INTEGRATION OF ARTIFACT DETECTION IN CLINICAL DECISION SUPPORT SYSTEMS

by

SHERMEEN NIZAMI

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in
partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL AND COMPUTER ENGINEERING

Ottawa-Carleton Institute of Electrical and Computer Engineering

Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario, Canada

© 2015

Shermeen Nizami

ABSTRACT

Artifact detection (AD) is a critical component of Clinical Decision Support Systems (CDSS) to assess the quality of data and issue clinical alarms with high specificity and low false alarm rates. An in-depth review of the literature reveals that, although many advanced AD algorithms have been developed, very few have been evaluated in a real-time environment or reached actual clinical implementation. Furthermore, no single AD algorithm is necessarily best-suited to serve the varying needs of a CDSS. This identifies the need to develop and evaluate a framework that supports the interoperability of a range of configurable AD algorithms that could be integrated with CDSS.

This thesis develops an AD framework to address six different shortcomings identified by the literature review. This framework supports the following six features: *(f.1) Flexibility to serve the needs of patient populations from different types of Critical Care Units (CCU) through generalization and customizability; (f.2) Reusability across multiple types of physiologic data harvested by different Original Equipment Manufacturer (OEM) monitors; (f.3) Standardized definitions of Signal Quality Indicators (SQI) that promote interoperability and comparison between independently developed components; (f.4) Reusability and scalability by mixing and matching several AD, Parameter Derivation (PD), and Clinical Event Detection (CED) components in various combinations; (f.5) Customizability to evaluate and compare the performance of multiple combinations of independently developed components on offline and potentially real-time patient data when integrated with clinical workflows; and (f.6) Standardized component interfaces that can potentially support real-time clinical implementation of AD.*

The developed framework provides for a unique test bed with the ability to create new AD composition pipelines by mixing and matching independently developed

or de-coupled AD components. The performance of these pipelines can be evaluated in integration with CDSS. The AD composition pipeline most suited to a particular clinical need can then be hard-coded and integrated into the clinical workflow for real-time implementation. A novel contribution of this research is a Common Reference Model (CRM) that defines standard interfaces to allow for interoperability of framework components. The CRM defines patient data attributes including data types and their SQI.

The framework is validated in a clinical research study. Four experiments are conducted to evaluate different AD composition pipelines. Each pipeline is evaluated in integration with an SpO₂ alarms generation CDSS. It is shown that by using the framework the lowest achievable false alarm rate is 39% at a sensitivity of 80%, when compared across all four experiments. The experiments serve to demonstrate and validate the six features supported by the AD Framework developed here. This evaluation remains independent of the performance of the particular CED used in the experiments.

ACKNOWLEDGEMENTS

I begin with praise to God Almighty, Allah, who has made this journey possible for me. I would like to express profound gratitude towards my thesis supervisors, Professor James R. Green and Professor Carolyn McGregor. They have both provided comprehensive guidance and tremendous support during my doctoral journey at Carleton. I am grateful to my thesis advisory committee Dr. Adrian Chan, Dr. Ian Marsland and Dr. Liam Peyton. Their valuable advice and mentoring helped shape my research and guide me through this important stage in my academic career. I am indebted to my research collaborators at the Children's Hospital of Eastern Ontario, Kim Greenwood, Dr. JoAnn Harrold, and Dr. Nick Barrowman. They dedicated many hours to discuss the engineering, clinical and statistical aspects of this research. I would like to acknowledge the generous and selfless support of my friends and colleagues, Mohamed Abdelazez, Amna Basharat, Georgina Blasetti, Mina S. Dehghanian, Sonia Hammouda, Maryam Kaka, Jessica Keats, Aly Khedr, Sylvia Reitmanova, and Romaissa Saadi. Each one of them has contributed to my success in unique and very special ways. Finally, I will remain forever indebted to my dearest family for their constant support and encouragement. My parents, Imrana Tariq and Khaled Nizami, empowered me with the love of learning. I dedicate this thesis to my sons, Hasan and Yahya, so that it inspires them to be lifelong learners.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	v
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ACRONYMS	xiii
STATEMENT OF ORIGINALITY	xiv
1 INTRODUCTION	1
1.1 RESEARCH SIGNIFICANCE	1
1.2 RESEARCH MOTIVATION	3
1.3 RESEARCH PROBLEM	4
1.4 RESEARCH METHOD	6
1.5 RESEARCH CONTRIBUTIONS	7
1.6 RESEARCH VALIDATION	8
1.7 DELIMITATIONS AND LIMITATIONS	9
1.8 THESIS ORGANIZATION	11
2 LITERATURE REVIEW	12
2.1 BACKGROUND	12
2.1.1 ARTIFACT DETECTION	12
2.1.2 CLINICAL EVENT DETECTION	14
2.1.3 CLINICAL DECISION SUPPORT SYSTEMS (CDSS)	16

2.2	METHODOLOGICAL REVIEW	19
2.2.1	THEME I. CRITICAL CARE UNIT.....	19
2.2.2	THEME II. PHYSIOLOGIC DATA SOURCE	20
2.2.3	THEME III. HARVESTED DATA	23
2.2.4	THEME IV. DATA ANALYSIS	24
2.2.5	THEME V. CLINICAL EVALUATION	27
2.2.6	THEME VI. CLINICAL IMPLEMENTATION	29
2.3	RESULTS	33
2.4	CONCLUSIONS	34
2.5	SUMMARY	36
3	ARTIFACT DETECTION FRAMEWORK	37
3.1	COMPONENT-BASED SYSTEM DESIGN	37
3.1.1	FRAMEWORK.....	37
3.1.2	COMPONENT	38
3.1.3	INTERFACE.....	38
3.2	AD FRAMEWORK DEVELOPMENT	39
3.2.1	FRAMEWORK.....	39
3.2.2	COMPONENTS	40
3.2.3	INTERFACES	41
3.3	COMPONENT CATALOGUE	47
3.3.1	AD _{DIL} DE-INTERLACE COMPONENT	47
3.3.2	AD _{THRESH} THRESHOLD COMPONENT.....	48

3.3.3	AD _{CONVERTSQI} CONVERT SQI COMPONENT	50
3.3.4	AD _{FUSESQI} FUSE SQI COMPONENT	51
3.3.5	AD _{DIFF} DIFFERENTIAL COMPONENT.....	52
3.3.6	AD _{MEDFILT} MEDIAN FILTER COMPONENT	53
3.4	RESEARCH CONTRIBUTIONS	54
3.5	ADVANTAGES AND CHALLENGES	56
3.6	CHAPTER SUMMARY	57
4	RESEARCH VALIDATION	59
4.1	DATA ACQUISITION METHODS.....	61
4.1.1	DATASET	61
4.1.2	DATA ANNOTATION	62
4.1.3	ALARM CATEGORIZATION	63
4.2	AD FRAMEWORK VALIDATION	64
4.2.1	PERFORMANCE EVALUATION	65
4.2.2	PDA COMPONENT FOR FRAMEWORK VALIDATION	65
4.2.3	CED COMPONENT FOR SpO ₂ ALARM GENERATION	66
4.2.4	EXPERIMENT 1	69
4.2.5	EXPERIMENT 2	70
4.2.6	EXPERIMENT 3	70
4.2.7	EXPERIMENT 4	71
4.3	RESULTS	75
4.3.1	DATA CHARACTERISTICS.....	75

4.3.2	EXPERIMENT 1 RESULTS	77
4.3.3	EXPERIMENT 2 RESULTS	81
4.3.4	EXPERIMENT 3 RESULTS	85
4.3.5	EXPERIMENT 4 RESULTS	90
4.4	DISCUSSION	98
4.5	CONCLUSIONS	101
5	CONCLUSIONS	103
5.1	RESEARCH SIGNIFICANCE	103
5.2	RESEARCH CONTRIBUTIONS	103
5.3	FUTURE WORK	105
APPENDIX A: REVIEW OF THEMES II – VI		107
APPENDIX B: AD COMPONENT INTERFACES		118
APPENDIX C: MASIMO VS. NELLCOR SPO ₂ COMPARISON		120
REFERENCES		126

LIST OF TABLES

TABLE I: THEME I. CRITICAL CARE UNIT (CCU).	22
TABLE II: NICU PATIENT DEMOGRAPHICS.	76
TABLE III: NICU OEM MONITOR SETTING.	76
TABLE IV: EXPERIMENT 1 RESULTS	79
TABLE V: EXPERIMENT 2 RESULTS FOR MedWW OF 5 AND 10.....	81
TABLE VI: EXPERIMENT 2 RESULTS FOR MedWW OF 12, 15 AND 20	82
TABLE VII: EXPERIMENT 3 RESULTS WHERE ADDIFF COMPONENT IS CONFIGURED WITH SQTHRESH = 6	85
TABLE VIII: EXPERIMENT 3 RESULTS WHERE ADDIFF COMPONENT IS CONFIGURED WITH SQTHRESH = 12	86
TABLE IX: EXPERIMENT 3 RESULTS WHERE ADDIFF COMPONENT IS CONFIGURED WITH SQTHRESH = 18	86
TABLE X: EXPERIMENT 4 RESULTS WHERE AD _{DIFF} COMPONENT IS CONFIGURED WITH SQTHRESH = 6	90
TABLE XI: EXPERIMENT 4 RESULTS WHERE AD _{DIFF} COMPONENT IS CONFIGURED WITH SQTHRESH = 12	91
TABLE XII: EXPERIMENT 4 RESULTS WHERE AD _{DIFF} COMPONENT IS CONFIGURED WITH SQTHRESH = 18.....	92
TABLE XIII: THE BEST POSSIBLE (Sn (% CHANGE IN Sn), FAR (% CHANGE IN FAR)) ACHIEVED IN ALL FOUR EXPERIMENTS, WHERE $Sn \geq 80\%$	100
TABLE XIV: THE BEST POSSIBLE (Sn (% CHANGE IN Sn), FAR (% CHANGE IN FAR)) ACHIEVED IN ALL FOUR EXPERIMENTS, WHERE $Sn \geq 75\%$	100
TABLE XV: AD COMPONENT INTERFACES USED IN EXPERIMENT 1	118
TABLE XVI: AD COMPONENT INTERFACES USED IN EXPERIMENT 2	118
TABLE XVII: AD COMPONENT INTERFACES USED IN EXPERIMENT 3	118
TABLE XVIII: AD COMPONENT INTERFACES USED IN EXPERIMENT 4	119

LIST OF FIGURES

FIGURE 1-1: PROGRESSIVE TIMELINE SHOWING ARTEMIS RESEARCH COLLABORATIONS	4
FIGURE 2-1: PATIENT DATA ACQUISITION BY AN OEM PHYSIOLOGIC MONITOR WITH BUILT-IN PREPROCESSING INCLUSIVE OF AD, FOLLOWED BY POST-PROCESSING AD, EITHER STANDALONE OR COUPLED WITH CED AND/OR PD, WITH OUTPUT VARIABLES RESULTING FROM THE ENTIRE DATA ANALYSIS.	13
FIGURE 2-2: HIGH-LEVEL VIEW OF THE ARTEMIS FRAMEWORK.	18
FIGURE 2-3: OVERVIEW OF THE SIX THEMATIC TAXONOMIES	21
FIGURE 3-1: STATIC MODEL OF THE MULTIVARIATE COMPONENT-BASED AD FRAMEWORK.	40
FIGURE 3-2: CRM SCHEMA FOR REQUIREMENT AND PROVISION INTERFACES.	45
FIGURE 3-3: SQL FIELDS DEVELOPED AS PART OF THE CRM	46
FIGURE 3-4: COLOR-CODED LEGEND OF THE CRM TUPLE	46
FIGURE 3-5: LOW-LEVEL CODE FOR THE AD _{DIL} COMPONENT	48
FIGURE 3-6: LOW-LEVEL CODE FOR THE AD _{THRESH} COMPONENT	49
FIGURE 3-7: LOW-LEVEL CODE FOR THE AD _{CONVERTSQL} COMPONENT	50
FIGURE 3-8: ILLUSTRATION OF AD _{FUSESQL} INTEGRATED WITH AD _{CONVERTSQL}	51
FIGURE 3-9: LOW-LEVEL CODE FOR THE AD _{FUSESQL} COMPONENT	51
FIGURE 3-10: LOW-LEVEL CODE FOR THE AD _{DIFF} COMPONENT	52
FIGURE 4-1: LOW-LEVEL COMPONENT CODE FOR CLINICAL EVENT DETECTION AND GENERATION OF SpO ₂ ALARMS	68
FIGURE 4-2: EXPERIMENT 1 FLOWCHART	71
FIGURE 4-3: EXPERIMENT 2 FLOWCHART	72
FIGURE 4-4: EXPERIMENT 3 FLOWCHART	73
FIGURE 4-5: EXPERIMENT 4 FLOWCHART	74
FIGURE 4-6: EXPERIMENT 1 RESULTS OF SN AND FAR PLOTTED AS PERCENTAGES FOR TUNABLE PARAMETER CED _{DT}	79
FIGURE 4-7: RADAR PLOT FOR EXPERIMENT 1 SHOWS SN VS. FAR FOR VARIOUS CED DECISION THRESHOLDS (CED _{DT}). EACH PAIR AROUND THE PERIMETER REPRESENTS (FAR, CED _{DT}). FAR VALUES DECREASE IN A COUNTER-CLOCKWISE DIRECTION.	80

FIGURE 4-8: EXPERIMENT 2 RESULTS OF SN AND FAR PLOTTED AS PERCENTAGES FOR EACH COMBINATION OF TUNABLE	
PARAMETERS (CED_{DT} , MED_{WW})	83
FIGURE 4-9: RADAR PLOT FOR EXPERIMENT 2 SHOWS SN VS. FAR FOR VARIOUS MEDIAN FILTER WIDTHS (MED_{WW}) AND CED	
DECISION THRESHOLDS (CED_{DT}). EACH 3-TUPLE AROUND THE PERIMETER REPRESENTS (FAR , MED_{WW} , CED_{DT}).....	84
FIGURE 4-10: EXPERIMENT 3 RESULTS OF SN AND FAR PLOTTED AS PERCENTAGES FOR EACH VALUE OF THE TUNABLE	
PARAMETER CED_{DT} WITH $SQTHRESH$ VALUE OF 12.	87
FIGURE 4-11: EXPERIMENT 3 RESULTS OF SN AND FAR PLOTTED AS PERCENTAGES FOR EACH VALUE OF THE TUNABLE	
PARAMETER CED_{DT} WITH $SQTHRESH$ VALUE OF 6.	87
FIGURE 4-12: EXPERIMENT 3 RESULTS OF SN AND FAR PLOTTED AS PERCENTAGES FOR EACH VALUE OF THE TUNABLE	
PARAMETER CED_{DT} WITH $SQTHRESH$ VALUE OF 18.	88
FIGURE 4-13: RADAR PLOT FOR EXPERIMENT 3 SHOWS SN VS. FAR FOR VARIOUS CED DECISION THRESHOLDS (CED_{DT}) WHERE	
$SQTHRESH$ IS 6. EACH PAIR AROUND THE PERIMETER REPRESENTS (FAR , CED_{DT}).	88
FIGURE 4-14: RADAR PLOT FOR EXPERIMENT 3 SHOWS SN VS. FAR FOR VARIOUS CED DECISION THRESHOLDS (CED_{DT}) WHERE	
$SQTHRESH$ IS 18. EACH PAIR AROUND THE PERIMETER REPRESENTS (FAR , CED_{DT}).	89
FIGURE 4-15: RADAR PLOT FOR EXPERIMENT 3 SHOWS SN VS. FAR FOR VARIOUS CED DECISION THRESHOLDS (CED_{DT}) WHERE	
$SQTHRESH$ IS 12. EACH PAIR AROUND THE PERIMETER REPRESENTS (FAR , CED_{DT}).	89
FIGURE 4-16: EXPERIMENT 4 RESULTS OF SN AND FAR PLOTTED AS PERCENTAGES FOR EACH COMBINATION OF TUNABLE	
PARAMETERS (CED_{DT} , MED_{WW}) WITH $SQTHRESH = 12$	93
FIGURE 4-17: EXPERIMENT 4 RESULTS OF SN AND FAR PLOTTED AS PERCENTAGES FOR EACH COMBINATION OF TUNABLE	
PARAMETERS (CED_{DT} , MED_{WW}) WITH $SQTHRESH = 6$	93
FIGURE 4-18: EXPERIMENT 4 RESULTS OF SN AND FAR PLOTTED AS PERCENTAGES FOR EACH COMBINATION OF TUNABLE	
PARAMETERS (CED_{DT} , MED_{WW}) WITH $SQTHRESH = 18$	94
FIGURE 4-19: RADAR PLOT FOR EXPERIMENT 4 SHOWS SN VS. FAR FOR VARIOUS MEDIAN FILTER WIDTHS (MED_{WW}) AND CED	
DECISION THRESHOLDS (CED_{DT}) WITH $SQTHRESH = 6$. EACH 3-TUPLE AROUND THE PERIMETER REPRESENTS (FAR ,	
MED_{WW} , CED_{DT}).	95

FIGURE 4-20: RADAR PLOT FOR EXPERIMENT 4 SHOWS SN VS. FAR FOR VARIOUS MEDIAN FILTER WIDTHS (MED_{WW}) AND CED DECISION THRESHOLDS (CED_{DT}) WITH $SQTHRESH = 12$. EACH 3-TUPLE AROUND THE PERIMETER REPRESENTS (FAR, MED_{WW} , CED_{DT}).	96
---	----

FIGURE 4-21: RADAR PLOT FOR EXPERIMENT 4 SHOWS SN VS. FAR FOR VARIOUS MEDIAN FILTER WIDTHS (MED_{WW}) AND CED DECISION THRESHOLDS (CED_{DT}) WITH $SQTHRESH = 18$. EACH 3-TUPLE AROUND THE PERIMETER REPRESENTS (FAR, MED_{WW} , CED_{DT}).	97
---	----

LIST OF ACRONYMS

Artifact Detection	AD
Apnoea Of Prematurity	AOP
Blood Pressure	BP
Clinical Event Detection	CED
Clinical Decision Support System	CDSS
Common Reference Model	CRM
Critical Care Unit	CCU
Electrocardiograph	ECG
False Alarm	FA
False Alarm Rate	FAR
False Discovery Rate	FDR
Heart Rate	HR
First-of-a-kind	FOAK
Impedance Respiratory Wave	IRW
Intensive Care Unit	ICU
Late Onset Neonatal Sepsis	LONS
Missed Alarm	MA
Neonatal Intensive Care Unit	NICU
Original Equipment Manufacturer	OEM
Saturation of Peripheral Oxygen	SpO ₂
Parameter Derivation	PD
Pediatric Intensive Care Unit	PICU
Physiologic Data Acquisition	PDA
Respiratory Rate	RR
Streams Processing Language	SPL
True Alarm	TA

STATEMENT OF ORIGINALITY

The research presented in the thesis has been conducted originally by the author at Carleton University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical and Computer Engineering. Research results presented in this thesis have been published in journals and conferences as categorized below in chronological order. This statement summarizes the author's contribution as well as the allocation of each publication within this thesis.

REFEREED JOURNAL PUBLICATIONS

[17] **S. Nizami**, J. R. Green and C. McGregor, "Implementation of artifact detection in critical care: a methodological review," *Biomedical Engineering, IEEE Reviews In*, vol. 6, pp. 127-142, 2013.

The literature review conducted in Chapter 2 has been published in part in the journal *IEEE Reviews in Biomedical Engineering*. I researched, identified and synthesized pertinent literature. I developed the six review themes in consultation with my thesis supervisors who are the two co-authors. I wrote the critique and made all necessary revisions based on feedback from the co-authors and the journal editor.

[234] **S. Nizami**, K. Greenwood, N. Barrowman and J. Harrold, "Performance Evaluation of New-Generation Pulse Oximeters in the NICU: Observational Study," *Cardiovascular Engineering and Technology*, Springer, vol. 6, pp. 383-391, 2015.

This research is validated using data collected in a clinical study that was conducted at the Neonatal Intensive Care Unit (NICU) of the Children's Hospital of Eastern Ontario (CHEO), Ottawa, Canada. The Research Ethics Board (REB) at CHEO approved the neonatal data collection at the patients' bedside and subsequent analyses of these data. A crossover observational study consisting of primary analysis of the acquired

physiologic dataset has been published in the *Cardiovascular Engineering and Technology* journal. Parts of this publication are reported in Chapter 4 and in Appendix C: Masimo vs. Nellcor SpO2 Comparison.

I initiated this study in my capacity as a research associate at CHEO. I was responsible for surveying the literature, designing the study, drafting the application for research ethics approval, setting up the research equipment and recording data at patient bedside, annotating data in consultation with the bedside nurse and the neonatologist, de-identifying data, reverse engineering the structure of the OEM data packets, retrospectively analyzing data, computing performance metrics, and authoring, submitting and revising manuscripts for research dissemination. The research was conducted independent of OEM involvement and bias.

REFEREED CONFERENCE PUBLICATIONS BASED ON FULL PAPER

SUBMISSIONS

[160] **S. Nizami**, J. R. Green, J. M. Eklund and C. McGregor, "Heart disease classification through HRV analysis using Parallel Cascade Identification and Fast Orthogonal Search," *Medical Measurements and Applications Proceedings (MeMeA), IEEE International Workshop On*, pp. 134-139, 2010.

I conducted the background research on heart rate variability analysis, and wrote the introduction section. I presented the results of this paper at the IEEE International Workshop MeMeA 2010. This research was published in the conference proceedings. This publication is part of the reviewed literature in Chapter 2.

[128] **S. Nizami**, J. R. Green and C. McGregor, "Service oriented architecture to support real-time implementation of artifact detection in critical care monitoring," *in Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 4925-4928.

In this publication, I conducted the background research and developed an early prototype of the Artifact Detection (AD) framework. I designed the case study to determine the requirement and provision interfaces for artifact detection and clinical event detection components in the case of late-onset neonatal sepsis (LONS). I developed an eXtensible Markup Language (XML) schema as a preliminary example of component instantiation and interfacing that conforms to the Common Reference Model (CRM) developed in this thesis in Chapter 3. I prepared the manuscript and made all necessary revisions based on feedback from the co-authors and the reviewers. I presented a poster on this research at the EMBC 2011 conference. This research was published in the conference proceedings. This publication is part of the reviewed literature in Chapter 2.

[219] **S. Nizami**, J. R. Green and C. McGregor, "An artifact detection framework for clinical decision support systems," in *World Congress on Medical Physics and Biomedical Engineering, Toronto, Canada, June 7-12, 2015*, pp. 1393-1396.

The Artifact Detection (AD) framework developed in Chapter 3 has been published in part in the conference proceedings of the *World Congress*. I developed the framework and wrote the manuscript as the main author of the publication. I made all necessary revisions based on feedback from my thesis supervisors, who are the two co-authors, and the reviewers. I also presented this research during an oral session at the conference.

1 INTRODUCTION

This chapter gives an overview of the thesis. It introduces the reader to the scope of the thesis. Following subsections highlight the significance of this research, the research problem, and the motivation behind this research. This chapter also highlights the research methods used to address the problem, the contributions made by this research in addressing the problem, validation of the research, its delimitations as well as limitations. Lastly, it outlines thesis organization.

1.1 RESEARCH SIGNIFICANCE

Physiologic signals exhibit trends, dynamics and correlations reflecting the complexity of underlying patient physiology [1-3]. Continuous monitoring of physiologic data assists clinicians in making diagnoses and prognoses in Critical Care Units (CCU), including Intensive Care (ICU), Paediatric Intensive Care (PICU), Neonatal Intensive Care Units (NICU), and the Operating Room (OR) [4]. Within the scope of this thesis, Clinical Decision Support Systems (CDSS) are computerized healthcare analytic systems with the functionality to integrate patient data and analyze those using Parameter Derivation (PD), which derive clinically useful low frequency parameters from high frequency input data, and Clinical Event Detection (CED) algorithms, which identify clinically significant events and early onset indicators of various pathophysiologies. Past research in CDSS has shown the potential to lower patient mortality and morbidity rates when integrated in critical care workflows [5-9]. However, the inherent impact of signal artifacts on quality of physiologic data presents a challenge that affects the reliability and accuracy of analytical results [10]. The ECRI Institute, a Pennsylvania-based patient-safety organization, issues an annual report of the top 10 Health Technology Hazards. Leading up to and including 2015, ECRI has

reported medical device alarms to be the top most hazard (<https://www.ecri.org/Pages/2015-Hazards.aspx>). In the 2015 report, data integrity follows as the second most reported hazard due to incorrect or missing data in electronic healthcare systems. Both hazards are highly correlated. Compromising the quality of data, and hence its integrity, would cause CDSS to issue inappropriate alarms, which are then typically overridden by clinicians [11, 12]. Such alarms, termed as nuisance or false alarms, result in increased noise levels, staff disruption, and staff desensitization in busy critical care environments [13-16]. This induces alarm fatigue in clinicians, thereby compromising the quality of care at the patient bedside. Integrity and quality of data are crucial to the success of any analytic system. Therefore, it is important to design and implement CDSS that assess the quality of data and issue relevant alarms with high specificity and low false alarm rates.

Clinicians are increasingly interested in developing and implementing real-time CDSS, though most CCU lag behind in effective computerization of their medical systems and workflows [9]. However, each PD and CED algorithm within one CDSS can have different needs in terms of type, frequency, and quality of data [17]. Therefore, no single Artifact Detection (AD) algorithm is necessarily best suited to serve each PD and CED algorithm within a CDSS. Although there is an increased trend in the development and use of Signal Quality Indicators (SQI) as representative of signal quality in AD algorithms such as [18-27], no framework exists to uniformly deliver, compare or combine these SQI. Moreover, contemporary research underlines the importance of clinical verification and validation of such algorithms across varying populations as a pre-requisite to clinical implementation [28, 29]. Therefore, the development of a framework to support the interoperability of a range of configurable AD across different patient populations and CCU domains is required. In addition, such a framework must also support the evaluation of the AD algorithm(s) that are best suited to the given data needs of PD and/or CED algorithms within a CDSS. Moreover,

the framework should exploit the availability of SQI to improve the analytic power of the CDSS. To implement real-time processing architectures in CCUs, the first step would be then to standardize interoperability of AD, PD and CED within the CDSS [30]. In the context of this research, AD includes the detection, suppression, and remediation of artifacts within physiologic data. Here, the term *detection* is used in its broad sense similar to how it presents in the literature.

1.2 RESEARCH MOTIVATION

Artemis is a neonatal CDSS established through an IBM first-of-a-kind (FOAK) award [31]. It was first deployed at the Neonatal Intensive Care Unit (NICU) at The Hospital for Sick Children, (SickKids) in Toronto, Canada in 2009. Thereafter, the Artemis network grew rapidly connecting several NICUs at hospitals located worldwide [32, 33]. Recent network expansions include cloud-based implementation [34, 35], and remote monitoring applications [36]. Artemis has the ability to capture, store and analyze multivariate physiologic data streams for simultaneous diagnosis and prediction of multiple clinical events [31]. It is a Big Data analytics platform leveraged by concurrent multi-institutional and multidisciplinary research at the University of Ontario Institute of Technology (UOIT) and Carleton University. Figure 1-1 shows the timeline as research collaborations progressed in the Artemis project. Artemis has the functionality to simultaneously implement multiple CED and PD algorithms. Ongoing clinical research in Artemis includes CED for: LONS [37, 38]; AOP [39-41]; Retinopathy of Prematurity (ROP) [42]; pain management [43], and cerebral monitoring to identify useful neurodevelopmental indicators [44]. At present, Artemis lacks a systematic mechanism to identify artifacts associated with each of the acquired physiological data streams [45]. The motivation to improve the quality of decision support provided by Artemis is what drives this research. Rather than developing AD techniques in an *ad hoc* fashion, as is the current practice [17], this thesis develops a novel multivariate

AD framework that can be systematically integrated within Artemis and other CDSS. The objective of this research is to facilitate the interfacing of several AD techniques that can serve the differing data needs of PD and CED components within CDSS such as Artemis. The standardized framework developed in this research will allow for physiologic data analytics and signal processing by leveraging systematic integration to create novel compositions of state-of-the-art AD and CED algorithms, thereby, improving the quality of clinical decision support that is provided by Artemis.

1.3 RESEARCH PROBLEM

Several Artifact Detection (AD) algorithms have been published to assess the quality of physiologic data and minimize the impact of artifacts prior to PD and CED. However, a methodological literature review by the author, Nizami *et al.* [17], shows that the majority of AD algorithms have the following shortcomings (s.1)-(s.6):

Figure 1-1: Progressive timeline showing Artemis research collaborations.
Adapted from <http://hir.uoit.ca/cms/?q=node/24>

- (s.1) They are developed for patient populations from a specific type of CCU. Their flexibility to serve the needs of patient populations from other types of CCUs have not been assessed.
- (s.2) They are hard-coded to process particular types of physiologic data harvested by specific OEM monitors. Since different monitor models from the same or different OEM may have different built-in proprietary signal pre-processing. Therefore, they may differ in their representation of data and artifacts. However, the reusability of AD algorithms in processing different types of physiologic data harvested by different OEM monitors has not been demonstrated.
- (s.3) They are not equipped to generate unambiguous standardized signal quality indicators (SQI). The SQI generated by each algorithm are highly customized and cannot be meaningfully compared or integrated with those from another algorithm.
- (s.4) They are designed and evaluated in isolation within a standalone application; or in a coupled application with one or more specific PD or CED algorithms. The standalone algorithms are neither reused nor the coupled algorithms decoupled for use in other applications in the same domain. This indicates a lack of generalization.
- (s.5) They are evaluated offline. Since, real-time evaluation is an absolute requirement for integration of AD in clinical workflows. Therefore, this highlights the need of a standardized, scalable and customizable AD framework that allows for evaluation of AD algorithms on both offline and real-time patient data; and that can be integrated with clinical workflows.
- (s.6) They are not commonly implemented in clinical practice. No framework with standardized component interfaces is documented to support real-time clinical implementation of AD within a CDSS.

1.4 RESEARCH METHOD

This thesis applies the constructive research methodology to address the six shortcomings (s.1)-(s.6). According to Beverly Pasian, *"The constructive research process involves the following: (1) selecting a practically relevant problem; (2) obtaining a comprehensive understanding of the study area; (3) designing one or more applicable solutions to the problem; (4) demonstrating the solution's feasibility; (5) linking the results back to the theory and demonstrating their practical contribution; and (6) examining the generalizability of the results"* [46]. As an applicable solution to (s.1)-(s.6), this thesis employs component-based system design to develop an AD framework with interface definitions that standardize interoperability of AD, PD and CED within the CDSS. The advantages of component-based system design include reusability, flexibility, scalability, adaptability, and cost cutting [47-49]. A standardized framework offers generalization as well as customization. Component-based design is also used in application frameworks. *"A framework is a reusable software element that provides generic functionality, addressing recurring concerns across a range of applications"* [50]. Such frameworks provide underlying structure to support the development of software applications for a particular problem domain. Application frameworks are widely used in software systems, particularly for graphical user interfaces. The component-based design used in this thesis, including the standardization of component interfaces, could be considered as creating an application framework.

This thesis creates pipelines of components to address specific problems. In a pipeline architecture, data flows in a strictly forward direction through components, such that the solution is distributed across multiple components that operate independently [51]. The term "stream" refers to the flow of information between subsequent components. As an alternative to pipelining, the Blackboard architecture

consists of a collection of components all operating independently on shared data structures in a common space referred to as the blackboard [52]. Although the components work cooperatively to solve the problem, they do not call each other. Instead, a central control component moderates the actions all components to ensure a final solution is achieved. Components operate using an opportunistic strategy, running when the state of the solution permits their application.

Other researchers have developed component-based frameworks in the clinical domain for disease tracking and prevention [53], bioinformatics [54, 55], and health information networks (HIN) [56]. However, this research differs from the above in its unique contribution of integrating AD in CDSS as discussed below.

1.5 RESEARCH CONTRIBUTIONS

Novel contributions of this thesis include: (1) a static framework model with four different types of components; (2) a Common Reference Model (CRM) that defines standard interfaces; (3) a catalogue of AD components; and (4) a clinical validation study illustrating framework composition.

The framework is a unique test bed that can mix and match independently developed AD components to evaluate their performance in combination with different types of PD and CED components. The combination most suited to a clinical requirement can then be hard-coded into a workflow for real-time application in CCUs. AD, PD and CED algorithms, published in the past or future, whether standalone or coupled, may be used as CDSS components with modifications as needed.

A critical element of the framework is a Common Reference Model (CRM) that defines standard interfaces among AD, CED, and PD components. A thorough review of the literature in Chapter 2 identified common physiologic data attributes that capture the functionality of existing AD algorithms. These attributes include: (1) type;

(2) frequency; and (3) signal quality [17]. Each of these attributes is defined in the CRM.

As mentioned above, a review of the literature identified six shortcomings (*s.1-s.6*) that currently limit the widespread implementation of AD algorithms within CDSS. To address these shortcomings, the framework leverages six features (*f.1*) – (*f.6*) that promote its use as an independent multivariate CDSS, or part thereof. These features are:

(f.1) Flexibility to serve the needs of patient populations from different types of Critical Care Units (CCU) through generalization and customizability;

(f.2) Reusability across multiple types of physiologic data harvested by different OEM monitors;

(f.3) Standardized definitions of Signal Quality Indicators (SQI) that promote interoperability and comparison between independently developed components;

(f.4) Reusability and scalability by mixing and matching several AD, PD and CED components in various combinations;

(f.5) Customizability to evaluate and compare the performance of multiple combinations of independently developed components on offline and potentially real-time patient data when integrated with clinical workflows; and

(f.6) Standardized component interfaces that can potentially support real-time clinical implementation of AD.

1.6 RESEARCH VALIDATION

The framework is validated within a multivariate CDSS with AD, PD and CED components. The author conducted the clinical validation study at the NICU, Children's Hospital of Eastern Ontario (CHEO) in Ottawa, Canada. The hospital's Research Ethics Board (REB) approved the study. This study is fully described in Chapter 4. The study demonstrates how use of the AD framework supports the six features (*f.1*) – (*f.6*).

1.7 DELIMITATIONS AND LIMITATIONS

This section explicitly states the delimitations and limitations of this research.

"Delimitations of a study are those characteristics that limit the scope (define the boundaries) of the inquiry as determined by the conscious exclusionary and inclusionary decisions that were made throughout the development of the proposal [57]". In other words, delimitations are factors that are controlled by the researcher to define and limit the scope of the study [58]. Some of the delimitations of this research are as follows.

The research objective is to design and evaluate a framework that can integrate multiple AD techniques with CDSS. The objective is not to evaluate the performance of existing AD algorithms. The interested reader may find quantitative studies comparing performance metrics of AD algorithms elsewhere, such as in [24, 59-63].

Proprietary filters and built-in pre-processing algorithms OEM physiological monitors produce a certain bias in the output data. Furthermore, as shown in the literature review in Chapter 2, these algorithms lack standardization and are typically not published by the manufacturer in sufficient detail to be replicated by researchers. Although, the framework is constrained to accept OEM monitor data at its input which has been pre-processed for artifacts by the monitor itself, however, the configuration interface of each component is designed to take into account the variability in monitor outputs. This also makes the framework generalizable for reuse in different clinical settings and with different OEM monitors.

Component-based system design focuses on the capabilities of the components as expressed by their external interfaces rather than implementation of their source code [64]. Therefore, it is of irrelevance how components may be built using various programming languages such as MATLAB, C++, and Java.

Originally, it was proposed to validate this framework via two CED case studies from the Artemis research namely: the Late Onset Neonatal Sepsis (LONS) study that utilises heart rate variability and respiration rate variability data as detailed in [38], and the Apnoea of Prematurity (AOP) study [39]. However, unforeseen delays impacted the commencement of these studies, subsequently delaying the availability of clinical data precluding their use in this thesis. Therefore, it was decided instead to utilise companion research underway at the NICU at CHEO. This comparative study of alternate pulse oximeters was undertaken by the author of this thesis. The study provided the opportunity for rigorous artifact annotation that was necessary for use in the validation experiments utilising multiple AD compositions.

Validation of this research with one CED provides a limited demonstration of the framework's ability to simultaneously support multiple CED from multiple CCU domains. However, the inclusion of multiple AD compositions demonstrates the principle that AD components can be mixed and matched to form complex pipelines to service potentially different CED requirements, possibly across different CCU domains.

"The limitations of the study are those characteristics of design or methodology that set parameters on the application or interpretation of the results of the study; that is, the constraints on generalizability and utility of findings that are the result of the devices of design or method that establish internal and external validity " [57]. Limitations are factors that are not under control of the researcher [58]. Some of the limitations of this research are as follows.

The overall performance enhancement of a CDSS that may be achieved through use of this framework is limited by the performance of the AD, CED and PD algorithms that may be used as components. Standardization of the CRM implies that all component interfaces must conform to the CRM to be integrated successfully within this framework.

At present, there are limited ways for OEM patient monitors to intelligently discern the source of an artifact. For example, some built-in AD techniques can filter low frequency artifacts due to motion or 60 Hz power line interference. However, this also removes part of the physiological signal occurring at the same frequencies. Furthermore, there is no guarantee that removed outliers are not real events. In fact, they may be indicative of unsuccessful clinical management of an actual clinical event requiring change in treatment [65]. Since the framework inputs data directly from the OEM monitors, AD and CED algorithms implemented through this framework are unable to discriminate between signal changes caused by artifacts and those caused by underlying patient pathophysiology. The exception to this would be when a clinician electronically annotates the event in real-time for the analytical system to clearly find out what is happening. While research has proposed to annotate artifacts using an iPod touch [66], such real-time annotations are not yet the norm as they would take away precious point of care time from the attendants [67].

It is typical for each AD algorithm to track certain types of artifacts on specific types of physiological streams. Therefore, the framework will be limited in capturing only those types of artifacts for which AD components are available [68]. However, the framework can grow by inclusion of a larger variety of AD components through future work. This thesis does not investigate processing power and financial costs incurred for implementing CDSS in healthcare settings.

1.8 THESIS ORGANIZATION

This thesis consists of five chapters. Chapter 2 thematically reviews published AD techniques. Chapter 3 develops the multivariate AD framework, and produces a catalogue of AD components. Chapter 4 presents a research validation study conducted on clinically acquired neonatal data. Chapter 5 concludes the thesis with a discussion on research contributions and future work.

2 LITERATURE REVIEW

This chapter conducts a thematic review of Artifact Detection (AD) techniques that are published for use in Critical Care Units (CCU). This methodological review has been in part published as S. Nizami, J. R. Green and C. McGregor, "Implementation of Artifact Detection in Critical Care: A Methodological Review", *IEEE Reviews in Biomedical Engineering*, vol. 6, pp. 127-142, 2013 [17]. The six themes developed for this critique are: I. Critical Care Unit; II. Physiologic Data Source; III. Harvested data; IV. Data Analysis; V. Clinical Evaluation; and VI. Clinical Implementation. The review identifies research gaps and open areas for research in each of the themes.

In this chapter, Section 2.1 sets the background for this thesis. Section 2.2 thematically reviews literature on AD algorithms. Section 2.3 presents results of the review. Section 2.4 concludes the review by highlighting open research problems. It also provides specific recommendations for new research directions in promoting implementation of AD in real-time clinical workflows. Section 2.5 summarizes the chapter.

2.1 BACKGROUND

2.1.1 ARTIFACT DETECTION

Original Equipment Manufacturer (OEM) patient monitors typically come with a black box approach to signal preprocessing from data acquisition to logging. Clinicians debate the reliability of these data as they deduce that some built-in preprocessing algorithms employ a reductionist approach that oversimplifies complex human physiology [15]. Longitudinal studies [16, 69-71] infer that OEM monitors have relatively simplistic built-in data preprocessing for AD. Therefore, data logged at their output remains impacted by artifacts [72-74]. Artifacts can mimic physiologic data

[75-77], adding to the challenge of distinguishing between the two. This increases false alarm rates in monitors [61, 74, 78], leading to staff desensitization [13-16]. Clinicians cannot rely on analyzing artifact-laden data [79]. This has previously resulted in incorrect diagnoses [77, 80]; unnecessary therapy; surgery and iatrogenic diseases [76]. Independent research groups have developed a variety of post processing techniques to address the problem of artifacts in OEM monitor data.

Figure 2-1 shows the reviewed AD configurations. The purpose of AD is identical in any configuration, i.e., to assess and enhance the quality of physiologic data. Figure 2-1 also depicts Signal Quality Indicators (SQI) among other output variables. Latest AD research shows increasing interest in devising SQI [24, 81-83]. This chapter later analyzes the current state of development of SQI, and their utility in clinical workflows.

2.1.1.1 Past Reviews of AD Techniques

Historically, AD has been reviewed with significantly different scope and context from this work, which makes this review a novel contribution in this research space. Borowski *et al.* [84] have conducted a comprehensive review on alarm generation in medical devices. Their primary focus is ergonomics and human factors engineering, although, they briefly synthesize some AD algorithms and, as a result, recommend using multivariate alarm systems. Schettlinger *et al.* [62] have largely synthesized their own research in developing univariate filters for outlier, trend and level shift detection in various ICU data types. They extensively describe filter development and compare advantages and disadvantages of different filter designs. Chambrin [85] infers that multivariate techniques can reduce false alarm rates in ICUs. Numerous publications, such as [24, 59-63], provide comparative numeric summaries of performance metrics of AD algorithms. This review neither reiterates algorithmic details nor performance comparison as that falls outside its scope. Takla *et al.* [71] note that while AD techniques developed by independent researchers may have higher specificity than built-in algorithms in OEM monitors, extensive studies are required to evaluate their accuracy prior to real-time implementation in clinical environments. Siebig *et al.* [78] demonstrate agreement amongst clinical staff, including intensivists, that integrative monitoring through data fusion can potentially yield better results as compared to simpler univariate threshold detection methods. Imhoff *et al.* [61, 63] emphasize that methodological research is needed for integrating multivariate AD algorithms in real-time clinical systems. However, they do not present a conceptual framework to conduct such research.

2.1.2 CLINICAL EVENT DETECTION

A common goal of CED algorithms is to generate an alarm to alert clinical staff when physiologic signals exceed normal ranges. As discussed above, signal quality issues

can lead to high false alarm rates. A number of recent studies have proposed approaches to suppressing false alarms using annunciation delays (a delay between when an alarm threshold has been crossed and when the alert is sounded or displayed) and relaxed alarm thresholds. Burgess *et al.* [86] study the trade-off between sensitivity (Sn) and false alarm rates (FAR) by testing different alarm thresholds and introducing annunciation delays of 30 and 60 sec for HR and RR streams respectively. No quantitative performance analysis conducted, however their qualitative analysis suggests that these measures may be applicable in non-CCU environments where both the staff-to-patient ratio and the average disease severity are lower. Similarly, Taenzer *et al.* [87] suggests the use of 30 sec annunciation delays and discusses the tradeoff between Sn and FAR. They suggest that careful calibration of alarm thresholds is essential in finding a balance between Sn and FAR, which is patient-centered and can be decided by the clinician based on severity of disease. As with Burgess *et al.*, no quantitative evaluation of Sn vs. FAR is provided.

In 2011, Welch, vice president of Masimo Corp., published methods for reducing SpO₂ alarms based on: (i) averaging data over a fixed window size of 8 seconds, (ii) lowering alarm thresholds, for example, a clinical threshold at 90% would be set at 88-87%, and (iii) introducing a 15 sec alarm annunciation delay [88]. Welch's study quantifies the reduction in overall alarm generation rates without evaluating performance metrics of Sn or FAR. In 2015, the same group replicated their earlier study on SpO₂, HR, RR, systolic, diastolic and mean blood pressure (SBP, DBP, MAP) streams [89]. They evaluate the percentage reduction in total alarm rates for a delay ranging from 5-120 sec. Again, no quantification of Sn vs. FAR is included. The CED component developed in Section 4.2.3 to validate the framework leverages the concepts of introducing annunciation delays and relaxing the alarm thresholds.

Not all clinicians agree with the proposed strategy of trading off Sn or timeliness for the reduction in FAR. Lynn *et al.* [15] frames the clinician's perspective as follows:

“every competent clinician appreciates the lethality of sepsis and the importance of early detection for successful resuscitation...so how and why can it be argued reasonably, our setting threshold alarms for this disease at limits that would purposefully delay its detection?” These concerns highlight the need to quantify, not only the benefit of reduced FAR, but also the commensurate reduction in Sn. This motivates the rigorous quantitative performance analysis in Chapter 4.

2.1.3 CLINICAL DECISION SUPPORT SYSTEMS (CDSS)

Clinical Decision Support Systems (CDSS) are computerized healthcare analytic systems that allow for automated application of expert clinical knowledge to patient-specific information to aid clinicians in the decision making process at point-of-care [90]. Although, CDSS have the potential to reduce clinicians’ workload, increase the efficiency of point-of-care delivery, and improve the quality of patient outcomes in terms of reduced morbidity and mortality rates [91]. However, to achieve that, CDSS need to be effectively integrated into clinical workflows [90].

Past reviews confer that integrating CDSS into clinical workflows generally requires unique customization to locally implemented processes, minimal disruption to the clinician’s cognitive workflow, seamless integration with patient monitoring systems and health records, and integration of data in a meaningful format for consumption by the CDSS [12]. Furthermore, widespread use of CDSS necessitates the development and evaluation of scalable, standardized models that are portable across clinical settings [91, 92]. Recently, the Agency for Healthcare Research and Quality at the U.S. Department of Health and Human Services conducted a technology assessment in state of the art in CDSS [90]. They note that most of the research published on CDSS has focused on a single or limited set of conditions [90]. Thus, studies are needed to determine how CDSS can be designed and deployed to provide

decision support in diverse combinations of comorbid conditions. Secondly, they note that clinical recommendations need to be prioritized to avoid alarm fatigue [90].

In the context of this thesis, a CDSS is a healthcare analytic system with the functionality to integrate patient data and analyze those using Parameter Derivation (PD) and Clinical Event Detection (CED) algorithms. Clinical Event Detection (CED) techniques analyze these data to identify clinically significant events and early onset indicators of various pathophysiologies as in [38, 40, 42, 43, 93-101]. Parameter Derivation (PD) techniques derive clinically useful low frequency parameters from high frequency input data as in [102-108]. Past research in CDSS has shown the potential to lower patient mortality and morbidity rates when integrated in critical care workflows [6-9]. However, the inherent impact of signal artifacts on quality of physiologic data presents a challenge that affects the reliability and accuracy of analytical results [10]. Several Artifact Detection (AD) algorithms have been published to assess the quality of physiologic data and minimize the impact of artifacts prior to PD and CED. However, a methodological literature review by the author has found six shortcomings in the majority of AD algorithms as part of the preceding review.

Although clinicians are increasingly interested in developing and implementing real-time CDSS, most Critical Care Units (CCU) lag behind in effective computerization of their medical systems and workflows [9]. Moreover, the authors' review identifies that each PD and CED algorithm within one CDSS can have different needs in terms of type, frequency, and quality of data [17]. Therefore, no single AD algorithm is necessarily best-suited to serve each PD and CED algorithm within a CDSS. To implement real-time processing architectures in CCUs, the first step would be then to produce expert level documentation to standardize interoperability of system components within the CDSS [30].

2.1.3.1 Artemis

One particular CDSS, Artemis [109], is described in some detail here, as it was developed by one of the co-supervisors of this research, and serves as the motivation for the AD framework development in this thesis, as previously discussed in Section 1.2. Figure 2-2 shows the high-level view of the Artemis framework. The remainder of this section describes the hardware implementation of the Artemis CDSS. Artemis is integrated with Capsule Tech modules (Capsule Tech Inc.) to interface with and synchronize the myriad of critical care monitoring devices. The Capsule Tech module has the hardware and software capability to capture data from more than 450 different types of devices from all of the major vendors. All data are forwarded to a Capsule DataCaptor Interface Server (CDIS) that can support up to 500 simultaneously connected DataCaptor terminals. Using the Capsule application-programming interface (API) and software development tool kit, a server-based function can be implemented to extract only those data streams necessary for analysis. These streams are sent to a configurable Medical Data Hub system that receive the aggregated data object from the CDIS and creates concurrent data streams for streaming analysis. Artemis then

**Figure 2-2: High-level view of the Artemis framework.
Reprinted with first author's permission from [93].**

applies clinical rule-based reasoning to these incoming data streams within the Online Analytics component. Nonetheless, a systematic approach is required to isolate possible artifacts associated with physiological data streams acquired by the Capsule Tech modules in Artemis.

2.2 METHODOLOGICAL REVIEW

Papers of interest were located in Scopus, IEEEExplore and PubMed databases using the keywords artifact, artefact, artifact detection, alarm, physiologic monitoring, intensive care, neonatal intensive care, clinical decision support systems, operating room and patient monitors amongst others. No constraints were applied on the publication year. Google Scholar was also used. Reviewed work on AD included algorithms for the detection, suppression, and remediation of artifacts within physiologic data. Table I lists the reviewed publications.

This methodological review introduces six themes for critically analysing state-of-the-art in AD. This section develops a unique taxonomy for each of the six themes. Figure 2-3 depicts an overview of each themes and its taxonomy. Appendix A: Review of Themes II – VI tabulates the results from reviewed themes II to VI. For a better understanding, the reader may consider the author name and citation in each row to be the row header. The cell adjacent to the right contains the Theme II results for that citation. Each cell in the third column has results from two themes, Theme III in the top half of the cell and Theme IV in the bottom half of the cell. Similarly, each cell in the fourth column has results from two themes, Theme V in the top half of the cell and Theme VI in the bottom half of the cell.

2.2.1 THEME I. CRITICAL CARE UNIT

This theme catalogues domain specific AD research to examine its reusability and portability across other CCUs. The taxonomy developed for this theme is: ICU, PICU,

NICU, OR and Other relevant studies. The synthesis shown in Table I reveals that almost all techniques are evaluated in a single CCU domain, with the exception of the work of Feldman and Ebrahim [108] that is evaluated in the ICU, PICU and OR [107]. In general, trends in physiologic signals display similar dynamics across CCU domains [110]. This drives the hypothesis that AD algorithms could be modified for use across different critical care settings. However, domain specific information such as types and frequency of data, and patient demographics such as range of age, weight and medical condition may be directly, or indirectly, hard coded in the algorithm. Therefore, it is important to consider the original domain of clinical application prior to attempting application elsewhere.

2.2.2 THEME II. PHYSIOLOGIC DATA SOURCE

This theme synthesizes the methodologies used to source physiologic data. The theme taxonomy is: Physiologic Monitor, Inclusion/Exclusion Criteria and Sample Size, as tabulated in the second column of Appendix A: Review of Themes II – VI. The types, frequency, numeric values and quality of data produced by patient monitors (and probes) differs between models from the same or different OEMs. This is due to different built-in proprietary signal preprocessing, inclusive of AD [71, 111]. For example, comparative studies show characteristic discrepancies in neonatal data acquired using various OEM Pulse Oximeters (PO) including Masimo SET Radical (Masimo Corp., Irvine, CA, USA), Datex-Ohmeda TruSat (GE Healthcare, Chalfont St Giles, UK), Siemens SC7000 (Siemens UK, Frimley, UK), Nonin 7500 (Nonin Medical, Plymouth, MN, USA), [112]; Nellcor OxiMax N-600x (Covidien-Nellcor, Boulder, CO, USA) [112, 113]; Philips FAST MP50 [113] and Philips Intellivue MP70 (Philips, Germany) [114]. Therefore, knowledge of the monitor OEM and model is important for recognizing different biases introduced in the data. Some studies may collect data under certain inclusion and exclusion criteria. Sample size is usually given in the

literature. The discretion generally lies with the researcher to determine if enough data is available to draw a statistically and clinically sound decision. Sample size may be deduced based on trial, availability of resources and study requirements.

Table I: Theme I. Critical Care Unit (CCU).
Reprinted with permission from the author's publication [17].

<i>ICU</i>	<i>PICU</i>	<i>NICU</i>	<i>OR</i>	<i>Other relevant studies</i>
Bradley, 2012 [25] Hu, 2012 [115] Li, 2012, 2009, 2008 [23, 105, 106] Sun, 2012 [116] Scalzo, 2012 [117] Borowski, 2011 [118] Siebig, 2010 [70] Schettlinger, 2010 [60] Charbonnier, 2010, 2004 [119, 120] Otero, 2009 [121] Blum, 2009 [122] Aboukhalil, 2008 [26] Otero, 2007 [123] Sieben, 2007 [124] Zong, 2004 [27] Jakob, 2000 [125] Schoenberg, 1999 [111] Ebrahim, 1997 [107] Feldman, 1997 [108]	Zhang, 2008 [126] Tsien, 2000, 1997 [127-129] Ebrahim, 1997 [107] Feldman, 1997 [108]	Monasterio, 2012 [130] Hoshik, 2012 [131] Nizami, 2011 [132] Salatian, 2011 [69] Belal, 2011 [133] Blount, 2010 [67] Quinn, 2009 [134] Zhang, 2008 [126] Portet, 2007 [135] Walls-Esquivel, 2007 [75] Moorman, 2006 [136] Tsien, 2001, 2000 [127, 137, 138] McIntosh, 2000 [139] Cao, 1996 [140]	Karlen, 2012, 2011 [141, 142] Koeny, 2012 [143] Schmid, 2011 [74] Yang, 2010, 2009 [68, 110, 144] Ansermino, 2009 [145] Hoare, 2002 [146] Gostt, 2002 [147] Ebrahim, 1997 [107] Feldman, 1997 [108] Sittig, 1990 [148] Navabi, 1989 [149]	Clifford, 2012 [72] Redmond, 2012 [81] Reisner, 2012 [150] Martinez-Tabares, 2012 [151] Silva, 2012 [152] Hayn, 2012 [153] Jekova, 2012 [59] Di Marco, 2012 [154] Addison, 2012 [103] Johannesen, 2012 [155] Lázaro, 2012 [102] Otero, 2012 [156] Bsoul, 2011 [157] Xia, 2011 [158] Zaunseder, 2011 [159] Kužilek, 2011 [160] Sukor, 2011, [161] Acharya, 2011 [162] Khandoker, 2011 [163] Nizami, 2010 [164] Nemati, 2010 [165] Alvarez, 2010 [166] Gil, 2008 [167, 168] Chen, 2008 [169] Kostic, 2007 [170] Yu, 2006 [171] Tarassenko, 2006 [172]

2.2.3 THEME III. HARVESTED DATA

The taxonomy developed under this theme is: Data Type and Acquisition/Sampling or Storage Frequency. In Appendix A: Review of Themes II – VI, the top half of each cell in column three summarizes the types and frequencies of data harvested in each study. Knowledge of the OEM is also a convenient indicator of this information. Routinely harvested or monitored physiologic data types in critical care include: Electrocardiograph (ECG), Heart Rate (HR), Breathing or Respiratory Rate (RR), Impedance Respiratory Wave (IRW), Noninvasive Oxygen Saturation (SpO₂), Invasive or noninvasive Arterial Oxygen Saturation (SaO₂), Temperature (Temp), the set of Blood Pressure (BP) measurements (namely Systemic Artery, Pulmonary Artery, Central Venous, Systolic (SBP), Diastolic (DBP) And Mean (MBP) Pressures, Arterial (ABP)), Maximal Airway Pressure (MAP), Expired Air Volume (EV), Minute Ventilation (MV) and Transcutaneous Partial Pressures of Carbon Dioxide (TpCO₂) and Oxygen (TpO₂). Frequency of the data is restricted by analog to digital sampling capability of the monitor [68]. Storage frequency depends on data logging capabilities of both the OEM monitor and the hardware and software mechanism used for storing study data. Review results in Appendix A: Review of Themes II – VI demonstrate that physiologic data are acquired, derived, sampled and stored at varying frequencies. It is common for AD algorithms to be hard coded to input particular types of data streams having specific frequencies. High frequency signals, such as continuous waveforms of ECG, ABP and PPG, are sampled at 100 Hz or more. Data that form low frequency time series, such as HR, SBP, DBP and SpO₂, are either time-averaged at a rate of once every second to once every minute from high frequency signals; or measured intermittently every half hour to an hour such as temperature and non-invasive BP.

2.2.4 THEME IV. DATA ANALYSIS

Analytic aspects of AD algorithms are reviewed using the following thematic taxonomy: Dimensionality, Focus, Signal Quality and Clinical Contribution. Theme findings are summarised in Appendix A: Review of Themes II – VI in the bottom half of each cell in column three. Dimensionality represents the number of variables or data types that an algorithm is capable of analyzing. According to Imhoff *et al.* [61], both the clinical problem and the approach to solving it are (a) univariate when a single feature of a specific data stream is analysed; or (b) multivariate when results are derived from simultaneous analysis of multiple variables; and in between these two approaches lays (c) the univariate clinical problem solved using multivariate data. A new definition for multivariate analysis has emerged in recent research, and as such should be appended to the above as: (d) an algorithm is multivariate if it analyses different metrics derived from the same physiologic signal. Examples of multivariate type (d) research on the PPG is found in [23] and on multi-lead ECG in [59, 72, 117, 153-155]. A uniquely different approach to multivariate analysis is found in [139, 140], where two different data streams were acquired from the same probe making their physiologic correlation easier to exploit.

The Data Analysis theme characterizes the focus of each algorithm as: (1) stream; (2) patient; and/or (3) disease-centric. A stream-centric algorithm aims to indicate, quantify or improve signal quality of a specific data stream for increased reliability. Although the term patient-centric has broad implications in health care, it means here that the algorithm was trained on patient specific data and is therefore heavily tailored to each sample patient in the study. For example, baseline data from a particular patient may be required to instantiate an algorithm. BioSign [172] is an example of a real-time, automated, stream and patient-centric system. It produces a single-parameter representation of patient status by fusing five types of vital sign data.

A disease-centric methodology focuses on identifying or predicting a specific disease or a clinically significant outcome. The bradycardia detector in [135] and the prognostic tool for Late Onset Neonatal Sepsis (LONS) in [173] are both examples of stream and disease-centric approaches.

The clinical contribution taxonomy reveals two unique configurations of AD: (i) Standalone AD and (ii) Coupled AD. Both configurations are shown in Figure 2-1. Standalone AD techniques typically output filtered or original physiologic data, annotations and SQI. Coupled AD techniques are coded as part of algorithms that process artifacts similar to standalone AD techniques with the additional ability for CED or PD.

Signal Quality is a key element in this taxonomy. Missing segments, error, noise and artifacts inevitably affect data quality thus adversely impacting analytic accuracy and reliability [169]. To address this issue, Clifford *et al.* [65] recommend that an SQI calibrated to provide a known error rate be made available for each datum. However, we infer in [132] that it suffices for SQI to be available at a frequency relative to the requirement of another AD, CED or PD application that consumes the SQI. The latter is particularly relevant when down sampled data is required by CED or PDs and is described in Section 3.3.3.

It has long been hypothesized that the performance of post processing AD algorithms can improve by consuming streaming SQI output in parallel for each data stream by OEM patient monitors [146]. However, use and delivery of SQI is not yet standardised across OEM monitors. For example, in case of a lead disconnection, the Philips IntelliVue (Philips, Germany) monitor outputs a random value above 8 million in data streams such as the ECG, IRW, HR, RR, SpO₂ and BP. This also generates a corresponding alert type of 'medium priority technical alarm' logged at a value of "2". However, this value is qualitative and not quantitative. The Delta, Gamma, Vista, Kappa and SC6002 - SC9000 monitor series (Dräger Medical Systems, Lubeck,

Germany) output an SQI value between 0-100% for the electroencephalography (EEG) channel. This SQI is calculated using sensor impedance data, artifact information and other undisclosed variables. The same monitors also output a fixed label 'ARTF' indicating artifact on any monitored data stream [174, 175]. For example, the monitor classifies QRS complexes only at ECG values > 0.20 mV for widths > 70 ms. An artifact condition 'ARTF' may be declared when a QRS complex is detected but the ECG signal fails to meet these minimum criteria. Although OEM monitors may output signal quality information, there is no logical way to compare the SQI produced by different monitors. This confounds clinicians' and researchers' ability to determine how patient data have been modulated from acquisition to logging. There are four factors leading to this inference: (1) different built-in signal processing which results in different data representation between different OEM monitors as shown by [112-114, 176]; (2) difference in qualification of signal artifacts [177]; (3) difference in quantification of SQI, for example in [24, 81-83]; and (4) lack of literature on proprietary signal preprocessing. As a result, a post processing AD algorithm may need to be strictly matched to input data sourced from a particular OEM monitor as shown in Figure 2-1.

Comparative studies led by Masimo [178, 179] declare that its 510 (k) FDA approved RADICAL SET PO technology has the highest SpO_2 quality, as measured by their *Performance Index (PI)*, when compared to 19 other OEM PO. In these publications, Masimo defines and calculates *PI* as the percentage of time during which a PO displayed a current SpO_2 value that was within 7% of the simultaneous control value. However, the Masimo SET technology does not automatically evaluate or log this quantitative SQI. Another quality measure called *Dropout rate (DR)* was calculated in [179], which equals the percentage of measurement time during which no current SpO_2 values are displayed. Although Masimo SET showed equal or worse *DR* than two Datex-Ohmeda PO, the reasons for this data loss are not discussed by Masimo. Independent research groups that compared OEM PO in [112-114] neither researched

the effect of the difference in data characteristics on SQI nor did they mention if the PO output SQI. Through a recent discourse in [180] on historic developments of the Masimo SET technology, its OEM has replied to Van Der Eijk *et al.* [113], claiming greater accuracy in unstable conditions, such as motion artifact and low perfusion, leading to lower false alarm rates. However, [180] does not describe any SQI that can be consumed meaningfully by other AD, CED or PD applications. This review recommends that AD algorithms that produce SQI, such as *PI* and *DR* amongst others, be evaluated upon data acquired in [112-114] to contribute towards future AD research

It follows that AD algorithms designed to post process OEM monitor data must also consume and deliver standardised SQI. In this way, another AD algorithm or a CED or PD mechanism can make informed choices concerning data quality and validity. Review results in Appendix A: Review of Themes II – VI show the increasing trend in SQI development. However, no framework exists to uniformly deliver, compare or combine these SQI for integration with clinical workflows.

2.2.5 THEME V. CLINICAL EVALUATION

Patient safety requires clinical evaluation of algorithms prior to real-time clinical implementation. This theme reviews clinical evaluation methods bases on this taxonomy: Data Annotation, Mode and Performance Metric. Results are Appendix A: Review of Themes II – VI in the top half of each cell of column four. There are no set gold standards for clinical evaluation of AD performance. Each study sets its own gold standard against which its performance is evaluated. This includes evaluations of OEM monitors.

Annotated physiologic data, where available, typically serve as the gold standard for validation studies. Events of interest in the data, such as artifacts and clinically significant events, are marked in real-time or retrospectively. The onus of

perceiving what constitutes an “event” is on the expert reviewers, who identify the event to the best of their knowledge. Inter-reviewer variability is the significant [181] or subtle [70, 182] difference known to arise when the same dataset is annotated by different reviewers. Retrospective data annotation has been supported by video monitoring in [70, 74, 75]. Video monitoring is only useful when the event is visually perceptible such as sleep movements, certain seizures and routine care. However, it cannot capture crucial physiologic changes such as HR deceleration or BP elevation. The advantage of real-time annotations is recording of richer and more accurate content with input from staff on duty [67]. However, this can be costly and requires cooperation from busy staff.

Study data are either acquired in real-time (online) mode from patient monitors or in retrospective (offline) mode for secondary analysis from existing databases. The majority of AD techniques were validated using offline patient data and very few were tested in real-time CCU environments. Appendix A: Review of Themes II – VI also documents the types of performance metrics used in each research. It shows the common trends that will help future researchers to design and compare different algorithms by evaluating them using the same metrics. Numeric comparison of these performance metrics can be found elsewhere in [24, 59-63]. Performance metrics must be interpreted carefully since statistical significance, or absence thereof, is not always representative of clinical significance, or absence thereof. For example, one missed clinical event may not signify a statistical difference in the sensitivity of one OEM monitor over another. However, the same event could be very important clinically and crucial that it not be missed even once. Theoretically, a missed event or a false alarm may be caused by artifacts of various types; for example, motion artifact and power line or optical noise induced in an attached or detached sensor.

Studies that collected real-time annotated data or video monitored data, such as [67, 70, 74, 75, 121, 124, 129, 136, 170, 183] among others, can utilise the same

data sets to develop and validate SQI. Compatible SQI can be used to compare performance of different AD algorithms and OEM monitors. Performance metrics, for example sensitivity and specificity in alarm studies, can be re-evaluated taking into consideration the SQI at each alarm instance.

2.2.6 THEME VI. CLINICAL IMPLEMENTATION

The theme of Clinical Implementation reviews whether or not AD techniques have been put into clinical practice. Theme results are synthesized in Appendix A: Review of Themes II – VI, in the bottom half of each cell in column four. This theme reveals that the vast majority of AD techniques that are published have not been put into clinical practice.

Interesting correlations arise between results of this theme and those of the previous theme, which comprised data annotation and mode of clinical evaluation. The review in Section 2.2.5 shows that the majority of the AD algorithms have been evaluated retrospectively or offline. Theme VI adds that most of the AD techniques evaluated retrospectively or offline have not reached clinical implementation. Some of these form parts of larger research projects that are not yet fully integrated in clinical workflows. Examples include work done by Portet *et al.* [135] that contributed towards the BabyTalk project in [184], the author's own work in [132] that serves as preliminary research in Artemis [31], and work done by Seely *et al.* [25, 185-187] that formed part of the CIMVA tool (this project is now closed; it was previously funded by TMS Inc., Ottawa, Canada). Research in which data were both annotated and evaluated in real-time include [126, 129, 170]; however, none of these algorithms were put into actual clinical practice. Correlating the results from Themes V and VI also identifies the research in which data were annotated and evaluated in real-time and the work was commercialized thus reaching clinical implementation. These include work done by Schoenberg *et al.* [111] that was implemented as iMDsoft Clinical

Information Systems [188], past and ongoing research by Moorman *et al.* [136] that formulated the HeRO™ system [189], and research contributions made by Addison *et al.* [103] that led to commercializing the RR_{oxi} by Covidien-Nellcor (Boulder, CO, USA).

It is clear that to reach clinical implementation methods must be fully validated for use within a real-time setting. Furthermore, most of these methods have been developed in isolation as stand-alone applications with little thought given to interoperability among the many AD, PD, and CED components that are required within a complex clinical environment. Therefore, we gather from all six themes in this review that faster and easier adoption and increased robustness of such techniques can be achieved by implementing them within a flexible framework to promote interoperability and real-time testing across multiple patient populations and CCU domains.

The remainder of this section reviews a number of commercialized OEM CDSS and the very few CDSS developed by independent research groups that have made their way into clinical workflows. As will be shown, none of these systems provides a flexible AD framework supporting multiple AD components.

The Philips IntelliVue monitoring system (Royal Philips Electronic, Netherlands) features *Guardian Early Warning Score* (EWS) that allows each hospital to choose its own scoring criteria; *Neonatal Event Review* which detects apnoea, bradycardia and desaturation; *Oxy-cardiorespirography* (Oxy-CRG) that presents compressed trends of a neonate's HR, RR, and SpO₂; and *ProtocolWatch* that is claimed to reduce sepsis mortality rates. Presumably Intellivue preprocesses patient data for artifacts prior to CED or PD, however, no validation studies or algorithm details of this system are published.

The GE IntelliRate™ monitor (Milwaukee, WI, USA) issues asystole, bradycardia and tachycardia alerts by fusing ECG, ABP and PO data. It was evaluated by GE on a small population of 55 CCU patients in 2002 [190]. The evaluation was critiqued in 2008 in [26] for lack of description of patient demographics and algorithm

specifications. GE has republished the exact same study in 2010 [191]. The Sapphire clinical decision support system [192] uses IntelliRate™ technology, but does not evaluate it.

Multi-lead ECG arrhythmia detection is deployed by GE in their Datex-Ohmeda Bedside Arrhythmia Monitoring system (GE, Milwaukee, WI, USA), the MARS Ambulatory ECG system and the MARS Enterprise system (GE, Freiburg, Germany). MARS uses the OEM's EK-Pro Arrhythmia Detection Algorithm which has been evaluated in over 2000 monitored hours spanning at least 100 patients. These techniques may fall under the category of AD coupled with CED and PD. However, the literature lacks comparison between different models marketed by the same or different OEMs.

OEM Covidien-Nellcor (Boulder, CO, USA) has developed RR_{oxi} , a coupled AD and PD technology that derives RR from PO. RR_{oxi} has been validated in real-time on 139 healthy subjects in [103]. The OEM is commended for this substantial evaluation. However, healthy subjects are not representative of patient populations which the device is intended to monitor. RR_{oxi} has been validated retrospectively in 12 patients with congestive heart failure, by evaluating 20 minutes of data from each patient [104]. However, larger studies that investigate patient populations with several different pathophysiologies are likely required to convince clinicians to adopt another patient monitoring technology in their workflows. Fidelity 100 is a wireless ECG monitor developed by Signalife (Studio City, CA, USA). It was evaluated in real-time in 54 patients undergoing percutaneous coronary intervention [170].

Although OEM monitors come with different settings applicable for use in different types of CCUs, validation studies on population data from all application domains are not found. There are a growing number of online open source physiologic databases, such as Capnabase [141, 142], FDA ECG Warehouse [193], hemodynamic parameter database [194], and PhysioNet [195]. This review recommends that these

databases be used to compare and validate OEM monitors of different makes and models.

The remainder of the section reviews clinical implementation of AD technologies developed by independent research groups. Artemis was reviewed separately in Section 2.1.3.1 above as it motivates the research in this thesis. Otero *et al.* have implemented TRACE, a graphical tool which allows clinicians to edit monitoring rules and criteria in real-time. Coupled AD and CED algorithms based on fuzzy set theory input these customized criteria to generate patient alarms in [121, 123]; and detect sleep apnoea in [156]. Given its promising results, evaluation of TRACE against similar CDSS is recommended. The research conducted in 1999 by Schoenberg *et al.* [111] was integrated as part of the commercially available iMDsoft Clinical Information System [188]. However, algorithmic details and evaluation were never published. The coupled AD and CED algorithm for Bradycardia by Portet *et al.* [135] was evaluated on offline NICU data with the intent of integration with the BabyTalk project. BabyTalk's proof of concept has been described in several publications [196-200]. However, latest research by Hunter *et al.* in 2012 [184] infers that a long road lies ahead, including necessary clinical trials, before BabyTalk could be implemented in real-time clinical workflows. Several new AD and CED algorithms can be tested to improve outcomes of this project by incorporating formalised interfaces as recommended in this review. The patented HeRO™ system (Medical Predictive Science Corp., Charlottesville, VA, USA) that scores neonatal Heart Rate Variability (HRV) for predicting LONS is developed using coupled AD and CED algorithms [173]. It conducts a multivariate type (d) analysis of the ECG with multiple coefficients [136]. The algorithms in this 510 (k) FDA approved device have been extensively described and evaluated both offline and in real-time by Moorman *et al.* [136, 189, 201-210]. This pioneering research has shown promising reduction in neonatal mortality by 2% in a randomized control trial on 3003 preterm babies across nine NICUs in the US [189]. The drawback of this trial was a

10% increase in blood work and 5% more days on antibiotics in the monitored infants. Ironically, this constitutes the original problem this research set out to resolve in [207]. It is recommended that other variability measures, such as those of RR as in [38] and PPG as in [211], be evaluated in comparison with the HeRO™ HRV score to come up with (a) individual scores for each data type; and (b) a composite score that exploits sensor fusion for improved outcomes. BioSign [172] has been evaluated retrospectively and in real-time in a number of clinical studies including randomized control trials in Europe and the US before 2006. However, no later publications could be located.

2.3 RESULTS

As briefly stated in section 1.3, this methodological literature review shows that the majority of AD algorithms have the following shortcomings (s.1) - (s.6):

- (s.1) They are developed for patient population from a specific type of CCU. Their flexibility to serve the needs of patient populations from other types of CCUs has not been assessed.*
- (s.2) They are hard-coded to process particular types of physiologic data harvested by specific OEM monitors. Since different monitor models from the same or different OEM may have different built-in proprietary signal pre-processing. Therefore, they may differ in their representation of data and artifacts. However, the reusability of AD algorithms in processing different types of physiologic data harvested by different OEM monitors has not been demonstrated.*
- (s.3) They are not equipped to generate unambiguous standardized SQI. The SQI generated by each algorithm are highly customized and cannot be meaningfully compared or integrated with those from another algorithm.*
- (s.4) They are designed and evaluated in isolation within a standalone application; or in a coupled application with one or more specific PD or CED algorithms. The*

standalone algorithms are neither reused nor the coupled algorithms decoupled for use in other applications in the same domain. This indicates a lack of generalization.

(s.5) They are largely evaluated offline despite the fact that real-time evaluation is an absolute requirement for integration of AD in clinical workflows. Although offline evaluation is important for demonstrating the validity of an algorithm, it must also be evaluated in a real-time environment prior to deployment in a true clinical setting. This highlights the need of a standardized, scalable and customizable AD framework that allows for evaluation of AD algorithms on both offline and real-time patient data; and that can be integrated with clinical workflows.

(s.6) They are not commonly implemented in clinical practice. No framework with standardized component interfaces is documented to support real-time clinical implementation of AD.

2.4 CONCLUSIONS

This section derives conclusions from this review and provides new research directions for promoting integration of AD in critical care. Post processing AD techniques are highly domain specific. This necessitates modification for validation and reuse in a different CCU domain. Algorithms may be hard coded to input OEM specific data types and frequency. This limits their use with different OEM monitor data. They may be validated under certain inclusion/exclusion criteria which need to be considered when applying the techniques in other contexts. Acquisition and sampling frequency play an important role in patient management since a critically ill patient's condition may deteriorate to a life threatening extent within seconds. Reusability is deterred when such implicit limitations are not expressed. Therefore, a standardised structured approach for design and reporting of AD algorithms is worth adopting. Conformity to common definitions and syntax for input and output interfaces will ensure presence of

all pertinent information. These formalised interfaces will allow the expression of data type, frequency, value and SQI.

Results from the first three themes are useful in selecting one or more AD algorithms that fulfill data requirements of given CED or PD techniques. From amongst this selection, AD algorithms can be mixed and matched to discover optimal compositions for varying CED and PD requirements.

OEM monitors marketed for use across different CCUs have undisclosed built-in preprocessing algorithms, inclusive of AD. Moreover, studies validating their use in different CCU domains and patient population are scarce. The resulting unknown bias imparted in OEM data leads to inevitable variance in analytic results which can effect clinical decisions. This variability can be decreased if monitors output comparable standardised SQI. As of yet, SQI do not conform to any standards and are derived differently in each publication, whether it be SQI delivered by OEM patient monitors or by AD algorithms developed by independent research groups. Interestingly, none of the reviewed algorithms reported using SQI provided by OEM monitors. Clinical utility of SQI can be enhanced by using formalised definitions such that SQI output by different preprocessing and post processing AD algorithms are comparable as well as compatible. An SQI matched to the same set of definitions is also proposed as a requirement at the input of AD, CED or PD algorithms. The objective is to enable these algorithms to compare the incoming SQI generated by one or more AD techniques with their required SQI type and frequency. The AD, CED or PD algorithm may then accept or reject incoming data segments based on fulfillment of the required SQI. In conclusion, standardised SQI are vital to allow informed clinical choices concerning use and validity of physiologic data.

Results of the Clinical Evaluation theme show that majority of AD techniques are validated on offline data and very few have been evaluated in real-time CCU environments. The Clinical Implementation theme reveals that AD techniques

developed by independent research groups have rarely found their way into clinical implementation. This leads to the inference that a gap exists between research efforts in AD and their utilization in real-time clinical workflows.

Whereas real-time clinical implementation of AD algorithms is noticeably lacking, there is growing interest amongst clinicians to use CED and PD for automated clinical decision support in CCU, such as in [6, 31, 96, 212-222]. Physiologic signal quality assessment through integration of AD can improve the outcome, reliability and accuracy of CED and PD research.

2.5 SUMMARY

This methodological review introduces the following six unique thematic taxonomies to synthesize over 80 AD algorithms: I. Critical Care Unit; II. Physiologic Data Source; III. Harvested data; IV. Data Analysis; V. Clinical Evaluation; and VI. Clinical Implementation. This review identifies six shortcomings of state-of-the-art AD techniques, listed as (s.1)-(s.6) in Section 2.3. To address these shortcomings, it is recommended that AD algorithms conform to generic input and output interfaces with commonly defined data: (1) type; (2) frequency; and (3) SQI. This shall promote generalization, standardization, customization, reusability, flexibility, scalability and adaptability of AD algorithms across different CCUs. The conclusions and recommendations of this review provide new research direction for promoting integration of AD in real-time clinical workflows.

3 ARTIFACT DETECTION FRAMEWORK

This chapter develops a novel multivariate Artifact Detection (AD) framework for integration with Clinical Decision Support Systems (CDSS). Parts of this chapter have been published and presented in conference proceedings as S. Nizami, J. R. Green and C. McGregor, "An artifact detection framework for clinical decision support systems," *in World Congress on Medical Physics and Biomedical Engineering, Toronto, Canada, June 7-12, 2015*, pp. 1393-1396 [223].

This thesis applies a constructive research approach as previously elaborated in Section 1.4. Section 3.1 introduces component-based system design as a method to build an applicable solution to the research problem stated in Section 1.3. Section 3.2 develops the framework with four types of components, with interfaces defined by a Common Reference Model (CRM). The CRM is an original contribution of this thesis. Section 3.4 describes how the AD framework promotes the six desirable features relevant to each theme reviewed in Chapter 2. Section 3.5 discusses advantages and challenges of the research methodology. Section 3.6 summarizes the chapter.

3.1 COMPONENT-BASED SYSTEM DESIGN

Component-based system design allows for framework construction and extension through interfacing of independent modular components [49, 224, 225]. The terms framework, component, and interface are defined as follows. Component-based frameworks developed for clinical applications that fall outside the scope of this thesis are found in [53, 55, 56].

3.1.1 FRAMEWORK

A *framework* is a high level hypothetical description of a complex entity or process. It ties in components and interfaces, providing an environment that supports end-to-end

system functionality. From an electrical engineering perspective, a framework is analogous to a circuit board in which empty slots can be filled with components [49]. The wires connecting the components represent interfaces that create a working instance at run-time. The low level details of how a component, for example an integrated circuit (IC), implements its role in the framework are hidden from the user and other components in the framework. In developing a framework, the low-level source code implementation can be realised using middleware technology [225], such as MATLAB®, Java, C++ and SPL amongst others. For example, this thesis implements the low-level component source code using MATLAB®.

3.1.2 COMPONENT

A component can be broadly defined as a unique and self-contained entity with a clear and well-defined function within the framework. More precisely, it is packaged software that contains the source code; it is both replaceable and reusable as a stand-alone module. Components can collaborate with each other to render composite behaviour that fulfills goals at more complex levels of system hierarchy [225].

3.1.3 INTERFACE

Components interact with other components and with the system through one or more interfaces to realise system goals. An *interface* is defined as a means to communicate with or access a component [226]. Clearly defined uniform interfaces enable components to make explicit their own functional requirements as well as to enact according to specifications of other collaborating components. Therefore, defining interfaces using unambiguous formalisms with semantics commonly understood by all components within the framework is key for achieving system goals.

Interfaces stipulate prerequisites, provisions and constraints of component operation. A component can have one or more interfaces, selectively instantiated at

run-time depending on the component's role in a particular composition. As described in [225], an interface can be categorized as: 1) requirement, 2) provision, and 3) configuration. Each component has its own operational requirements, specified by its requirement interface, that define what the system or other components in the system must provide in order for the component to function [226]. The provision interface makes explicit what a component can provide either to another component or as a contribution to the system output. The configuration interface incorporates user-defined functionality. It allows the user to define run-time parameters for a particular application. A configuration interface can be part of the user interface designed for a clinician to interact with system settings.

3.2 AD FRAMEWORK DEVELOPMENT

According to Larsen, *"Designing, building, and testing component-based frameworks require considering several things including static models that illustrate component structure, dynamic models that illustrate component collaboration, as well as the technology to implement the component framework"* [227].

3.2.1 FRAMEWORK

This section develops a static model consisting of framework components and interfaces. Figure 3-1 shows the static model developed for the AD framework. This static model describes how the four different types of framework components might interconnect within a given CDSS. The arrows in the figure indicate the direction of data flow between component interfaces. The Common Reference Model (CRM) describes the fields for the data exchange in the interfaces. The validation study developed in Chapter 4 shall illustrate component collaboration through dynamic modelling of the framework.

3.2.2 COMPONENTS

As shown in Figure 3-1, there are four types of components in this framework: (1) Patient Data Acquisition (PDA); (2) Artifact Detection (AD); (3) Parameter Derivation (PD); and (4) Clinical Event Detection (CED). The PDA component inputs patient data. Some examples of data sources include, but are not limited to, OEM patient monitors, clinical data entry, lab results, physician's order, and patient demographics on the Hospital Information System (HIS). The PDA translates these data to the schema defined by the CRM. It is important to note that the PDA component is able to input an OEM-generated SQI stream, if available, corresponding to any given input data stream. This would require mapping the input SQI to an SQI type available within the CRM. For available SQI types, refer to Sec. 3.2.3.1. The SQI stream would then pass through the provisions interface of the PDA component. If the OEM monitor does not provide an SQI stream, then the SQI at the provisions interface is set to a default *SQvalue* of 1. The provisions interface of the PDA component then supplies all data to

Figure 3-1: Static model of the multivariate component-based AD framework.
Reprinted with permission from the author's publication [219].

the AD component. It can also provide these data directly to the CED/PD components where required.

The framework is a unique test bed that allows mixing and matching of multiple independently developed AD, CED and PD algorithms. Any AD, CED, and PD algorithm, published previously or in future, standalone or coupled, can be used as a framework component. For example, Figure 3-1 depicts a pool of AD algorithms, $AD_1 \dots AD_x$, that exist as independent AD components in the framework. Similarly, $CED_1 \dots CED_n$, are a pool of CED algorithms slotted in the framework as independent CED components. The model also consists of a pool of PD algorithms, $PD_1 \dots PD_m$, which are independent PD components. These components can be used either as standalone algorithms or in cascade with the same or different type of component.

Previously coupled algorithms can be explicitly decoupled and slotted as independent components. Decoupling would include splitting the algorithm into two or more independent AD, CED and/or PD components, each with interfaces conforming to the CRM. Hence, the CED/PD algorithm no longer remains bound to use the results produced by its originally coupled or built-in AD algorithm. The performance of a given CED/PD technique can then be evaluated in composition with any other available AD component that fulfills all the data requirements of the CED/PD. The framework facilitates testing various compositions of AD, CED and PD techniques developed by different researchers. The most suited combinations can then be hard-coded for implementation in CDSS.

3.2.3 INTERFACES

This subsection develops the three types of interfaces for the AD framework.

3.2.3.1 Requirement and Provision Interfaces: The Common Reference Model

A novel contribution of this research is the development of a Common Reference Model with common data standards defined for the requirement and provision interfaces. Results of the thematic review, particularly of themes III and IV in Section 2.3, allows for the creation of an abstract or common reference model for interfaces that captures the functionality of existent and potential AD, CED and PD components. Signal quality is a key element in this reference model. The CRM provides for the quantification and implementation of standardized Signal Quality Indicators (SQI) to enhance the reliability of CDSS.

The CRM comprises metadata which are intended to establish a common understanding of the meaning or semantics of the data exchanged between component interfaces. This allows components to communicate in the same language regardless of their underlying low-level logic. As an example, in our earlier work [132], we employed eXtensible Markup Language (XML) documentation to standardize data exchange between components.

The hierarchical schema developed for the requirement and provision interfaces in the CRM is shown in Figure 3-2. Metadata and variables are described as follows in italics.

- i) *PatientData* is a class of objects representing patient data exchanged between components. The schema, shown in Figure 3-2, consists of three properties: *PatientID*, *DeviceID* and *Data*. *PatientID* is self-explanatory; it identifies the patient with whom the data are associated. It can be any type of patient identifier used by the hospital's information system. Some examples are the patient's admission reference number and the patient's name with date of birth. *DeviceID* represents the hospital or OEM identifier for the patient monitor or

other device from which the data are being acquired. The more complex *Data* property is described as follows.

- ii) The *Data* field has the following attributes: *Type*, *TimeStamp*, *Value* and *SQI*. *Data.Type* takes on a value from a controlled vocabulary, such as 'ECG', 'PPG', and 'HR'. This may be implemented using an *enumerator* (enum, similar to their definition in Java). An enum variable is a string or numeric data type that defines a list of valid constants or predefined values. In any given instantiation, the enum can take on only one value from this list. For dynamic modelling of this framework, the list may be created using definitions from a controlled ontology, such as Systematized Nomenclature of Medicine--Clinical Terms (SNOMED CT)¹. *Data.TimeStamp* is the time at which each datum is logged. A component has specific data exchange and processing rates, thus requires data at specific frequencies. Therefore, it is important to know the OEM physiologic monitor's data logging frequency. Input and output data logging frequencies of cascaded components must be also compatible. *Data.Value* contains the actual numeric or string value of each datum. The low-level component source code may deliver an *SQI* associated with each datum described as follows.
- iii) The *SQI* field quantifies the quality or confidence for each datum. It is further described by these attributes: *SQType*, and *SQValue*, as shown in Figure 3-2. *SQType* is an enum, such as: 'Binary', 'Rank', and 'Continuous'. Figure 3-3 depicts some examples of the *SQType* field developed in this CRM. Additional *SQI* types can be introduced to this set in future. The range of *SQValue* is dependent on *SQType*. For example, if *SQType* is 'Binary', then *SQValue* can be either 0 or 1. Here, and with other types, a higher/larger value indicates increased quality. As detailed in Chapter 2, only a small number of newer AD

¹ https://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html

algorithms produce an explicit SQI. For example, the algorithms in [19, 228] generate binary SQI for ECG and PPG data which map naturally to the *SQType* 'Binary'. Quesnel *et. al* [20, 21] have developed an algorithm that estimates the signal-to-noise (SNR) ratios of ECG data in the range of -20dB to 20dB, in steps of 5dB. These estimated SNR values can be mapped to the *SQType* 'Rank'. Another example of a 'Rank' *SQType* is found in Table 1 of [171], where the authors examine the agreement between ECG and PPG waveforms to generate a 4-level 'Rank' type SQI.

Based on the literature review in Chapter 2, one can infer that although OEM monitors may output signal quality information, there is no methodological way to interpret an SQI value or to compare the SQI produced by different monitors. There are four factors that lead to the above inference: (1) different built-in signal processing which results in different data representation between different OEM monitors as shown by [112-114, 176]; (2) difference in qualification of signal artifacts [177]; (3) difference in quantification of SQI, for example, see [24, 81-83] where different SQI quantification levels are used for the same signal type; and (4) lack of literature on proprietary signal preprocessing. For example, the Delta, Gamma, Vista, Kappa and SC6002 - SC9000 monitor series (Dräger Medical Systems, Lubeck, Germany) output a continuous numeric SQI value between 0-100% for the electroencephalography (EEG) channel.

The OEM simply states that this SQI is calculated using sensor impedance data, artifact information and other undisclosed variables [174, 175]. These reasons confound clinicians' and researchers' ability to interpret and utilize SQI values for purposes such as clinical decision support.

The CRM addresses this lack of interpretability of SQI values by defining a number of *SQTypes* with well-defined semantics. Input SQI sourced from a particular OEM monitor must be strictly matched to a valid *SQType* in the CRM before

consumption by any framework component. Missing segments, error, noise and artifacts inevitably affect data quality thus adversely impacting analytic accuracy and

reliability [169]. To address this issue, Clifford *et al.* [65] recommend that an SQI calibrated to provide a known error rate for a given value of the SQI be made available for each datum.

In some cases, a single SQI may apply to multiple data streams. For example, a single *SQI* stream may be applicable to both the Pulse Rate (PR) and Arterial Oxygen Saturation (SpO₂), since both are sourced by the same sensor. In this case, the SQI would be associated with each data type separately, since SQI is an attribute of Data. In this thesis, a tuple is used to represent both data and SQI information for each data

type. Figure 3-4 illustrates the colour-coded legend for the tuple as used in subsequent figures.

It may be necessary to convert between *SQTypes*. Consider that a clinician cognitively perceives signal quality of a given data stream to be along a continuum, interval or ratio. Therefore, the clinician may wish to know when a signal is of "Good", "Fair", "Poor" or "Unreliable" quality. In this case, a 'continuous' *SQType* can be converted to a 'rank' *SQType*. For example, a continuous SQI of 75-100% can be used

to represent a "Good" signal, followed by 50-74% being called a "Fair" signal, 25-49% SQI would equal a "Poor" signal, and below 25% would be an "Unreliable" signal deemed unsuitable for clinical analysis. Considering that different ranges would be appropriate for different CED applications, these ranges should be configurable.

The configuration interface of a component allows the user to set parameters for a particular application prior to run-time. For example, this interface could allow the user to set minimum and maximum threshold values of physiologic variables that

could vary greatly between neonatal and adult populations [229]. AD blocks become more powerful by including tunable parameters. For a given composition of AD and CED blocks, the configuration could be changed to optimize performance of each

component, whether in standalone or cascaded mode of operation. The configuration interface may be different for each component, depending on what parameters are required for the low level logic to operate successfully in a given component. For each of the example AD blocks in the following component catalogue, both the internal logic and also the configuration interfaces are defined.

3.3 COMPONENT CATALOGUE

The literature reviewed in Chapter 2 includes a wide variety of AD algorithms. This section demonstrates how the proposed framework can accommodate such diverse AD functionality. A non-exhaustive catalogue of exemplar AD components is developed here, emanating from the following six diverse functional AD groups: (1) de-interlacing of data and SQI; (2) data threshold; (3) SQI conversion; (4) SQI fusion; (5) data fusion; and (6) data smoothing. Although, each exemplar component differs in its low-level source code, all of the components conform to the same requirement and provision interfaces as defined by the CRM. Low-level logic and configuration interfaces are described for each component description below.

3.3.1 AD_{DIL} DE-INTERLACE COMPONENT

Some monitors produce a single output stream that is, in fact, a combination of data and SQI. The AD_{DIL} component is designed to de-interlace (DIL) these two information streams by allowing the user to define a set of symbols (*artSyms*) to be associated with corresponding SQI values. Typically, *artSyms* would be a list of OEM-specified artifact indicators, either numeric or string values. The low-level code for the AD_{DIL} component is shown in Figure 3-5.

The configuration interface of the AD_{DIL} component would specify the Data.Type to be examined, *artSyms*, and the corresponding set of *SQValue* (SQI_{Match}). This AD

component produces a 'rank' *SQType*, with 'binary' being a special case of 'rank', where $SQI_{Match}=1$.

For example, for Infinity monitors (Dräger Medical Systems, Lubeck, Germany), the set of *artSyms* would include {NaN, ^^, 5}; where NaN is substituted for any missing datum, ^^ is an artifact indicator, and 5 is an alarm state (i.e., part of the alarms stream) in the case of leads disconnection. In a different example, Philips Intellivue MP70 monitors (Philips, Germany) generate a value of "2" in the alarms data stream in case of leads disconnection. However, with the alarms data stream connected to the input of the AD_{DIL} component, the value "2" can be identified by the component as an artifact. In such a way, the component can de-interlace the alarms stream, and generate a corresponding binary SQI stream. Other examples include the Delta, Gamma, Vista, Kappa and SC6002 - SC9000 monitor series (Dräger Medical Systems, Lubeck, Germany) which output a fixed label 'ARTF' indicating artifact on any monitored data stream [174, 175]. For example, the monitor classifies QRS complexes only at ECG values > 0.20 mV for widths > 70ms. An artifact condition 'ARTF' may be declared when a QRS complex is detected but the ECG signal fails to meet these minimum criteria. The user can configure the AD_{DIL} such that a value of 'ARTF' produces an SQI of zero. Multiple instances of this component are integrated in the AD framework in the validation study in Chapter 4.

3.3.2 AD_{THRESH} THRESHOLD COMPONENT

The AD_{Thresh} component compares a univariate input data stream to a user-configured

```

if (Data.Value(i) ∈ artSyms)
    SQIout(i) = SQIMatch(Data.Value(i))
end

```

Figure 3-5: Low-level code for the AD_{DIL} component

threshold to generate a corresponding SQI stream. The configuration interface of the

$A_{DThresh}$ component would specify the *Data.Type* to be examined; the number of thresholds, n_{Thresh} , the set of thresholds ($thresh_j; j=1:n_{Thresh}$); the operator(a,b) to be applied, such as $[a \geq b]$, $[a = b]$, $[a < b]$; and the set of n_{Thresh} *SQValues* (SQI_j) corresponding to each threshold. The $A_{DThresh}$ component produces a 'rank' *SQType*, with 'binary' being a special case of 'rank', where $n_{Thresh} = 1$ and $SQI_j=0$. The low-level logic is described in Figure 3-6.

For example, Philips Intellivue MP70 monitors (Philips, Germany) produce a random integer above 8 million in the case of leads disconnection or *sensor off*. An $A_{DThresh}$ component could be used to identify these artifacts and annotate the SQI stream. Here an $A_{DThresh}$ component that produces a 'binary' *SQType* would be configured with the operator $[a \geq b]$; $n_{Thresh} = 1$ and $SQI_j=0$. Then, for each datum in the input stream *Data.Value* that is equal or greater than 8×10^6 , the $A_{DThresh}$ provisions interface provides a binary *SQValue* of 0. Another possible use of the $A_{DThresh}$ component is in the threshold-based ectopic heart beat removal process described in [230, 231]. In this process, whenever the magnitude of an ECG R-wave exceeds a preset threshold, that beat is considered to be ectopic and, hence, an artifact. In this case, an $A_{DThresh}$ component would be configured with the percentage threshold for identifying ectopic beats, thereby, issuing a 'binary' *SQType* of *SQValue* = 0 every time an ectopic beat that surpasses the threshold is detected.

```

for j=1: nThresh
{
    if (operator(Data.Value(i),thresh_j)
        SQI_out(i) = SQI_j
    }
end

```

Figure 3-6: Low-level code for the $A_{DThresh}$ component

3.3.3 AD_{CONVERTSQI} CONVERT SQI COMPONENT

The AD_{ConvertSQI} component facilitates conversion between the different types of SQI required or provided by different framework components. Conversions can occur at different levels of abstraction. Conversions from a more dense to a less dense *SQType* are valid, e.g., 'continuous' to 'rank', but not the converse. Consider, for example, the Dräger Medical Systems EEG monitor reviewed in Section 2.2.4. These monitors produces a continuous SQI stream in the range 0 to 100%. Should another AD, PD, or CED component require an *SQType* of 'binary' or 'rank' instead, the user may wish to map this OEM-generated 'continuous' SQI to the desired type. The low-level code for the AD_{ConvertSQI} is shown in Figure 3-7. The configuration interface would specify the *output SQType* to be produced; the number of thresholds, n_{Thresh} , the ordered set of thresholds (thresh_j ; $j=1:n_{\text{Thresh}}$); and the set of $n_{\text{Thresh}}+1$ *SQValues* (SQI_j) corresponding to each threshold with the additional *SQValue* for the default case ($\text{SQI}_{\text{default}}$).

To map the 'continuous' SQI stream between 0 to 100% to a 'rank' *SQType*, the configuration interface would specify the *output SQType* as 'rank'; the number of thresholds, $n_{\text{Thresh}}=3$, the ordered set of thresholds ($\text{thresh}_j=\{25, 50, 75\}$); and the set of $n_{\text{Thresh}}+1$ *SQValues* (SQI_j) would be $[0,1,2,3]$ where $\text{SQI}_{\text{default}} = 3$ would be applied in the case where the SQI datum is greater than all thresholds.

```
SQIout (i)= SQIdefault
for j=1: nThresh
{
    if (SQIIN (i)≤threshj)
    {
        SQIout(i) = SQIj
        break
    }
}
end
```

Figure 3-7: Low-level code for the AD_{ConvertSQI}

3.3.4 AD_{FUSESQI} FUSE SQI COMPONENT

AD_{FuseSQI} component accepts more than one data stream at its requirements interface, along with the respective SQI of each stream. This component combines N incoming SQI inputs to generate a single fused SQI. The fused SQI value is equal to the operator, i.e., minimum, maximum, or average SQI value from all the input SQI data at any given instant. Since all input *SQTypes* must be the same; AD_{ConvertSQI} components may be required as part of the workflow to convert all input *SQTypes* to a common desired type. This is exemplified in Figure 3-8. The configuration interface of the AD_{FuseSQI} component would define N , the required input *SQType* (same as output), the operator (min, max, avg), to be applied to all input SQI values. Additionally, the configuration interface can specify which data types to forward at the provision interface, as only a subset of the two input streams may be required beyond this component. The low-level logic is illustrated in Figure 3-9 below. This is a relatively simple illustration of data fusion. Previous literature on data fusion shows that fusion can occur at different levels of abstraction, can require more complex combination of operators, and may require weighting [232].

$$SQI_{out}(i) = \text{operator}(SQI_1, SQI_2, \dots, SQI_N)$$

Figure 3-9: Low-level code for the AD_{FuseSQI} component

3.3.5 AD_{DIFF} DIFFERENTIAL COMPONENT

In the case where one is provided with two independent measurements of the same physiological parameter then one can derive an SQI that exploits data fusion. For example, consider the work on wearable devices and systems published by Da He *et al.* in 2015 [233]. The study collects ballistocardiogram (BCG), ECG and PPG data synchronously, and suggests that checking if all three physiological signals measure the same HR can be used to ensure that the data are not corrupted [233]. However, the study does not demonstrate how and if it actually checks for the data quality. Such a system would benefit from the use of the AD_{Diff} component.

The AD_{Diff} component calculates an absolute differential error function between two input data streams, Data₁ and Data₂. It then compares this error value to a configured threshold. The input 'binary' SQI streams are examined to ensure that both input data streams are valid prior to computing the difference. If either stream has poor signal quality, the output *SQValue* = 0. The configuration interface would specify the *output SQType* to be produced; the *Data.Type* of Data₁ and Data₂; the number of SQI thresholds, *n_{Thresh}*, to be applied to the difference; the ordered set of thresholds

```
SQIout (i)= SQIdefault
if (Data1.SQI.SQValue(i)==0)|| (Data2.SQI.SQValue(i)==0)
    return;
diff = |Data1.Value(i)- Data2.Value(i)|
for j=1: nThresh
{
    if (diff ≤ SQthreshj)
    {
        SQIout(i) = SQIj
        break
    }
}
end
```

Figure 3-10: Low-level code for the AD_{Diff} component

($SQthresh_j$; $j=1:n_{Thresh}$); and the set of $n_{Thresh}+1$ *SQValues* (SQI_j) corresponding to each threshold with the additional *SQValue* for the default case ($SQI_{default}$). As with the $AD_{FuseSQI}$ component, the configuration interface can specify which data types to forward at the provision interface. The low-level logic for this component is illustrated in Figure 3-10.

For example, an AD_{Diff} component can derive a 'rank' *SQType* stream from HR and PR streams by configuring the component to have *output SQType* set to 'rank'; the *Data.Type* of $Data_1=HR$ and $Data_2=PR$; $n_{Thresh}=3$; $SQthresh=\{6,12,18\}$; and $SQI_j = \{3,2,1,0\}$, where the $SQI_{default} = 0$. An AD_{Diff} component configured in this way is used in the validation study in experiments 3 and 4 described in Sections 4.2.6 and 4.2.7 respectively.

3.3.6 $AD_{MEDFILT}$ MEDIAN FILTER COMPONENT

$AD_{MedFilt}$ component implements a median filter. It is used for smoothing a data stream to mediate transient abrupt artifacts. The configuration interface defines the size of sliding window, Med_{WW} , to use when computing the median value. Its requirement interface inputs a single data type and its corresponding SQI stream. Each datum in the output data stream is equal to the median of the past Med_{WW} input data samples. Only a subset of these Med_{WW} may actually be used in computing the median because the $AD_{MedFilt}$ component only includes the data within the sliding window for which the input SQI are acceptable. The SQI stream is passed through this component unchanged. By comparing the filtered and unfiltered data using an AD_{Diff} component, one could compute an SQI proportional to the degree of smoothing applied to each point.

3.4 RESEARCH CONTRIBUTIONS

This section discusses how the AD framework developed in this chapter supports each of the six framework features (f.1) - (f.6) originally defined in section 1.5. Each feature is restated below, followed by a discussion of how the framework, CRM, and exemplar AD catalogue support each feature:

(f.1) *Flexibility to serve the differing needs of patient populations from different types of CCUs through generalization and customizability.* The CRM includes several fields to generalize and customize each component, for example, *Data.Type*, *Data.SQType*,

(f.2) *Reusability across multiple types of physiologic data harvested by different OEM monitors.* The configuration interface of each component permits setting OEM-specific and *Data.Type*-specific values such that the same component may be applied to various physiological data types arising from different OEM monitors. For example, the *artSyms* configuration parameter allows the AD_{DIL} component to identify artifacts flagged by different OEM monitors.

(f.3) *Standardized definitions of SQI that promote interoperability between independently developed components* For example, the CRM defines standardized types of SQI, such as, 'continuous', 'rank', and 'binary'; and the framework allows for arbitrary SQI streams to be converted to these standard types.

(f.4) *Reusability and scalability by cascading, and mixing and matching several AD, PD and CED components in various combinations.* By requiring all component interfaces to conform the standardized CRM, interoperability is promoted allowing for component reuse and the creation of highly complex pipelines leveraging simple and well-tested components. As demonstrated in Chapter 4, the same set of components from the catalog above may be reused, cascaded, mixed and matched and evaluated in various framework configurations.

(f.5) Customizability to evaluate and compare the performance of multiple combinations of independently developed components on offline and potentially real-time patient data when integrated with clinical workflows. The framework facilitates real-time validation of components and complex analysis pipelines. As found in the literature review in Chapter 2, algorithms are typically developed and validated in offline environments. Therefore, testing in a real-time environment would require potentially error-prone reformulation of the algorithm which invokes the need of subsequent re-validation. Conversely, algorithms developed within the framework would be developed and tested in a simulated real-time environment. In this way, transition to real-time clinical implementation and validation is facilitated.

(f.6) Standardized component interfaces that can potentially support real-time clinical implementation of AD. If independent research and OEM groups choose to implement their algorithms within the context of the framework (adhering to the CRM), then it is more likely that these algorithms will reach clinical implementation because the CRM supports interoperability between all components. Furthermore, the framework simplifies IT requirements for hospitals since it provides a unified functional environment in which all AD, PD, and CED components required by multiple CCUs can be supported and executed. Lastly, the framework facilitates testing and validation of new algorithms across different clinical settings, populations, CCUs, pathologies, etc. This will make the system more robust and therefore more likely to be adopted [49]. As stated in Chapter 2, most algorithms described in the literature have been developed and evaluated only on offline data. One hurdle to adoption of this research in a clinical setting is the requirement to transform the complex algorithms for real-time implementation. The framework is implicitly designed to run components in a real-time streaming environment. Through the exercise of implementing the algorithm within the framework, the algorithm will be made suitable for execution in real-time.

In this way, the framework facilitates the composition and evaluation of complex real-time analysis pipelines, as demonstrated in the validation experiments in Chapter 4.

3.5 ADVANTAGES AND CHALLENGES

This section discusses the advantages and challenges of component-based system design as they pertain to the developed AD framework.

The prime worth of developing a framework lies in its reusability [48], which also leads to cost cutting [47]. Reusability means that the components of the framework are scalable and adaptable for use within similar applications requiring only minor modifications [49]. For example, AD or CED algorithms designed for a neonatal environment could be adapted for use in the OR or adult ICU by modifying their configuration. Or components designed for single patient or single stream analysis could be cascaded within this framework to multiple patient or multivariate analysis. The multivariate framework developed in this thesis is scalable in that it can integrate multiple AD and CED/PD components to support simultaneous detection of multiple diseases. Standards developed for this framework can be generalized for reuse in application areas that require AD outside of the clinical domain. The drawback of reusability is that it requires a higher level of abstraction when designing the model. This allows for generalization and development of a simplistic model at the cost of reduced flexibility in low-level component design [49]. This forces new components to comply with pre-defined standards, i.e., the CRM, for successful integration with the existing framework. However, this challenge is overcome by employing XML as the CRM definition vehicle. XML has the inherent capability of standardizing data portability across applications due to its self-describing nature. It is recognized for its efficacy in representing healthcare ontologies when used to promote interoperability and knowledge sharing [234, 235]. XML is also extensible wherein new data type definitions can be added by new applications [236]. Although, a drawback associated

with using XML is the overhead transmission cost. However, there is ongoing research in the field of XML compaction to reduce this data bloat [237, 238].

Frameworks also offer customizability. This means that custom designed components can be added to modify or extend the system without affecting the interaction between existing components. For example, new AD and CED/PD components can be integrated in the framework provided that their interfaces are specified according to the CRM proposed in this thesis. Kaisler [48] notes that adhering to industry or organizational standards helps in reducing development costs. He states that as more applications are created using the framework, the base mechanism becomes more robust thus increasing the likelihood of the framework becoming an industry standard. Hence, as greater reuse of this framework is created, it will expand the opportunities of formalised clinical implementation of AD, CED and PD techniques.

It is universally understood that framework development is a time consuming and arduous effort [49], therefore it should be attempted where multiple opportunities exist for its implementation [48]. The multivariate AD framework developed in this thesis is the first of its kind. Integrating the framework with clinical decision support systems in different clinical organizations worldwide can potentially lead to a sophisticated enterprise in healthcare.

3.6 CHAPTER SUMMARY

This chapter introduced component-based system design as the research methodology in this thesis. A multivariate AD framework was developed for integration with health care analytic systems. The framework was based on four types of components: (1) PDA; (2) AD; (3) CED; and (4) PD. A CRM was developed to define the requirements and provisions interfaces for all four component types. As part of the CRM, I have defined a standardized way for components to exchange and transmit SQI information. This included three illustrative examples of distinct SQI types with a discussion of how

these could be transformed from one to the other. An AD catalogue was developed in this chapter which is validated in experiments developed in Chapter 4.

4 RESEARCH VALIDATION

In this chapter, I present four different experiments to validate the six framework features (f.1) - (f.6). A number of AD composition pipelines are constructed and evaluated in the context of an SpO₂ alarms generation study. This study demonstrates the stated benefits of the AD framework developed in this thesis.

Physiologic patient data for this research was collected in a clinical study conducted at the Neonatal Intensive Care Unit (NICU) of the Children's Hospital of Eastern Ontario (CHEO), Ottawa, Canada. The Research Ethics Board (REB) at CHEO approved the neonatal data collection at the patients' bedside and subsequent analyses of these data. Arterial oxygen saturation (SpO₂), Pulse Rate (PR), and Heart Rate (HR) data streams were acquired along with the corresponding alarms issued by Original Equipment Manufacturer (OEM) monitors. The research was conducted independent of OEM involvement and bias.

I initiated this study in my capacity as a research associate at CHEO. I was responsible for surveying the literature, designing the study, drafting the application for research ethics approval, setting up the research equipment and recording data at patient bedside, annotating data in consultation with the bedside nurse and the neonatologist, de-identifying data, reverse engineering the structure of the OEM data packets, retrospectively analyzing data, computing performance metrics, and authoring, submitting and revising manuscripts for research dissemination. A crossover observational study consisting of primary analysis of the acquired physiologic dataset has been published as Nizami S., Greenwood K., Barrowman N., and Harrold J., "Performance evaluation of new-generation pulse oximeters in the NICU: Observational study", *Cardiovascular Engineering and Technology*, June 2015:1-9 [239].

The primary analysis evaluated the performance of two new-generation pulse oximeter technologies, namely Nellcor OXIMAX and Masimo SET SmartPod. Study results provide performance metrics for Nellcor OXIMAX and Masimo SET OEM monitors in terms of Sensitivity (S_n), Positive Predictive Value (PPV), and False Discovery Rate (FDR) of arterial oxygen saturation (SpO_2) alarms. The term False Alarm Rate (FAR) is commonly substituted for FDR in literature pertaining to alarm generation. A detailed comparison of Masimo and Nellcor SpO_2 datasets from the published study can be found in Appendix C: Masimo vs. Nellcor SpO_2 Comparison. Ultimately, the primary analysis shows no statistically or clinically significant difference between these two sensor technologies. Therefore, only Masimo data is analyzed in subsequent experiments described below.

In this chapter, I develop an SpO_2 alarms generation study to validate the AD framework developed in this thesis. This study consists of four different experiments. Each experiment uniquely comprises a different composition pipeline of AD components that are integrated with PDA and CED components for SpO_2 alarms generation, developed as part of the study. These compositions are orchestrated by mixing and matching different AD components that were developed in the AD catalogue in Section 3.3. Thereby, the four experiments orchestrate various compositions within a CDSS that generates SpO_2 alarms at its output. The alarms generation performance in each experiment is evaluated using S_n and FAR.

The remaining chapter is organized as follows. Section 4.1 describes the data acquisition methods for the crossover observational study conducted at CHEO. Section 4.2 describes the research methods used in the validation study. This includes development of the PDA component, the CED component for SpO_2 alarms generation and the four different composition pipelines of AD components used in each of the experiments. Section 4.3 presents results of this study, including data characteristics and experimental results. Section 4.5 discusses the results and concludes the chapter.

4.1 DATA ACQUISITION METHODS

All patients admitted to the NICU at CHEO were eligible to enroll in this study, with the following exclusion criteria: (a) parent's refusal, (b) moribund patients, (c) patients with patent ductus arteriosus (PDA) or cyanotic congenital heart disease. Each infant was simultaneously monitored non-invasively by Masimo and Nellcor pulse oximeters using weight-appropriate foot sensors from the respective OEM applied to each foot. The selection of the right or left limb to apply either Masimo or Nellcor sensor was randomized at the beginning of each recording. To minimize bias due to limb selection, the bedside nurse switched each sensor to the other foot midway through the recording. Blue posies were wrapped around both sensors to prevent optical crosstalk between them. Three ECG electrodes by Dräger were appropriately placed on the infant's chest and abdomen.

4.1.1 DATASET

I set up the research equipment and recorded the data at patient bedside. The following time-stamped data streams and corresponding alarms were collected simultaneously from each infant at a frequency of one reading every two seconds (0.5 Hz): PR and SpO₂ from Masimo SET SmartPod Model # MS16356 (Masimo Corp., Irvine, CA, USA) integrated with an Infinity Delta monitor; PR and SpO₂ from Nellcor OxiMax SmartPod Model # MS23997 (Covidien-Nellcor, Boulder, CO, USA) integrated with another Infinity Delta monitor; ECG derived Heart Rate (HR) and Respiratory Rate (RR) from one of the Infinity Delta monitors. The HR, sourced by independent electrical sensors, is used as the gold standard for evaluating the performance of both pulse oximeters that use optical sensors. Masimo and Nellcor SmartPods do not display any Signal Quality Indicators (SQI). The raw output includes only rudimentary artifact information in the form of symbols embedded within the data stream itself. The

exhaustive list of indicators $\{\text{NaN}, \text{'^'}, \text{'5'}\}$ is leveraged in the four validation experiments below. Significant effort was required to reverse engineer the structure of the physiological data streams logged at the output of the OEM monitors. In particular, extended correspondence and interaction with the Dräger global sales and engineering teams were required to decipher the information encoded in the output data streams.

The averaging mode of both SmartPods was set to *Normal*. This setting is less sensitive to artifact and is therefore slower to alarm. Alarm thresholds on the monitors were set according to clinical guidelines in this NICU as follows: Lower and upper alarm limits for PR and HR were 100 and 200 beats per minute (bpm) respectively. Similar to past recommendations [240], lower SpO₂ limits ranged between 85 and 88%; and upper SpO₂ limits ranged between 94 and 100%. Limits set for each patient depended on the clinical context including age, weight and medical condition. Recordings on each patient lasted approximately four hours. RS232 serial ports on both Infinity Delta monitors were connected through Digi International Edgeport 4 (Digi International, Minnetonka, MN) hardware to a USB port on a computer. Eltima Port Monitor Professional Edition Software v4.x (Eltima Software, Frankfurt, Germany) was installed on that computer for reading and logging data transmitted by each monitor in real-time.

4.1.2 DATA ANNOTATION

During data collection, I annotated clinically significant events, perceptible artifacts and monitor alarms in real-time in consultation with the infant's bedside nurse. Annotations include oxygen desaturation, clinical intervention, patient motion, routine care, sleep stages, feeding, sensor off, and any other events that triggered an alarm. I asked the bedside nurse to assess the patient and report the clinical validity of each alarm whenever feasible. The neonatologist and I then reviewed the data

retrospectively in conjunction with all real-time annotations to categorize SpO₂ alarms generated by each pulse oximeter. For the purpose of this study, a real clinical event implies a physiologic change in the patient's condition as determined in real-time by the bedside nurse or retrospectively by the neonatologist. This includes both true desaturation (hypoxia) and violation of the upper saturation limit (hyperoxia). All alarms or clinical events occurring within a 1-minute period were counted only once as they were assumed to be related to the same physiological event. As a corollary, alarms or clinical events occurring consecutively within one minute intervals were pooled together as one long event lasting up to five minutes for example. Previous studies have considered longer event durations such as 5 minutes in the Pediatric Intensive Care Unit [183] and 2 minutes in the Intensive Care Unit [74]. There is no consensus upon an optimal definition for event duration to discriminate between linked and independent events [241]. Clinically significant events of hypoxia or hyperoxia in NICU patients have been observed to last anywhere between 4-76 sec [242, 243]. Hence, the selection of a one-minute interval was considered appropriate for this study.

4.1.3 ALARM CATEGORIZATION

This section describes how each SpO₂ alarm generated by the Masimo and Nellcor pulse oximeters was categorized. An alarm associated with a real clinical event was classified as a *true positive* (TP); silence during a real clinical event was classified as a *false negative* (FN); an alarm without a clinically significant event associated with it was termed as *false positive* (FP). FP alarms include not only those induced by motion artifact and other nonphysiological conditions; they also include true, but brief, breaches of alarm limits without clinical significance [73]. These clinically irrelevant alarms have previously been described as *clinically false* [63] or *nuisance* alarms that add to alarm fatigue [88, 244]. The ECG-derived HR and its accompanying alarms

were taken as gold standard measurements. If there was an alarm associated with a low HR, then an accompanying oxygen desaturation would be considered a true event. During data collection, if the nurse determined that there was a clinical event, but the OEM monitor did not alarm, then this was considered to be a missed alarm (FN). One benefit of having two independent pulse oximeters is that such events were often missed by one OEM monitor and captured by the other due to the difference in their preprocessing algorithms, thereby facilitating the detection of missed alarms. The neonatologist visually compared the difference in the HR and PR values when determining the validity of an alarm retrospectively. A difference ≤ 12 bpm was considered plausible. A larger difference cast doubt on the signal quality of the pulse oximeter.

4.2 AD FRAMEWORK VALIDATION

This section develops four experimental methods for framework validation. The component analysis pipeline is described in detail for Experiment 1, and then more briefly for the remaining three experiments. Since the focus is to validate the use of the AD components integrated in the framework, all experiments comprise the same PDA and CED components. Therefore, the PDA and CED components are described in separate subsections below.

In each experiment, AD components from the AD catalogue developed in Sec. 3.3 are selected for step-wise integration in the framework. This demonstrates the mixing and matching feature of the framework where independently developed AD and CED components can be combined in series and/or parallel to discover an optimal composition for a given clinical use case. Since the specific configurations of the AD components vary between experiments, the configured parameters for each experiment are included in Appendix B: AD Component Interfaces. The physiologic dataset acquired at CHEO is streamed through the integrated framework in each

experiment. This simulates real-time integration of the AD framework in clinical workflows. MATLAB® R2013a was used for all simulations.

4.2.1 PERFORMANCE EVALUATION

In each experiment, the performance of the OEM alarm generation algorithm and the four composition pipelines were evaluated by labelling each SpO₂ alarm as one of TP, FP, or FN. Results are computed for Masimo data pooled across all patients. The two performance measures determined for each pulse oximeter are:

- i. Sensitivity (Sn): ratio of true alarms and total number of true events = $TP/(TP+FN)$
- ii. False Alarm Rate (FAR): ratio of false positive alarms and all alarms = $FP/(TP+FP)$

The principal aim of each pipeline is to reduce the number of false SpO₂ alarms that are generated by the CDSS while maintaining high sensitivity. By selecting performance criteria, the framework facilitates selection of optimal CDSS compositions, which can be then hard coded for the clinical use in the NICU. The benefits of the CRM are illustrated in the successful composition of complex analysis pipelines using diverse components not otherwise explicitly designed for interoperability.

4.2.2 PDA COMPONENT FOR FRAMEWORK VALIDATION

The PDA component for the experiments comprises the Digi International Edgeport 4 hardware and Eltima Port Monitor Professional Edition Software for data logging. Custom software was written to conform the OEM-generated data streams to the CRM. Specifically, the *Data.Type* (SpO₂, HR, PR, alarm status) and corresponding *Data.Values* were extracted from each interleaved OEM data packet. Each packet was produced by the monitor every two seconds. The low-level code of the PDA component

interpolated and synchronized the data streams at one sample every second. Since the OEM monitors fail to provide an explicit SQI stream for any of the data types, a default SQI stream with a *SQType*='binary' and *SQValue*=1 was generated by the PDA component for each data type.

4.2.3 CED COMPONENT FOR SPO₂ ALARM GENERATION

I develop a novel CED component that generates SpO₂ alarms. To make this contribution, I devised the following approach: (a) developing comprehension of the clinical context, (b) capturing expert clinical knowledge, and (c) translating clinical knowledge into computational algorithms. This approach resulted in translating clinical rules into low-level logic to create the CED component with a requirements interface that conforms to the CRM.

As found in the literature review in Chapter 2, alarm generation studies in [86-89] suggest (i) reducing alarm thresholds, and/or (ii) introducing annunciation delays, i.e., a delay between when an alarm threshold is crossed and when the alert is sounded or displayed. These studies test annunciation delays anywhere from 5s to 120s for a variety of physiologic data types. However, none of these studies quantify the trade-off between Sn and FAR resulting from with their suggested alarm generation algorithms. I develop a CED component that incorporates both suggestions, (i) and (ii), as follows:

- (i) A reduction of the lower SpO₂ alarm threshold and an increase of the upper SpO₂ alarm threshold. Both limits were adjusted by 3% which corresponds to the OEM-specified margin of error in the accuracy of the sensor reading. This implies that within the CED low-level component code:
 - a) The low alarm threshold, $Thresh_{Lo}$, is breached if the SpO₂ value is lower than the OEM monitor's alarm threshold by at least 3%.

- b) The upper alarm threshold, $Thresh_{HI}$, is breached if the SpO_2 value is higher than the OEM monitor's alarm threshold by at least 3%.
- (ii) The introduction of annunciation delays that are tunable between 5 to 60 seconds.

Figure 4-1 shows the flowchart of the low-level source code of the CED component. As shown, the CED continuously compares the SpO_2 value with the lower and upper limits, $Thresh_{LO}$ or $Thresh_{HI}$ respectively. A history of threshold breaches are stored in circular buffers, $error_{LO}$ and $error_{HI}$ respectively. These breaches are summed over a sliding window such that total error is a function of both the magnitude and duration of the threshold breaches. The integrated error is continuously compared to tunable lower and upper decision thresholds DT_{LO} and DT_{HI} . These decision thresholds are set proportional to the CED_{DT} value, which is set at the configuration interface of the CED component. Specifically, DT_{LO} is set equal to CED_{DT} and DT_{HI} is set to twice CED_{DT} since high SpO_2 alarms are not clinically deemed to be as dangerous as low SpO_2 alarms. Therefore, the CED waits twice as long to generate a high SpO_2 alarms as compared to a low SpO_2 alarm. If the integrated error exceeds either DT_{LO} or DT_{HI} , then the CED generates an alarm at the output.

The resulting CED component is analogous to the integrative aspects of a Proportional-Integral-Difference (PID) controller [245], in that the sum of errors over time, here between measured SpO_2 and set thresholds, forms the basis for the control signal, which in this case is the alarm generation. Extending this analogy, a PID controller is a broadly useful controller since it relies only on the measured process variable, and not on knowledge of the underlying process [246]. This makes it suitable for component design in this framework, where the CED component relies only on the

measured process variable, and not on knowledge of the underlying physiological process.

Figure 4-1: Low-level component code for clinical event detection and generation of SpO₂

4.2.4 EXPERIMENT 1

Experiment 1 constitutes the simplest of the four framework compositions considered in this study. Figure 4-2 shows the flowchart for Experiment 1. This flowchart has three functional horizontal swim lanes, each describing the PDA, AD, and CED stages of the pipeline. As described in Section 3.2.3.1 and depicted in Figure 3-4, each data stream is represented by a tuple with both data and SQI information. The fourth swim lane in the flowchart illustrates the colour-coded legend for the tuple. The AD framework in this experiment inputs data from the SpO_2 sensor only, consuming two data types, SpO_2 and SpO_2 alarm status (SpO_2Alarm). The low-level logic inside the PDA component, as described in section 4.2.2, maps the incoming values to its respective data type (SpO_2 , SpO_2Alarm) and assigns a default $SQValue$ of 1 to each datum of each *Data.Type* since an SQI value is not provided by the OEM monitor in this case.

The PDA then provides SpO_2 and its associated SQI stream to one instantiation of AD_{DIL} , and provides SpO_2Alarm and its SQI stream to the second instantiation of AD_{DIL} . As described in Sec. 3.3.1, the low-level code of the AD_{DIL} component de-interlaces OEM-specified artifact values. Here the user-set configuration interface includes $ArtSyms = \{NaN, \wedge\wedge, 5\}$, $SQI_{noMatch}=1$, and $SQI_{Match}=0$. The 'NaN' string implies missing data and the ' $\wedge\wedge$ ' symbol represents OEM-specified artifact values in the SpO_2 stream, whereas '5' is interlaced within the SpO_2Alarm to imply that the SpO_2 sensor is off. Hence, the use of the two AD_{DIL} components would provide the original data streams of *Data.Type* SpO_2 and SpO_2Alarm , along with their respective SQI streams with $SQValue = 0$ wherever the *Data.Value* is equal to any one of the *ArtSyms*. These two data streams and their associated SQI streams are then input to the requirements interface of an $AD_{FuseSQI}$ component. As described in Sec. 3.3.4, the low-level code of the $AD_{FuseSQI}$ component fuses two or more incoming SQI inputs to generate a single fused SQI . In this experiment, the operator is set to *min*, hence, it

provides an output *SQValue* that is the minimum of the two input *SQValue*. As shown in Figure 4-2, this output SQI stream is associated with the original SpO_2 stream, as that is required by the CED component described in Section 4.2.3.

4.2.5 EXPERIMENT 2

Experiment 2 builds on the composition used in Experiment 1 above. Here, an $AD_{MedFilt}$ component is added to pass SpO_2 through a median filter in order to reduce transient artifacts. Figure 4-3 shows the flowchart for Experiment 2. As described in Sec. 3.3.6, the low-level code of the $AD_{MedFilt}$ component would produce a median filtered SpO_{2Med} data stream and its associated SQI stream, which are then passed to the requirements interface of the CED component.

4.2.6 EXPERIMENT 3

Experiment 3 leverages data fusion to derive an estimate of signal quality for SpO_2 . Here, an AD_{Diff} component computes the difference between *PR* and *HR*. The difference between these two estimates of the same physiological parameter is a proxy for overall signal quality for both the SpO_2 . Figure 4-4 illustrates the flowchart for Experiment 3. As described in Sec. 3.3.5, the low-level code of the AD_{Diff} component computes the difference between the instantaneous *HR* and *PR* values. By comparing that difference to a threshold, it generates a 'binary' *SQType* which is then passed to the requirements interface of the CED component. The configuration interface is set as per the illustrative example in Sec. 3.3.5, however, a single threshold is used here to produce a 'binary' *SQType*. The SQI threshold (*SQthresh*) is varied in the range {6,12,18} to examine its effect and results are reported separately for each. These configurations are detailed in Appendix B: AD Component Interfaces.

4.2.7 EXPERIMENT 4

Experiment 4 builds on the composition used in Experiment 3 above. Here, an $AD_{MedFilt}$ component is added to pass SpO_2 through a median filter in order to reduce transient artifacts. Figure 4-5 shows the flowchart for Experiment 4. As described in Sec. 3.3.6, the low-level code of the $AD_{MedFilt}$ component would produce a median filtered SpO_{2Med} data stream and its associated SQI stream, which are then passed to the requirements interface of the CED component.

Figure 4-2: Experiment 1 Flowchart

Figure 4-3: Experiment 2 Flowchart

Figure 4-4: Experiment 3 Flowchart

Figure 4-5: Experiment 4 Flowchart

4.3 RESULTS

The data streams used in the research validation study include HR acquired from the Dräger ECG monitor, and SpO2 and PR acquired from the Masimo pulse oximeter. This section presents the characteristics of the dataset and the experimental results pooled across all eleven patients. Sections 4.3.2 to 4.3.5 present the results of all four experiments using three types of graphical organizations: (i) Tables of Sn and FAR values for the three tunable parameters, namely CEDDT, MedWW, and SQThresh when applicable; (ii) Linear plot of FAR in decreasing order, with a corresponding linear plot of Sn for all possible tunable parameters, and (iii) Radar plots of Sn vs. FAR for each possible combination of tunable parameters. Where applicable, table cells are highlighted using three color codes as follows: (1) light blue for Masimo OEM performance (Sn, FAR) = (85, 46); (2) green for the best achievable (Sn, FAR) where $Sn \geq 80\%$; and (3) orange for the best achievable (Sn, FAR) where $Sn \geq 75\%$. These Sn thresholds were chosen arbitrarily, where other threshold values may be chosen depending on clinical needs. As the configuration parameters of the AD and CED components are varied (tuned), the total number of alarms that are generated also varies. By reporting the performance metrics of Sn and FAR as percentages allows for comparison of results across experiments.

4.3.1 DATA CHARACTERISTICS

Data were collected from eleven preterm infants with diverse pathologies between February and June of 2010. A total of 79,200 samples of each data type were rendered for analysis. To synchronize data collected from the three OEM monitors, these samples were interpolated to obtain one sample every second, resulting in 158,400 samples of each data type. Patient demographics including weight and gestational age

distributions are listed in Table II. The ranges for OEM monitor settings used during data collection are provided Table III.

Table II: NICU Patient Demographics.

Reprinted with permission from the author's publication [239].

Demographics (<i>N</i> = 11)	Median (range)
Gender (M:F)	7:4
Birth weight (g)	1197 (515-2240)
Weight at recording (g)	2390 (677-3410)
Gestational age (weeks)	28 (25-35)
Corrected gestational age at recording (weeks)	36 (27-42)

Table III: NICU OEM Monitor Setting.

Reprinted with permission from the author's publication [239].

OEM Monitor Settings (<i>N</i> = 11)	Range or Value
Lower SpO ₂ alarm limit (%)	85-92
Upper SpO ₂ alarm limit (%)	95-100
Lower Heart Rate alarm limit (bpm)	100
Upper Heart Rate alarm limit (bpm)	200

4.3.2 EXPERIMENT 1 RESULTS

The AD composition pipeline in Experiment 1 consisted of two AD_{DIL} and one $AD_{FuseSQI}$ components. The AD_{DIL} components de-interlaced the OEM generated artifacts while the $AD_{FuseSQI}$ component combined the SQI streams from the two AD_{DIL} components. This process produced 145,309 clean samples of data from the original 158,400 samples.

Table IV shows results from Experiment 1 where no median filter is used. The left-most column contains all tunable values of the CED_{DT} parameter used in configuring the CED component as described earlier in Section 4.2.3. FAR values (%) are shown in the remaining column headers with their corresponding S_n values shown in the cells within the rows. The best achievable result $(S_n, FAR) = (80, 41)$ is obtained when $CED_{DT} = 12$, where $S_n \geq 80\%$. If S_n is only required to be $\geq 75\%$, then the best achievable performance becomes $(S_n, FAR) = (76, 40)$ when $CED_{DT} = 15$.

Figure 4-6 shows a linear plot of FAR in decreasing order, with a corresponding linear plot of S_n for all possible CED_{DT} values used in Experiment 1. Figure 4-7 is a Radar plot that shows S_n vs. FAR for all possible CED_{DT} values used in Experiment 1. Each pair around the perimeter represents (FAR, CED_{DT}) . FAR values decrease in a counter-clockwise direction. Figure 4-6 and Figure 4-7 present unique graphic views of the tabulated data from Table IV. The linear plots in Figure 4-6 visualize S_n and FAR separately, as functions of the decision threshold, whereas trajectory in the radar plot in Figure 4-7 permits one to explicitly visualize the trade-off between S_n and FAR. For example, if the performance requirement is to maintain $S_n \geq 80\%$, then a valid operational range of FAR is observed by moving counter-clockwise along the radar plot until its trajectory falls within the circular axis for $S_n = 80\%$.

Table IV: Experiment 1 results

	FAR (%)									
CED _{DT}	28	30	34	35	36	40	41	43	46	51
									OEM (85)	
5										85
8									83	
10								82		
12							80			
15						76				
20					73					
25			70							
30				66						
35		66								
60	60

Figure 4-7: Radar Plot for Experiment 1 shows Sn vs. FAR for various CED decision thresholds (CED_{DT}). Each pair around the perimeter represents (FAR, CED_{DT}). FAR values decrease in a counter-clockwise direction.

4.3.3 EXPERIMENT 2 RESULTS

Table VI show results of Experiment 2 for different Med_{WW} values. The first column lists all values used for tuning the median filter width, Med_{WW} , of the $AD_{MedFilt}$ component. The second column contains all tunable values of the CED_{DT} parameter used in configuring the CED component. FAR values (%) are shown in the remaining column headers with their corresponding Sn values shown in the cells within each row.

Table V: Experiment 2 results for Med_{WW} of 5 and 10

Med_{WW}	CED_{DT}	FAR (%)														
		23	27	28	29	30	32	35	36	39	40	41	42	43	46	51
															OEM (85)	
5	5															85
	8														83	
	10													82		
	12											80				
	15										76					
	20								73							
	25							69								
	30							66								
	35					66										
	60			60												
10	5													80		
	8												80			
	10									80						
	12								78							
	15						74									
	20				72											
	25					66										
	30					65										
	35		64													
	60	59														

Table V shows that the best achievable result (Sn , FAR) = (80, 39) is obtained when $Med_{WW} = 10$ and $CED_{DT} = 10$, where $Sn \geq 80\%$. If Sn is only required to be $\geq 75\%$, then the best achievable performance becomes (Sn , FAR) = (75, 32) when $Med_{WW} = 15$ and $CED_{DT} = 12$, as shown in Table VI. It is clear from these two tables

that $S_n \geq 80\%$ is achievable only when Med_{ww} is 10 or less. Figures Figure 4-8 Figure 4-9 show the linear and radar plots of S_n and FAR for this experiment respectively.

Table VI: Experiment 2 results for Med_{ww} of 12, 15 and 20

Med_{ww}	CED_{DT}	FAR (%)															
		20	21	23	24	25	26	27	29	30	31	32	34	35	36	39	40
12	5																76
	8																76
	10															75	
	12												75				
	15										72						
	20									67							
	25									62							
	30									61							
	35						61										
	60				54												
15	5													75			
	8														75		
	10													75			
	12											75					
	15									71							
	20							64									
	25									59							
	30									58							
	35					59											
	60			55													
20	5											65					
	8											64					
	10											64					
	12										64						
	15								62								
	20										58						
	25										57						
	30						54										
	35	55															
	60		49

Figure 4-9: Radar Plot for Experiment 2 shows Sn vs. FAR for various median filter widths (Med_{WW}) and CED decision thresholds (CED_{DT}). Each 3-tuple around the perimeter represents (FAR, Med_{WW}, CED_{DT}).

4.3.4 EXPERIMENT 3 RESULTS

Table VII shows results from Experiment 3 where AD_{Diff} component is configured with $SQ_{thresh} = 6$. AD_{Diff} component configuration was described earlier in Section 3.3.5. No median filter is used in this experiment. When $Sn \geq 80\%$, the best achievable result for $(Sn, FAR) = (86, 52)$ with $CED_{DT} = 5$. In this table, the orange cell highlight where $Sn \geq 75\%$ overlaps with the green cells since the best possible $(Sn, FAR) = (86, 52)$ when $CED_{DT} = 5$. It is clear from this table that the CDSS performance is worse at any other CED_{DT} threshold. Figure 4-11 and Figure 4-13 show the linear and radar plots of Sn and FAR for this experiment respectively.

Table VII: Experiment 3 results where AD_{Diff} component is configured with $SQ_{thresh} = 6$

$SQ_{thresh} = 6$	FAR (%)						
CED_{DT}	40	44	45	46	52	53	54
				OEM (85)			
5					86		
8						83	
10							81
12						80	
15						78	
20						77	
25						76	
30							74
35						72	
60							70

Table VIII shows results from Experiment 3 where AD_{Diff} component is configured with $SQ_{thresh} = 12$. With $Sn \geq 80\%$, the best achievable result of $(Sn, FAR) = (82, 50)$ is obtained when $CED_{DT} = 12$. In this table, the orange cell highlight where $Sn \geq 75\%$ overlaps with the green cells since the best possible $(Sn, FAR) = (82, 50)$ when $CED_{DT} = 12$. It is clear from this table that the CDSS performance is worse at any other CED_{DT} threshold. Figure 4-10 and Figure 4-14 show the linear and radar plots of Sn and FAR for this experiment respectively.

Table IX shows results where AD_{Diff} component is configured with $SQ_{thresh}=18$. With $Sn \geq 80\%$, the best achievable result for $(Sn, FAR) = (80, 50)$ with $CED_{DT} = 12$. With $Sn \geq 75\%$, the best achievable result for $(Sn, FAR) = (78, 47)$ with $CED_{DT} = 20$. Figure 4-12 and Figure 4-15 show the linear and radar plots of Sn and FAR for this experiment respectively.

Table VIII: Experiment 3 results where AD_{Diff} component is configured with $SQ_{thresh} = 12$

SQthresh = 12		FAR (%)										
CED_{DT}		40	41	42	45	46	47	48	49	50	51	52
						OEM (85)						
5												85
8												83
10											83	
12										82		
15										80		
20										78		
25										76		
30											74	
35										74		
60										71		

Table IX: Experiment 3 results where AD_{Diff} component is configured with $SQ_{thresh} = 18$

SQthresh = 18		FAR (%)											
CED_{DT}		40	41	42	44	45	46	47	48	49	50	51	52
							OEM(85)						
5													85
8												82	
10												81	
12											80		
15											77		
20								78					
25								75					
30										73			
35								72					
60									68

Figure 4-15: Radar Plot for Experiment 3 shows Sn vs. FAR for various CED decision thresholds (CED_{DT}) where SQThresh is 12. Each pair around the perimeter represents (FAR, CED_{DT}).

Figure 4-14: Radar Plot for Experiment 3 shows Sn vs. FAR for various CED decision thresholds (CED_{DT}) where SQThresh is 18. Each pair around the perimeter represents (FAR, CED_{DT}).

4.3.5 EXPERIMENT 4 RESULTS

Table X shows results from Experiment 4 where AD_{Diff} component is configured with $SQ_{thresh} = 6$. The first column lists all values used for tuning the median filter width, Med_{WW} , of the $AD_{MedFilt}$ component. When $Sn \geq 80\%$, the best achievable result for $(Sn, FAR) = (84, 52)$ when $Med_{WW} = 5$ and $CED_{DT} = 5$. When $Sn \geq 75\%$, the best achievable $(Sn, FAR) = (75, 40)$ when $Med_{WW} = 10$ and $CED_{DT} = 12$. Med_{WW} greater than 10 resulted in lower (Sn, FAR) values where $Sn < 75$, hence, those results are omitted from the table. Figure 4-17 and Figure 4-19 show the linear and radar plots of Sn and FAR for this experiment respectively.

Table X: Experiment 4 results where AD_{Diff} component is configured with $SQ_{thresh} = 6$

SQThresh = 6		FAR (%)								
Med _{WW}	CED _{DT}	38	39	40	41	42	46	52	53	54
							OEM(85)			
5	5							84		
	8								80	
	10								79	
	12								78	
	15								77	
	20								76	
	25								74	
	30									73
	35								73	
	60									70
10	5					77				
	8					77				
	10				75					
	12			75						
	15			74						
	20		73							
	25			72						
	30				70					
	35			69						
	60				67					

Table XI shows results from Experiment 4 where AD_{Diff} component is configured with $SQ_{thresh} = 12$. When $Sn \geq 80\%$, the best achievable result for $(Sn, FAR) = (82, 49)$ with $Med_{WW} = 5$ and $CED_{DT} = 12$. When $Sn \geq 75\%$, the best achievable $(Sn, FAR) = (75, 37)$ is obtained when $Med_{WW} = 12$ and $CED_{DT} = 12$. Med_{WW} greater than 12 resulted in lower (Sn, FAR) values where $Sn < 75$, hence, those results are omitted from the table. Figure 4-16 and Figure 4-20 show the linear and radar plots of Sn and FAR for this experiment respectively.

Table XI: Experiment 4 results where AD_{Diff} component is configured with $SQ_{thresh} = 12$

SQThresh = 12		FAR (%)															
Med _{WW}	CED _{DT}	35	36	37	38	39	40	41	42	45	46	47	48	49	50	51	52
											OEM(85)						
5	5																85
	8															83	
	10														83		
	12													82			
	15													80			
	20														78		
	25														76		
	30															74	
	35														74		
	60														71		
10	5													74			
	8												74				
	10											74					
	12										74						
	15									74							
	20								73								
	25								72								
	30							72									
	35							72									
	60							69									
12	5							75									
	8							75									
	10					75											
	12			75													
	15	73															
	20	71															

Table XII shows results from Experiment 4 where AD_{Diff} component is configured with $SQ_{thresh} = 18$. When $Sn \geq 80\%$, the best achievable result (Sn, FAR) = (80, 44) is obtained when $Med_{WW} = 10$ and $CED_{DT} = 5$. When $Sn \geq 75\%$, the best achievable (Sn, FAR) = (76, 36) is obtained when $Med_{WW} = 10$ and $CED_{DT} = 15$. Med_{WW} greater than 12 resulted in lower (Sn, FAR) values where $Sn < 75$, hence, those results are omitted from the table. Figure 4-18 and Figure 4-21 show the linear and radar plots of Sn and FAR for this experiment respectively.

Table XII: Experiment 4 results where AD_{Diff} component is configured with $SQ_{thresh} = 18$

SQThresh = 18		FAR (%)															
Med _{WW}	CED _{DT}	32	33	34	36	39	40	41	42	43	44	46	47	49	50	51	52
												OEM (85)					
5	5																85
	8															82	
	10														81		
	12														80		
	15													77			
	20											78					
	25												75				
	30													73			
	35												72				
	60													68			
10	5										80						
	8									79							
	10								79								
	12					77											
	15				76												
	20		74														
	25		73														
	30			72													
	35			72													
	60	68															
12	5							76									
	8						75										
	10						73										
	12				75												
	15			72													
	20		69

Figure 4-17: Experiment 4 results of Sn and FAR plotted as percentages for each combination of tunable parameters (CED_{DT}, Med_{WW}) with SQThresh = 6

Figure 4-16: Experiment 4 results of Sn and FAR plotted as percentages for each combination of tunable parameters (CED_{DT}, Med_{WW}) with SQThresh = 12

Figure 4-18: Experiment 4 results of Sn and FAR plotted as percentages for each combination of tunable parameters (CED_{DT}, Med_{WW}) with SQThresh = 18

Figure 4-19: Radar Plot for Experiment 4 shows Sn vs. FAR for various median filter widths (Med_{WW}) and CED decision thresholds (CED_{DT}) with SQThresh = 6. Each 3-tuple around the perimeter represents (FAR, Med_{WW}, CED_{DT}).

Figure 4-20: Radar Plot for Experiment 4 shows Sn vs. FAR for various median filter widths (Med_{WW}) and CED decision thresholds (CED_{DT}) with SQThresh = 12. Each 3-tuple around the perimeter represents (FAR, Med_{WW}, CED_{DT}).

Figure 4-21: Radar Plot for Experiment 4 shows Sn vs. FAR for various median filter widths (Med_{ww}) and CED decision thresholds (CED_{DT}) with SQThresh = 18. Each 3-tuple around the perimeter represents (FAR, Med_{ww}, CED_{DT}).

4.4 DISCUSSION

The dataset used in these experiments was well suited to optimizing AD component composition pipelines. There were a total of 119 alarms across the eleven patients in this study. It is rare to have a dataset with such detailed annotations of artifacts, clinical events, clinical interventions, and observations. For example, a large volume of physiological data is available through resources such as PhysioNet [195], however, those data typically lack such high quality annotations. The patients in this study represent a neonatal population with varying disease severity, weight, and gestational age. Although, such a wide range of patients provides for the development of widely applicable rules, however, as discussed above, many of the decision thresholds are required to be patient-centric. For example, patient 6 was far more ill than the other 10 patients, with 32% of associated clinical events. Other limitations of the dataset include possible ambiguity in categorizing alarms as true versus false. Especially in cases where the SpO₂ reading hovers around the OEM monitor's alarms threshold setting. In the present study, such indeterminate alarms were categorized as false alarms. The study sample size was limited due to hospital logistics and resources. In future, a larger sample size could facilitate sub-group analyses with division based on clinical characteristics, weight, and gestational and chronological ages of infants.

The four experiments in this study facilitated step-wise integration and cascading of various AD components developed in Section 3.3. Each experiment comprised of a different composition pipeline of AD components. The performance of each pipeline was evaluated based on two performance metrics, Sn and FAR. The performance of each pipeline was also evaluated in terms of the percentage change in Sn and FAR as compared to the OEM monitor's Sn and FAR respectively, according to the following definitions. Negative values of percentage change indicate reduction and positive values indicate increment.

(i) Percentage change in Sn as compared to the OEM Sn (%) =

$$(Sn - OEM_Sn)/OEM_Sn \times 100 = (Sn - 85)/85 \times 100$$

(ii) Percentage change in FAR as compared to the OEM FAR (%) =

$$(FAR - OEM_FAR)/OEM_FAR \times 100 = (FAR - 46)/46 \times 100$$

Table XIII presents the best achievable (Sn (percentage change in Sn), FAR (percentage change in FAR)) in all four experiments, under the criterion that Sn $\geq$ 80%. Hence, it summarizes the results from the green cells in Table VI to Table XII. The configured values for Med_{ww}, CED_{DT}, and SQThresh are also stated for each experiment. This allowed us to determine which AD composition pipeline was best suited to the CED of SpO₂ alarms.

Table XIV presents the best achievable (Sn (percentage change in Sn), FAR (percentage change in FAR)) in all four experiments, under the criterion that Sn $\geq$ 75%. Hence, it summarizes the results from the orange cells in Table IV to Table XII. From Table XIV, we conclude that Experiment 2 results in the best achievable performance of Sn = 75% and FAR = 32%. Although a considerable reduction in Sn is observed (11.7%), this parameter combination results in a significant reduction in FAR (30.4%). Therefore, the optimal parameter combination is considered to be Med_{ww} = 15 and CED_{DT} = 12.

From Table XIII and Table XIV, we conclude that Experiment 2 results in the best possible performance of Sn = 80% and FAR = 39%, representing percentage reductions of 5.8% and 15.2% for Sn and FAR respectively. This is considered optimal since we observe the greatest reduction in FAR while maintaining Sn $\geq$ 80%. Therefore, the optimal parameter combination for this condition was Med_{ww} = 10 and CED_{DT} = 10.

Table XIII: The best possible (Sn (% change in Sn), FAR (% change in FAR)) achieved in all four experiments, where Sn $\geq$ 80%

Sn > 80%		AD _{Diff}	
		Yes	No
AD _{MedFilt}	Yes	<p>Experiment 4</p> <p>Med_{WW} 5, CED_{DT} 5, SQThresh 6: (84 (-1.1%), 52 (13%))</p> <p>Med_{WW} 5, CED_{DT} 12, SQThresh 12: (82 (-3.5%), 49 (6.5%))</p> <p>Med_{WW} 10, CED_{DT} 5, SQThresh 18: (80 (-5.8%), 44 (-4.3%))</p>	<p>Experiment 2</p> <p>Med_{WW} 10, CED_{DT} 10: (80 (-5.8%), 39 (-15.2%))</p>
	No	<p>Experiment 3</p> <p>CED_{DT} 5, SQThresh 6: (86 (1.1%), 52 (13%))</p> <p>CED_{DT} 12, SQThresh 12: (82 (-3.5%), 50 (8.6%))</p> <p>CED_{DT} 12, SQThresh 18: (80 (-5.8%), 50 (8.6%))</p>	<p>Experiment 1</p> <p>CED_{DT} 12: (80 (-5.8%), 41 (-10.8%))</p>

Table XIV: The best possible (Sn (% change in Sn), FAR (% change in FAR)) achieved in all four experiments, where Sn $\geq$ 75%

Sn > 75%		AD _{Diff}	
		Yes	No
AD _{MedFilt}	Yes	<p>Experiment 4</p> <p>Med_{WW} 10, CED_{DT} 12, SQThresh 6: (75 (-11.7%), 40 (-15%))</p> <p>Med_{WW} 12, CED_{DT} 12, SQThresh 12: (75 (-11.7%), 37 (-19.5%))</p> <p>Med_{WW} 10, CED_{DT} 15, SQThresh 18: (76 (-10.5%), 36 (-21.7%))</p>	<p>Experiment 2</p> <p>Med_{WW} 15, CED_{DT} 12: (75 (-11.7%), 32 (-30.4%))</p>
	No	<p>Experiment 3</p> <p>CED_{DT} 5, SQThresh 6: (86 (1.1%), 52(13%))</p> <p>CED_{DT} 12, SQThresh 12: (82 (-3.5%), 50 (8.6%))</p> <p>CED_{DT} 20, SQThresh 18: (78 (-8.2%), 47 (21.7%))</p>	<p>Experiment 1</p> <p>CED_{DT} 15: (76 (-10.5%), 40 (-15%))</p>

4.5 CONCLUSIONS

Experiment 1 provided a baseline performance. Experiment 2 evaluated the use of an $AD_{MedFilt}$ component which implemented median filtering as a means to remediate artifacts in the SpO_2 data stream. Experiment 3 exploited data fusion to combine redundant estimators of heart rate by combining ECG-derived HR with PPG-derived PR using an AD_{Diff} component. It was hypothesised that AD detection and suppression could be achieved by detecting significant divergence in these two estimates. Experiment 4 combines the AD components used in both Experiments 2 and 3.

The results of Experiment 1 illustrated the effect of varying the CED decision threshold (CED_{DT}) on the performance of the CDSS. By adjusting this threshold, the system could be made to be more conservative or permissive leading to an explicit trade-off between Sn and FAR. This CED_{DT} is patient-centered and may be adjusted depending on the severity of disease and clinical resources available (e.g. the nurse-to-patient ratio may differ in the NICU vs. the general ward).

Comparing results from Experiments 1 and 2 indicated that use of $AD_{MedFilt}$ significantly improved both the Sn and FAR of the CDSS. As expected, increasing the median filter width reduced both Sn and FAR, since the median filter smoothed out transient SpO_2 values. A range of median filter widths were evaluated in combination with a range of CED_{DT} values seeking the combination that provided that greatest decrease in FAR while maintaining a $Sn \geq 80\%$ or 75% . These latter Sn thresholds were somewhat arbitrary, but reflect the need to detect the majority of true clinical events. As mentioned above, these thresholds are also expected to be patient-centered.

Experiments 3 and 4 leveraged data fusion, via an AD_{Diff} component, to identify periods of low signal quality. Clifford *et al.* [65] have recommended that an SQI be generated for each datum when a known error rate is available for calibration.

Following this, we hypothesized that by computing an error rate from the combined information from two different sensor modalities, PR from SpO₂ and HR from ECG, an SQI signal could be generated and increased performance would be achievable. Results over three different AD_{Diff} threshold values failed to demonstrate improved performance. In fact, the frequency of all three types of alarms, true, missed, and false, increased with the use of AD_{Diff}. Close inspection of the generated alarms revealed fragmentation of previously contiguous alarms into more alarms of shorter duration. This was due to instantaneous masking of individual SpO₂ values due to transient disparities between HR and PR, not necessarily associated with prolonged periods of low signal quality.

Through future work, the CED algorithm may be modified to process the SQI in a variety of ways that may lead to improved performance. Suggestions for future exploration include retaining the previous alarm state during periods of low signal quality, or appraising cumulative SQI values instead of instantaneous ones.

5 CONCLUSIONS

This thesis presented a novel framework to standardize interoperability of AD, PD and CED within CDSS. The framework addresses six different shortcomings identified by the literature review. It provides a unique test bed with the ability to create new AD composition pipelines by mixing and matching independently developed or de-coupled AD components. The framework is validated in a clinical research study.

5.1 RESEARCH SIGNIFICANCE

Integrity and quality of data are crucial to the success of any analytic system. Therefore, it is important to design and implement CDSS that assess the quality of data and issue relevant alarms with high specificity and low false alarm rates. Additionally, each PD and CED algorithm within one CDSS can have different needs in terms of type, frequency, and quality of data [17]. Therefore, no single AD algorithm is necessarily best suited to serve each PD and CED algorithm within a CDSS. This identifies the need to develop and evaluate a framework that supports the interoperability of a range of configurable AD algorithms.

5.2 RESEARCH CONTRIBUTIONS

This section discusses how the experiments conducted in the previous chapter validate each of the six framework features (*f.1*) - (*f.6*) which were originally defined in Section 1.5. Each feature is restated below, followed by a discussion of how the experiments validate each feature:

(f.1) Flexibility to serve the differing needs of patient populations from different types of CCUs through generalization and customizability. Although the data in these experiments all arose from the NICU, the inherent flexibility of the framework can accommodate varying types of data streams captured in other types of CCU. Similarly,

the component-based nature of this framework allows for the creation of CED components relevant to different types of CCU domains and their integration with the most appropriate available AD components. As a result, these experiments and analysis pipelines are not restricted in application to the NICU. This could be demonstrated using future experiments based on data from other units, whether gathered specifically for this research or taken from repositories such as Physio Net [195].

(f.2) Reusability across multiple types of physiologic data harvested by different OEM monitors. Some of the AD components selected from the catalog are used to process different physiologic streams acquired by different OEM monitors in the various experiments in this chapter. For example, the AD_{DIL} component is used to process HR from the Dräger OEM monitor, and SpO₂ and PR from the Masimo OEM pulse oximeter.

(f.3) Standardized definitions of SQI that promote interoperability between independently developed components. These experiments employ multiple components to generate SQI. These components were developed based on current algorithms identified in the literature review. For example, the AD_{Diff} component is derived from the work of Yu *et al.* [171], and applied to the HR and PR streams in experiments 3 and 4, while the CED component leverages the ideas of delays alarm annunciation and threshold reduction that were introduced in [86-89]. These experiments demonstrate the integration of components that were developed independently and whose interoperability is facilitated through use of standardized SQI.

(f.4) Reusability and scalability by cascading, and mixing and matching several AD, PD and CED components in various combinations. Each of the four experiments represents a different composition of components. The analysis pipelines developed here vary in scale through reuse and cascading of components. This mixing and matching is made possible by each component's adherence to the CRM.

(f.5) Customizability to evaluate and compare the performance of multiple combinations of independently developed components on offline and potentially real-time patient data when integrated with clinical workflows. As discussed in Chapter 2, a number of studies have suggested the introduction of delays in alarm annunciation to reduce false alarm rates. Clearly, such a strategy would be expected to reduce the false alarm rate. However, there is a lack of quantitative evaluation in terms of the impact of such a strategy on Sn and PPV. The framework developed here promotes and enables such a quantitative study design as demonstrated by the experiments developed here. In fact, it was found that such strategies failed to suppress false alarms while maintaining sufficiently high Sn.

(f.6) Standardized component interfaces that can potentially support real-time clinical implementation of AD. The review of the literature found a paucity of CDSS having reached real-time clinical implementation. The framework is implicitly designed to run components in a real-time streaming environment. Consequently, the highly complex analysis pipelines in each of the experiments were evaluated using a simulated real-time environment. As a result, with negligible reformulation, the optimal framework composition resulting from this evaluation can be integrated within clinical workflows.

5.3 FUTURE WORK

This thesis is motivated by the need to integrate an AD framework in the Artemis CDSS. The AD framework is now ready for integration in the Artemis environment. Future work shall realize such integrated dynamic modelling of the AD framework. The two ongoing case studies originally identified present excellent opportunities for future validation of the AD framework for integration with Artemis. These are (i) CED for Late-Onset Neonatal Sepsis (LONS), and (ii) CED of Apnoea of Prematurity (AOP). These two studies are suitable for evaluation since they have different data requirements due to the different clinical rules used in their low-level logic.

Artemis is built using IBMTM middleware for data stream processing called InfoSphere Stream Computing System (Streams). The Streams computing platform comes equipped with a high-level declarative programming language called Streams Processing Language (SPL) [247]. SPL contains built-in stream processing operators, many of which are suitable for implementation of online healthcare analytics. SPL can be extended through user-defined operators [248]. That includes implementing the PDA, AD, PD and CED components source code in MATLAB®, Java, C++, or IBM Streams Processing Language (SPL) as exemplified in [39].

The dynamic modelling of the framework integrated in the two Artemis case studies shall include implementation of the AD and CED component interfaces defined by the CRM schemas. An example of component instantiation and interfacing using eXtensible Markup Language (XML) can be found in the authors' work in [132]. Preparation of a manuscript is in progress that describes how the AD framework shall be integrated with the Artemis CDSS in the AOP case study.

Additional future work should include the implementation of a wide range of AD, PD, and CED components to further validate the expressive power of the CRM. The CRM was developed following an extensive review of the literature summarizing the requirements, provisions and configurations for a large number of existing AD algorithms. However, it is expected that the CRM will continue to evolve as a wider range of algorithms are implemented as components within the framework. The true validation of the framework will be its adoption by independent research groups. The true benefit of this research to patients and health care practitioners will be realized through integration in clinical decision support systems.

APPENDIX A: REVIEW OF THEMES II – VI

Reprinted with permission from the author's publication [17].

Author	Theme II: Physiologic Data Source Data Monitor/ Inclusion (I)/ Exclusion (E) criteria/ Sample size	Theme III: Harvested Data Data Type/Acquisition(A)/ Sampling or Storage (S) Frequency	Theme V: Clinical Evaluation Data Annotation/ Mode/ Performance Metric
		Theme IV: Data Analysis Dimensionality/ Focus/ Signal Quality/ Clinical Contribution	Theme VI: Clinical Implementation
Bradley [25]	Phillips Intellivue MP70 (Philips Healthcare, Andover, Massachusetts)/ I: respiratory and/or cardiac failure, enrollment in the study within 36 hours of ICU admission, expected period on study greater than 72 hours/ E: chronic atrial fibrillation, transfer from another ICU/ 34 patients	ECG, EtCO ₂ / A: ECG at 500 Hz; S: 125 Hz	Retrospective/Offline/ Percentage data loss
		Univariate/Stream/ Developed SQI/ AD	CIMVA (TMS Inc., Ottawa, Canada)
Hu [115]	Monitor not specified, data storage system BedMasterEx™/ (i) I: code blue adult patients > 18 yr (ii) I: control patients with same: (1) All Patient Refined or Medicare Diagnosis Related Group; (2) age ± 5 years; (3) gender; (4) same CCU/ (ii) E: had code blue; experienced an unplanned ICU transfer/ (i) 223 patients (ii) 1768 patients	Alarms/Not applicable	Retrospective/ Offline & Simulated real-time / Se, FA
		Multivariate/Stream+ disease-centric/ No SQI / AD + CED	None
Li [23]	MIMIC II database/ I: asystole & ventricular tachycardia/ 104 patients, total 1055x6s	PPG/ A & S: 125 Hz	Retrospective/ Offline/ Acc, ROCC
		Univariate/ Stream + Patient-centric/ Developed SQI/ AD	None
Sun [116]	i - MIMIC II database, ii – BIOPAC PPG 100C module, TSD200 PPG reflective transducer, iii - BIOPAC OXY 100C module with TSD123A SpO ₂ finger transducer / i -87 seconds, ii & iii –3 healthy volunteers x 20 sec	PPG, ECG/ A & S : 125 Hz for MIMIC II datasets;Not specified for BIOPAC data	(i) Retrospective (ii) By real-time observer/ Offline/ RMSE
		Multivariate/ Stream- centric/ No SQI / AD	None
Scalzo [117]	I: intracranial hypertension/ 108 patients	Intracranial Pressure (ICP) & ECG waveforms, ICP alarms/ A: ICP at 240 Hz	Retrospective/ Offline/ AUROC, TPR, FPR
		Multivariate (d)/ Stream- centric/ No SQI/ AD	None
Borowski [118]	Infinity patient monitoring system, Dräger Medical, (Lubeck, Germany)/ 1245:52:28 hr	SBP, MAP, HR, SpO ₂ , alarms / S: 250 Hz	Retrospective/ Offline/ Se, FARR
		Univariate/ Stream-centric/ No SQI/ AD	None
Siebig [70]	Infinity patient monitoring system & full-disclosure data logging software eData, (Dräger Medical, Lübeck, Germany)/ 38 patients, total 515 hr	At least HR, IABP, SpO ₂ , alarms/ S: 1 Hz	Retrospective with video/ Offline/ Relevant and FA
		Univariate/ Stream-centric/ No SQI/ AD	None

Author	Theme II: Physiologic Data Source Data Monitor/ Inclusion (I)/ Exclusion (E) criteria/ Sample size	Theme III: Harvested Data Data Type/Acquisition(A)/ Sampling or Storage (S) Frequency	Theme V: Clinical Evaluation Data Annotation/ Mode/ Performance Metric
		Theme IV: Data Analysis Dimensionality/ Focus/ Signal Quality/ Clinical Contribution	Theme VI: Clinical Implementation
Schettlinger [60]	1 hr of SBP; 30 mins of HR	SBP, HR/ A & S: 1 Hz	Not specified/ Offline/ Not specified
		Univariate/ Stream-centric/ No SQI/ AD	None
Charbonnier [119]	Datex-Engstrom monitor with a multi- parameter module/ I: various disorders with specific clinical contexts such as mechanical ventilation and cessation of sedative drug administration/ 14 patients, total 50 hr	SpO ₂ ,SBP, DBP,MBP, HR, MAP, RR,EV, MV, maximal airway flow A: 100 Hz, S: 1 Hz	Retrospective/ Offline/ Se, Sp
		Multivariate/ Stream- centric/ No SQI / AD	None
Li [105, 106]	1MIMIC II database/ 6000 hr	ECG, ABP/ A: 500 Hz; S: 125 Hz	Retrospective/ Offline/ True and False Alarms, RMSE
		Multivariate/ Stream- centric/ Developed SQI / AD + PD	None
Otero [121]	Monitor not specified, data logging software SUTIL/ 78 patients, total 196 hr	HR, RR, BP, SpO ₂ / Not specified	Real-time/ Real- time,TRACE [249]/ CDR, FPR
		Multivariate/ Stream + disease -centric/ No SQI / AD + CED	None
Blum [122]	Solar 9000 monitors and Monitor Capture Server data logging software (GE Healthcare, UK)/ 28 days, total 293,049 alarms	SBP, MAP, CVP, Chest Impedance (CI) and their alarms/ Not applicable	Retrospective/ Offline/ Sp
		Multivariate/ Stream + patient -centric/ No SQI / AD	None
Aboukhalil [26]	1MIMIC II database/ I: critical ECG arrhythmia alarm in the presence of one channel of ECG and an ABP waveforms/ E: 49 patients with active intra-aortic balloon pumps / 5386 alarms from 447 patients, total 41,301 hr	ECG, ABP/ A: 500 Hz; S: 125 Hz	Retrospective /Offline / True and False Alarm Reduction Rates
		Multivariate/ Stream + disease -centric/ Specified SQI / AD + CED	None
Otero [123]	OEM not specified / 71 patients, total 175 hr	HR, RR, BP, SpO ₂ / S: 1 Hz	Not specified / Offline, TRACE [249]/ CDR, FPR, FNR
		Multivariate/ Stream- centric/ No SQI / AD	None

Author	Theme II: Physiologic Data Source Data Monitor/ Inclusion (I)/ Exclusion (E) criteria/ Sample size	Theme III: Harvested Data Data Type/Acquisition(A)/ Sampling or Storage (S) Frequency	Theme V: Clinical Evaluation Data Annotation/ Mode/ Performance Metric
		Theme IV: Data Analysis Dimensionality/ Focus/ Signal Quality/ Clinical Contribution	Theme VI: Clinical Implementation
Sieben [124]	Not specified	RR, SpO ₂ , arrhythmia indicator, HR, PR, premature ventricular contraction, SBP, DBP, MBP, temperature, thresholds, alarms/ Not specified	Real-time / Offline/ Se, FARR
		Multivariate/ Stream- centric/ Developed SQI but insufficient details given / AD	None
Charbonnier [120]	OEM Not specified / 18 patients, total 36 hr	HR, SBP, DBP, MBP, SAO ₂ , MAP / A: 100 Hz, SAO ₂ at 0.2 Hz; S: 1 Hz	Real-time/ Simulated data/ Number of False, True & Technical Alarms
		Univariate/ Stream-centric/ No SQI/ AD	None
Zong [27]	³ MIMIC database / I: at least multi-lead ECG, ABP / 46 patients, total 1890 hr	ECG at 500 Hz, ABP at 125 Hz	Retrospective / Offline/ TA, FA
		Multivariate/ Stream- centric/ Calculated SQI / AD	None
Jakob [125]	Datex-Ohmeda, (Helsinki, Finland)/ I: systemic and pulmonary artery catheters after coronary artery bypass grafting/ E: clinical signs of heart failure/ 41 patients	HR, systemic (SAP) & pulmonary (PAP) artery pressures, SpO ₂ , central venous pressure (CVP), peripheral & central temp/ A: 10-s median / S: 2-min median values	Retrospective/ Offline/ Se, Sp, PPV
		Univariate/ Stream-centric/ No SQI/ AD	None
Schoenberg [111]	OEM Not specified / 6 patients, total 337 hr	HR, SBP, DBP, SpO ₂ / Not specified	Real-time/ Real-time/ Se, PPV
		Multivariate/ Stream- centric/ Unknown output score in case of missing data / AD	iMDsoft Clinical Information Systems [188]
Zhang [126]	HP Viridia neonatal component monitoring system (Hewlett Packard)/ 11 patients, total 196 hr	HR, PR, RR, SBP, DBP, MBP, SAO ₂ , venous O ₂ saturation, O ₂ perfusion/ A: 1 Hz/ S: 1 Hz and derived 1-min averages	Real-time/ Real-time/ Se, Sp, PPV, Acc
		Multivariate/ Stream + patient -centric/ No SQI / AD	None
Tsien [127]	SpaceLabs monitor (SpaceLabs Medical, Redmond, WA, USA)/ I: monitored data containing all five signals of interest	HR, RR, SpO ₂ , Invasive SBP, DBP, MBP / A: 1 sample every 5 s	Real-time/ Offline/ AUROC
		Multivariate/ Stream- centric/ No SQI / AD	None

Author	Theme II: Physiologic Data Source Data Monitor/ Inclusion (I)/ Exclusion (E) criteria/ Sample size	Theme III: Harvested Data Data Type/Acquisition(A)/ Sampling or Storage (S) Frequency	Theme V: Clinical Evaluation Data Annotation/ Mode/ Performance Metric
		Theme IV: Data Analysis Dimensionality/ Focus/ Signal Quality/ Clinical Contribution	Theme VI: Clinical Implementation
Tsien [129]	SpaceLabs monitor (SpaceLabs Medical, Redmond, WA, USA)/ E:cardiac PICU patients/ 35118 minutes	HR, ECG lead number, RR, SpO ₂ , Invasive SBP, DBP, MBP/A & S: 1 sample every 5-6 s	Real-time/ Real-time/ TA, FA
		Not specified/ Stream- centric/ No SQI/ Annotated data for AD	None
Monasterio [130]	¹ MIMIC II database / I: SpO ₂ data / 1616 events annotated on 27 patients	Two ECG leads, Impedance Pneumogram (IP), PPG each at 125 Hz; HR, SpO ₂ at 1 Hz	Retrospective/ Offline/ Acc, Sp, Se, PPV, NPV
		Multivariate/ Stream- centric/ Specified SQI / AD	None
Hoshik [131]	GE Solar 8000M and I and Dash 3000 (GE Healthcare, Milwaukee, USA)/ 1100 patients	Three ECG leads at 240 Hz, Chest Impedance (CI) at 60 Hz, PPG,HR,RR, SPO ₂ each at 0.5 Hz; alarms, respiratory support, demographics, apnea-bradycardia documentation	Retrospective/ Offline/ FPR, FNR
		Multivariate/ Stream + disease-centric/ Discussed but not specified/ AD + CED	None
Nizami [132]	Not mentioned	ECG, PPG	Not evaluated
		Multivariate/Stream + disease-centric/SQI utilized/ AD + CED	Artemis [31]
Salatian [69]	OEM Not specified/ 1 data segment	BP at 1 Hz	Not discussed/ Offline/ Graphical display
		Univariate/ Stream-centric/ No SQI/ AD	None
Belal [133]	Hewlett Packard Merlin M1064a 1176a/ I: neonates < 2 months old/ 54 patients, total 2426 hr	HR, RR, SpO ₂ / S: 1 Hz	Partially retrospective & automated markup/ Offline/ Acc, Sp, Se, ROCC
		Multivariate/ Stream + disease-centric/ No SQI/ AD + CED	None
Blount [67]	OEM Not specified/ 4 patients	Nursing interventions causing artifacts	Real-time/ Offline/ Not specified
		Not specified/ Stream- centric/ No SQI/ AD	None
Quinn [134]	OEM Not specified/ I: 24-29 weeks gestation premature babies in their 1st week of life/ 15 patients x 24 hr	Core and Peripheral temp, DBP, SBP, HR, SpO ₂ , TpCO ₂ , TpO ₂ / A & S: 1 Hz	Retrospective/ Offline/ AUROC, EER
		Multivariate/ Stream + disease-centric/ Noticed but not discussed/ AD + CED	None

Author	Theme II: Physiologic Data Source Data Monitor/ Inclusion (I)/ Exclusion (E) criteria/ Sample size	Theme III: Harvested Data Data Type/Acquisition(A)/ Sampling or Storage (S) Frequency	Theme V: Clinical Evaluation Data Annotation/ Mode/ Performance Metric
		Theme IV: Data Analysis Dimensionality/ Focus/ Signal Quality/ Clinical Contribution	Theme VI: Clinical Implementation
Portet [135]	OEM Not specified/ I: preterm infants/ 13 patients x 24 hr	Core and Peripheral temp, DBP, SBP, HR, SpO ₂ , TpCO ₂ , TpO ₂ , humidity of incubator/ A & S: 1 Hz	Retrospective/ Offline / Se, Sp, Acc, k
		Univariate/ Stream + disease-centric/ No SQI/ AD + CED	BabyTalk project [184]
Walls- Esquivel [75]	OEM Not specified/ I: preterm infants < 30 wks gestation	EEG, ECG, EMG, RR	Real-time with video/ Real-time/ Not applicable
		Multivariate/ Stream- centric/ No SQI / AD	Not applicable
Moorman [136]	Not specified	ECG at S: 4 kHz	Real-time / Real-time/ AUROC, Positive Predictive Accuracy
		Multivariate (d)/ Stream + disease-centric/ No SQI/ AD + CED + PD	HeRO™ system [189]
Tsien [137]	OEM Not specified/ 274 hr	HR, IBP, TpCO ₂ , TpO ₂ / A & S: 1 Hz and 1min averages	Retrospective/ Offline/ AUROC
		Multivariate/ Stream- centric/ No SQI / AD	None
Tsien [138]	OEM Not specified/ E: less than four monitored signals / 200 hr	HR, BP, TpCO ₂ , TpO ₂ / S: 1 sample per min	Retrospective/ Offline/ AUROC
		Multivariate/ Stream- centric/ No SQI / AD	None
McIntosh [139]	Hewlett Packard 78344A (South Queensferry, U.K.)/ I: pneumothorax/ E: infants with birth asphyxia, persistent pulmonary hypertension, requiring inotropic support/ 42 patients	TpCO ₂ , TpO ₂ / A: 1 Hz & S: 1-min average	Retrospective/ Offline/ AUROC, PPV, NPV
		Multivariate/ Stream + disease –centric/ No SQI /AD + CED	None
Cao [140]	Hewlett Packard 78344 multichannel neonatal monitors/ I: preterm infants/ 10 patients x 10 hr, total 6000 values of PO ₂ , PCO ₂	TpCO ₂ , TpO ₂ / A: 1 Hz & S: 1- min average	Retrospective/ Offline/ Se, Sp
		Multivariate/ Stream- centric/ No SQI / AD	None
Karlen [141]	(i) Capnabase database using Nellcor pulse oximeter & S/5 Collect software (Datex-Ohmeda, Finland) (ii) Complex System Laboratory database/ (i) I: general anesthesia/ (i) 124 patients x 120 sec + 42 patients x 480 sec (ii) 2 patients	PPG/ (i) A: 100 Hz & S: 300 Hz	Retrospective/ Offline/ Se, PPV
		Univariate/ Stream-centric/ Developed SQI/ AD	None
Koeny [143]	OEM Not specified/ 17 patients	HR, NIBP/ A: 0.33 Hz; S: 0.0167 Hz	Real-time/ Offline/ Graphical display
		Univariate/ Stream-centric/ No SQI/ AD	None

Author	Theme II: Physiologic Data Source Data Monitor/ Inclusion (I)/ Exclusion (E) criteria/ Sample size	Theme III: Harvested Data Data Type/Acquisition(A)/ Sampling or Storage (S) Frequency	Theme V: Clinical Evaluation Data Annotation/ Mode/ Performance Metric
		Theme IV: Data Analysis Dimensionality/ Focus/ Signal Quality/ Clinical Contribution	Theme VI: Clinical Implementation
Karlen [142]	Capnabase database using S/5 Collect software (Datex-Ohmeda, Finland)/ I: general anesthesia/ 42 patients	ECG, PPG, RR/ A: ECG at 300 Hz, PPG at 100 Hz; RR at 25 Hz; S: 300 Hz	Retrospective/ Offline/ Error, Power, Robustness
		Multivariate/ Stream- centric/ Discussed but not specified/ AD	None
Schmid [74]	Kappa XLT monitor (Dräger), Zeus anesthesia workstation (Dräger), Nortis MedLink, (Lubeck, Germany), Erasmus MC eData TapeRec, (Rotterdam, The Netherlands)/ I: anesthesia for elective cardiac surgery (aortocoronary bypass grafting and valve surgery)/ 25 patients	ECG, PPG, SBP, MAP, DBP, CVP, LAP, SpO ₂ , Temp, MV, RR, PAW, CO _{2 exp} , CO _{2 insp} , Isoflurane _{insp} , Isoflurane _{exp} , alarms/Not specified	Retrospective with video/ Offline/ Alarm Validity
		Univariate/ Stream-centric/ No SQI/ AD	None
Yang [110]	GE S/5 Monitor/ 40 patients	EtCO ₂ , MAP, MV exp, NIBP mean/ A & S: 1 sample every 5 sec	Retrospective/ Offline/ TPR, FPR
		Univariate/ Stream-centric/ No SQI/ AD	None
Yang [68]	OEM Not specified/ 10 patients	NIBP mean/ A & S: 1 sample every 5 sec	Retrospective/ Offline/ TPR, FPR
		Univariate/ Stream-centric/ No SQI/ AD	None
Ansermino [145]	GE S/5 Monitor/ 47 surgeries x 1 hr (19 children, 28 adults)	HR, NIBP mean, SpO ₂ , EtCO ₂ , MV exp, RR/ Not specified	Real-time/ Offline/ Se, PPV, NPV, TPR, FPR
		Univariate/ Stream-centric/ No SQI/ AD	None
Yang [144]	GE S/5 Monitor/ E: fibrillation and other instances of abnormal heart rhythm/ 2 patients	HR, PR, IBP/ A: 1 sample every 5 sec	Retrospective/ Simulated data & [145]/ RMSE
		Multivariate/ Stream- centric/ No SQI / AD	None
Hoare [146]	Local hospital database, OEM Not specified / 245 cases	HR/ A & S: 1 sample every 30 sec	Retrospective/ Offline/ AUROC, FPR, PPV
		Univariate/ Stream-centric/ No SQI/ AD	None
Gostt [147]	Datex AS/3 Anaesthesia Monitor, Nellcor N-200/ 9 paediatric and 11 adult patients	HR, PR/ A: 1 sample every 5 sec, but not always due to data capture issues	Retrospective/ Real- time/ Acc, Se, Sp
		Multivariate analysis for univariate problem/ Stream + disease-centric/ No SQI/ AD + CED	None

Author	Theme II: Physiologic Data Source Data Monitor/ Inclusion (I)/ Exclusion (E) criteria/ Sample size	Theme III: Harvested Data Data Type/Acquisition(A)/ Sampling or Storage (S) Frequency	Theme V: Clinical Evaluation Data Annotation/ Mode/ Performance Metric
		Theme IV: Data Analysis Dimensionality/ Focus/ Signal Quality/ Clinical Contribution	Theme VI: Clinical Implementation
Sittig [148]	OEM Not specified / I: during open heart surgery/ 1 patient	HR, HR from ABP, Pulmonary artery mean pressure, MAP/ A: 1 sample every 2 min S: 1 sample repeated every sec	Not specified/ Real- time/ Graphical evaluation
		Univariate/ Stream-centric/ No SQI/ AD	None
Navabi [149]	5 kinds of Datascope monitors, Nellcor PO, Critikon VRP respiratory monitor/ 21 surgical cases	HR, EtCO ₂ , Inspired O ₂ , NIBP, SAO ₂ , HR, Inspired & expired lung volumes, RR, MAP, MV/ A: EtCO ₂ , Inspired O ₂ , NIBP, HR, SAO ₂ : 0.2 Hz, Inspired & expired lung volumes, RR, MAP, MV: 0.0167 Hz	Not specified/ Real- time/ CDR, FARR
		Multivariate/ Stream- centric/ No SQI / AD	None
Ebrahim [107] Feldman [108]	SpaceLabs Medical Gateway, SpaceLabs Medical PC2/ 12 OR, 60 adult ICU, 13 PICU patients, each ICU record is 4 hours long	HR, BP, PR, SpO ₂ / A & S: HR, PR reading every 3- 5 s	Real-time/ Real-time in [108]/ False and missed alarms
		Multivariate analysis for univariate problem/ Stream centric/ No SQI/ AD+PD	None
Clifford [72]	Sana Project database / 30,000 x 10 sec	Twelve lead ECG/ A & S: 500 Hz	Retrospective/ Offline/ Acc, ROCC
		Multivariate type (d)/ Stream-centric/ Developed SQI / AD	None
Redmond [81]	TeleMedCareHealth Monitor (TeleMedCare Pty Ltd Sydney, Australia)/ I: home dwelling patients with chronic obstructive pulmonary disease and/or congestive heart failure/ 288 patients, total 300 recordings	Single Lead I-ECG/ S: 500 Hz	Retrospective/ Offline/ Se, Sp, Acc, k
		Univariate/ Stream-centric/ SQI can be developed/ AD	None
Reisner [150]	Propaq 206EL monitors (Protocol Systems, Beaverton, Ore)/ I: prehospital trauma adult patients/ 671 patients	ECG, RR, HR, SBP, DBP/ A & S: ECG at 182 Hz; HR, RR at 1 Hz	Retrospective/ Offline/ AUROC
		Multivariate / Stream + disease-centric/ No SQI/ AD + CED	None

Author	Theme II: Physiologic Data Source Data Monitor/ Inclusion (I)/ Exclusion (E) criteria/ Sample size	Theme III: Harvested Data Data Type/Acquisition(A)/ Sampling or Storage (S) Frequency	Theme V: Clinical Evaluation Data Annotation/ Mode/ Performance Metric
		Theme IV: Data Analysis Dimensionality/ Focus/ Signal Quality/ Clinical Contribution	Theme VI: Clinical Implementation
Martinez- Tabares [151]	(i) INNOVATEC S.L. (Spain), CELBIT LTDA. ELECTRODOCTOR (Colombia), Lionheart 1 (BIO-TEK) simulator, QRS-Card (Pulse Biomedical Inc) (ii) ² Sana Project database/ (ii) 1000 x 12 sec	Twelve lead ECG/ A & S: 500 Hz	Retrospective/ Offline/ Se, Sp, Acc, ROCC
		Multivariate / Stream- centric/ Developed SQI / AD	None
Silva [152]	¹ MIMIC II database/ 1361 epochs	Multi-lead ECG, PPG, ABP, Resp, CVP/ S: 125 Hz	Retrospective/ Offline/ AUROC
		Multivariate/ Stream- centric/ Developed SQI / AD	None
Hayn [153]	² Sana Project database / 2000 x 10 sec	Twelve lead ECG/ A & S: 500 Hz	Retrospective/ Offline/ Se, Sp, Acc
		Multivariate type (d)/ Stream-centric/ Developed SQI / AD	None
Jekova [59]	² Sana Project database / 1500 x 10 sec	Twelve lead ECG/ A & S: 500 Hz	Retrospective/ Offline/ Se, Sp, ROCC
		Multivariate type (d)/ Stream-centric/ Developed SQI / AD	None
Di Marco [154]	² Sana Project database / 1498 x 10 sec	Twelve lead ECG/ A & S: 500 Hz	Retrospective/ Offline/ Acc, ROCC
		Multivariate type (d)/ Stream-centric/ Developed SQI / AD	None
Johannesen [155]	² Sana Project database / 1500 x 10 sec	Twelve lead ECG/ A & S: 500 Hz	Retrospective/ Offline/ Acc, ROC curves
		Multivariate type (d)/ Stream-centric/ Developed SQI / AD	None
Addison [103]	Nell-1 oximeter module, (Covidien-Nellcor, Boulder, CO, USA) with a Nellcor Max-A disposable probe, Datex-Ohmeda CardioCap/S5/ 139 healthy volunteers x 8 min	PPG, EtCO ₂ / A: PPG at 75.7 Hz	Real-time/ Real-time/ Root Mean Square Difference (RMSD)
		Univariate/ Stream-centric/ No SQI/ AD + PD	RR _{OxI} (Covidien- Nellcor,Boulder,CO, USA)
Lázaro [102]	Biopac OXY100C,ECG100C, RSP100C sensor, TSD201 transducer & Finometer / 17 subjects x 9 min	PPG at 250 Hz, ECG leads I, III & precordials at 1000 Hz, respiratory signal r(n) at 125 Hz, BP at 250 Hz	Not specified/ Offline/ Inter-subject mean & standard deviation
		Multivariate/ Stream- centric/ No SQI / AD + PD	None

Author	Theme II: Physiologic Data Source Data Monitor/ Inclusion (I)/ Exclusion (E) criteria/ Sample size	Theme III: Harvested Data Data Type/Acquisition(A)/ Sampling or Storage (S) Frequency	Theme V: Clinical Evaluation Data Annotation/ Mode/ Performance Metric
		Theme IV: Data Analysis Dimensionality/ Focus/ Signal Quality/ Clinical Contribution	Theme VI: Clinical Implementation
Otero [156]	Polysomnographic device, (Nicolet Biomedical Inc.)/ I: sleep study/ 10 patients, total 59 hr 10 min	Respiratory airflow, SpO ₂ / A: 68.25 Hz; S: 4 Hz	Retrospective/ Offline, TRACE [249]/ CDR, FPR
		Multivariate / Stream + disease-centric/ No SQL/ AD + CED	None
Bsoul [157]	PhysioNet Apnea-ECG database/ 35 subjects	Polysomnography and ECG/ A: ECG at 100 Hz, S: ECG at 250 Hz	Retrospective/ Offline/ F-measure, Se, Acc,
		Multivariate / Stream + disease-centric/ No SQL/ AD + CED + PD	None
Xia [158]	² Sana Project database/ 1000 x 10 sec	Twelve lead ECG/ S: 500 Hz	Retrospective/ Offline/ Acc, Se, Sp
		Multivariate type (d)/ Stream-centric/ Discussed SQL / AD	None
Zaunseeder [159]	² Sana Project database/ 1000 x 10 sec	Twelve lead ECG/ S: 500 Hz	Retrospective/ Offline/ Acc
		Multivariate type (d)/ Stream-centric/ Developed SQL / AD	None
Kužilek [160]	² Sana Project database/ Not clear	Twelve lead ECG/ S: 500 Hz	Retrospective/ Offline/ Not clear
		Multivariate type (d)/ Stream-centric/ No SQL / AD	None
Sukor [161]	MLT1020FC Reflection mode infrared finger probe, Differential bio-amplifier ST4400, PowerLab data acquisition system (ADInstruments, Sydney, Australia)/ 13 healthy subjects	PPG, ECG/ A: 1 kHz	Retrospective & Real-time/ Offline/ Se, Sp, Acc, k
		Multivariate validation of a univariate classifier/ Stream-centric/ SQL can be developed/AD + PD	None
Acharya [162]	Sleep staging signals Grass amplifiers (Astro-Med Inc., USA), other OEM not specified/ 25 subjects with suspected disease, 14 normal subjects	ECG/ A & S: 256 Hz	Real-time PSG annotations by a clinician/ Offline/ Acc, Se, Sp, PPV using confusion matrix
		Univariate/ Stream + disease-centric/ No SQL/ AD + CED	None

Author	Theme II: Physiologic Data Source Data Monitor/ Inclusion (I)/ Exclusion (E) criteria/ Sample size	Theme III: Harvested Data Data Type/Acquisition(A)/ Sampling or Storage (S) Frequency	Theme V: Clinical Evaluation Data Annotation/ Mode/ Performance Metric
		Theme IV: Data Analysis Dimensionality/ Focus/ Signal Quality/ Clinical Contribution	Theme VI: Clinical Implementation
Khandoker [163]	OEM Not specified/ I: sleep apnoea & E: cardiac history/ 29 patients	EEG, electrooculograms, ECG, leg movements, body positions, thoracic & abdominal wall expansion, oronasal airflow, SpO ₂ / S: 250 Hz for ECG, S: 32 Hz for thoracic and abdominal wall expansion	Polysomnography annotations/ Offline/ Acc, Se, Sp
		Multivariate validation of a univariate classifier/ Stream + disease-centric/ No SQI/ AD + CED	None
Nizami [164]	PhysioNet databases: nsrdb; MIT-BIH: nsr2db, mitdb, sldb; BIDMC chfdb, chf2db/ I: various records with normal sinus rhythm, arrhythmia or congestive heart failure/100 patient records	RR Interval / Not applicable	Retrospective/ Offline/ Acc
		Univariate/ Stream + disease-centric/ No SQI/ AD + CED	None
Nemati [165]	Peripheral Arterial Tonometry (PAT) measurement (Itamar Medical, Isreal), other OEMs not specified/ I: apnoea- hypopnoea index (AHI) of 0 – 69.3 events/hour 30 patients x 6-8 hr	ECG, 4 channels of respiratory rate (from chest & abdomen plethysmograph, nasal & oral thermistor, nasal pressure), PAT / A & S: 500 Hz, except PAT at 100 Hz	Polysomnography annotations/ Offline/ Signal-to-Noise Ratio
		Multivariate/ Stream- centric/ Developed and implemented SQI/ AD	None
Alvarez [166]	Polysomnograph (Alice 5, Respironics, Philips Healthcare, The Netherlands), Nonin PureSAT PO (Nonin Medical, Plymouth, MN, USA)/ I: daytime hypersomnolence, loudsnoring, nocturnal choking and awakening, apneic events/ E: other sleep disorders e.g., insomnia, parasomnia, narcolepsy/ 148 patients x 8 hr	SaO ₂ / S: 1 Hz	Real-time Polysomnography annotations by a clinician/ Offline/ Acc, Se, Sp, ROCC
		Univariate/ Stream + disease-centric/ No SQI/ AD + CED	None
Chen [169]	Propaq EncoreR 206EL monitor (ProtocolR Systems Inc.)/ I: hemorrhaged patients receiving blood with documented injury/ E: insufficient documentation of injury / 823 patients	at least one of HR, RR, DBP, SBP, SaO ₂ A & S: 1 Hz, except BP taken intermittently	Retrospective/ Offline/ AUROC
		Multivariate / Stream + disease-centric/ Developed SQI / AD + CED	None

Author	Theme II: Physiologic Data Source Data Monitor/ Inclusion (I)/ Exclusion (E) criteria/ Sample size	Theme III: Harvested Data Data Type/Acquisition(A)/ Sampling or Storage (S) Frequency	Theme V: Clinical Evaluation Data Annotation/ Mode/ Performance Metric
		Theme IV: Data Analysis Dimensionality/ Focus/ Signal Quality/ Clinical Contribution	Theme VI: Clinical Implementation
Kostic [170]	ECG recorders: Signalife Fidelity 100, NorthEast DR180+, Mortara ELI 200, HP Page Writer 1700A & GE MAC 5000/ I: percutaneous coronary intervention (PCI)/ 54 patients	ECG/ S: 720 Hz	Real-time/ Real-time/ Qualitative, graphical assessment
		Univariate/ Stream + disease-centric/ No SQL/ AD + CED	None
Yu [171]	Propaq EncoreR 206EL monitor (ProtocolR Systems Inc.)/ I: trauma patients during helicopter transport to a hospital/ 726 patients x 25 mins	ECG waveform at 182 Hz, PPG waveform at 91 Hz, HR, PR at 1 Hz	Retrospective/ Offline/ AUROC
		Multivariate / Stream + disease-centric/ Developed SQL / AD + CED	None
Gil [167, 168]	(i) ECG Apnea Data Base from Physionet; (ii) Digital polygraph EGP800 (Bitmed), COSMO EtCO ₂ /SpO ₂ Monitor(Novamatrix, Medical Systems)/ (i) I: children suspected of having Obstructive sleep apnea syndrome (OSAS)/ (ii) E: lack of manual annotations, unacceptable respiratory flow signal quality, doubtful clinical diagnosis/ (i) 70 adults (ii) 26 children	Polysomnography data: a chin electromyogram, 6 EEG & 2 electro-oculogram channels, ECG, air flow, respiratory plethysmography, PPG, SaO ₂ , CO ₂ / S: all signals at 100 Hz	Retrospective/ Offline/ Se, Sp, PPV
		Multivariate / Stream + disease-centric/ No SQL / AD + CED	None
Tarassenko [172]	OEM not specified / I: patients monitored for at least 24 hr after a myocardial infarct and again for a few hours 5 days later; patients with severe heart failure; patients with acute respiratory problems; elderly patients with hip fracture/ 150 patients x 24 hr	HR, RR,SpO ₂ ,temp, BP/ A & S: 1 sample every min	Retrospective & Real- time/ Real time, BioSign system/ Percentage of True alarms
		Multivariate/Stream + patient-centric/ No SQL/ AD	None

¹PhysioNet MIMIC II database recorded using Philips Intellivue MP-70 Medical Systems; ²PhysioNet database: Sana Project, OEM not specified; ³PhysioNet's MIMIC database recorded using Hewlett Packard CMS "Merlin" bedside monitor. *Abbreviations and acronyms*: Accuracy (Acc); Sensitivity (Se); Specificity (Sp); Receiver Operating Characteristic Curve (ROCC); Area Under the ROC Curve (AUROC); Positive Predictive Value (PPV); Negative Predictive Value (NPV); Correct Detection Rate (CDR); False Positive Rate (FPR); False Negative Rate (FNR); True Alarm Rate (TA); False Alarm Rate (FA); Root Mean Square Error (RMSE); False Alarm Reduction Rate(FARR); Equal Error Rate (EER); Cohen's kappa coefficient (κ); Electrooculograms (EOG).

APPENDIX B: AD COMPONENT INTERFACES

Requirements, provisions and configuration interfaces for each AD component used in each of the four experiments are tabulated in this appendix.

Table XV: AD component interfaces used in Experiment 1

Component	Requirements Interface	Provisions Interface	Configuration Interface
AD _{Dil}	[SpO ₂ Alarms , SQL]	[SpO ₂ Alarms, SQL]	<i>artSyms</i> {NaN, ^^, 5}
AD _{Dil}	[SpO ₂ , SQL]	[SpO ₂ , SQL]	<i>artSyms</i> {NaN, ^^, 5}
AD _{FuseSQL}	[SpO ₂ Alarms, SQL]; [SpO ₂ , SQL]	[SpO ₂ , SQL]	<i>N</i> = 2; operator (min); <i>SQType</i> = 'binary'

Table XVI: AD component interfaces used in Experiment 2

Component	Requirements Interface	Provisions Interface	Configuration Interface
AD _{Dil}	[SpO ₂ Alarms , SQL]	[SpO ₂ Alarms, SQL]	<i>artSyms</i> {NaN, ^^, 5}
AD _{Dil}	[SpO ₂ , SQL]	[SpO ₂ , SQL]	<i>artSyms</i> {NaN, ^^, 5}
AD _{FuseSQL}	[SpO ₂ Alarms, SQL]; [SpO ₂ , SQL]	[SpO ₂ , SQL]	<i>N</i> = 2; operator (min); <i>SQType</i> = 'binary'
AD _{MedFilt}	[SpO ₂ , SQL]	[SpO ₂ Med , SQL]	<i>Med_{ww}</i> = {5,10,20,25,30,35,60}

Table XVII: AD component interfaces used in Experiment 3

Component Name	Requirement Interface	Provision Interface	Configuration Interface
AD _{Dil}	[HR , SQL]	[HR , SQL]	<i>artSyms</i> {NaN, ^^, 5}
AD _{Dil}	[PR , SQL]	[PR , SQL]	<i>artSyms</i> {NaN, ^^, 5}
AD _{Dil}	[SpO ₂ , SQL]	[SpO ₂ , SQL]	<i>artSyms</i> {NaN, ^^, 5}

Component Name	Requirement Interface	Provision Interface	Configuration Interface
AD _{Dil}	[SpO ₂ Alarms, SQI]	[SpO ₂ Alarms, SQI]	<i>artSyms</i> {NaN, ^^, 5}
AD _{FuseSQI}	[SpO ₂ Alarms, SQI] ; [SpO ₂ , SQI]	[SpO ₂ , SQI]	<i>N</i> = 2; operator (min); <i>SQType</i> = 'binary'
AD _{Diff}	[HR, SQI] ; [PR, SQI]	[PR, SQI]	<i>Data1.Type</i> = 'HR'; <i>Data2.Type</i> = 'PR'; <i>SQType</i> = 'binary'; <i>SQThresh</i> = {6,12,18}; <i>SQI_{default}</i> = 0
AD _{FuseSQI}	[SpO ₂ , SQI]; [PR, SQI]	[SpO ₂ , SQI];	<i>N</i> = 2; operator (min); <i>SQType</i> = 'binary'

Table XVIII: AD component interfaces used in Experiment 4

Component Name	Requirement Interface	Provision Interface	Configuration Interface
AD _{Dil}	[HR, SQI]	[HR, SQI]	<i>artSyms</i> {NaN, ^^, 5}
AD _{Dil}	[PR, SQI]	[PR, SQI]	<i>artSyms</i> {NaN, ^^, 5}
AD _{Dil}	[SpO ₂ , SQI]	[SpO ₂ , SQI]	<i>artSyms</i> {NaN, ^^, 5}
AD _{Dil}	[SpO ₂ Alarms, SQI]	[SpO ₂ Alarms, SQI]	<i>artSyms</i> {NaN, ^^, 5}
AD _{FuseSQI}	[SpO ₂ Alarms, SQI] ; [SpO ₂ , SQI]	[SpO ₂ , SQI]	<i>N</i> = 2; operator (min); <i>SQType</i> = 'binary'
AD _{Diff}	[HR, SQI] ; [PR, SQI]	[PR, SQI]	<i>Data1.Type</i> = 'HR'; <i>Data2.Type</i> = 'PR'; <i>SQType</i> = 'binary'; <i>SQThresh</i> = {6,12,18}; <i>SQI_{default}</i> = 0
AD _{FuseSQI}	[SpO ₂ , SQI]; [PR, SQI]	[SpO ₂ , SQI];	<i>N</i> = 2; operator (min); <i>SQType</i> = 'binary'
AD _{MedFilt}	[SpO ₂ , SQI]	[SpO _{2Med} , SQI]	<i>Med_{ww}</i> = {5,10,20,25,30,35,60}

APPENDIX C: MASIMO VS. NELLCOR SPO₂

COMPARISON

Reprinted with permission from the author's publication [239].

C.1 DATA CHARACTERISTICS

This appendix describes the methodology and results of data comparison between the three OEM monitors. Datasets for each pulse oximeter were pooled across all patients and were explored as follows. The relative frequency plots of SpO₂ data of both pulse oximeters and their box plots were generated. Such exploration can reveal unknown differential characteristics between the two datasets, which can otherwise go unnoticed in higher order statistical abstractions. SpO₂ data from both pulse oximeters were tested for normality using Lilliefors test. Both datasets failed the normality test at the 5% significance level. Therefore, the nonparametric Kruskal-Wallis test was selected to compare ranks of the median of SpO₂ samples from both pulse oximeters. The relative frequency plots of the Dräger HR, and Nellcor and Masimo PRs were generated. These data were similarly tested for normality and failed. The empirical cumulative distribution function (CDF) for the Dräger HR was explored to determine the cause of this failure. The failure of the normality test led to the selection of the nonparametric Kruskal-Wallis test to compare ranks of the median HR and PRs.

Figure C-1 depicts the relative frequency of SpO₂ data of both pulse oximeters. Masimo produced reduced frequency of SpO₂ values between 87-90%, similar to results shown by earlier research [112, 114, 250]. Its algorithm calibrates values in this range by shifting them to values above 90%; hence the rise in frequency between 97-99% as compared to Nellcor. This difference is further explored in Figure C-2 through the box plot for SpO₂ data of each pulse oximeter. The median SpO₂ values

are Nellcor (97%) and Masimo (98%). The medians are similar; however, their ranks are different as shown by the range of the box plots. The Kruskal-Wallis test applied to the ranks of these medians results in rejecting the null hypothesis that no difference exists ($p < 0.0001$) Figure C-3 compares the relative frequency of HR and PR data of all three OEM monitors, exhibiting some discrepancies at the low and high values. Box plots in Figure C-4 explore the differences in the HR and PR values. Their median values are 161 beats per minute (bpm) for Dräger HR, and 162 bpm for both Nellcor and Masimo PR. The medians are similar; however, their ranks are different as shown by the range of these box plots. The Kruskal-Wallis test on the ranks of the median HR and PR results in ($p < 0.0001$, $\alpha = 0.05$). 95% CIs for the difference in the ranks of their medians are: [-11693, -9505] for HR and Nellcor PR; [-8017, -5829] for HR and Masimo PR; and [2583, 4770] for Nellcor PR and Masimo PR. The CDF of the HR, $F(x)$ in Figure C-, does not satisfy the 68-95-99 rule for a Gaussian (normal) distribution [251]. Taken together, these results indicate that there is no statistically significant difference in the SpO_2 estimates provided by the two OEM monitors.

Spearman's correlation coefficient $\rho = 0.65$ for SpO_2 data of both pulse oximeters; $\rho = 0.81$ for PR data of both pulse oximeters; $\rho = 0.93$ for HR and Nellcor PR; $\rho = 0.79$ for HR and Masimo PR.

C.2 DISCUSSION

Statistically significant differences were found in the Nellcor and Masimo SpO_2 data with a difference of 1% in their medians. Interestingly, this research led to the discovery that the Masimo SmartPod used at CHEO housed a discontinued version of the calibration algorithm and not the latest version which the OEM had been supplying to other hospitals since 2008 [250]. Therefore, the 1% higher median of Masimo may be attributable to its calibration algorithm causing an increase in the frequency of values between 97-99%. The revised calibration algorithm is shown to return a

distribution of values that is similar to Nellcor OxiMax oximeters [112]. Therefore, it may be inferred that the discrepancy in the medians may be essentially non-existent. Moreover, since this difference did not impact patient management in real-time, it was deemed clinically unimportant. Ranks of the median PRs differed statistically from each other and from that of the gold standard HR; however, a difference of 1 bpm in the actual median value is not clinically relevant. Nellcor PR exhibited the highest correlation with the HR. The McNemar test found no statistically significant difference between the performances of the two pulse oximeters since ($p = 0.22$) > ($\alpha/2 = 0.025$).

Despite competing claims of higher accuracy made by OEMs, this research produces evidence that there is no major advantage to using one technology over the other in preterm infants in the NICU. The high false alarm rates in both Nellcor and Masimo pulse oximeters result in low PPV and accuracy highlighting the need for improved AD and false alarm suppression. The majority of *FP* alarms in this study were issued when physiologic data values hovered close to the lower or higher SpO₂ limit. This is a limitation of the simplistic threshold-based alarm generation algorithms in current pulse oximeters. For greater clinical effectiveness, these algorithms need modification by including more complex clinical algorithms. This may lead to increasing their Sn and decreasing their FAR, while decreasing alarm fatigue and staff desensitization in hospitals. Similar to earlier studies [112-114], another limitation is the slightly different averaging times applied by each OEM technology to smooth the data. Masimo-supported research [176] describes the effect of averaging times on SpO₂ readings acquired by Masimo SET.

C.3 CONCLUSION

While the SpO₂ and PR data had statistically different characteristics between the two monitors, the differences were not observed to be clinically important. Contrary to

claims made by OEM-supported research, this study did not find statistically or clinically significant differences in the sensitivity, PPV and accuracy of SpO₂ alarms generated by Nellcor OXIMAX and Masimo SET pulse oximeter technologies. This study produced evidence that both pulse oximeters have high false alarm rates when used in the preterm neonatal population with diverse pathologies. Performance evaluation plays an essential role for hospitals during purchase decisions. Unfortunately, it is often not possible for individual healthcare organizations to conduct detailed side by side evaluation of medical devices during the procurement process to validate vendor performance claims. Most hospitals have to rely on OEM and third party evaluation data in conjunction with their brief high level product assessments. As a result of this research, CHEO is considering updating the pulse oximeter SmartPod technology integrated with the Infinity Delta monitors in its NICU.

Figure C-4: Box plot for pooled HR and PR data from Dräger, Nellcor and Masimo across all patients

REFERENCES

- [1] B. A. Sullivan, S. M. Grice, D. E. Lake, J. R. Moorman and K. D. Fairchild, "Infection and other clinical correlates of abnormal heart rate characteristics in preterm infants," *J. Pediatr.*, vol. 164, pp. 775-780, 2014.
- [2] F. Musial, "Why should we integrate biomarkers into complex trials?" *Forsch. Komplementarmed.*, vol. 19, pp. 232-233, 2012.
- [3] A. Bravi, A. Longtin and A. J. E. Seely, "Review and classification of variability analysis techniques with clinical applications," *BioMedical Engineering Online*, vol. 10, 2011.
- [4] J. Fackler and M. Spaeder, "Why doesn't healthcare embrace simulation and modeling? What would it take?" *Proceedings - Winter Simulation Conference*, pp. 1137-1142, 2011.
- [5] P. Stehlik, A. Bahmanpour, Y. A. Sekercioglu, P. Dārziņš and J. L. Marriott, "Fundamental elements identified for success of disease state management clinical decision support systems," *Electronic Journal of Health Informatics*, vol. 9(1), 2015.
- [6] A. Minutolo, M. Esposito and G. De Pietro, "A pattern-based approach for representing condition-action clinical rules into DSSs," *Lecture Notes in Electrical Engineering*, vol. 152 LNEE, pp. 777-789, 2013.
- [7] M. Pryor, D. White, B. Potter and R. Traill, "Evaluating online diagnostic decision support tools for the clinical setting." *Stud. Health Technol. Inform.*, vol. 178, pp. 180-185, 2012.
- [8] J. Horsky, G. D. Schiff, D. Johnston, L. Mercincavage, D. Bell and B. Middleton, "Interface design principles for usable decision support: A targeted review of best practices for clinical prescribing interventions," *J. Biomed. Inform.*, vol. 45, pp. 1202-1216, 2012.
- [9] W. H. Drummond, "Neonatal informatics-dream of a paperless NICU: Part four: Integrating caregiving, automated process management, and clinical decision support," *NeoReviews*, vol. 11, pp. 174-183, 2010.
- [10] M. McShea, R. Holl, O. Badawi, R. Riker and E. Silfen, "The eICU Research Institute - a collaboration between industry, health-care providers, and academia," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 29, pp. 18-25, 2010.
- [11] V. Chopra and L. F. McMahon. Redesigning hospital alarms for patient safety: Alarmed and potentially dangerous. *JAMA, American Medical Association* 311(12), pp. 1199-1200. 2014.

- [12] E. S. Berner, "Clinical decision support systems: state of the art," *AHRQ Publication no. 09-0069-EF. Rockville, Maryland: Agency for Healthcare Research and Quality*, June 2009.
- [13] J. M. Solet and P. R. Barach, "Managing alarm fatigue in cardiac care," *Prog. Pediatr. Cardiol.*, vol. 33, pp. 85-90, 2012.
- [14] M. Cvach, "Monitor alarm fatigue: An integrative review," *Biomedical Instrumentation and Technology*, vol. 46, pp. 268-277, 2012.
- [15] L. A. Lynn and J. P. Curry, "Patterns of unexpected in-hospital deaths: A root cause analysis," *Patient Safety in Surgery*, vol. 5, pp. Article number 3-24 pages, 11 February 2011.
- [16] K. E. Raymer, J. Bergström and J. M. Nyce, "Anaesthesia monitor alarms: a theory-driven approach," *Ergonomics*, vol. 55, pp. 1487-1501, 2012.
- [17] S. Nizami, J. R. Green and C. McGregor, "Implementation of artifact detection in critical care: a methodological review," *Biomedical Engineering, IEEE Reviews in*, vol. 6, pp. 127-142, 2013.
- [18] M. Abdelazez, P. X. Quesnel, A. D. C. Chan and H. Yang, "Signal quality Indices for ambulatory electrocardiograms used in myocardial ischemia monitoring," in *World Congress on Medical Physics and Biomedical Engineering, Toronto, Canada, 2015*, pp. 1076-1079.
- [19] C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance and L. Tarassenko, "Signal-Quality Indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring," *Biomedical and Health Informatics, IEEE Journal of*, vol. 19, pp. 832-838, 2015.
- [20] P. X. Quesnel, A. D. C. Chan and H. Yang, "Signal quality and false myocardial ischemia alarms in ambulatory electrocardiograms," in *Medical Measurements and Applications (MeMeA), 2014 IEEE International Symposium on*, 2014, pp. 1-5.
- [21] P. X. Quesnel, A. D. C. Chan and H. Yang, "Real-time biosignal quality analysis of ambulatory ECG for detection of myocardial ischemia," *IEEE International Symposium on Medical Measurements and Applications Proceedings (MeMeA)*, 2013.
- [22] Q. Li and G. D. Clifford, "Signal quality and data fusion for false alarm reduction in the intensive care unit," *J. Electrocardiol.*, vol. 45, pp. 596-603, 2012.
- [23] Q. Li and G. D. Clifford, "Dynamic time warping and machine learning for signal quality assessment of pulsatile signals," *Physiol. Meas.*, vol. 33, pp. 1491, 2012.
- [24] G. D. Clifford and G. B. Moody, "Signal quality in cardiorespiratory monitoring," *Physiol. Meas.*, vol. 33, 2012.
- [25] B. Bradley, G. C. Green, I. Batkin and A. J. E. Seely, "Feasibility of continuous multiorgan variability analysis in the intensive care unit," *J. Crit. Care*, vol. 27, pp. 218.e9-218.e20, 2012.

- [26] A. Aboukhalil, L. Nielsen, M. Saeed, R. G. Mark and G. D. Clifford, "Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform," *J. Biomed. Inform.*, vol. 41, pp. 442-451, 2008.
- [27] W. Zong, G. Moody and R. Mark, "Reduction of false arterial blood pressure alarms using signal quality assessment and relationships between the electrocardiogram and arterial blood pressure," *Medical and Biological Engineering and Computing*, vol. 42, pp. 698-706, 2004.
- [28] Y. Bai, D. H. Do, P. R. E. Harris, D. Schindler, N. G. Boyle, B. J. Drew and X. Hu, "Integrating monitor alarms with laboratory test results to enhance patient deterioration prediction," *J. Biomed. Inform.*, vol. 53, pp. 81-92, 2015.
- [29] S. Raihana, D. Dunsmuir, T. Huda, G. Zhou, Q. S. Rahman, A. Garde, M. Moinuddin, W. Karlen, G. A. Dumont, N. Kissoon, S. El Arifeen, C. Larson and J. M. Ansermino, "Development and Internal Validation of a Predictive Model Including Pulse Oximetry for Hospitalization of Under-Five Children in Bangladesh," *PLoS ONE*, 2015.
- [30] W. H. Drummond, "Neonatal informatics dream of a paperless NICU: Part five: Putting "IT" all together," *NeoReviews*, vol. 11, pp. 277-289, 2010.
- [31] M. Blount, M. R. Ebling, J. M. Eklund, A. G. James, C. McGregor, N. Percival, K. P. Smith and D. Sow, "Real-time analysis for intensive care: Development and deployment of the Artemis analytic system," *IEEE Engineering in Medicine and Biology Magazine*, vol. 29, pp. 110-118, 2010.
- [32] C. McGregor, "Big Data in Neonatal Intensive Care," *Computer*, vol. 46, pp. 54-59, 2013.
- [33] N. Bressan, A. James and C. McGregor, "Trends and opportunities for integrated real time neonatal clinical decision support," in *Biomedical and Health Informatics (BHI), IEEE-EMBS International Conference on*, 2012, pp. 687-690.
- [34] H. Khazaei, N. Mench-Bressan, C. McGregor and J. E. Pugh, "Health infomatics for neonatal intensive care units: an analytical modeling perspective," *Translational Engineering in Health and Medicine, IEEE Journal of*, pp. 1-1, 2015.
- [35] H. Khazaei, C. McGregor, M. Eklund, K. El-Khatib and A. Thommandram. Toward a big data healthcare analytics system: A mathematical modeling perspective. IEEE World Congress on Services. 2014.
- [36] R. Greer, C. Olivier, J. E. Pugh, J. M. Eklund and C. McGregor, "Remote, real-time monitoring and analysis of vital signs of neonatal graduate infants," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, 2014, pp. 1382-1385.
- [37] C. McGregor, C. Catley, J. Padbury and A. James, "Late onset neonatal sepsis detection in newborn infants via multiple physiological streams," *J. Crit. Care*, vol. 28, pp. e11-e12, 2, 2013.

- [38] C. McGregor, C. Catley and A. James, "Variability analysis with analytics applied to physiological data streams from the neonatal intensive care unit," in *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*, 2012, pp. 1-5.
- [39] A. Thommandram, J. M. Eklund, C. McGregor, J. E. Pugh and A. G. James, "A rule-based temporal analysis method for online health analytics and its application for real-time detection of neonatal spells," in *Big Data (BigData Congress), 2014 IEEE International Congress on*, 2014, pp. 470-477.
- [40] C. Catley, K. Smith, C. McGregor, A. James and J. M. Eklund, "A framework for multidimensional real-time data analysis: a case study for the detection of apnoea of prematurity," *International Journal of Computational Models and Algorithms in Medicine*, vol. 2, pp. 16-37, 2011.
- [41] C. Catley, K. Smith, C. McGregor, A. James and J. M. Eklund, "A framework to model and translate clinical rules to support complex real-time analysis of physiological and clinical data," in *Proceedings of the 1st ACM International Health Informatics Symposium*, Arlington, Virginia, USA, 2010, pp. 307-315.
- [42] J. Cirelli, B. Graydon, C. McGregor and A. James, "Analysis of continuous oxygen saturation data for accurate representation of retinal exposure to oxygen in the preterm infant," *Enabling Health and Healthcare through ICT*, IOS Press, pp. 126-131, Feb, 2013.
- [43] N. Bressan, C. McGregor, M. Blount, M. Ebling, D. Sow and A. A. James, "Identification of noxious events for newborn infants with a neural network," *4th Congress of the European Academy of Paediatric Societies*, 2012.
- [44] J. M. Eklund, N. Fontana, E. Pugh, C. McGregor, P. Yelder, A. James, M. Keyzers, C. Hahn and P. McNamara, "Automated sleep-wake detection in neonates from cerebral function monitor signals," in *Computer-Based Medical Systems (CBMS), 2014 IEEE 27th International Symposium on*, 2014, pp. 22-27.
- [45] J. Percival, C. McGregor, N. Percival and A. James, "Enabling the integration of clinical event and physiological data for real-time and retrospective analysis," *Information Systems and e-Business Management*, pp. 1-19, 2014.
- [46] B. Pasian, *Designs, Methods & Practices for Research of Project Management*. Gower Publishing, 2014.
- [47] A. Tolk, "Interoperability and composability," in *Modeling and Simulation Fundamentals: Theoretical Underpinnings and Practical Domains*, J. A. Sokolowski and C. M. Banks, Eds. John Wiley & Sons, Inc, 2010, pp. 403-433.
- [48] S. H. Kaisler, *Software Paradigms*. Hoboken, N.J., USA: Wiley-Interscience, 2005.
- [49] I. Crnkovic and M. Larsson, *Building Reliable Component-Based Software Systems*. Boston : Artech House, 2002.

- [50] H. Cervantes, P. Velasco-Elizondo and R. Kazman, "A Principled Way to Use Frameworks in Architecture Design," *Software, IEEE*, vol. 30, pp. 46-53, March, 2013.
- [51] C. V. Ramamoorthy and H. F. Li, "Pipeline Architecture," *ACM Comput.Surv.*, vol. 9, pp. 61-102, mar, 1977.
- [52] F. Buschmann, R. Meunier, H. Rohnert, P. Sommerlad and M. Stal, *Pattern-Oriented Software Architecture: A System of Patterns*. New York, NY, USA: John Wiley & Sons, Inc, 1996.
- [53] M. Serhani, A. Benharref and E. Badidi, "Towards dynamic non-obtrusive health monitoring based on SOA and cloud," *Health Information Science*, vol. 7798, pp. 125-136, 2013.
- [54] C. M. Machado, F. Couto, A. R. Fernandes, S. Santos, N. Cardim and A. T. Freitas, "Semantic characterization of hypertrophic cardiomyopathy disease," in *Bioinformatics and Biomedicine Workshops (BIBMW), 2010 IEEE International Conference on*, 2010, pp. 432-439.
- [55] J. Etienne, B. Wachmann and L. Zhang, "A component-based framework for knowledge discovery in bioinformatics," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia, PA, USA, 2006, pp. 916-921.
- [56] G. Juneja, B. Dournaee, J. Natoli and S. Birkel, "SOA in healthcare - part II," *SOA Magazine*, vol. March 23, 2009.
- [57] D. Cline and D. Clark, "A writer's guide to research and development proposals," 2000.
- [58] C. M. Roberts, *The Dissertation Journey: A Practical and Comprehensive Guide to Planning, Writing and Defending Your Dissertation*. Corwin Publishers, 2010.
- [59] I. Jekova, V. Krasteva, I. Christov and R. Ab, "Threshold-based system for noise detection in multilead ECG recordings," *Physiol. Meas.*, vol. 33, pp. 1463, 2012.
- [60] K. Schettlinger, R. Fried and U. Gather, "Real-time signal processing by adaptive repeated median filters," *Int J Adapt Control Signal Process*, vol. 24, pp. 346-362, May 2010.
- [61] M. Imhoff, S. Kuhls, U. Gather and R. Fried, "Smart alarms from medical devices in the OR and ICU," *Best Practice & Research Clinical Anaesthesiology*, vol. 23, pp. 39-50, 2009.
- [62] K. Schettlinger, R. Fried and U. Gather, "Robust filters for intensive care monitoring: Beyond the running median," *Biomed. Tech.*, vol. 51, pp. 49-56, 2006.
- [63] M. Imhoff and S. Kuhls, "Alarm algorithms in critical monitoring," *Anesth. Analg.*, vol. 102, pp. 1525-1537, 2006.

- [64] Wenguang Wang, Weiping Wang, Yifan Zhu and Qun Li, "Service-oriented simulation framework: An overview and unifying methodology," *Simulation*, vol. 87, pp. 221-252, 2011.
- [65] G. D. Clifford, W. J. Long, G. B. Moody and P. Szolovits, "Robust parameter extraction for decision support using multimodal intensive care data," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, pp. 411-429, 2009.
- [66] J. Percival, C. McGregor, N. Percival, R. Kamaleswaran and S. Tuuha, "A framework for nursing documentation enabling integration with HER and real-time patient monitoring," in *Computer-Based Medical Systems (CBMS), 2010 IEEE 23rd International Symposium on*, 2010, pp. 468-473.
- [67] M. Blount, C. McGregor, A. James, D. Sow, R. Kamaleswaran, S. Tuuha, J. Percival and N. Percival, "On the integration of an artifact system and a real-time healthcare analytics system," in *IHI'10 - Proceedings of the 1st ACM International Health Informatics Symposium*, 2010, pp. 647-655.
- [68] P. Yang, G. A. Dumont and J. M. Ansermino, "Online pattern recognition based on a generalized hidden Markov model for intraoperative vital sign monitoring," *Int J Adapt Control Signal Process*, vol. 24, pp. 363-381, May 2010.
- [69] A. Salatian and L. Oborkhale, "Filtering of ICU monitor data to reduce false alarms and enhance clinical decision support," *International Journal of Bio-Science and Bio-Technology*, vol. 3, pp. 49-55, 2011.
- [70] S. Siebig, S. Kuhls, M. Imhoff, J. Langgartner, M. Reng, J. Schölmerich, U. Gather and C. E. Wrede, "Collection of annotated data in a clinical validation study for alarm algorithms in intensive care-a methodologic framework," *J. Crit. Care*, vol. 25, pp. 128-135, 2010.
- [71] G. Takla, J. H. Petre, D. J. Doyle, M. Horibe and B. Gopakumaran, "The problem of artifacts in patient monitor data during surgery: a clinical and methodological review," *Anesth. Analg.*, vol. 103, pp. 1196-1204, 2006.
- [72] G. D. Clifford, J. Behar, Q. Li and I. Rezek, "Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms," *Physiol. Meas.*, vol. 33, pp. 1419-1433, 2012.
- [73] A. Konkani, B. Oakley and T. J. Bauld, "Reducing hospital noise: A review of medical device alarm management," *Biomedical Instrumentation and Technology*, vol. 46, pp. 478-487, 2012.
- [74] F. Schmid, M. S. Goepfert, D. Kuhnt, V. Eichhorn, S. Diedrichs, H. Reichensperner, A. E. Goetz and D. A. Reuter, "The wolf is crying in the operating room: Patient monitor and anesthesia workstation alarming patterns during cardiac surgery," *Anesth. Analg.*, vol. 112, pp. 78-83, 2011.
- [75] E. Walls-Esquivel, M. F. Vecchierini, C. Héberlé and F. Wallois, "Electroencephalography (EEG) recording techniques and artefact detection in early

premature babies," *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 37, pp. 299-309, 2007.

[76] M. F. Ma´rquez, L. Coli´n, M. Guevara, P. Iturralde and A. G. Hermosillo, "Common electrocardiographic artifacts mimicking arrhythmias in ambulatory monitoring," *Am. Heart J.*, vol. 144, pp. 187-197, 2002.

[77] C. Chase and W. J. Brady, "Artifactual electrocardiographic change mimicking clinical abnormality on the ECG," *Am. J. Emerg. Med.*, vol. 18, pp. 312-316, 2000.

[78] S. Siebig, W. Sieben, F. Kollmann, M. Imhof, T. Bruennler, F. Rockmann, U. Gather and C. E. Wrede, "Users' opinions on intensive care unit alarms - A survey of German intensive care units," *Anaesth. Intensive Care*, vol. 37, pp. 112-116, 2009.

[79] M. H. Hooper, L. Weavind, A. P. Wheeler, J. B. Martin, S. S. Gowda, M. W. Semler, R. M. Hayes, D. W. Albert, N. B. Deane, H. Nian, J. L. Mathe, A. Nadas, J. Sztipanovits, A. Miller, G. R. Bernard and T. W. Rice, "Randomized trial of automated, electronic monitoring to facilitate early detection of sepsis in the intensive care unit," *Crit. Care Med.*, vol. 40, pp. 2096-2101, 2012.

[80] H. Xia, G. A. Garcia and X. Zhao, "Automatic detection of ECG electrode misplacement: a tale of two algorithms," *Physiol. Meas.*, vol. 33, pp. 1549-1561, 2012.

[81] S. J. Redmond, Y. Xie, D. Chang, J. Basilakis and N. H. Lovell, "Electrocardiogram signal quality measures for unsupervised telehealth environments," *Physiol. Meas.*, vol. 33, pp. 1517-1533, 2012.

[82] S. J. Redmond, Q. Y. Lee, Y. Xie and N. H. Lovell, "Applications of supervised learning to biological signals: ECG signal quality and systemic vascular resistance," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 57-60, 2012.

[83] I. Silva, G. B. Moody and L. Celi, "Improving the quality of ECGs collected using mobile phones: The PhysioNet/Computing in Cardiology Challenge 2011," *Computers in Cardiology*, vol. 38, pp. 273-276, 2011.

[84] M. Borowski, M. Grges, R. Fried, O. Such, C. Wrede and M. Imhoff, "Medical device alarms," *Biomed. Tech.*, vol. 56, pp. 73-83, 2011.

[85] M. Chambrin, "Alarms in the intensive care unit: How can the number of false alarms be reduced?" *Critical Care*, vol. 5, pp. 184-188, 2001.

[86] L. P. A. Burgess, T. H. Herdman, B. W. Berg, W. W. Feaster and S. Hebsur, "Alarm limit settings for early warning systems to identify at-risk patients," *J. Adv. Nurs.*, vol. 65, pp. 1844-1852, 2009.

[87] A. H. Taenzer, J. B. Pyke, S. P. McGrath and G. T. Blike, "Impact of pulse oximetry surveillance on rescue events and intensive care unit transfers: A before-and-after concurrence study," *Anesthesiology*, vol. 112, pp. 282-287, 2010.

- [88] J. Welch, "An evidence-based approach to reduce nuisance alarms and alarm fatigue," *Biomedical Instrumentation and Technology*, vol. 45, pp. 46-52, 2011.
- [89] J. Welch, B. Kanter, B. Skora, S. McCombie, I. Henry, D. McCombie, R. Kennedy and B. Soller, "Multi-parameter vital sign database to assist in alarm optimization for general care units," *J. Clin. Monit. Comput.*, pp. 1-6, 2015.
- [90] D. Lobach, G. D. Sanders and T. J. Bright, "Enabling health care decision making through clinical decision support and knowledge management," Agency for Healthcare Research and Quality (US), Rockville (MD), Tech. Rep. 203, Apr. 2012.
- [91] E. V. Murphy, "Clinical Decision Support: Effectiveness in Improving Quality Processes and Clinical Outcomes and Factors That May Influence Success," *Yale J. Biol. Med.*, vol. 87, pp. 187-197, 06/06, 2014.
- [92] T. J. Bright, A. Wong, R. Dhurjati, E. Bristow, L. Bastian, R. R. Coeytaux, G. Samsa, V. Hasselblad, J. W. Williams, M. D. Musty, L. Wing, A. S. Kendrick, G. D. Sanders and D. Lobach, "Effect of clinical decision-support systems: a systematic review," *Annals of Internal Medicine*, vol. 157, pp. 29-43, July 3, 2012.
- [93] A. Thommandram, J. E. Pugh, J. M. Eklund, C. McGregor and A. G. James, "Classifying neonatal spells using real-time temporal analysis of physiological data streams: Algorithm development," in *Point-of-Care Healthcare Technologies (PHT), 2013 IEEE*, 2013, pp. 240-243.
- [94] A. T. Cruz, E. A. Williams, J. M. Graf, A. M. Perry, D. E. Harbin, E. R. Wuestner and B. Patel, "Test characteristics of an automated age- and temperature-adjusted tachycardia alert in pediatric septic shock," *Pediatr. Emerg. Care*, vol. 28, pp. 889-894, 2012.
- [95] T. Smith, D. Den Hartog, T. Moerman, P. Patka, E. M. M. Van Lieshout and N. W. L. Schep, "Accuracy of an expanded early warning score for patients in general and trauma surgery wards," *Br. J. Surg.*, vol. 99, pp. 192-197, 2012.
- [96] A. Jalali, P. Ghorbanian, A. Ghaffari and C. Nataraj, "A novel technique for identifying patients with ICU needs using hemodynamic features," *Advances in Fuzzy Systems*, 2012.
- [97] C. McGregor, C. Catley and A. James, "A process mining driven framework for clinical guideline improvement in critical care," *Learning from Medical Data Streams LEMEDS'11, 13th Conference on Artificial Intelligence in Medicine*, 2012.
- [98] K. A. Batchelder, P. D. Mannheimer, R. S. Mecca and J. M. Ojile, "Pulse oximetry saturation patterns detect repetitive reductions in airflow," *J. Clin. Monit. Comput.*, vol. 25, pp. 411-418, 2011.
- [99] A. M. Sawyer, E. N. Deal, A. J. Labelle, C. Witt, S. W. Thiel, K. Heard, R. M. Reichley, S. T. Micek and M. H. Kollef, "Implementation of a real-time computerized sepsis alert in nonintensive care unit patients," *Crit. Care Med.*, vol. 39, pp. 469-473, 2011.

- [100] P. Ordóñez, T. Armstrong, T. Oates and J. Fackler, "Classification of patients using novel multivariate time series representations of physiological data," in *Proceedings - 10th International Conference on Machine Learning and Applications, ICMLA*, 2011, pp. 172-179.
- [101] M. Altuve, G. Carrault, A. Beuchee, P. Pladys and A. I. Hernández, "On-line apnea-bradycardia detection using hidden semi-Markov models," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 4374-4377, 2011.
- [102] J. Lázaro, E. Gil, R. Bailón, A. Mincholé and P. Laguna, "Deriving respiration from photoplethysmographic pulse width," *Medical and Biological Engineering and Computing*, pp. 1-10, 2012.
- [103] P. S. Addison, J. N. Watson, M. L. Mestek and R. S. Mecca, "Developing an algorithm for pulse oximetry derived respiratory rate (RR_{oxi}): A healthy volunteer study," *J. Clin. Monit. Comput.*, vol. 26, pp. 45-51, 2012.
- [104] M. Mestek, P. Addison, A. Neitenbach, S. Bergese and S. Kelley, "Accuracy of Continuous Noninvasive Respiratory Rate Derived From Pulse Oximetry in Congestive Heart Failure Patients " *CHEST*, pp. 113A, 2012.
- [105] Q. Li, R. Mark and G. Clifford, "Artificial arterial blood pressure artifact models and an evaluation of a robust blood pressure and heart rate estimator," *BioMedical Engineering OnLine*, vol. 8, pp. 13, 2009.
- [106] Q. Li, R. G. Mark and G. D. Clifford, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter," *Physiol. Meas.*, vol. 29, pp. 15-32, 2008.
- [107] M. H. Ebrahim, J. M. Feldman and I. Bar-Kana, "A robust sensor fusion method for heart rate estimation," *J. Clin. Monit.*, vol. 13, pp. 385-393, 1997.
- [108] J. M. Feldman, M. H. Ebrahim and I. Bar-Kana, "Robust sensor fusion improves heart rate estimation: Clinical evaluation," *J. Clin. Monit.*, vol. 13, pp. 379-384, 1997.
- [109] C. McGregor, C. Catley, A. James and J. Padbury, "Next generation neonatal health informatics with Artemis," *Medical Informatics Europe (MIE 2011)*, Oslo, Norway, pp. 115-119, 2011.
- [110] P. Yang, G. Dumont and J. M. Ansermino, "A Cusum-based multilevel alerting method for physiological monitoring," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, pp. 1046-1052, 2010.
- [111] R. Schoenberg, D. Z. Sands and C. Safran, "Making ICU alarms meaningful: a comparison of traditional vs. trend-based algorithms." *Proceedings Annual Symposium AMIA*, pp. 379-383, 1999.
- [112] E. D. Johnston, B. Boyle, E. Juszczak, A. King, P. Brocklehurst and B. J. Stenson, "Oxygen targeting in preterm infants using the Masimo SET Radical pulse oximeter,"

Archives of Disease in Childhood: Fetal and Neonatal Edition, vol. 96, pp. F429-F433, 2011.

[113] A. C. Van Der Eijk, S. Horsch, P. H. C. Eilers, J. Dankelman and B. J. Smit, "'New-generation' pulse oximeters in extremely low-birth-weight infants: How do they perform in clinical practice?" *Journal of Perinatal and Neonatal Nursing*, vol. 26, pp. 172-180, 2012.

[114] D. Li, V. Jeyaprakash, S. Foreman and A. M. Groves, "Comparing oxygen targeting in preterm infants between the Masimo and Philips pulse oximeters," *Archives of Disease in Childhood: Fetal and Neonatal Edition*, vol. 97, pp. F311-F312, 2012.

[115] X. Hu, M. Sapo, V. Nenov, T. Barry, S. Kim, D. H. Do, N. Boyle and N. Martin, "Predictive combinations of monitor alarms preceding in-hospital code blue events," *J. Biomed. Inform.*, vol. 45, pp. 913-921, 2012.

[116] X. Sun, P. Yang, Y. Li, Z. Gao and Y. T. Zhang, "Robust heart beat detection from photoplethysmography interlaced with motion artifacts based on empirical mode decomposition," *Proceedings - IEEE-EMBS International Conference on Biomedical and Health Informatics: Global Grand Challenge of Health Informatics, BHI 2012*, pp. 775-778, 2012.

[117] F. Scalzo, D. Liebeskind and X. HU, "Reducing False Intracranial Pressure Alarms using Morphological Waveform Features," *Biomedical Engineering, IEEE Transactions on*, vol. PP, pp. 1-1, 2012.

[118] M. Borowski, S. Siebig, C. Wrede and M. Imhoff, "Reducing false alarms of intensive care online-monitoring systems: An evaluation of two signal extraction algorithms," *Computational and Mathematical Methods in Medicine*, vol. 2011, .

[119] S. Charbonnier and S. Gentil, "On-line adaptive trend extraction of multiple physiological signals for alarm filtering in intensive care units," *Int J Adapt Control Signal Process*, vol. 24, pp. 382-408, 2010.

[120] S. Charbonnier, G. Becq and L. Biot, "On-line segmentation algorithm for continuously monitored data in intensive care units," *Biomedical Engineering, IEEE Transactions on*, vol. 51, pp. 484-492, 2004.

[121] A. Otero, P. Félix, S. Barro and F. Palacios, "Addressing the flaws of current critical alarms: a fuzzy constraint satisfaction approach," *Artif. Intell. Med.*, vol. 47, pp. 219-238, 2009.

[122] J. M. Blum, G. H. Kruger, K. L. Sanders, J. Gutierrez and A. L. Rosenberg, "Specificity improvement for network distributed physiologic alarms based on a simple deterministic reactive intelligent agent in the critical care environment," *J. Clin. Monit. Comput.*, vol. 23, pp. 21-30, 2009.

[123] A. Otero, P. Félix, F. Palacios, C. Pérez-Gandía and C. O. S. Sorzano, "Intelligent alarms for patient supervision," in *IEEE International Symposium on Intelligent Signal Processing, WISP*, 2007, pp. 1-6.

- [124] W. Sieben and U. Gather, "Classifying alarms in intensive care - Analogy to hypothesis testing," *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4594 LNAI, pp. 130-138, 2007.
- [125] S. Jakob, I. Korhonen, E. Ruokonen, T. Virtanen, A. Kogan and J. Takala, "Detection of artifacts in monitored trends in intensive care," *Comput. Methods Programs Biomed.*, vol. 63, pp. 203-209, 2000.
- [126] Y. Zhang and P. Szolovits, "Patient-specific learning in real time for adaptive monitoring in critical care," *J. Biomed. Inform.*, vol. 41, pp. 452-460, 2008.
- [127] C. L. Tsien, "Event discovery in medical time-series data." in *Proc AMIA Symp.* 2000, pp. 858-862.
- [128] C. L. Tsien and J. C. Fackler, "Poor prognosis for existing monitors in the intensive care unit," *Crit. Care Med.*, vol. 25, pp. 614-619, 1997.
- [129] C. Tsien and J. Fackler, "An annotated data collection system to support intelligent analysis of Intensive Care Unit data," *Advances in Intelligent Data Analysis Reasoning about Data*, pp. 111-121, 1997.
- [130] V. Monasterio, F. Burgess and G. D. Clifford, "Robust classification of neonatal apnoea-related desaturations," *Physiol. Meas.*, vol. 33, pp. 1503, 2012.
- [131] H. Lee, C. G. Rusin, D. E. Lake, M. T. Clark, L. Guin, T. J. Smoot, A. O. Paget-Brown, B. D. Vergales, J. Kattwinkel, J. R. Moorman and J. B. Delos, "A new algorithm for detecting central apnea in neonates," *Physiol. Meas.*, vol. 33, pp. 1, 2012.
- [132] S. Nizami, J. R. Green and C. McGregor, "Service oriented architecture to support real-time implementation of artifact detection in critical care monitoring," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 4925-4928.
- [133] S. Y. Belal, A. J. Emmerson and P. C. W. Beatty, "Automatic detection of apnoea of prematurity," *Physiol. Meas.*, vol. 32, pp. 523, 2011.
- [134] J. A. Quinn, C. K. I. Williams and N. McIntosh, "Factorial switching linear dynamical systems applied to physiological condition monitoring," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, pp. 1537-1551, 2009.
- [135] F. Portet, G. Feng, J. Hunter and S. Sripada, "Evaluation of on-line bradycardia boundary detectors from neonatal clinical data," *Engineering in Medicine and Biology Society, 29th Annual International Conference of the IEEE*, pp. 3288-3291, 2007.
- [136] J. R. Moorman, D. E. Lake and M. P. Griffin, "Heart rate characteristics monitoring for neonatal sepsis," *Biomedical Engineering, IEEE Transactions on*, vol. 53, pp. 126-132, 2006.
- [137] C. L. Tsien, I. S. Kohane and N. McIntosh, "Building ICU artifact detection models with more data in less time." in *Proc AMIA Symp.* 2001, pp. 706-710.

- [138] C. L. Tsien, I. S. Kohane and N. McIntosh, "Multiple signal integration by decision tree induction to detect artifacts in the neonatal intensive care unit," *Artif. Intell. Med.*, vol. 19, pp. 189-202, 2000.
- [139] N. McIntosh, J. Becher, S. Cunningham, B. Stenson, I. A. Laing, A. J. Lyon and P. Badger, "Clinical diagnosis of pneumothorax is late: Use of trend data and decision support might allow preclinical detection," *Pediatr. Res.*, vol. 48, pp. 408-415, 2000.
- [140] C. Cao, N. McIntosh, I. S. Kohane and K. Wang, "Artifact detection in the PO₂ and PCO₂ time series monitoring data from preterm infants," *J. Clin. Monit. Comput.*, vol. 15, pp. 369-378, 1999.
- [141] W. Karlen, K. Kobayashi, J. M. Ansermino and G. A. Dumont, "Photoplethysmogram signal quality estimation using repeated Gaussian filters and cross-correlation," *Physiol. Meas.*, vol. 33, pp. 1617-1629, 2012.
- [142] W. Karlen, C. J. Brouse, E. Cooke, J. M. Ansermino and G. A. Dumont, "Respiratory rate estimation using respiratory sinus arrhythmia from photoplethysmography," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 1201-1204.
- [143] M. Koeny, S. Leonhardt, M. Czaplik and R. Rossaint, "On the road to predictive smart alarms based on a networked operating room," *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, pp. 1-4, 2012.
- [144] P. Yang, G. Dumont and J. Ansermino, "Sensor fusion using a hybrid median filter for artifact removal in intraoperative heart rate monitoring," *J. Clin. Monit. Comput.*, vol. 23, pp. 75-83, 2009.
- [145] J. M. Ansermino, J. P. Daniels, R. T. Hewgill, J. Lim, P. Yang, C. J. Brouse, G. A. Dumont and J. B. Bowering, "An evaluation of a novel software tool for detecting changes in physiological monitoring," *Anesth. Analg.*, vol. 108, pp. 873-880, 2009.
- [146] S. W. Hoare, D. Asbridge and P. C. W. Beatty, "On-line novelty detection for artefact identification in automatic anaesthesia record keeping," *Med. Eng. Phys.*, vol. 24, pp. 673-681, 2002.
- [147] R. K. Gostt, G. D. Rathbone and A. P. Tucker, "Real-time pulse oximetry artifact annotation on computerized anaesthetic records," *J. Clin. Monit. Comput.*, vol. 17, pp. 249-257, 2002.
- [148] D. F. Sittig and M. Factor, "Physiologic trend detection and artifact rejection: a parallel implementation of a multi-state Kalman filtering algorithm," *Comput. Methods Programs Biomed.*, vol. 31, pp. 1-10, 1990.
- [149] M. J. Navabi, K. C. Mylrea and R. C. Watt, "Detection of false alarms using an integrated anesthesia monitor," *Engineering in Medicine and Biology Society, International Conference of the IEEE*, vol. 11 pt 6, pp. 1774-1775, 1989.

- [150] A. T. Reisner, L. Chen and J. Reifman, "The association between vital signs and major hemorrhagic injury is significantly improved after controlling for sources of measurement variability," *J. Crit. Care*, vol. 27, pp. 533.e1-533.e10, 2012.
- [151] F. J. Martinez-Tabares, J. Espinosa-Oviedo and G. Castellanos-Dominguez, "Improvement of ECG signal quality measurement using correlation and diversity-based approaches," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 4295-4298, 2012.
- [152] I. Silva, J. Lee and R. G. Mark, "Signal quality estimation with multichannel adaptive filtering in intensive care settings," *IEEE Transactions on Biomedical Engineering*, vol. 59, pp. 2476-2485, 2012.
- [153] D. Hayn, B. Jammerbund and G. Schreier, "QRS detection based ECG quality assessment," *Physiol. Meas.*, vol. 33, pp. 1449-1461, 2012.
- [154] L. Y. Di Marco, W. Duan, M. Bojarnejad, M. Zheng, S. King, A. Murray and P. Langley, "Evaluation of an algorithm based on single-condition decision rules for binary classification of 12-lead ambulatory ECG recording quality," *Physiol. Meas.*, vol. 33, pp. 1435, 2012.
- [155] L. G. Lars Johannesen and, "Automatic ECG quality scoring methodology: mimicking human annotators," *Physiol. Meas.*, vol. 33, pp. 1479, 2012.
- [156] A. Otero, P. Félix, S. Barro and C. Zamarrón, "A structural knowledge-based proposal for the identification and characterization of apnoea episodes," *Applied Soft Computing Journal*, vol. 12, pp. 516-526, 2012.
- [157] M. Bsoul, H. Minn and L. Tamil, "Apnea MedAssist: Real-time Sleep Apnea Monitor Using Single-Lead ECG," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 15, pp. 416-427, 2011.
- [158] Henian Xia, G. A. Garcia, J. C. McBride, A. Sullivan, T. De Bock, J. Bains, D. C. Wortham and Xiaopeng Zhao, "Computer algorithms for evaluating the quality of ECGs in real time," *Computing in Cardiology*, pp. 369-372, 2011.
- [159] S. Zaunseder, R. Huhle and H. Malberg, "CinC challenge - Assessing the usability of ECG by ensemble decision trees," *Computers in Cardiology*, vol. 38, pp. 277-280, 2011.
- [160] J. Kužílek, M. Huptych, V. Chudaíček, J. Spilka and L. Lhotská, "Data driven approach to ECG signal quality assessment using multistep SVM classification," *Computers in Cardiology*, vol. 38, pp. 453-455, 2011.
- [161] J. A. Sukor, S. J. Redmond and N. H. Lovell, "Signal quality measures for pulse oximetry through waveform morphology analysis," *Physiol. Meas.*, vol. 32, pp. 369-384, 2011-03-01, 2011.
- [162] U. R. Acharya, E. C. Chua, O. Faust, T. Lim, L. Feng and B. Lim, "Automated detection of sleep apnea from electrocardiogram signals using nonlinear parameters," *Physiol. Meas.*, vol. 32, pp. 287, 2011.

- [163] A. Khandoker and M. Palaniswami, "Modeling Respiratory Movement Signals During Central and Obstructive Sleep Apnea Events Using Electrocardiogram," *Annals of Biomedical Engineering*, vol. 39, pp. 801-811, 2011.
- [164] S. Nizami, J. R. Green, J. M. Eklund and C. McGregor, "Heart disease classification through HRV analysis using Parallel Cascade Identification and Fast Orthogonal Search," *Medical Measurements and Applications Proceedings (MeMeA), IEEE International Workshop on*, pp. 134-139, 2010.
- [165] S. Nemati, A. Malhotra and G. D. Clifford, "Data fusion for improved respiration rate estimation," *EURASIP J. Adv. Signal Process*, pp. 1-10, 2010.
- [166] D. Álvarez, R. Hornero, J. V. Marcos and F. del Campo, "Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis," *Biomedical Engineering, IEEE Transactions on*, vol. 57, pp. 2816-2824, 2010.
- [167] E. Gil, J. María Vergara and P. Laguna, "Detection of decreases in the amplitude fluctuation of pulse photoplethysmography signal as indication of obstructive sleep apnea syndrome in children," *Biomedical Signal Processing and Control*, vol. 3, pp. 267-277, 2008.
- [168] E. Gil, M. Mendez, J. M. Vergara, S. Cerutti, A. M. Bianchi and P. Laguna, "Detection of obstructive sleep apnea in children using decreases in the amplitude fluctuations of PPG signal and HRV," *Engineering in Medicine and Biology Society, Annual International Conference of the IEEE*, pp. 3479-3482, 2008.
- [169] L. Chen, T. M. McKenna, A. T. Reisner, A. Gribok and J. Reifman, "Decision tool for the early diagnosis of trauma patient hypovolemia," *J. Biomed. Inform.*, vol. 41, pp. 469-478, 2008.
- [170] M. N. Kostic, S. Fakhar, T. Foxall, B. S. Drakulic and M. W. Krucoff, "Evaluation of novel ECG signal processing on quantification of transient ischemia and baseline wander suppression," *Engineering in Medicine and Biology Society, 29th Annual International Conference of the IEEE*, pp. 2199-2202, 2007.
- [171] C. Yu, Z. Liu, T. McKenna, A. T. Reisner and J. Reifman, "A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms," *Journal of the American Medical Informatics Association*, vol. 13, pp. 309-320, 2006.
- [172] L. Tarassenko, A. Hann and D. Young, "Integrated monitoring and analysis for early warning of patient deterioration," *Br. J. Anaesth.*, vol. 97, pp. 64-68, 2006.
- [173] Medical Predictive Science Corporation, Charlottesville, VA, USA, "HeRO Technical Publications," 2004.
- [174] Dräger Medical Systems Inc Telford, U.S.A., "Instructions for use: Infinity Delta Series," 2008.
- [175] Dräger Medical Systems Inc Telford, U.S.A., "Infinity gateway VF5.0 protocol handbook," 2007.

- [176] J. Vagedes, C. F. Poets and K. Dietz, "Averaging time, desaturation level, duration and extent," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 98, pp. F265-F266, May 01, 2013.
- [177] R. Sahni, "Noninvasive monitoring by photoplethysmography," *Clin. Perinatol.*, vol. 39, pp. 573-583, 2012.
- [178] N. Shah, H. B. Ragaswamy, K. Govindugari and L. Estanol, "Performance of three new-generation pulse oximeters during motion and low perfusion in volunteers," *J. Clin. Anesth.*, vol. 24, pp. 385-391, 2012.
- [179] S. J. Barker, "'Motion-resistant' pulse oximetry: A comparison of new and old models," *Anesth. Analg.*, vol. 95, pp. 967-972, 2002.
- [180] M. O'Reilly, "Reply to "'new-generation' pulse oximeters in extremely low-birth-weight infants"," *Journal of Perinatal and Neonatal Nursing*, vol. 26, pp. 282-283, 2012.
- [181] M. Verduijn, N. Peek, N. F. de Keizer, E. van Lieshout, A. J. M. de Pont, M. J. Schultz, E. de Jonge and B. A. J. M. de Mol, "Individual and Joint Expert Judgments as Reference Standards in Artifact Detection," *J. Am. Med. Inform. Assoc.*, vol. 15, pp. 227-234, 2008.
- [182] S. Cunningham, A. Symon and N. McIntosh, "The practical management of artifact in computerised physiological data," *Int. J. Clin. Monit. Comput.*, vol. 11, pp. 211-216, 1994.
- [183] S. T. Lawless, "Crying wolf: False alarms in a pediatric intensive care unit," *Crit. Care Med.*, vol. 22, pp. 981-985, 1994.
- [184] J. Hunter, Y. Freer, A. Gatt, E. Reiter, S. Sripada and C. Sykes, "Automatic generation of natural language nursing shift summaries in neonatal intensive care: BT-Nurse," *Artif. Intell. Med.*, vol. 56, pp. 157-172, 2012.
- [185] A. J. E. Seely, G. C. Green and A. Bravi, "Continuous multiorgan variability monitoring in critically ill patients — complexity science at the bedside," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 5503-5506.
- [186] A. Bravi, G. Green, A. Longtin and A. J. E. Seely, "Monitoring and identification of sepsis development through a composite measure of heart rate variability," *PLoS ONE*, vol. 7, pp. e45666, 2012.
- [187] S. Ahmad, T. Ramsay, L. Huebsch, S. Flanagan, S. McDiarmid, I. Batkin, L. McIntyre, S. R. Sundaresan, D. E. Maziak, F. M. Shamji, P. Hebert, D. Fergusson, A. Tinmouth and A. J. E. Seely, "Continuous multi-parameter heart rate variability analysis heralds onset of sepsis in adults," *PLoS ONE*, vol. 4, pp. e6642-e6642, 2009.
- [188] iMDsoft Clinical Information Systems, "<http://www.imd-soft.com/>," last accessed Jan 2016.

- [189] J. R. Moorman, W. A. Carlo, J. Kattwinkel, R. L. Schelonka, P. J. Porcelli, C. T. Navarrete, E. Bancalari, J. L. Aschner, M. Whit Walker, J. A. Perez, C. Palmer, G. J. Stukenborg, D. E. Lake and T. Michael O'Shea, "Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: A randomized trial," *J. Pediatr.*, vol. 159, pp. 900-906.e1, 2011.
- [190] D. A. Sitzman, "Executive summary: IntelliRate technology for patient monitoring," *GE Medical Systems Information Technologies*, 2002.
- [191] GE Healthcare, "Intellirate technology for patient monitoring," *GE Medical Systems Information Technologies*, 2010.
- [192] SAPHIRE Consortium, "IST- 27074 SAPHIRE: Intelligent healthcare monitoring based on a semantic interoperability platform," 2008.
- [193] P. Kligfield and C. L. Green, "The cardiac safety research consortium ECG database," *J. Electrocardiol.*, vol. 45, pp. 690-692, 0, 2012.
- [194] J. Havlík, L. Kučerová, I. Kohút, J. Dvořák and V. Fabián, "The database of the cardiovascular system related signals," *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7451 LNCS, pp. 169-170, 2012.
- [195] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. Peng and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet : Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, pp. e215-220, 2000.
- [196] G. Feng, S. Yaji, J. Hunter and F. Portet, "Using temporal constraints to integrate signal analysis and domain knowledge in medical event detection," *Artif. Intell. Med.*, pp. 46-55, 2009.
- [197] F. Portet, E. Reiter, A. Gatt, J. Hunter, S. Sripada, Y. Freer and C. Sykes, "Automatic generation of textual summaries from neonatal intensive care data," *Artif. Intell.*, vol. 173, pp. 789-816, 2009.
- [198] A. Gatt and F. Portet, "Text content and task performance in the evaluation of a Natural Language Generation system," *International Conference Recent Advances in Natural Language Processing, RANLP*, pp. 107-112, 2009.
- [199] A. Gatt, F. Portet, E. Reiter, J. Hunter, S. Mahamood, W. Moncur and S. Sripada, "From data to text in the neonatal intensive care Unit: Using NLG technology for decision support and information management," *AI Communications*, vol. 22, pp. 153-186, 2009.
- [200] J. Hunter, Y. Freer, A. Gatt, R. Logie, N. McIntosh, M. van der Meulen, F. Portet, E. Reiter, S. Sripada and C. Sykes, "Summarising complex ICU data in natural language." in *Annual Symposium Proceedings, AMIA*, 2008, pp. 323-327.
- [201] J. R. Moorman, C. E. Rusin, L. Hoshik, L. E. Guin, M. T. Clark, J. B. Delos, J. Kattwinkel and D. E. Lake, "Predictive monitoring for early detection of subacute

potentially catastrophic illnesses in critical care," in *Engineering in Medicine and Biology Society, Annual International Conference of the IEEE, EMBC*, 2011, pp. 5515-5518.

[202] J. R. Moorman, J. B. Delos, A. A. Flower, H. Cao, B. P. Kovatchev, J. S. Richman and D. E. Lake, "Cardiovascular oscillations at the bedside: early diagnosis of neonatal sepsis using heart rate characteristics monitoring," *Physiol. Meas.*, vol. 32, pp. 1821, 2011.

[203] Y. Xiao, M. P. Griffin, D. E. Lake and J. R. Moorman, "Nearest-Neighbor and Logistic Regression Analyses of Clinical and Heart Rate Characteristics in the Early Diagnosis of Neonatal Sepsis," *Medical Decision Making*, vol. 30, pp. 258-266, March/April 2010.

[204] M. P. Griffin, D. E. Lake, T. M. O'Shea and J. R. Moorman, "Heart rate characteristics and clinical signs in neonatal sepsis," *Pediatr. Res.*, vol. 61, pp. 222-227, 2007.

[205] M. P. Griffin, D. E. Lake, E. A. Bissonette, F. E. Harrell Jr., T. M. O'Shea and J. R. Moorman, "Heart rate characteristics: Novel physiomarkers to predict neonatal infection and death," *Pediatrics*, vol. 116, pp. 1070-1074, 2005.

[206] B. P. Kovatchev, L. S. Farhy, H. Cao, M. P. Griffin, D. E. Lake and J. R. Moorman, "Sample asymmetry analysis of heart rate characteristics with application to neonatal sepsis and systemic inflammatory response syndrome," *Pediatr. Res.*, vol. 54, pp. 892-898, 2003.

[207] M. P. Griffin, T. M. O'Shea, E. A. Bissonette, F. E. Harrell Jr., D. E. Lake and J. R. Moorman, "Abnormal heart rate characteristics preceding neonatal sepsis and sepsis-like illness," *Pediatr. Res.*, vol. 53, pp. 920-926, 2003.

[208] D. E. Lake, J. S. Richman, M. P. Griffin and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability," *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, vol. 283, pp. R789-R797, 09/01, 2002.

[209] M. P. Griffin and J. R. Moorman, "Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis," *Pediatrics*, vol. 107, pp. 97-104, 2001.

[210] M. P. Griffin, D. F. Scollan and J. R. Moorman, "The dynamic range of neonatal heart rate variability," *J. Cardiovasc. Electrophysiol.*, vol. 5, pp. 112-124, 1994.

[211] P. M. Middleton, C. H. H. Tang, G. S. H. Chan, S. Bishop, A. V. Savkin and N. H. Lovell, "Peripheral photoplethysmography variability analysis of sepsis patients," *Medical and Biological Engineering and Computing*, vol. 49, pp. 337-347, 2011.

[212] D. J. Meredith, D. Clifton, P. Charlton, J. Brooks, C. W. Pugh and L. Tarassenko, "Photoplethysmographic derivation of respiratory rate: A review of relevant physiology," *Journal of Medical Engineering and Technology*, vol. 36, pp. 1-7, 2012.

- [213] K. Steurbaut, K. Colpaert, B. Gadeyne, P. Depuydt, P. Vosters, C. Danneels, D. Benoit, J. Decruyenaere and F. De Turck, "COSARA: Integrated service platform for infection surveillance and antibiotic management in the ICU," *J. Med. Syst.*, vol. 36, pp. 3765-3775, 2012.
- [214] A. Minutolo, M. Esposito and G. De Pietro, "KETO: A knowledge editing tool for encoding condition - Action guidelines into clinical DSSs," *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7208 LNAI, pp. 352-364, 2012.
- [215] N. W. Chbat, W. Chu, M. Ghosh, G. Li, M. Li, C. M. Chiofolo, S. Vairavan, V. Herasevich and O. Gajic, "Clinical knowledge-based inference model for early detection of acute lung injury," *Ann. Biomed. Eng.*, vol. 40, pp. 1131-1141, 2012.
- [216] C. L. Green, P. Kligfield, S. George, I. Gussak, B. Vajdic, P. Sager and M. W. Krucoff, "Detection of QT prolongation using a novel electrocardiographic analysis algorithm applying intelligent automation: Prospective blinded evaluation using the Cardiac Safety Research Consortium electrocardiographic database," *Am. Heart J.*, vol. 163, pp. 365-371, 2012.
- [217] B. D. Bradley, G. Green, T. Ramsay and A. J. E. Seely, "Impact of sedation and organ failure on continuous heart and respiratory rate variability monitoring in critically ill patients: A pilot study," *Crit. Care Med.*, 2012.
- [218] V. Herasevich, "Informatics infrastructure for syndrome surveillance, decision support, reporting, and modeling of critical illness," *Mayo Clin. Proc.*, vol. 85, pp. 247-254, 2010.
- [219] K. Niemi, S. Geary, B. Quinn, M. Larrabee and K. Brown, "Implementation and evaluation of electronic clinical decision support for compliance with pneumonia and heart failure quality indicators," *American Journal of Health-System Pharmacy*, vol. 66, pp. 389-397, 2009.
- [220] L. Allart, C. Vilhelm, H. Mehdaoui, H. Hubert, B. Sarrazin, D. Zitouni, M. Lemdani and P. Ravauux, "An architecture for online comparison and validation of processing methods and computerized guidelines in intensive care units," *Comput. Methods Programs Biomed.*, vol. 93, pp. 93-103, 2009.
- [221] F. Wu, M. Williams, P. Kazanzides, K. Brady and J. Fackler, "A modular clinical decision support system clinical prototype extensible into multiple clinical settings," in *3rd International Conference on Pervasive Computing Technologies for Healthcare*, 2009, pp. 1-4.
- [222] E. Gronvall, L. Piccini, A. Pollini, A. Rullo and G. Andreoni, "Assemblies of Heterogeneous Technologies at the Neonatal Intensive Care Unit," *Ambient Intelligence*, pp. 340-357, 2007.
- [223] S. Nizami, J. R. Green and C. McGregor, "An artifact detection framework for clinical decision support systems," in *World Congress on Medical Physics and Biomedical Engineering, Toronto, Canada, June 7-12, 2015*, pp. 1393-1396.

- [224] S. Cerutti and C. Marchesi, *Advanced Methods of Biomedical Signal Processing*. Hoboken, Piscataway, NJ: Wiley; IEEE Press, 2011.
- [225] S. De Cesare, M. Lycett and R. Macredie, *Development of Component-Based Information Systems*. Armonk, N.Y.: M.E. Sharpe, 2006.
- [226] G. T. Leavens and M. Sitaraman, *Foundations of Component-Based Systems*. Cambridge, [England] ; New York : Cambridge University Press, 2000.
- [227] G. Larsen, "Component-based enterprise frameworks," *Commun ACM*, vol. 43, pp. 24-26, Oct, 2000.
- [228] J. Behar, A. Johnson, G. D. Clifford and J. Oster. A comparison of single channel fetal ECG extraction methods. *Ann. Biomed. Eng.* 42(6), pp. 1340-1353. 2014.
- [229] V. B. Sivarajan and D. Bohn, "Monitoring of standard hemodynamic parameters: heart rate, systemic blood pressure, atrial pressure, pulse oximetry, and end-tidal CO₂," *Pediatric Critical Care Medicine*, vol. 12, pp. S2-S11, 2011.
- [230] N. Akhter, H. Gite, S. Tharewal and K. V. Kale, "Computer based RR-interval detection system with ectopy correction in HRV data," in *Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on*, 2015, pp. 1613-1618.
- [231] N. Akhter, S. Tharewal, H. Gite and K. V. Kale, "Microcontroller based RR-interval measurement using PPG signals for heart rate variability based biometric application," in *Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on*, 2015, pp. 588-593.
- [232] A. D. C. Chan, K. B. Englehart, B. Hudgins and D. F. Lovely, "Multi-expert automatic speech recognition using acoustic and myoelectric signals," *IEEE Trans. on Biomedical Engineering*, vol. 53, pp. 676-685, 2006.
- [233] D. D. He, E. S. Winokur and C. G. Sodini, "An ear-worn vital signs monitor," *Biomedical Engineering, IEEE Transactions on*, vol. 62, pp. 2547-2552, 2015.
- [234] B. Jafarpour, S. Abidi, A. Ahmad and S. S. R. Abidi, "INITIATE: An intelligent adaptive alert environment," in *15th World Congress on Medical Informatics (MEDINFO)*, 2015, .
- [235] S. S. R. ABIDI and C. Yu-N, "An ontology-based approach using the extended markup language (XML)," in *Medical Infobahn for Europe*, A. Hasman, B. Blobel, J. Dudeck, R. Engelbrecht, G. Gell and H. Prokosch, Eds. Amsterdam: IOS Press, 2000, pp. 827-831.
- [236] C. Catley and M. Frize, "A prototype XML-based implementation of an integrated 'intelligent' neonatal intensive care unit," in *Information Technology Applications in Biomedicine, 2003. 4th International IEEE EMBS Special Topic Conference on*, 2003, pp. 322-325.

- [237] K. Nguyen and G. Xu, "Cachetor: Detecting cacheable data to remove bloat," in *Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering*, Saint Petersburg, Russia, 2013, pp. 268-278.
- [238] S. Yoshida, S. Nakashima, J. Odagiri and S. Itoh, "Development of an XML-document compaction method to improve data-processing performance," *Syst. Comput. Jpn.*, vol. 38, pp. 1-13, 2007.
- [239] S. Nizami, K. Greenwood, N. Barrowman and J. Harrold, "Performance Evaluation of New-Generation Pulse Oximeters in the NICU: Observational Study," *Cardiovascular Engineering and Technology, Springer*, vol. 6, pp. 383-391, 2015.
- [240] N. Finer and T. Leone, "Oxygen Saturation Monitoring for the Preterm Infant: The Evidence Basis for Current Practice," *Pediatr. Res.*, vol. 65, pp. 375-380, print, 2009.
- [241] D. E. Elder, A. J. Campbell and D. Galletly, "Current definitions for neonatal apnoea: Are they evidence based?" *J. Paediatr. Child Health*, vol. 49, pp. E388-E396, 2013.
- [242] N. M. Razi, M. DeLauter and P. B. Pandit, "Periodic breathing and oxygen saturation in preterm infants at discharge," *Journal of Perinatology*, vol. 22, pp. 442-444, 2002.
- [243] K. Lim, K. I. Wheeler, T. J. Gale, H. D. Jackson, J. F. Kihlstrand, C. Sand, J. A. Dawson and P. A. Dargaville, "Oxygen saturation targeting in preterm infants receiving continuous positive airway pressure," *J. Pediatr.*, vol. 164, pp. 730-736.e1, 2014.
- [244] K. C. Graham and M. Cvach, "Monitor alarm fatigue: Standardizing use of physiological monitors and decreasing nuisance alarms," *American Journal of Critical Care*, vol. 19, pp. 28-34, 2010.
- [245] S. P. Boyd, C. H. Barratt, S. P. Boyd and S. P. Boyd, *Linear Controller Design: Limits of Performance*. Prentice Hall Englewood Cliffs, NJ, 1991.
- [246] S. Bennett, *A History of Control Engineering, 1930-1955*. IET, 1993.
- [247] R. Rea, "IBM InfoSphere streams: Redefining real time analytic processing," IBM Corporation, 2010.
- [248] B. Gedik, H. Andrade, K. Wu, P. S. Yu and M. Doo, "SPADE: The system s declarative stream processing engine," in *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, Vancouver, Canada, 2008, pp. 1123-1134.
- [249] Intelligent Systems Group, "TRACE, Tool foR anAlyzing and disCovering pattErns, <https://www.gsi.dec.usc.es/trace>," *Department of Electronics and Computer Science, University of Santiago De Compostela, La Coruña*, last accessed Jan 2016.

[250] The BOOST II United Kingdom, Australia, and New Zealand Collaborative Groups, "Oxygen saturation and outcomes in preterm infants," *N. Engl. J. Med.*, vol. 368, pp. 2094-2104, 2013.

[251] A. Weaver and S. Goldberg, *Clinical Biostatistics made Ridiculously Simple*. Miami, FL, USA: MedMaster Inc., 2011.