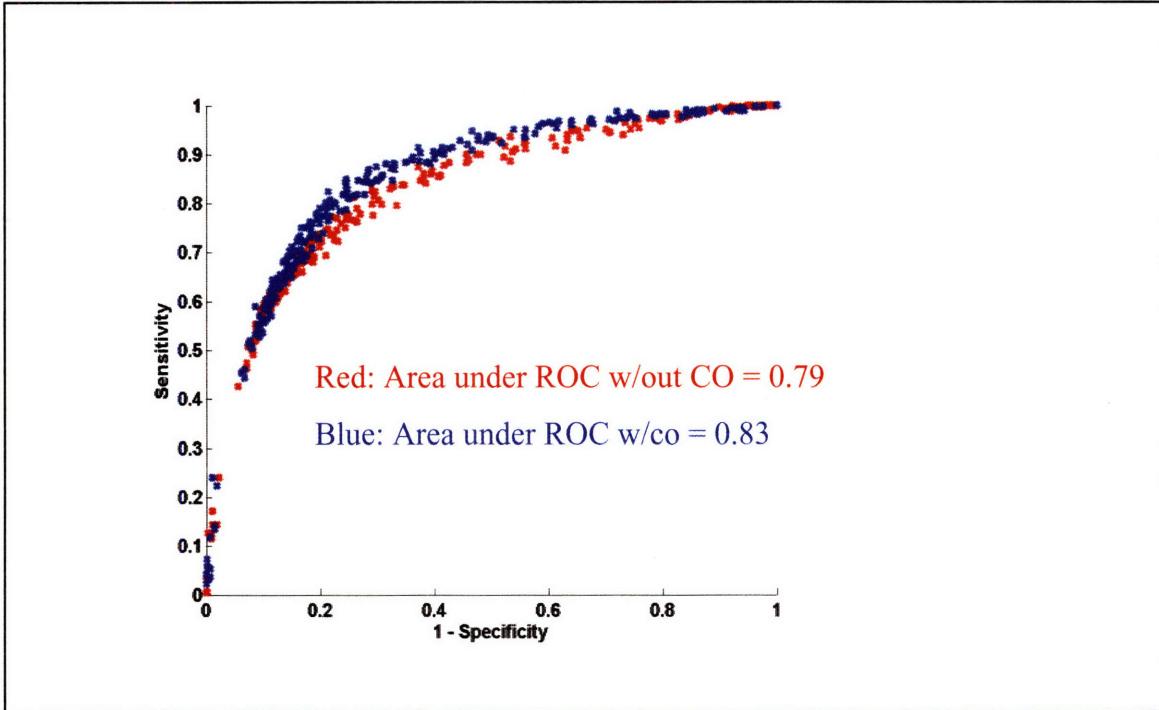
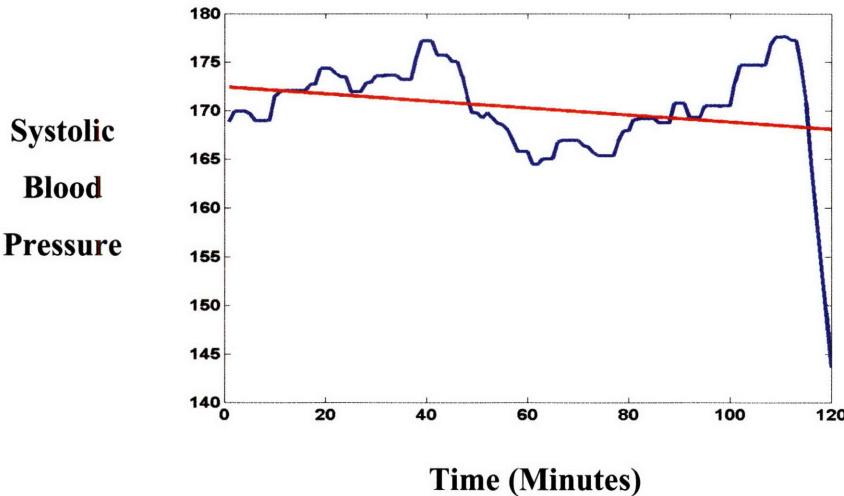


Figure 6-9: Effects of ECO on ROC

6.3.4. Comparison of physiologic similarity algorithm with conventional time domain feature descriptors

The physiologic similarity algorithm employs wavelet feature extraction to characterize a times series at multiple scales. The multiscale feature extraction technique captures the structure of the multiscale dynamics of a time series. Classical time domain techniques that can characterize a time series in terms of its mean and slope may be simpler alternatives. Thus, the test segments were characterized in terms of the trend mean and slope. The mean value was determined over a test segment using the sample trend mean over the window of interest (see Figure 6-10). The slope was found using least-squares regression fitting functions from the MATLAB statistical toolbox.



Blue line represent original systolic blood pressure in patient over 120 minutes, and the red line represents a least-squares regression best fit line.

Figure 6-10: Example of line regression fitting.

The mean and slope parameters were determined for the heart rate, blood pressure (systolic/mean/diastolic), and estimated cardiac output. The slope is defined as unit change per minute, and is denoted in this thesis as Δ . The class separation between the time domain parameters was assessed by comparing the reference library segments representative of stable and unstable hemodynamics. The definition of class separation, D between a parameter, P_u , from an unstable class and parameter, P_s , from a stable class is given as:

Equation 20

$$D(P_s, P_u) = \frac{abs(m_s - m_u)}{\sigma_s + \sigma_u}$$

In Equation 20, $m_{u,s}$ is defined as the mean value of the parameter for a given class (*U or S*), and $\sigma_{u,s}$ is the standard deviation of the parameter for that class. The class separation

of the different parameters is presented in Table 6-5. The most significant difference between the stable and unstable classes was seen with the mean systolic blood pressure (separation = 0.575). The mean blood pressure was also significantly different between the two classes. The mean heart rate was not as well separated between the two classes. The slopes of the mean and systolic blood pressures, (Δ SBP, Δ MBP), were significant different between the two classes. The slope of the estimated cardiac output (Δ ECO) had only weak separation between the two classes. It is important to consider that this slope is estimated over an entire two-hour window. Thus, there may be changes in a trend that occur over shorter or longer time scales that would not be captured by the two-hour slope estimate.

Table 6-5: Time Domain Parameter Characterization

	Unstable	Stable	Separation
HR	92.1 +/- 19.9	84.4 +/- 14.1	0.2274
ΔHR	-0.009 +/- 0.10	0.003 +/- 0.07	0.0715
SBP	104.12 +/- 19.43	126.26 +/- 19.04	0.5752
ΔSBP	-0.067 +/- 0.17	0.0197 +/- 0.13	0.2959
MBP	71.74 +/- 11.7	84.19 +/- 12.84	0.5076
ΔMBP	-0.04 +/- 0.11	0.02 +/- 0.08	0.2846
DBP	54.65 +/- 9.22	61.88 +/- 10.72	0.3625
ΔDBP	-0.02 +/- 0.08	0.01 +/- 0.08	0.2576
ΔECO	-0.03 +/- 0.11	-0.01 +/- 0.07	0.1425

To compare the performance of the physiologic similarity algorithm with a “simpler” time domain alternative, the time domain features (mean and slopes of respective parameters) were incorporated into an overall feature vector. The simple time domain feature vector replaced the wavelet symbol feature vectors for the queries as well as the reference libraries. In a manner similar to what was described in Figure 6-5, the N most

similar segments based upon time domain features are retrieved for a given query record and utilized to predict the likelihood of deterioration.

As can be seen in Figure 6-11, the time-domain “simplified” classifier demonstrated a sensitivity of 0.74 and specificity of 0.78 in predicting hemodynamic deterioration. The overall performance was moderately worse than the wavelet method (sensitivity of 0.78 and specificity of 0.82). The results suggest that simple time-domain techniques are capable of having reasonable performance in predicting hemodynamic deterioration two hours prior to the point of deterioration. The signal characteristics at a gross level in terms of mean and slope are informative in the predictive model that was evaluated. The superior performance of the wavelet technique in terms of both sensitivity and specificity suggests that different ICU records may share some signal dynamics at multiple time scales that are similar to one another. Such dynamics can be characterized using a wavelet representation since wavelets resolve a signal into several scales. The mean and slope of a trend at only one scale would be insufficient for such a characterization.

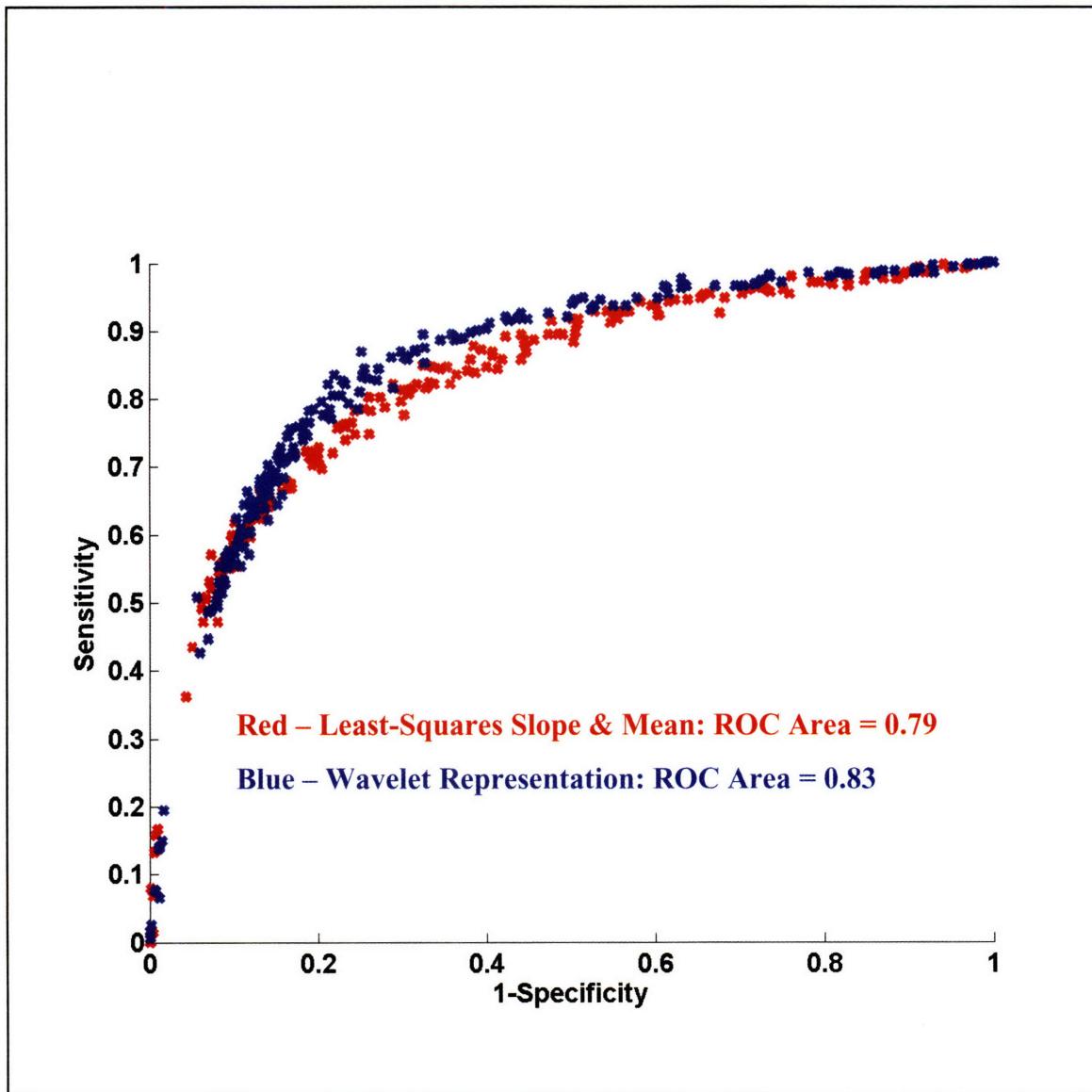


Figure 6-11: ROC of time-domain regression and wavelet-symbolic representations for the prediction of hemodynamic instability

In Figure 6-12, the wavelet-based instability index gradually increases from $t = 200$ minutes to $t = 330$ minutes. At $t = 360$ minutes, a vasopressor is started in response to the deterioration. The signal dynamics between $t=250$ minutes and $t=370$ minutes are similar to dynamics seen in the highest ranked reference segment retrieved from the unstable library. However, by examining the “best-fit” line approximations of the last 120 minutes of the systolic blood pressure trend from the query record and the retrieved record in Figure 6-13, we see that the best-fit line does not adequately preserve the similarity of these two signals.

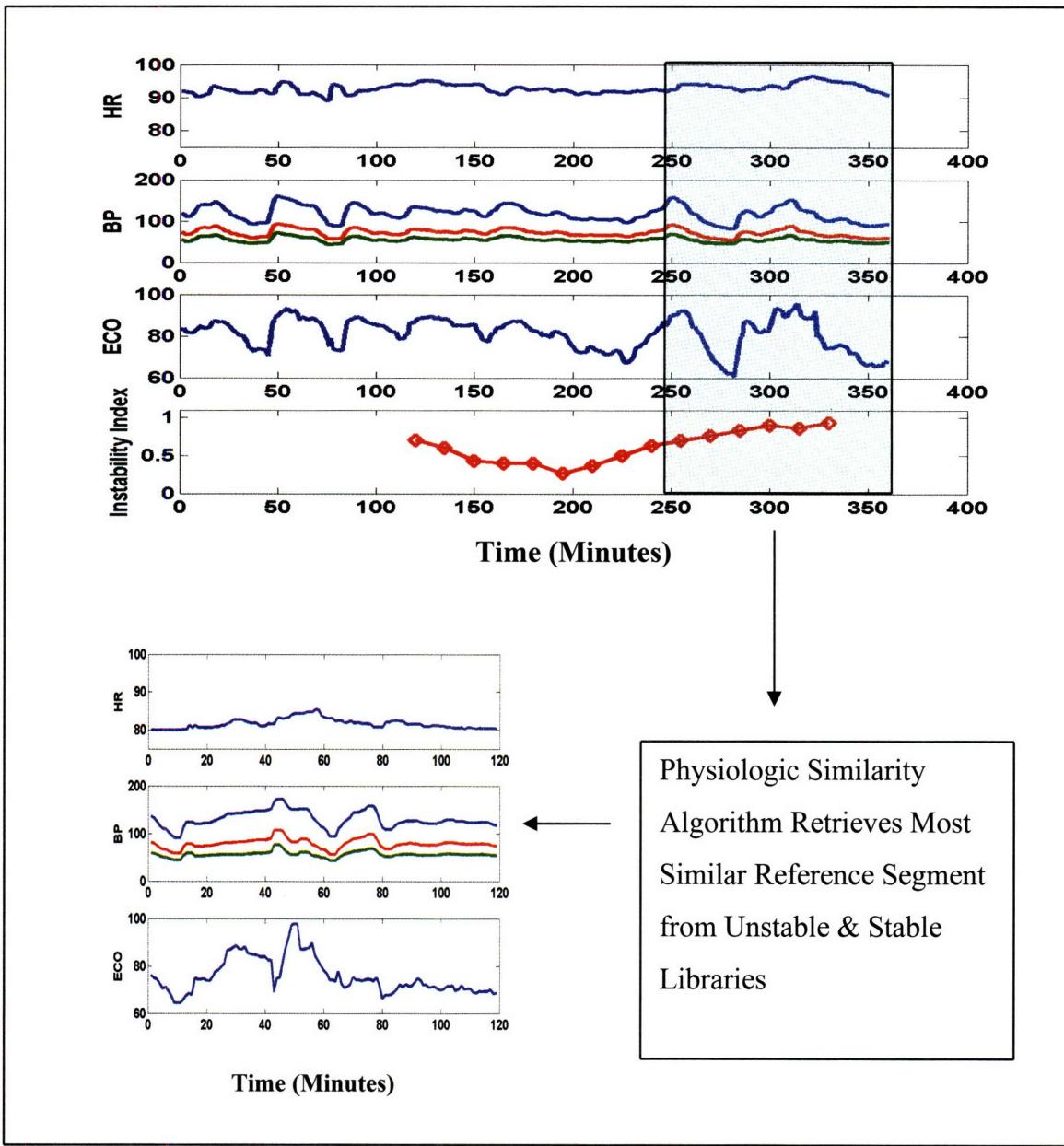


Figure 6-12: Example of similar multiscale dynamics in query and retrieved data segments

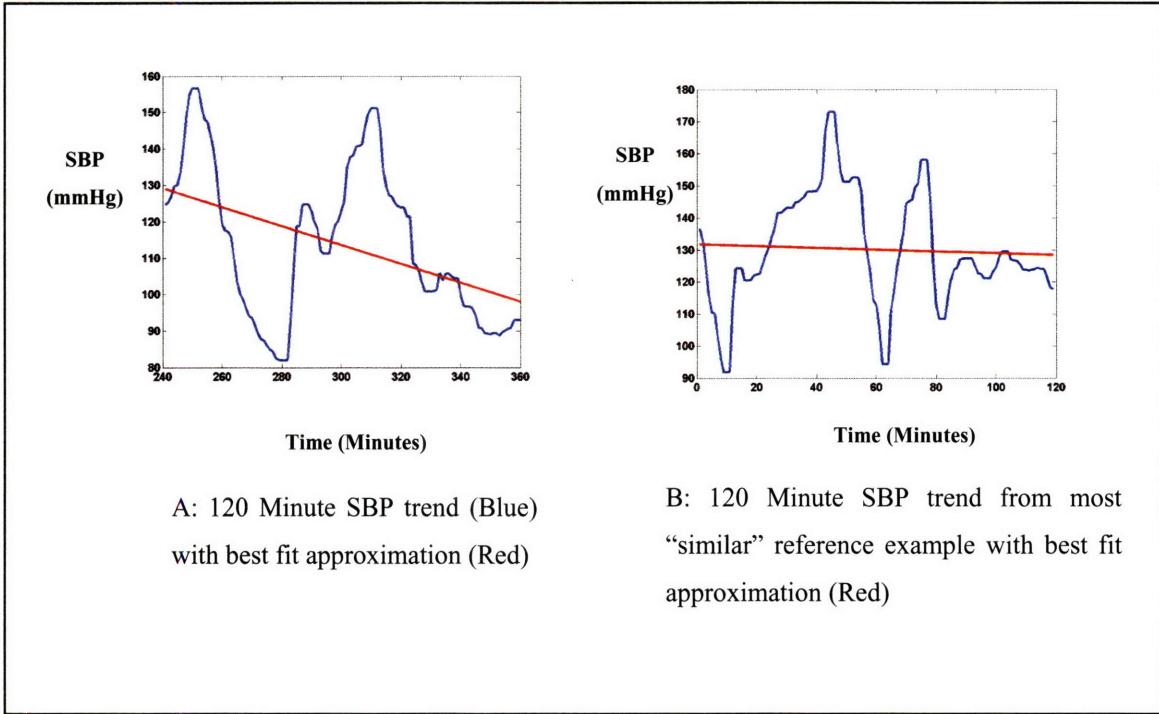


Figure 6-13: Best-fit lines representative of two different segments

Physiological Interpretation of Wavelet-Based Symbolic Feature Vector

The wavelet methodology used in this thesis characterizes physiologic trends at multiple time scales in terms of the trend dynamics at each respective time scale. Furthermore, the wavelet method also produces an approximation or “average” at one time scale, thus characterizing the static value of a trend. Physiologic homeostasis requires that the static values of certain physiologic variables not exceed certain bounds. For example, if the mean blood pressure is 30 mmHg, the body cannot adequately perfuse all end organs. Consequently, the static characterization of a physiologic trend would be clinically important to characterizing the hemodynamic stability of a patient. The relative importance in the static and dynamic wavelet symbols as input features for the physiologic similarity algorithm was evaluated using Equation 20. The class separation of the different wavelet symbols is plotted in Figure 6-14. As can be seen, certain wavelet symbols allow much greater class discrimination than others. In particular, the as can be inferred from Table 6-6 and Table 6-7, the static wavelet symbols have a greater class

discrimination (mean SBP and mean MBP) than the dynamic wavelet symbols. However, the dynamic wavelet symbols still do have significant class discrimination utility. The wavelet symbols that indicate dynamics of the ECO parameter had the best class separation among the dynamic wavelet symbols.

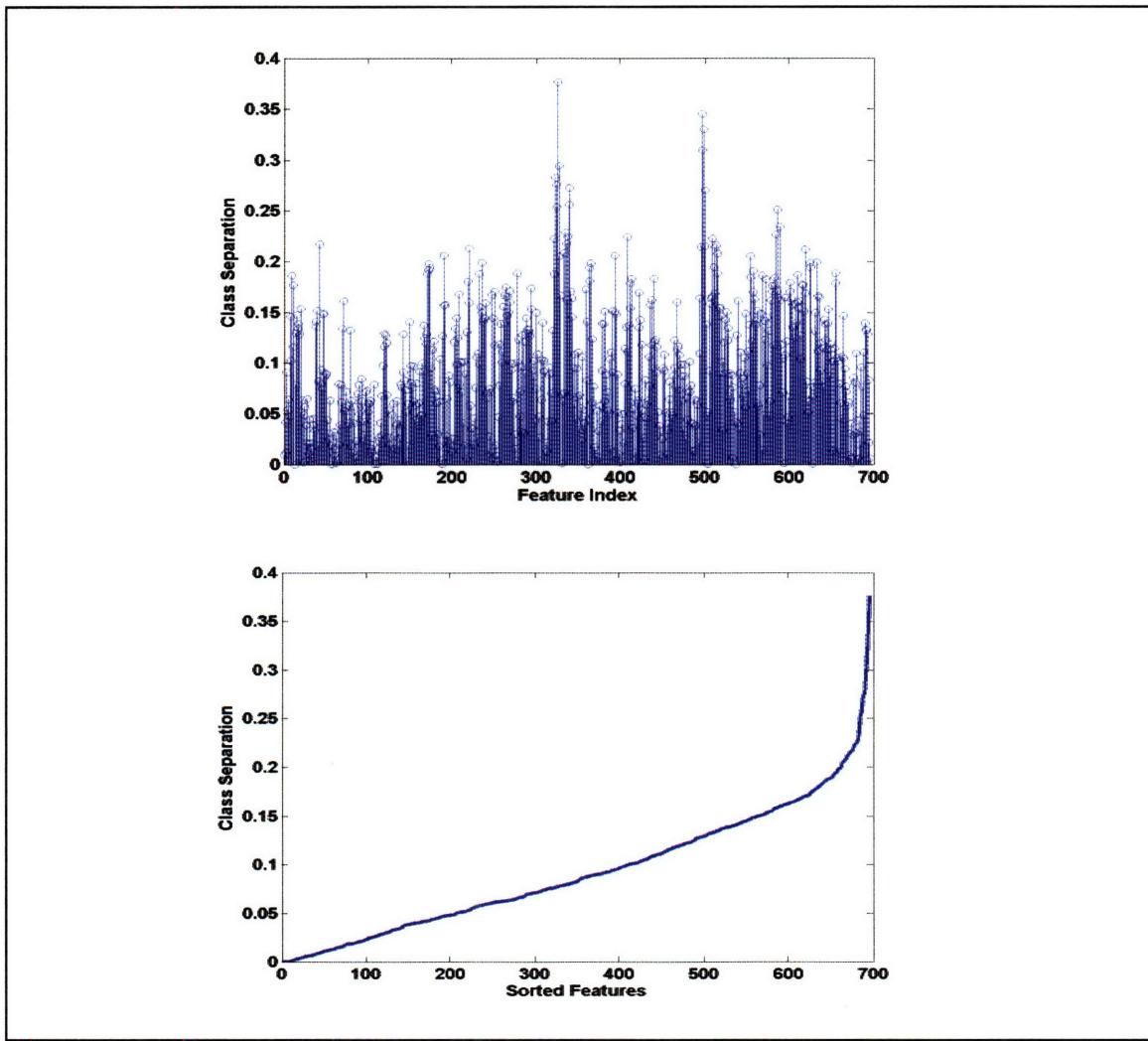


Figure 6-14: Wavelet symbol class separation

Table 6-6: Class separation of static(approximation) wavelet symbols

Class Separation (D)	Parameter	Average value over 60 minute window length
0.3762	SBP	89.9
0.3451	MBP	60.4
0.3300	MBP	62.8
0.3098	MBP	57.9
0.2945	SBP	93.8
0.2830	SBP	78.2
0.2756	SBP	86.0
0.2726	SBP	140.6
0.2704	MBP	65.2
0.2560	SBP	144.5

Table 6-7: Class separation of dynamic wavelet symbol

Class Separation	Wavelet Parameter	Scale (Length of wavelet in minutes)	Change over Wavelet Length
0.2512	ECO	16	-10.6000
0.2345	ECO	16	-7.3000
0.2268	ECO	16	-11.7000
0.2243	MBP	16	-16.3000
0.2174	HR	8	0.1000
0.2129	SBP	8	4.0000
0.2120	ECO	32	-7.3000
0.2061	MBP	8	1.7000
0.2059	SBP	4	2.2000
0.2053	ECO	8	-10.2000

In Figure 6-15, the hemodynamic instability index smoothly increases as the patient's hemodynamics deteriorate. This example is particularly useful to demonstrate that the percent change in the instability index is greater in magnitude than any one commonly measured parameter such as heart rate or blood pressure. The instability index changes from approximately 0.3 to 1.0 over 4 hours (200% increase). The blood pressure, heart rate, and estimated cardiac output have individual percent changes of less than 50% each. Thus, the instability index may be useful as a single derived numeric that is sensitive and specific for hemodynamic state changes. Fusion of several physiologic measurements into one clinically useful variable may help to alleviate information overload in the ICU.

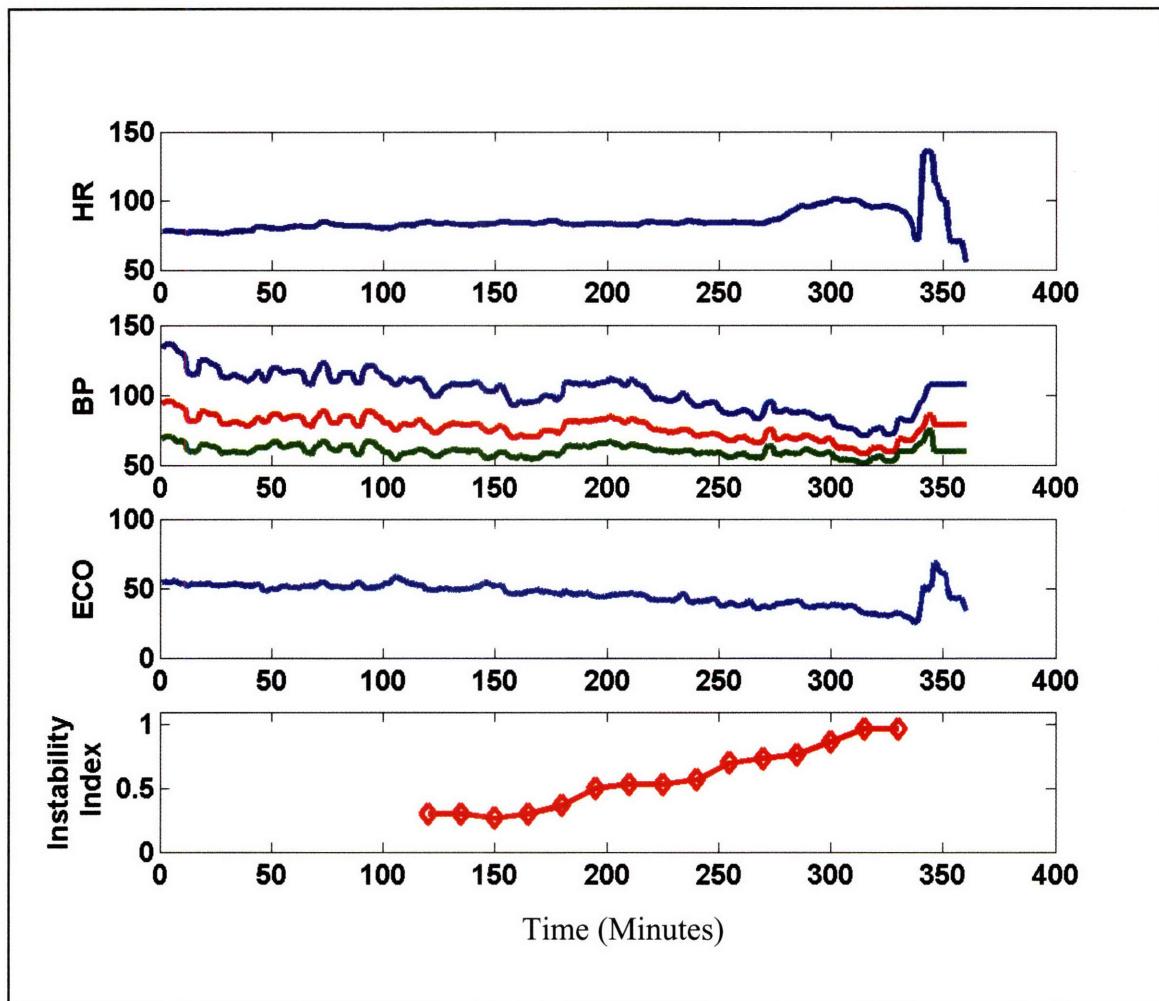


Figure 6-15: Hemodynamic instability index with progressive deterioration over four hours

7. Conclusions

In this final chapter, the major contributions of this thesis are summarized. In particular, we have introduced and characterized a major new ICU database, MIMIC-II, to support the development of next-generation patient monitoring systems. Examples of novel clinical studies that can be facilitated with MIMIC-II were also provided. A novel trend similarity algorithm based on classic information-retrieval models was introduced and evaluated. Finally, the trend similarity algorithm was utilized as a hemodynamic instability index that was predictive of the need for significant therapeutic interventions in unstable ICU patients. This chapter is organized in the following manner: in the next section, we review the main contributions of this thesis. Then, the final section of the thesis discusses future extensions of this work that may be explored.

7.1. Major Thesis Contributions

7.1.1. Development of a major new ICU database: MIMIC-II

The MIMIC-II (Multiparameter Intelligent Monitoring in Intensive Care) database was created through a partnership between industry and academia to support the development of advanced ICU patient monitoring systems that can substantially improve the efficiency, accuracy and timeliness of clinical decision making in intensive care. Previous efforts in developing physiologic and clinical databases have been limited by the selection of only a partial subset of all the data generated over a patient's stay. MIMIC-II is a far more comprehensive database that includes several continuous channels of high resolution physiologic waveforms such as ECG and blood pressures, vital signs, monitoring alarms, continuous therapeutic intervention profiles of each patient's stay, laboratory results, fluid balance, nursing progress notes and discharge summaries. MIMIC-II currently includes over 17000 ICU patient records (with clinical data from one

hospital. In Chapter 3, we described how MIMIC-II was created and characterized its contents. We have characterized the MIMIC-II ICU patient population with respect to patient acuity, availability of physiologic measurements, frequencies of significant interventions, demographics, and clinical problems (ICD-9 codes). Our research group is developing “gold-standard,” UMLS-based annotations of clinically significant hemodynamic events in patient records using clinical expertise from a committee of clinicians. We are also developing automated methods of de-identifying records per HIPAA requirements. The ultimate goal will be to disseminate significant portions of the MIMIC-II database so that other researchers can utilize this resource to support their research.

7.1.2. Clinical studies with the MIMIC-II database

In Chapter 4, we presented examples of two clinical studies that can be supported with MIMIC-II. In the first study, we compared the agreement between non-invasive (NIBP) and invasive blood pressure measurements that are simultaneously recorded in MIMIC-II patients. The availability of such a massive dataset that was acquired under realistic clinical conditions allows for a more comprehensive analysis of agreement between NIBP and ABP under different physiologic states. In particular, NIBP and ABP were assessed prior to the start of vasopressor therapy as well as prior to acute elevations of creatinine (as a marker of renal failure). The results of this retrospective study suggest that NIBP systolic blood pressure is an inadequate estimate of the true underlying arterial systolic blood pressure. The results were consistent with theoretical model-based studies of NIBP as well as animal studies.

The second clinical study involved analyzing the diurnal variations in hemodynamic variables (blood pressure, heart rate) of patients as well as hemodynamically significant interventions in the ICU. The results of the study suggest that there may be statistically significant differences in the response of clinicians to hemodynamic deterioration as a function of time of day in the ICU. In particular, patients’ blood pressures were allowed to deteriorate further in the early morning in comparison to the afternoon or evening. Furthermore, there were significant differences in the likelihood of starting a new

vasopressor as function of time of day. Thus, this clinical study has introduced a novel high throughput histogram-based method that may highlight clinically important patterns in ICU care that warrant further investigation.

7.1.3. Development of a time series similarity algorithm using wavelet-based symbolic representations

In Chapter 5, a new method for assessing similarities between multiparameter time series was described. We introduced a novel wavelet-based symbolic transformation that allows for the use of information retrieval algorithms that are popular in the document indexing research community. We characterized time series using wavelet feature extraction and then mapped the wavelet coefficients to symbolic representations. We then integrated the symbols into a term-frequency vector that could be processed using classical information-retrieval algorithms. We evaluated the performance of the novel time series similarity algorithm on several different synthetic data sets. The algorithm that was described has several attractive features: it can be implemented in an unsupervised framework, it is computationally efficient, and its accuracy is comparable to state-of-the-art algorithms that require significant training data.

7.1.4. Novel algorithm to predict hemodynamic instability in ICU patients

In Chapter 6, the wavelet-based similarity metric was utilized for identifying physiologic patterns that may be predictive of hemodynamic deterioration in ICU patients. The predictive algorithm identified multiscale dynamics in heart rate, blood pressure, and estimated cardiac output (ECO) as salient features to distinguish between hemodynamic stability and impending hemodynamic deterioration. The algorithm was able to “discover” unique multiparameter signatures in an unsupervised framework. The K-nearest neighbors algorithm was used to identify “similar” hemodynamic patterns for a library of ICU records with hemodynamic time series and their respective medication administration profile.

The classifier that was developed from the new similarity metric had a relatively high specificity (~0.8) and sensitivity (~0.8) in predicting the need for a vasopressor intervention at least two hours before it is given. The difficulty in attaining a higher sensitivity is also due to the predictive nature of the classifier. The performance analysis of the predictive algorithm was limited by the definition of “gold-standard” labels that were used to adjudicate algorithm decisions as “true” and “false.” The use of medication profiles for defining the “gold-standard” labels implies that clinicians are 100% sensitive to hemodynamic deterioration and do not initiate therapy with a vaso-active medication when it is not necessary to do so. However, as shown in Chapter 4, ICU clinical vigilance is suboptimal and medication administration decision may at times be inconsistent with the physiologic needs of patients. To our knowledge, there is little research in developing automated algorithms that can forecast a blood-pressure drop hours before it happens. It is hoped that other investigators can utilize MIMIC-II to develop similar algorithms and --in the process---refine the “gold-standard” label for hemodynamic “instability” and “stability.”

The framework presented here can accommodate additional physiologic and clinical signals that can perhaps improve the sensitivity. For example, fluid-balance time series as well as clinical laboratory measurements may further aid in identifying patterns of impending hemodynamic deterioration.

7.2. Future Work

7.2.1. Next-Generation clinical database development

Advances in hardware and software technologies have facilitated the development of the MIMIC-II database. The evolution in technology is likely to continue and will allow for the development of even larger, and more comprehensive databases in the future.

The development of massive storage systems that can store terabytes of data on a single desktop machine will facilitate the acquisition of high resolution physiologic data over

extended monitoring intervals. Furthermore, such data acquisition systems are likely to be networked to monitoring networks with gigabit data transfer rates; thus allowing for transmission of even more physiologic data on a single network. Archiving systems that can be networked to a remote clinical environment in different geographies may be possible in the future. Such advances may allow for the development of a richer database that is representative of several different patient populations at disparate clinical sites. A more diverse database is less likely to reflect the clinical practice of only one institution. Thus, physiologic and clinical patterns in ICU data that are detected may be more applicable to broad patient populations.

Other clinical environments outside of the ICU may pose their own unique challenges. For example, interpretation of physiologic and clinical data from the emergency department (ED) has its own unique challenges—patients are usually monitored with only non-invasive technologies, monitoring intervals are typically shorter, and the data may be far noisier. The operating room (OR) may be a data rich environment with a myriad of monitoring devices attached to a patient during a highly invasive surgery. Anesthesiologists are responsible for closely monitoring the physiologic stability of the patient at all times. Such a need motivates the development of automated algorithms that may support the clinical decision making in the OR. Thus, development of a database that can support algorithm development for OR and ED monitoring systems can particularly benefit from a database like MIMIC-II with a focus to those environments.

The bedside monitoring data in MIMIC-II was archived by a customized system supplied by a medical equipment vendor. Such collaboration is necessary in order to translate the proprietary native data formats into an open data standard such as WFDB. Creation of a common data standard would greatly facilitate the ability of researchers to utilize databases such as MIMIC-II in their algorithm development and validation process. By working with a common database, it would be possible to compare the relative merits of different algorithms for patient monitoring. Furthermore, common databases and standards also allow researchers to use common toolsets for visualizing and annotating a database. It has proven to be extremely challenging to annotate a large-scale database

such as MIMIC-II by a limited number of clinicians. Development of web-enabled visualization and annotation tools may allow for the creation of a greater community of researchers and clinicians with interests in annotating MIMIC-II data. Thus, it may be possible to leverage such a community in a successful manner similar to the open-source movement that has been responsible for innovative software, and more recently---the wikipedia.

Along with the aforementioned information storage and networking technologies, measurements technologies are likely to advance to the point of facilitating the acquisition of richer physiological datasets. For example, the MIMIC-II database contains commonly reported laboratory measurements such as blood chemistry and cell counts. In the era of bioinformatics, a patient's genomic, proteomic, and metabolomic profile may be measured with the ease in the future. For example, septic shock is thought to be a manifestation of a systemic inflammatory response to a bacterial infection. The dynamics of sepsis physiology may be dependent on a patient's inflammatory response pattern which may be in part determined by a have a certain genotype [69]. Identifying physiologically important trends based upon genomic, proteomic, and conventional physiologic measurements would be greatly facilitated by a database that contains such information on a significant number of ICU patients.

The future extensions to physiologic database development will certainly pose ethical and legal challenges with respect to dissemination of clinical data to a wide community of users while simultaneously protecting a patient's confidential medical information. Research in automatically de-identifying medical data is ongoing [12]. Ultimately, governmental agencies such as the NIH will need to set guidelines and standards that protect patients and their caregivers while facilitating access to data that researchers can use to develop technologies that advance medical care.

7.2.2. Future clinical studies with the MIMIC-II database.

In chapter 5, two clinical studies were described as examples of retrospective research that was enabled by the MIMIC-II database. The richness of data in MIMIC-II can

facilitate numerous other clinical studies. While the focus of this thesis has been in the domain of hemodynamic monitoring, other physiological systems can be studied. For example, the pulmonary system is of particular importance in the significant number of ICU patients that are ventilated. In [30], MIMIC-II was utilized to study acute respiratory distress syndrome (ARDS) in ICU patients. In particular, physiological and non-physiological factors, such as ventilator settings, were identified that may play causal roles in the development of ARDS.

One of the most unique aspects of MIMIC-II includes the availability of high resolution waveform data. In [72], we utilized the arterial blood pressure waveforms and intermittently measured cardiac output measurements in the MIMIC-II database to evaluate several different algorithms that derive CO from ABP waveforms. The availability of large volumes of data along with variability in physiologic states in MIMIC-II can facilitate several different research studies with physiologic waveforms. Besides cardiac output, there are several other invasive measurements made in the ICU that can perhaps be estimated with less invasive monitoring modalities. For example, the pulmonary artery wedge pressure as measure of left-heart filling is valuable in evaluating the volume status of a patient but typically requires invasive right heart catheterization. Utilization of the ABP waveform along with physiologic signal processing may allow for the development of a less-invasive, continuous PAWP estimate. Such a research study could be facilitated by the MIMIC-II database.

False arrhythmia alarms are among the most significant challenges plaguing the value of ICU monitors. The development of novel alarming algorithms that are both sensitive and specific requires access to a large database of annotated physiologic waveforms from ICU patients. The total number of monitoring hours available in MIMIC-II far exceeds any other database. Efforts are underway to utilize ECG waveforms in MIMIC-II to evaluate more advanced arrhythmia alarms. In particular, utilization of several physiologic waveforms such as ECG, arterial blood pressure, and photo-plethysmography may allow for the development of robust alarming algorithms.

7.2.3. Time series similarity metric

The time series similarity metric based on the wavelet-based information-retrieval model has many possible research extensions. In this thesis, the only wavelet basis function that was evaluated was the Haar function. The Haar function was chosen for its simple implementation and computational tractability. However, other basis functions with different smoothness properties may yield different results depending on the given application.

In this thesis, we focused on physiologic time-series trends with regularly sampling intervals (1 sample/minute). However, several clinical time series such as laboratory values are typically irregularly sampled. The symbolic information-retrieval model is still applicable to irregularly sampled time series. However, the wavelet feature extraction must be adapted to characterize irregularly sampled data. In an ICU monitoring application, this may be particularly challenging because the sampling frequency of a laboratory measurement may be suggestive of an underlying instability in a patient. Thus, if five arterial blood gasses (ABG) are sampled in a day in one patient, and only just ABG is sampled in another patient, one may infer that there are different concerns regarding the acid-base status of these two patients.

The symbols that have been generated reflect the value of a single wavelet coefficient derived from one time series. Furthermore, the symbol frequency feature vector that is generated reflects the frequency of occurrence of individual symbols. However, there are several real-world processes within and outside of the ICU that are best characterized by two or more time series. For example, the direction of change in a patient's estimated cardiac output while the pulmonary artery pressure is increasing may be more important in characterizing the physiologic state than characterizing each trend direction independently. Thus, an extension of our work may involve formulating a symbolic feature vector that characterizes the relative, simultaneous change in time series along with their individual dynamics. An advantage of using the information-retrieval model is that there is a rich literature in document processing for evaluating different term

frequency vector approaches. An analogy could be made to document retrieval methods where terms in the vector may consist of more than one word, rather than the simple one-word terms that are used as examples in Chapter 5.

The simple histograms and Parzen density estimators are considered non-parametric estimators of a probability density function. In utilizing the Parzen density estimators, widths for the Gaussian kernels were chosen based upon each wavelet coefficient's sample distribution. Suboptimal kernels may lead to inaccurate density estimates that over-fit the data. There are techniques for kernel variation to overcome such problems that can be explored [13].

A parametric model may allow for a more optimal estimation of the true wavelet coefficient density function. For example, a Gaussian mixture model of the wavelet coefficients can be estimated using the Expectation-Maximization (EM) algorithm. The EM algorithm is a technique for finding the maximum likelihood (ML) estimates of the parameters (mean, variance, prior probabilities) of a mixture model for a density function. The optimal number of mixtures can be determined using an information criterion such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion(BIC) [13]. Characterizing a distribution with a set of parameters may have important advantages. A parametric model may be more robust to outliers as well as allow for a significant data compression by storing feature vectors consisting only of model parameters.

7.2.4. Hemodynamic instability prediction

The similarity metric was used to develop an algorithm for the prediction of hemodynamic instability. There are several possible extensions in this area of research. In Chapter 6, only the one-minute heart rate and arterial blood pressure trends were utilized for predicting future deterioration. However, there are several other variables that may significantly improve the performance of the predictive algorithm. The fluid-balance data (IV fluid input and urine output) can contribute significant information about whether a patient may have underlying hypovolemia. Furthermore, developing a more

comprehensive model that includes other interventions may provide for a more complete physiologic characterization of a patient. For example, if a patient's blood pressure deteriorates after receiving an IV fluid bolus, one may infer that heart function (contractility) is poor. A more refined characterization of the patient's physiology may allow for the development of an algorithm that can also classify the instability pattern. For example, deterioration in blood pressure can be due to one of more of the following etiologies: sepsis, hypovolemia, or acute heart failure. Identifying the type of shock is important because the optimal therapy would be different. Thus, creation of "symbols" that correspond to physiologic responses to therapeutic interventions may allow for characterizing the similarity of ICU records in more novel and clinically meaningful frameworks. Real-time clinical decision support tools that can aid in timely therapy selection may ultimately improve outcomes in the ICU.

Bibliography

- [1] OT Abdala, G Clifford, M Saeed, A Reisner, GB Moody, I Henry, RG Mark: The Annotation Station: An Open-Source Technology for Annotating Large Biomedical Databases, *Computers in Cardiology*, 31: 681-684; IEEE Computer Society Press, September 2004.
- [2] Bishop, C.M. *Neural Networks for Pattern Recognition*, Oxford University Press. 1995.
- [3] Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1: 307–310, 1986.
- [4] A Bur, H Herkner, M Vlcek, C Woisetschlager, U Derhaschnig, G Delle Karth, AN Laggner, MM Hirschl. Factors influencing the accuracy of oscillometric blood pressure measurement in critically ill patients. *Critical Care Medicine*. 31(3):793-799, March 2003.
- [5] A Bur, M Hirschl, H Herkner, E Oschatz, J Kofler, C Woisetschlager, AN Laggner. Accuracy of oscillometric blood pressure measurement according to the relation between cuff size and upper-arm circumference in critically ill patients. *Critical Care Medicine*. 28(2):371-376, February 2000.
- [6] SD Carrigan, G Scott, M Tabrizian. Toward Resolving the Challenges of Sepsis Diagnosis. *Clin Chem*. 50: 1301-1314. 2004.
- [7] MC Chambrin, P Ravaux, C Chopin, J Mangalaboyi, P Lestavel, and F Fourrier. Computer-assisted evaluation of respiratory data in ventilated critically ill patients. *Int J Clin Monit Comput*, 6(4):211 {215, Dec 1989.
- [8] Cook, Suzanne F; WA Visscher, CL Hobbs, RL Williams, the Project IMPACT Clinical Implementation Committee. Project IMPACT: Results from a Pilot Validity Study of a New Observational Database. *Critical Care Medicine*. 30(12):2765-2770, December 2002.

- [9]Cropp AJ, Woods LA, Raney D, Bredle DL. Name that tone. The proliferation of alarms in the intensive care unit. *Chest*. 1994 Apr;105(4):1217-20.
- [10] Dawant, B.M. Uckun, S. Manders, E.J. Lindstrom, D.P. The SIMON project: model-based signal acquisition, analysis, and interpretation in intelligent patient monitoring. *IEEE Engineering in Medicine and Biology Magazine*. Volume 12, Issue 4. Pages 82-91. December 1993.
- [11] Donchin Y, Gopher D, Olin M, Badihi Y, Biesky M, Sprung CL, Pizov R, Cotev S. A look into the nature and causes of human errors in the intensive care unit. *Qual Saf Health Care*. Apr;12(2):143-7; 2003.
- [12] Douglass M, Clifford GD, Reisner A, Moody GB, Mark RG. Computer-Assisted De-Identification of the Free Text in the MIMIC II Database. In *Proceedings of Computers in Cardiology*. 2004.
- [13] RO Duda and PE Hart. *Pattern Classification*. John Wiley & Sons, 2000.
- [14] Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, 95, 14863–14868. 1998.
- [15] J Frassica, Frequency of Laboratory Test Utilization in the Intensive Care Unit and Its Implications for Large-Scale Data Collection Efforts. *J. Am. Med. Inform. Assoc.* 12:229-233. 2005.
- [16] Friesen GM, Jannett TC, Jadalloh MA, Yates SL, Quint SR, Nogle HT. A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Trans Biomed Eng*. 37:85-98, 1990
- [17] S Galhotra, MA DeVita, RL Simmons, A Schmid, members of the Medical Emergency Response Improvement Team (MERIT) Committee. Impact of patient monitoring on the diurnal pattern of medical emergency team activation. *Critical Care Medicine*. 34(6):1700-1706, June 2006.

- [18]LA Geddes, M Voelz, C Combs, D Reiner and CF Babbs. Characterization of the oscillometric method for measuring indirect blood pressure. *Annals of Biomedical Engineering*. Issue Volume 10, Number 6. Pages 271-280. November 1982.
- [19]A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, PCh. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C-K Peng, and H.E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet. Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, pp. e215-e220, 2000.
- [20]B Goldstein, J McNames, BA McDonald, M Ellenby, S Lai, Z Sun, D Krieger, RJ Sclabassi. Physiologic data acquisition system and database for the study of disease dynamics in the intensive care unit. *Critical Care Medicine*. 31(2):433-441, February 2003.
- [21]Gravlee GP, Quill TJ, Levine MI: Accuracy of four indirect methods of blood pressure measurement with hemodynamic correlations. *J Clin Monit*; 6:284–298, 1990.
- [22]Hall, A. & Graham Walton, G. Information overload within the health care system: a literature review. *Health Information and Libraries Journal*, 21,102–8. 2004.
- [23]JA Hanley, BJ McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29-36. 1982.
- [24]CW Hanson, BE Marshall. Artificial intelligence applications in the intensive care unit. *Critical Care Medicine*. 29(2):427-435, February 2001.
- [25]Hetland, M. L., “A survey of recent methods for efficient retrieval of similar time sequences”, In Mark Last, Abraham Kandel, and Horst Bunke, editors, *Data Mining in Time Series Databases*, World Scientific, 2004.
- [26]Imhoff M, Kuhls S. Alarm algorithms in critical care monitoring. *Anesth Analg*. 2006 May;102(5):1525-37.
- [27]Institute of Medicine. *To Err is Human; Building a Safer Health System*. Washington DC: National Academy Press; 1999.

- [28]Jalaleddine, S.M.S.; Hutchens, C.G.; Strattan, R.D.; Coberly, W.A. ECG data compression techniques-a unified approach. *IEEE Transactions on Biomedical Engineering*. Volume 37, Issue 4, Page(s):329 – 343, Apr 1990.
- [29]C. Jensema, MR Rovins. Word frequency in captioned television.
<http://www.cfv.org/caai/nadh138.pdf>
- [30]X Jia, M Saeed, RG Mark, A Malhotra, DS Talmor, Daniel. RISK FACTORS FOR ACUTE LUNG INJURY AND ACUTE RESPIRATORY DISTRESS SYNDROME IN PATIENTS MECHANICALLY VENTILATED > 48 HOURS IN THE ICU. *Critical Care Medicine*. 34(12) Abstract supplement:A126, December 2006.
- [31]D Jones , S Bates , S Warrillow , H Opdam, D Goldsmith, G Gutteridge, and R Bellomo. Circadian pattern of activation of the medical emergency team in a teaching hospital. *Critical Care*. Volume 9:R303-R306. 2005.
- [32]MA Kelley, D Angus, DB Chalfin, ED Crandall, D Ingbar, W Johanson, J Medina, CN Sessler, JS Vender. Critical Care Crisis in the United States. *Chest*. Volume 125, Issue 4. pages 1514-1517. April 2004.
- [33]E. Keogh. Exact indexing of dynamic time warping. In *Proceedings of Knowledge and Information Systems*. 2005.
- [34]Keogh, E., Chakrabarti, K., Pazzani, M. & Mehrotra, S. Locally adaptive dimensionality reduction for indexing large time series databases. In *Proceedings of ACM SIGMOD Conference on Management of Data*, pp 151-162. May 2001.
- [35]I Karhonen, M Van Gils, J Gade. The challenges in creating critical-care databases. *IEEE Engineering in Medicine and Biology Magazine*. Volume 20, Issue 3. Pages 58-62: May/Jun 2001.

[36]I Korhonen, J Ojaniemi, K Niieminen, M Van Gils, A Heikela, A Kari. Building the IMPROVE data library. *IEEE Engineering in Medicine and Biology Magazine*. Volume 16 Issue 6. Pages 25-32. Nov/Dec 1997.

[37]A Kumar, D Roberts, MD Daniel, KE Wood, B Light, JE Parrillo, S Sharma, R Suppes, D Feinstein, S Zanotti, L Taiberg, D Gurka, A Kumar, M Cheang. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical Care Medicine*. 34(6):1589-1596, June 2006.

[38]Tin H Kyaw. Master's Thesis. MIT. 2005.

[39]Landigan CP, Rothschild JM, Cronin JW, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. *N Engl J Med*. 351:1838-48; 2004

[40]ST Lawless. Crying wolf: false alarms in a pediatric intensive care unit. *Critical Care Medicine*. 22:981-985. 1994.

[41]JR Le Gall, P Loirat, A Alperovitch, P Glaser, C Granthil, D Mathieu, P Mercier, R Thomas, D Villers. A simplified acute physiology score for ICU patients. *Critical Care Med*. 12(11):975-7. November 1984.

[42]S. Lemeshow and J. R. Le Gall. *Modeling the severity of illness of ICU patients. A systems update*. JAMA, 272: 1049 - 1055. Oct 1994.

[43]J Lin, Experiencing SAX: a novel symbolic representation of time series. In *Proceedings of Data Mining and Knowledge Discovery*. 2007.

[44]Lin, J., Keogh, E., Lonardi, S., and Chiu, B. A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*. 2003.

[45]P Lio. Wavelets in bioinformatics and computational biology: state of art and perspectives. *Bioinformatics*. 19(1):2-9. January 2003.

- [46]P. Marino. The ICU Book. Lippincott Williams & Wilkins; 2nd edition 1998.
- [47]V Megalooikonomou, Q Wang, G Li, C Faloutsos. A multiresolution symbolic representation of time series. Proceedings of 21st International Conference on Data Engineering. Pages 668-679. 2005.
- [48]Meredith C, Edworthy J. Are there too many alarms in the intensive care unit? An overview of the problems. *Journal of Advanced Nursing*. 21:15-20. 1995.
- [49]Mitchell TM. Machine learning. McGraw-Hill,1997.
- [50]Moody GB, Mark RG, Zoccola A, Mantero S.. Derivation of Respiratory Signals from Multi-lead ECGs. In *Proceedings of Computers in Cardiology*, vol. 12, pp. 113-116 1985.
- [51]GB Moody, RG Mark. A database to support development and evaluation of intelligent intensive care monitoring. In *Proceedings of Computers in Cardiology*. Pages 657-660. 1996
- [52]FA Mora, G Passariello, G Carrault, JP Le Pichon. Intelligent patient monitoring and management systems: a review. *IEEE Engineering in Medicine and Biology Magazine*. Volume 12, Issue 4. Pages 23-33. December 1993.
- [53]Yves Meyer, editor. *Wavelets and Applications*. Springer-Verlag, 1992.
- [54] R. Mukkamala, AT Reisner, HM Hojman,, RG Mark, RJ Cohen. Continuous Cardiac Output Monitoring by Peripheral Blood Pressure Waveform Analysis. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, VOL. 53, NO. 3, MARCH 2006
- [55]C. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *International Conference on Computer Vision*, 1998.

- [56]T. Penzel, B Kemp, G Klosch, A Schlogl, J Hasan, A Varri, I Korhonen. Acquisition of biomedical signals databases. IEEE Engineering in Medicine and Biology Magazine. Volume 20, Issue 3. Pages 25-32. May/Jun 2001.
- [57]Pickering TG. The clinical significance of diurnal blood pressure variations: dippers and nondippers. *Circulation*. 81:700-702. 1990.
- [58]Rady MY, Smithline HA, et al. A comparison of the shock index and conventional vital signs to identify acute, critical illness in the Emergency Department. *Ann Emerg Med*. 24: 685-690; 1994
- [59]Razminia M, Trivedi A, Molnar J, et al. Validation of a new formula for mean arterial pressure calculation: the new formula is superior to the standard formula. *Catheter Cardiovasc Interv*. 63: 419–25. 2004.
- [60]Rennie, J. D. and Jaakkola, T. Using term informativeness for named entity detection. In *Proceedings of the 28th Annual international ACM SIGIR Conference on Research and Development in information Retrieval*. 2005.
- [61]Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, Peterson E, Early goal-directed therapy in the treatment of severe sepsis and septic shock. *New England Journal of Medicine*. 2001 Nov 8;345(19):1368-77.
- [62]Rothschild JM, Landrigan CP, Cronin JW, Kaushal R, Lockley SW, Burdick E, Stone PH, Lilly CM, Katz JT, Czeisler CA, Bates DW. The Critical Care Safety Study: The incidence and nature of adverse events and serious medical errors in intensive care. *Crit Care Med*. 33:1694–700. 2005
- [63]Saeed M, Lieu C, Raber G, Mark R G. MIMIC II: A massive temporal ICU patient database to support research in intelligent patient monitoring. *Computers in Cardiology*; 29:641–644. 2002.
- [64]Salton, G. and Buckley, C. Term-weighting approaches in automatic text retrieval. *Inf. Process. Manage.* 24, 5. Pages 513-523. August 1988.

- [65] Shabot MM, LoBue M, Chen J. Wireless Clinical Alerts for Physiologic, Laboratory and MedicationData. *Proc AMIA Symp*. pages 789-793. 2000.
- [66]MM Shabot. Automated data acquisition and scoring for JCAHO ICU core measures. In *Proceedings of American Medical Informatics Association Symposium*. Pages 674-678. 2005.
- [67]WC Shoemaker, CJ Wo, L Chan, E Ramicone, ES Kamel, GC Velmahos, H Belzberg. Outcome Prediction of Emergency Patients by Noninvasive Hemodynamic Monitoring. *CHEST*. Vol 120, Issue 2. pages 528-537. AUGUST, 2001
- [68] Schoenfeld, David A. PhD; Bernard, Gordon R. MD; for the ARDS Network. Statistical evaluation of ventilator-free days as an efficacy measure in clinical trials of treatments for acute respiratory distress syndrome. *Critical Care Medicine*. 30(8):1772-1777, August 2002.
- [69]L. Steinstraesser. Sepsis-New strategies with host defense peptides? *Critical Care Medicine*. 32(12):2555-2556, December 2004.
- [70]Stone HA, Stroock AD, Ajdari A. Engineering flows in small devices: Microfluidics toward a lab-on-a-chip. *ANNUAL REVIEW OF FLUID MECHANICS* 36: 381-411 2004.
- [71]Gilbert Strang and Truong Nguyen, *Wavelets and Filter Banks*. Wellesley-Cambridge Press. 1996.
- [72]Sun, J.X.; Reisner, A.T.; Saeed, M.; Mark, R.G. Estimating cardiac output from arterial blood pressurewaveforms: a critical evaluation using the MIMIC II database. In *Proceedings of Computers in Cardiology*, Pages 295–298. 2005.
- [73]Thoraval L, Carrault G, Schleich JM, Summers R, Van de Velde M, Diaz J. Data fusion of electrophysiological and haemodynamic signals for ventricular rhythm tracking. *IEEE Eng Med Biol Mag*. 1997 Nov-Dec;16(6):48-55.

[74]Tsien CL. TrendFinder: Automated detection of alarmable trends. Doctoral Thesis, Massachusetts Institute of Technology, July 2000.

[75]CL Tsien, JC Fackler. Poor prognosis for existing monitors in the intensive care unit. *Critical Care Medicine*. 25(4):614-619, April 1997.

[76]CL Tsien, JC Fackler. An annotated data collection system to support intelligent analysis of intensive care unit data. In: *Advances in intelligent data analysis*. Springer-Verlag, pp. 111-121. 1997.

[77]Ursino, M. Cristalli, C. A mathematical study of some biomechanical factors affecting the oscillometric blood pressure measurement. *IEEE Transactions on Biomedical Engineering*. Volume 43. Issue 8. Pages 761-778. August 1996.

[78]M Van Gils, H Jansen, K Nieminen, R Summers, PR Weller. Using artificial neural networks for classifying ICU patient states. *IEEE Engineering in Medicine and Biology Magazine*. Volume 16 Issue 6. Pages 41-47. Nov/Dec 1997.

[79]Vetterli, M. Herley, C. Wavelets and filter banks: theory and design. *IEEE Transactions on Signal Processing*. Volume 40, Issue 9. Pages 2207-2232. Sept. 1992.

[80]A. C. Yang, S. S. Hseu, H. W. Yien, A. L. Goldberger, and C. K. Peng, "Linguistic analysis of the human heartbeat using frequency and rank order statistics," *Phys. Rev. Lett.*, vol. 90, no. 10, Mar. 14, 2003.

[81]Y Zhang, CL Tsien. Prospective Trials of Intelligent Alarm Algorithms for Patient Monitoring. In *Proceedings of American Medical Informatics Association*. Page 1068. 2001.

[82]R Zhao, RG Mark, P Szolovits. A model-based expert system for interpretation of hemodynamic data from ICU patients. In *Proceedings of Computers in Cardiology*. Pages 585-588. 1996.

[83]Zimmerman, Jack E, D. Wagner, EA Draper, L Wright, C Alzola, WA Knaus. Evaluation of Acute Physiology and Chronic Health Evaluation III predictions of hospital mortality in an independent database. *Critical Care Medicine*. 26(8):1317-1326, August 1998.

[84]W Zong, GB Moody and RG Mark. Reduction of false arterial blood pressure alarms using signal quality assessment and relationships between the electrocardiogram and arterial blood pressure. *Journal Medical and Biological Engineering and Computing*. Volume 42, Number 5. Pages 698-706. September, 2004.

[85]W Zong, GB Moody, RG Mark. Reduction of False Blood Pressure Alarms by Use of Electrocardiogram Blood Pressure Relationships. In *Proceedings of Computers in Cardiology*. Pages 305-308. 1999.