QUANTIFYING TRUST IN WEARABLE MEDICAL DEVICES

QUANTIFYING TRUST IN WEARABLE MEDICAL DEVICES

By MINI THOMAS, M.Eng.

A Thesis Submitted to the Department of Computing & Software and the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (Software Engineering, 2024)

Hamilton, Ontario, Canada (Department of Computing & Software)

TITLE:	Quantifying Trust in Wearable Medical Devices
AUTHOR:	Mini Thomas M.Eng.(Electrical Engineering)
SUPERVISORS:	Dr. Reza Samavi and Dr. Antoine Deza
NUMBER OF PAGES:	xxiii, 205

Lay Abstract

In this thesis, two challenges in quantifying trust in wearable medical devices, are addressed. The first challenge is the identification of factors influencing trust which are inherently subjective and vary widely among users. To address this challenge, we conducted an extensive survey to identify and validate the trust factors. These factors are stepping stones for defining the specifications and quantifying trust in wearable medical devices.

The second challenge is to develop a precise method for quantification of trust while taking into account the uncertainty and variability of trust factors. We constructed a Bayesian network, that captures the complexities of trust as probabilities of the trust factors (identified from the survey) and developed a data-driven approach to estimate the parameters of the Bayesian network to compute the measure of trust.

The findings of this thesis are empirically and experimentally validated across multiple use cases, incorporating real and synthetic data, various testing conditions, and diverse Bayesian network configurations. Additionally, we developed a customizable, parameterized prototype that empowers users and healthcare providers to effectively assess and compare the trustworthiness of different wearable medical devices.

Abstract

Advances in sensor and digital communication technologies have revolutionized the capabilities of wearable medical device (WMD) to monitor patients' health remotely, raising growing concerns about trust in these devices. There is a need to quantify trust in WMD for their continued acceptance and adoption by different users. Quantifying trust in WMD poses two significant challenges due to their subjective and stochastic nature. The first challenge is identifying the factors that influence trust in WMD, and the second is developing a formal framework for precise quantification of trust while taking into account the uncertainty and variability of trust factors. This thesis proposes a methodology to quantify trust in WMD, addressing these challenges.

In this thesis, first, we devise a method to empirically validate dominant factors that influence the trustworthiness of WMD from the perspective of device users. We identified the users' awareness of trust factors reported in the literature and additional user concerns influencing their trust. These factors are stepping stones for defining the specifications and quantification of trust in WMD.

Second, we develop a probabilistic graph using Bayesian network to quantify trust in WMD. Using the Bayesian network, the stochastic nature of trust is viewed in terms of probabilities as subjective degrees of belief by a set of random variables in the domain. We define each random variable in the network by the trust factors that are identified from the literature and validated by our empirical study. We construct the trust structure as an acyclic-directed graph to represent the relationship between the variables compactly and transparently. We set the inter-node relationships, using the goal refinement technique, by refining a high-level goal of trustworthiness to lower-level goals that can be objectively implemented as measurable factors.

Third, to learn and estimate the parameters of the Bayesian network, we need access to the

probabilities of all nodes, as assuming a uniform Gaussian distribution or using values based on expert opinions may not fully represent the complexities of the factors influencing trust. We propose a data-driven approach to generate priors and estimate Bayesian parameters, in which we use data collected from WMD for all the measurable factors (nodes) to generate priors. We use non-functional requirement engineering techniques to quantify the impacts between the node relationships in the Bayesian network. We design propagation rules to aggregate the quantified relationships within the nodes of the network. This approach facilitates the computation of conditional probability distributions and enables query-based inference on any node, including the high-level trust node, based on the given evidence.

The results of this thesis are evaluated through several experimental validations. The factors influencing trust in WMD are empirically validated by an extensive survey of 187 potential users. The learnability, and generalizability of the proposed trust network are validated with a real dataset collected from three users of WMD in two conditions, performing predefined activities and performing regular daily activities. To extend the variability of conditions, we generated an extensive and representative synthetic dataset and validated the trust network accordingly. Finally, to test the practicality of our approach, we implemented a user-configurable, parameterized prototype that allows users of WMD to construct a customizable trust network and effectively compare the trustworthiness of different devices. The prototype enables the healthcare industry to adapt and adopt this method to evaluate the trustworthiness of WMD for their own specific use cases.

To my family

Acknowledgements

It is with immense pleasure and honor that I extend my heartfelt thanks to everyone who supported me throughout the development of this thesis. Although space allows mentioning only a few, each contribution has left a significant mark on my journey. First, I want to thank God for His grace and mercy that have been my foundation. It is through His enduring presence and guidance that I have managed to bring to completion this thesis.

I am grateful to my husband, Joseph, and my twins, Joe and Mervin. Their unwavering support and prayers have been the pillars of my strength. Their unconditional love and understanding during the challenging phases in this journey has been crucial to my success. I am especially thankful to my parents and siblings for their endless encouragement and prayers. I hold dear the memory of my late father, whose fervent hope was to see me complete this thesis; I am grateful of his aspirations which continue to inspire me.

I owe a tremendous debt of gratitude to my supervisor, Dr. Reza Samavi, for his innovative ideas, meticulous feedback, and passionate commitment, which were instrumental in shaping this thesis. His belief in my potential and his unwavering support through challenges were indispensable. His rigorous approach and commitment to excellence not only shaped this academic work but also profoundly influenced my growth. I would like to thank my co-supervisor Dr. Deza Antoine whose initial trust in me made it possible for me to embark on this journey. His thoughtful insights, and unwavering support were instrumental in helping me complete this thesis.

My thanks also extend to my thesis committee members, Dr. Frantisek Franek and Dr. Kai Huang, for their insightful contributions and immense support. I am grateful to Dr. Mark Chignell and Alyssa Iglar whose valuable feedback significantly enhanced the empirical study phase of my work; and Dr. Thomas Doyle who provided steadfast support for the experimental aspects of my

research.

I am grateful to all my labmates at the Trustworthy Artificial Intelligence Research Lab (TAILab) for fostering a supportive and encouraging environment. Special thanks to Hirad Daneshvar, whose exceptional assistance and thoughtful insights were pivotal throughout my journey and in shaping this thesis. I am also deeply appreciative of Dr. Omar Boursalie's support and expertise, which were vital to my success.

I would like to thank Elizabeth Pietrantonio for inspiring my path toward software engineering. I am thankful to Frosina Stojanovska for her guidance in mathematics and statistics. I am grateful to Dr. Elizabeth Martin for her encouragement and motivation. I also want to thank all my colleagues at work who showed great understanding and support during critical times. The constant prayers of my church family have been a source of strength and inspiration throughout this endeavor.

I thank my employer Mohawk College for supporting my academic pursuits. This thesis was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), and Canadian Department of National Defence: Innovation for Defence Excellence and Security (IDEaS) Program.

This journey has been a testament to the power of community, faith, and perseverance. I am deeply grateful to everyone who has been part of it.

Table of Contents

List of Figures

List of Tables

Notations, Abbreviations and Symbols

Notations

$G = (X, E)$	Bayesian Network G with X nodes and E edges
$X = \{X_1, ..., X_n\}$	Set of nodes in X
$X_{iNonDescendant}$	Non-descendant node in X
X_i	Descendant node in X
$P(X_i)$	Prior probability of node X_i
Pa_{X_i}	Set of parents of a node X_i
$P(X_i \mid Pa^G_{X_i})$	Conditional probability of X_i given the parents of X_i in BN graph G
$P(X_1 \dots X_n)$	Joint probability distribution of nodes X_1 to X_n
$q = 1 \dots m$	Number of non-discretized observed values
$S^q_i = S^1_i \dots S^m_i$	Raw samples of observations of node i for q values
D^q_i	Discrete values of S^q_i
$l_{(i,k)}$	Lower threshold for each discrete level of S^q_i
$u_{(i,k)}$	Upper thresholds for each discrete level of S^q_i
$O_i = \{min, max\}$	Set of the minimum and maximum operating user-defined threshold value

ΔH	Step size between the levels
$W^q_{Pa_{(i),j}}$	Weight coefficient for parent j of node i
T_i	Number of parents of descendant node i
$P(X_i^k)$	Probability of X_i for level k
$\boldsymbol{\theta}$	Set of parameters (conditional probabilities) of all nodes in a BN
θ_i	Conditional probability distributions of node i in a BN
$P(D_i^k[1] \dots D_i^k[m], \theta_i)$	Joint probability of k discrete valued data for node i given θ_i
$P(\theta_i \vert D_i^k[1] \dots D_i^k[m]$	Posterior probability of θ_i given the instances of D_i^k
$P(X_i^k \vert Pa_{X_i^k})$	Probability of node X_i at level k given parents of node X_i at level k

Symbols

$\mathbb{E}$	Expectation
$+$	Positive Impact
$++$	Strongly Positive Impact
$-$	Negative Impact
$--$	Strongly Negative Impact
μ	Mean
σ	Standard Deviation
σ^2	Variance

Abbreviations

AUC	Area Under the Curve
BN	Bayesian Network
CIA	Confidentiality, Integrity, and Availability
CPD	Conditional Probability Distributions
CPS	Cyber-physical system
DA	Device Accuracy
DDP	Device Data Privacy
DR	Device Reliability
DRD	Device-Related Determinants
DS	Device Security
DV	Device Validation
ED	External Determinants
EoU	Ease of Use
EU	European Union
FDA	Food and Drug Administration
FP	Functional Product
GAN	Generative Adversarial Networks
GPU	Graphics Processing Unit
GUI	Graphical User Interface

HC	Health Canada
HIPAA	Health Insurance Portability and Accountability Act
IEEE	Institute of Electrical and Electronics Engineers
IoHT	Internet of Health Things
IoMT	Internet of Medical Things
IoT	Internet of Things
ISO	International Organization for Standardization
IT	Information Technology
IUIPC	Internet Users Information and Privacy Concerns
MANET	Mobile Ad hoc Network
NFR	Non-Functional Requirement
NL	Network Loss
NQ	Network Quality
O	Operations
OD	Other Determinants
OHT	Ontario Health Team
OT	Operation Technology
PDF	Probability Distribution Function
PGM	Probabilistic Graph Model
PHR	Personal Health Records

PII	Personal Identifiable Information
R	Reliability
REC	Recommendations
RQ	Research Question
RSSI	Received Signal Strength Indication
RAM	Random Access Memory
SA	Sensor Accuracy
SDLC	Software Development Life Cycle
SQ	Sensor Quality
T	Trust
TC	Trust Concerns
TS	Technical Support
UC	Use Case
UD	User Demography
URD	User-Related Determinants
UPE	User Prior Experience
US	User Specific
UTAUT2	Unified Theory of Acceptance and Use of Technology 2
VAE	Variational Autoencoder
WMD	Wearable Medical Device

Declaration of Academic Achievement

The following is a declaration that the research described in this thesis was completed by Ms. Mini Thomas and recognizes the guidance and contribution of Dr. Reza Samavi. Mini Thomas contributed to the inception and design of the study. Mini Thomas was also responsible for the experimental testing protocols, design and development of the experiments and models, data collection, data analysis, and the writing of the manuscript. Dr. Reza Samavi contributed to the inception and design of the study and the review of the manuscript. The co-supervisor Dr. Deza Antoine, and supervisory committee members Dr. Frantisek Franek and Dr. Kai Huang have provided guidance and advice through the design and experimental phase.

Chapter 1

Introduction

In the past, health care was mainly delivered in the homes of patients by either family members or visiting doctors. Towards the end of the 19th century, the focus of healthcare delivery shifted predominantly to institutional settings like hospitals, driven by both socioeconomic changes and developments in medical science. While these advancements historically led to the institutionalization of health care, there has been a recent trend reversing back to home-based care with remote patient monitoring systems. This shift has been enabled by progress in scientific and digital technology, allowing healthcare providers to serve patients beyond traditional settings such as hospitals or clinics [17].

Advancements in sensor technology have significantly improved remote patient health monitoring, utilizing wearable medical devices (WMDs). Wearable medical devices are electronic devices worn by individuals to continuously monitor health and fitness-related metrics such as heart rate, blood glucose levels, and physical activities, particularly for chronic disease care [67]. The real-time data collected from the WMDs can be fed to autonomous medical advisory systems. Autonomous medical systems integrate intelligent sensors in devices (e.g., WMDs) and medical data sources (e.g., patient records, laboratory, or clinical records) to collect data, make decisions by applying various algorithms (e.g., machine learning to data), to provide health-related insights, and suggest preemptive actions and medical advice [111, 134]. In this thesis, our focus is on WMDs which is a subset of autonomous systems.

WMDs promise to revolutionize medicine by enabling widespread remote patient monitoring

and providing tools for mobile health (mHealth) care [145]. One key driver in the increasing adoption of WMDs has been the COVID-19 pandemic, with public health measures and restrictions significantly limiting the number of in-person interactions in healthcare settings [19]. According to Deloitte's 2023 Connectivity and Mobile Trends Survey report, the wearable market is expected to grow from US$35 billion in 2020 to nearly US$115 billion by 2028 [37]. The convergence between wearable fitness and wearable devices has evolved and WMDs are being used in various measurements, from fitness and calorie tracking to monitoring various health parameters. WMDs have been used in chronic illness monitoring [132] and for detecting mental disorders [35]. For example, mental health is highly prevalent in all countries, with about one in eight people in the world having a mental health disorder, with improvement in mental health being a sustainable development goal [149]. Depression is the leading cause of mental health and in over 50% of patients with depression, the disorder is not recognized or adequately treated [35]. WMDs are used to meet the needs in these challenging and demanding use cases to assist and provide health and mental wellness solutions [35, 149, 30]. The ability to continuously track the health status of patients is particularly advantageous for elderly individuals and those with chronic conditions who need consistent care [88]. Utilizing WMDs for remote monitoring enhances access to healthcare services, especially benefiting those who are economically disadvantaged or residing in isolated or rural areas [14]. Furthermore, remote monitoring using WMDs not only alleviates the need for frequent hospital visits and stays for patients dealing with chronic illnesses, but it also brings about considerable cost savings for healthcare systems [131].

1.1 Motivation and Thesis Objectives

Although the use of WMD is expected to grow and offer significant advantages to both patients and healthcare systems, numerous obstacles could impede its broader implementation. Trust is a critical element for the continued acceptance and adoption of WMD by patients, health professionals, and institutions. The primary goal of this research is to quantify trust in WMDs, thereby assisting various stakeholders in making informed decisions about adopting these devices. However, given the stochastic nature of trust, numerous challenges arise in achieving this goal.

The first challenge is to determine the factors that influence the trustworthiness of WMDs, a task complicated by the inherently subjective nature of trust. These factors are fundamental for defining the specifications and effectively quantifying the trust in WMDs.

Different people (stakeholders) may have differing propensities for trusting a WMD based on their past experiences with technology and levels of knowledge regarding how WMD data collection, storage, and transmission work [94, 148, 2]. WMD stakeholders, encompass of device manufacturers, device developers, regulatory policymakers, retailers, and end users (i.e., individuals who use WMDs for health-monitoring purposes). These stakeholders may have intra and intergroup-level differences in the nature of their trust in WMDs. For example, a patient may consider a device more trustworthy if it exhibits the highest privacy protection, while a manufacturer may prioritize the accuracy of the device as the indicator of the device's trustworthiness.

To describe the challenge further, consider the following motivating scenario. Assume Alex, an athlete, and his family, his trainer, and his physician (primary stakeholders) are considering different WMDs to monitor Alex's health during indoor and outdoor activities. Alex and his family's trust in the WMD may depend on the device's safety and privacy. At the same time, Alex's trainer and physician may select a WMD depending on the device's accuracy and quality of sensor measurements. Further complicating this decision-making process, they may encounter situations where two devices perform similar functions. Here, the challenge intensifies as they must discern which device is more trustworthy, adding another layer of complexity to their decision criteria and necessitating a robust method to compare and quantify trustworthiness effectively. The important question in this scenario is how these stakeholders, each with possibly different and subjective views of trustworthiness, select the best WMD for Alex.

To this end, our first objective in this thesis is to demonstrate a method to empirically validate factors that are likely to be important in evaluating the trustworthiness of WMDs. These factors can be stepping stones for specifying and quantifying trust in WMDs. In this thesis, we conduct a survey to validate the factors impacting the trust of a WMD user.

The second challenge towards achieving our goal of quantifying trust in WMDs is to establish a formal framework considering the uncertainty and subjectivity of the trust factors. Previous

research to quantify trust includes fuzzy logic systems [76] (handling vague and uncertain data), machine learning models [45] (offer robust pattern recognition and predictive capabilities), factor analysis and structural equation modeling [52] (delve into complex relationships between variables), Dempster-Shafer theory [155], and probabilistic models [142] (that manage uncertainty by integrating multiple pieces of evidence). However, these methods suffer from several limitations: they struggle with scalability, and are often non-transparent ("black-box"), or are mostly evaluated in social contexts such as a recommendation, reputation, or communication network. They also require extensive training data, which are computationally demanding, and provide limited support in transparent relationships, which restricts their effectiveness in supporting interpretable trust quantification scenarios. In models using probabilistic methods, the estimation of the parameters requires that priors for all variables be known. The selection of the prior is critical since the parameter estimation is heavily weighted by the prior distribution. Commonly, priors are generated through methods such as data simulation, typically using distributions such as Gaussian, or through expert input, which, while insightful, can risk model overfitting and introduce bias [16, 60]. The latter approach also tends to demand an impractically high number of expert-driven probabilistic estimates, leading to costly and potentially inconsistent results.

Our second objective in this thesis addresses this challenge and presents a method to quantify trust in WMD considering uncertainty and effectively managing the complexities. Using the Probabilistic Graph Model (PGM), the stochastic nature of trust is viewed in terms of probabilities, reflecting subjective degrees of belief by a set of random variables in the domain. Each random variable may define an important trust factor. For instance, trust may be refined into two key factors: the reliability of the sensors in WMD (to deliver accurate readings) and the robustness of the WMD (to perform consistently across varying conditions, such as indoor and outdoor environments). Each factor is represented as a distinct random variable having different states (e.g., binary level with low and high state). We use the Bayesian network (BN) of the PGM to represent the relationship between the variables based on new evidence in a compact way. Bayesian networks offer a transparent reasoning process, allowing stakeholders to see the impacts of different factors on trust, adapt dynamically to new data or system changes, and handle incomplete data effectively.

A BN is specified by *structure* (encoding a set of conditional independence relations between the random variables) and *parameters* (a set of local conditional probabilities for each variable given its parents in the graph).

We use a two-step process to develop a BN to quantify trust in WMD [134]. The first step of the two-step process is to construct the BN *structure* with nodes and edges, where each node in the network represents a trust factor and edges represent their relationships. The selection of these factors is informed by our empirical research alongside the concerns raised in existing literature and regulatory frameworks. The inter-node relationships are structured using the goal refinement technique from requirements engineering [156, 32] to refine a high-level goal into specific, measurable sub-goals (or tasks). The refinement process continues until the sub-goals reaches an actionable or measurable level. We borrow this technique to refine the high-level goal of trustworthiness to lower-level goals that can be objectively measured as trust factors.

The second step of the two-step process is to provide a data-driven approach to generate priors and estimate *parameters* in the BN, in which we use data collected from WMD for all the measurable factors (nodes) to generate priors. We use the Non-Functional Requirement (NFR) techniques [156, 78]. The NFRs are an integral part of the conceptual modelling process in system development, primarily focusing on defining the impact of relationships, and behaviours within a system or domain [32]. Propagation rules are formulated to consolidate these quantified relationships across the network nodes. This methodology enables the computation of conditional probability distributions (CPDs) and allows for query-based inference on any node, including the overall trust node, given the available evidence.

We would like to emphasize that our aim is to develop a customizable trust network structure for WMDs. However, to demonstrate the development process, we conducted an extensive study and identified the factors influencing trust in WMD. Although these factors may differe by device, application, and stakeholders, our approach is crucial because it shows that once these factors are identified, they can be used to create a succinct representation of the BN to quantify trust.

The findings of this thesis have been confirmed through multiple experimental validations. An in-depth survey involving 187 potential users empirically validated the factors that influence trust

in WMDs. Our data-driven, Bayesian-based network to quantify trust was rigorously validated using real and synthetic datasets. We evaluated the learnability and generalizability of our data-driven BN using real datasets gathered from three individuals using WMDs across two scenarios: performing predefined activities and engaging in regular daily activities to quantify trust. To enhance the range of conditions and incorporate different noise levels (e.g., none 0%, low 10%, medium 20%, and high 30%), we created a comprehensive and representative synthetic dataset. This dataset was used to validate the efficacy of the trust network under these varied conditions.

Furthermore, we successfully developed and implemented a user-configurable parameterized prototype to effectively compare the trustworthiness of different WMDs. This prototype features dynamic customization options, allowing users to tailor inputs such as node counts, network levels, and the intricacies of node relationships and their strengths. This capability facilitates a comprehensive comparison of the trustworthiness of various WMDs, coupled with visualizations of the results to enhance interpretability. We present a replicable framework that can be adopted for evaluating trust among different devices across the healthcare industry. This prototype enables healthcare organizations to customize and utilize this methodology for evaluating the trustworthiness of WMDs tailored to their specific use cases.

1.2 Major Contributions

The contributions of this thesis to the field of trust quantification in WMDs are as follows:

1. An empirical study grounded in a thorough literature review (Chapter 2) of factors influencing the multidimensional aspect of trust. This study addresses specific research questions and offers a series of recommendations in the domain of WMDs and remote patient monitoring (Chapter 3):

    - RQ1: What is the user's awareness of the literature-based trust factors of WMD?
    - RQ2: To what extent do the literature-based and other factors influence the adoption of WMDs?

2. The development and presentation of a hierarchical BN structure that incorporates empirically validated factors, refined into measurable factors using NFR techniques, to evaluate the relative trustworthiness of various WMDs under identical testing conditions (Chapter 4).

3. The formulation of a proof of concept, of a data-driven approach that applies NFR techniques and propagation rules to generate priors for all BN nodes and to iteratively learn the BN parameters to evaluate trust in a WMD (Chapter 5).

4. The development of a user-configurable, parameterized prototype to evaluate the trustworthiness of WMDs. This prototype offers dynamic customization options, enabling tailored user inputs providing an effective comparison of the trustworthiness of various WMDs with a relative measure of trust. The prototype not only facilitates a detailed comparison but also enhances user understanding through visual representations of the results. We have established a replicable framework that healthcare organizations can adopt and adapt to evaluate and compare the trustworthiness of WMDs specific to their operational needs (Chapter 6).

The prototype and the code for the experimental evaluation of the approach described in this thesis are publicly available on GitHub[1].

1.3 Thesis Outline

The structure of the thesis is organized as follows:

Chapter 2 reviews a broad spectrum of research relevant to trust in WMDs, divided into three literature groups:

1. A wide-ranging examination of the definition of trust from perspectives of general human beliefs, social sciences, computer science, digital technology, and WMDs. This assists in exploring the different dimensions of trust in WMD.

2. A focused exploration of factors that influence multidimensional trust, specifically related to WMDs, serves to outline trust specifications within this domain. This exploration shapes the

[1]https://github.com/tailabTMU/TrustQ_MD

design of our trust framework and lays the groundwork for empirical studies which include a detailed analysis of the trust concerns, attitudes, and behaviours of WMD users.

3. An investigation into formal methods and approaches for quantifying trust, examining the techniques employed, and identifying gaps and areas for improvement.

Chapter 3 presents an empirical study that explores trust concerns, attitudes, and behaviours among WMD users, aimed at healthcare advocates including academics, professionals, and patients who support remote health monitoring. Key aspects of this chapter include:

1. Building a trust model that delineates the subjective notion of trust in terms of several empirically measurable factors described in literature accompanied by several hypotheses that define the relationship between the concept of trust and its constituent factors.

2. Conducting a survey and carrying out in-depth investigative study on user awareness of literature-based trust factors, and the influence of these and other factors on users' trust concern and their behavioural intent of adopting WMDs.

3. Offer a set of recommendations that can assist WMD users in making informed decisions regarding trust.

Chapter 4 details the development of a BN to quantify the stochastic nature of trust. The main elements in this chapter include:

1. Proposal of a BN framework using validated factors from literature and using NFR techniques to quantify trust in WMDs.

2. Development of a method to evaluate relative trustworthiness among different WMDs under identical testing conditions using the BN framework.

3. Demonstration of the applicability and scalability of the BN framework with simulated data.

4. Estimation of relative trustworthiness and query-inferences on different factors of a WMD.

Chapter 5 describes a data-driven approach for parameter estimation within the BN framework for trust quantification. The core themes of this chapter are:

1. Presentation of a data-driven Bayesian method to quantify trust using propagation rules.

2. Proof-of-concept demonstration using real data from WMDs to update probabilities, learn parameters, estimate relative trust, and evaluate the BN.

3. Generation of datasets with augmented synthetic data, to validate the BN and evaluate trust.

Chapter 6 describes the steps taken for the development of the parameterized prototype for trust quantification. The central aspects in this chapter encompass:

1. Presentation of the complete development life cycle for a user-interfaced, parameterized prototype for quantifying trust in WMDs.

2. Implementation and integration of the different system components to bring the conceptual model to life, ensuring seamless interaction among all elements.

3. Comprehensive testing using real and synthetic data, covering various scenarios to verify the system's robustness and applicability.

Chapter 7 concludes the thesis, discussing open issues, providing recommendations, and suggesting directions for future research.

The structure of this thesis is summarized in Fig. 1.1. The complete questionnaire used for our empirical study is presented in Appendix A. Appendix B details the results of the statistical analysis of the survey, and Appendix C presents the results of the different test cases for our prototype. The work presented in Chapter 4 [134] and Chapter 5 [135, 136] has been published. These publications were made by the author of this thesis, as the lead author, in collaboration with her supervisor and faculty at McMaster University. The empirical study in Chapter 3 and the prototype development in Chapter 6 were designed and conducted specifically for this thesis and are currently being submitted for journal publications. The work presented in Sections 5.3.3 and 5.4.2 are original contributions exclusive to this thesis.

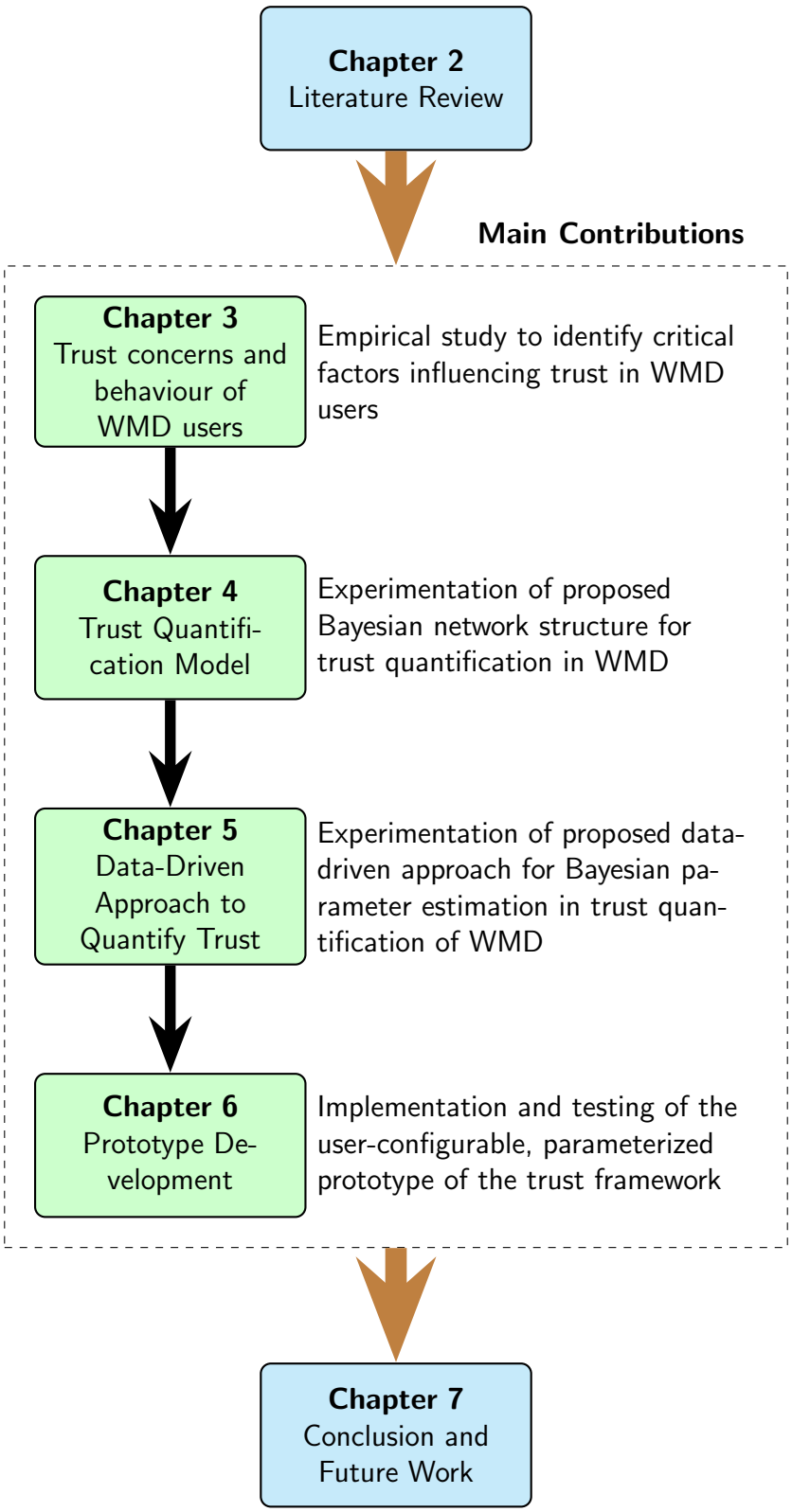

Figure 1.1: Thesis structure highlighting the flow and main contributions of the chapters, along with brief descriptions of each main chapter

Chapter 2

Literature Review

In this chapter, we begin by introducing the concepts and components of WMDs that are utilized throughout this thesis, in Section 2.1. Next, we present a review of the literature on related research in three areas: 1) definitions and perspectives of trust from different viewpoints, 2) concerns and factors influencing trust in WMD, and 3) formal methods and approaches for trust quantification. These three areas of research are examined in Sections 2.2, 2.3, and 2.4 respectively. Section 2.5 summarizes the literature review.

2.1 Wearable Medical Devices

Digital technology encompasses a broad range of electronic devices and systems that operate using digital signals and binary data. These technologies are foundational to modern computing and communication devices, facilitating the processing, storage, and manipulation of information across various platforms.

Within the scope of digital technology, the Internet of Things (IoT) represents a specific category involving devices interconnected via the internet, which enables them to collect, exchange, and process data. IoT extends digital technology's capabilities by embedding intelligence into everyday objects, enhancing their ability to communicate and function autonomously with a network of smart devices (sensors). This network transforms ordinary tools into integrated parts of a larger, digitally connected ecosystem, impacting industries from agriculture to healthcare and

beyond [63]. The IoT architecture usually consists of a network of heterogeneous sensors sensing various parameter data. The data sensed is collected through a centralized computer or gateway over the internet. The data is analyzed using machine learning algorithms and results are presented to the application layer [10].

Wearable medical devices are a niche within IoT focused specifically on health-related applications. These devices utilize digital technologies to monitor vital health metrics such as heart rate, blood glucose levels, and physical activity [67]. By collecting and transmitting data in real-time, WMDs exemplify how IoT can be specialized to improve healthcare delivery. They leverage both the foundational aspects of digital technology (like data processing and digital communication) and the connectivity framework of IoT, making them pivotal in the development of personalized healthcare solutions. Thus, WMDs fit within the digital technology landscape and illustrate the practical health applications of IoT, showcasing how digital advancements can directly contribute to improving health outcomes and patient care. Given the emerging nature of the field, our literature review considers various terms that are used interchangeably for WMDs, including wearable health devices, IoT-based wearable devices, Internet of Medical Things (IoMT), Internet of Health Things (IoHT), smart devices, digital health technologies, mHealth, telecare, and telehealth [100, 6], to broaden our search and study the factors of trust.

Wearable medical devices integrate intelligent sensors to collect data and monitor health parameters. WMDs have multiple layers such as data sensing (consists of a combination of heterogeneous sensors and devices), data preprocessing (deals with filtering and cleaning the raw data), data storage (in charge of representing the collected data in a scalable format), network layer (deals with the communication of data), and application layer (deals with decision making based on WMD readings) [34]. Hence, trust in WMDs is dependent on every layer, and it contributes to the trust of the entire device. If the system is vulnerable in any layer, it could generate inaccurate or unreliable data, thus jeopardizing the trustworthiness of the device.

2.2 Trust Definitions

Trust exhibits the factor of trustworthiness, whether in relation to humans, computers, machines, or systems. To understand the different aspects of trust, we investigated the definition of trust from literature across various contexts, including *general* human beliefs, sociology and behavioural psychology, *computer science*, *digital technology (IoT)*, and *WMD*.

The general definition, according to the Oxford Reference Dictionary, characterizes trust as a firm belief in the reliability, truth, or strength of an entity, highlighting trustworthiness as key. Meriam Webster defines trustworthiness as confidence and reliability within expected interactions. Trust, in general, is defined as reliance on integrity, ability, or character of a person or thing [84]. Three perspectives of trust are defined by Pal et al. (2019) [103]: from a social psychology viewpoint, trust is a form of faith between the trustor and the trustee that emerges from behaviours demonstrating generosity, integrity, and ability; from a personality theory viewpoint, trust is characterized as a belief deeply rooted in an individual's behaviour; and from a sociology and economics viewpoint, trust is described as a phenomenon that occurs within and between groups, organizations, and individuals. Grandison et al. (2000) [49], stated trust as the firm belief in the competence of an entity to act dependably, securely, and reliably within a specified context. In their view, an entity is deemed as trustworthy, subject to the probability that it will perform an action in a non-detrimental way. Similar opinions are expressed by Gambetta et al. (2000) [42], they defined trust as the subjective probability by which an individual, A, expects that another individual, B, performs a given action on which its welfare depends. Mui et al. (2002) [95], referred to trust based on past encounters, as reputation-based trust. From the above definitions, we see that trust is distinctively defined based on the reputation of peers and the evidence of being real to meet the expectations of the user.

In the field of computer science, particularly within a socio-economic framework, trust is essential in the applications developed by humans, as well as in the software, computer interactions, and communications that are employed to address problems and facilitate decision-making. A trusted system fulfills specified security requirements and sustains user confidence through high-quality performance [106]. A software is considered trustworthy if it has undergone thorough

development and analytical review, ensuring it functions as intended. Gligor et al. (2011) [44], discussed a framework for trust in human-computer networks, distinguishing between behavioural trust between individuals and organizations, and computational trust among devices, computers, and networks. Security principles such as simplicity, permission-based access, and transparent architecture are fundamental to designing secure and dependable systems [117].

In the domain of WMDs, trust integrates traditional device-specific factors with broader digital technology and IoT (including information and operational technology - IT and OT) considerations, forming a complex multidimensional framework. Trust considerations must include both data (software) and machine (hardware) components. The WMDs are specially designed to collect various types of physiological data with sensors. The sensor data collected through a gateway over the internet is analyzed, and results are presented to the application layer. Trust in WMDs depends on device *robustness*, including factors like interoperability and network security [36]. Furthermore, technical trust is fundamental in WMDs, focusing on the *operationality* and *functionality* of the device [6].

Security is also particularly important in WMDs given the sensitivity of health information [36]. According to a survey of enterprise customers reported in the Bain paper, security was identified as among the top three barriers, to the adoption of medical IoT applications [36]. WMDs exhibit a vast diversity in hardware, software, and communication protocols compared to IT devices, which typically operate on a limited number of standardized operating systems, hardware architectures, and protocol stacks. The WMDs are also often highly constrained in aspects such as computational power, memory, storage, communications bandwidth, and electrical power availability due to their small form factor. Hence, *it is difficult to apply over-the-shelf IT security techniques for WMD (IoT) implementations. Therefore, it is essential to consider the security triangle of confidentiality, integrity, and availability (CIA)* [106] to trust a WMD. Within the CIA triangle, confidentiality is the ability of a system to ensure that an asset is viewed only by authorized parties. Integrity is the ability of a system to ensure that an asset is modified only by authorized parties. Availability is the ability of a system to ensure that an asset can be used only by authorized parties. Ensuring CIA in WMD is very critical to avoid abuse or damage of the device.

Another important dimension of WMD is *data privacy* due to the substantial volume of data they generate from various sources, which are then transmitted to the cloud/server for storage and processing [36]. These devices produce diverse types of data, often encompassing personal information. Personal data or personal identifiable information (PII) is defined as any information relating to an identified or identifiable natural person. IoT devices can generate sensitive data such as racial or ethnic origin, health, genetic and biometric data. *Privacy in handling PII data is important to prevent misuse and potential harm.* For example, data from autonomous systems and WMDs can be misused and made to administer dangerous drug doses or electric shocks to patients [36]. Anonymization of PII is an approach that can be used to ensure that data held or disclosed is not personally identifiable.

2.2.1 Research Gaps in Trust Definitions for WMD

Given the diverse definitions of trust across different contexts, trust emerges as an inherently complex construct, influenced by multifaceted dimensions such as hardware, software, social interactions, and individual behaviors. Despite the multidimensional complexity of trust, current research often addresses these dimensions in isolation rather than holistically. This fragmented approach fails to capture the full spectrum of trust as it manifests in the context of WMDs. Consequently, a gap exists in comprehensively defining trust in WMDs that integrates social influences, user characteristics, device functionality, and operational procedures. In this thesis, we address this research gap by proposing a framework that integrates a multidimensional approach, thereby offering a more comprehensive understanding of the different dimensions of trust in WMDs in Chapter 4.

2.3 Trust Factors

The multidimensional nature of trust is influenced by a variety of factors. Trust influencing factors currently considered for WMDs are mostly based on regulatory standards [139], device design, development, and implementation specifications [119]. We conducted a literature review to study

different factors that influence trust in different dimensions.

According to the Office of the Privacy Commissioner of Canada's research report entitled "Wearable Computing - Challenges and Opportunities for Privacy Protection by the Office of the Privacy Commissioner of Canada" [108], "rapid technological innovations are fueling the development and adoption of a new generation of wearable devices." While people gain increasing benefits from using WMDs in monitoring chronic illness, post-operative care, and detection of specific conditions, there are growing concerns about the trustworthiness of WMDs. This has led to the implementation of regulatory standards and policies governing WMDs [139]. The Food and Drug Administration (FDA) defines a trustworthy device as containing hardware, software, or programmable logic "that is reasonably secure from cybersecurity intrusion and misuse; provides a reasonable level of availability, reliability, and correct operation; is reasonably suited to performing its intended functions and adheres to generally accepted security procedures" [139]. Health Canada (HC) and the European Union (EU) also have similar requirements [92, 133]. However, the resulting regulatory policies assume due diligence from the user and do not directly reflect users' trust concerns. The trustworthiness of a WMD depends on factors such as the understandability and experience of different users in dynamic and uncertain environments. The deployment of WMDs for health monitoring should not only ensure safe and accurate operations but also meet ethical standards and uphold human values.

Research focused on developing a framework for trust in the context of medical technology suggests that issues related to trust in medical devices are qualitatively distinct from those concerning general trust in technology [94]. Previous research has considered several factors that underlie trust. Adjekum et al (2018) [6], conducted a scoping review of the literature and identified enablers of (e.g., altruism, fair data access, and interoperability) and impediments to (e.g., limited accessibility, sociodemographic factors, and fear of data exploitation) trust in digital health applications. Samhale (2022) [119], examined trust factors influencing user engagement in the healthcare-centered IoT but did not explicitly consider factors related to user integration and system communication. Liu et al. (2023) [80], presented different constructs, such as performance and expert expectancy, that influence adoption and trust in mHealth technologies among older adults.

However, it is unclear how well Liu et al.'s results generalize beyond the group of older adults that they studied in Hong Kong. Wilkowska et al. (2019) [148], described four major trust factor categories related to health-oriented technologies: reliability, trustworthiness, operability, and ease. These studies considered the factors influencing trust based on the critical challenges in developing and implementing devices but without explicitly considering user attitudes and experiences.

The robust performance of the sensors in a diverse and challenging environment (e.g., continuous monitoring with physical activities) is important for WMDs [121, 44, 7]. The collected data are extremely sensitive, therefore, the access to the sensors should be secured [124, 125, 129] and the collected data should also be treated with privacy in mind [63, 27, 65]. Therefore, the main dimensions identified in the literature that a trustworthy WMD should possess include robustness, security, and privacy. Identifying the different factors of trust in WMDs is crucial for effectively stating the specification of trustworthiness.

2.3.1 Trust Specification

The trustworthiness of a WMD has a major influence on its adoption and use. Many factors can affect trust in WMD users, including various device functionality (e.g., device accuracy), social aspects (e.g., cultural influences), and personal characteristics (e.g., age or education). There are various methods for demonstrating the trustworthiness of a system, including synthesis, which constructs a system to meet predefined trust specifications; formal design, in which mathematical methods are employed to verify trust properties; monitoring/auditing during a system operation to ensure they behave as intended; and empirical testing, where the trustworthiness of a system is determined by experimental assessments [2]. In all these techniques, there is a need to provide specifications. A specification is a detailed formulation that offers a definitive description of a system for the purpose of developing or validating the system.

In software engineering, accurately specifying trust is crucial to satisfying user expectations, needs, and wants. Specifying trust is challenging as one must go beyond conventional functionality and safety aspects to meet the needs of the system users. For users to trust a system or device, it's crucial that they feel a sense of safety when using the product. Trust is not an inherently objective

construct and is contextually defined by the goals, needs, and intentions of people who work with or use systems. Considering users is especially important given that human-in-the-loop and real-time interaction is often required. Different individuals may exhibit varying levels of trust in a WMD based on their past experiences with technology and levels of knowledge regarding how WMD data collection, storage, and transmission work [94, 6, 148]. WMD stakeholders encompass device manufacturers, device developers, regulatory policymakers, retailers, and end users (i.e., individuals who use WMDs for health-monitoring purposes). These stakeholders might display differences in their trust in WMDs, both within and between groups. For instance, while a patient might value a device more highly if it demonstrates superior accuracy, a manufacturer could consider the device's data security and privacy capabilities as primary indicators of its trustworthiness.

It can be challenging to delineate system requirements when human behaviours, abilities, and preferences are highly variable. Thus, specifications for application designs typically need to be acquired interactively and iteratively, for specifying the multi-disciplinary trust [2]. Since trust is multidimensional, factors from different dimensions need to be considered to provide an understanding of and to specify trust.

2.3.2 Research Gaps in Identification of Trust Factors

Research on trust in WMDs has primarily focused on factors associated with regulatory standards and device characteristics, while the significant influence of user factors has not been emphasized. To the best of our knowledge, there is no established scale for measuring trust in WMD for the major dimensions identified including user factors. In this thesis, we address this gap, by conducting our own empirical study to investigate the trust concerns of the WMD users, i.e. the people who use these devices to monitor their own, a patients or a family member's health (reported in Chapter 3).

2.4 Trust Quantification

Next, we performed a literature review to examine the contributions to quantifying trust within specific domains. Trust has been mainly studied in communication networks or social networks. Yan et al. (2010) [153], emphasizes the trust management of a component-based software system, where different components such as service, network, and user-interface components; interact with each other to provide an application. In such component-based healthcare application systems, trust not only depends on each component's trustworthiness, but it is also important to ensure that all components cooperate well to satisfy trust requirements. The authors propose a trust management framework considering various direct factors (such as quality attributes of the trustee), and indirect factors (such as the system competence's influence on trust) affecting trust, using adaptive fuzzy logic criteria. Lee et al. (2016) [76], present a trust assessment model that considers reputation and knowledge properties. The reputation property is based on third-party information, incorporating the opinions of other entities, while knowledge is derived from first-party information provided by the trustee. A fuzzy-based algorithm for knowledge trust metrics is proposed. This work considers the social-relational aspect to provide trustworthy services.

Yan et al. (2014) [154], present a holistic view of the different properties that need to be considered for inter-layer and cross-layer IoT trust management systems, and for providing intelligent and trustworthy IoT applications/services based on the social trust relationship. Although this paper does not explicitly provide a mathematical model or quantification, it lays the foundational understanding of a complete IoT system and the associated research published on single/multi-layer trustworthy systems. Nitti et al. (2013) [101], present trustworthiness management for IoT, integrating the concepts of social networking. The authors have proposed subjective and objective approaches to compute trust using a static weighted sum approach. The subjective trustworthiness is computed based on the node's direct and indirect experience that is stored locally in each node, while the objective trustworthiness is obtained from a peer-to-peer scenario.

Chen et al. (2015) [29], present a design for adaptive trust management for social IoT systems, in which social relationships evolve dynamically among the owners of IoT devices. Trust is treated as the combination of an overall probabilistic assessment from direct interaction and the

social similarity in a recommendation system. Al- Hamadi et al. (2017) [7], propose a decision-making protocol that uses trust-based information sharing among health IoT devices, so that a collective knowledge base can be built to rate the environment and decide whether or not to visit a place/environment for health reasons. They use the probability method to calculate trust as aggregated user ratings. Saied et al. (2013) [116], presents a context-aware trust management system, to manage trust in a heterogeneous IoT architecture considering different resource constraints and capabilities to establish trust. The authors propose a dynamic weighted sum approach to find trust scores on cooperating nodes based on the interactions, to build confidence.

Govindrajan et al. (2011) [48] and Sahoo et al. (2016) [115], in their works demonstrate a weighted average approach for quality of service such as forwarding data packets or routing protocols in a communication network. Yu et al. (2008) [155], provides a trustworthy model for agents in large and open networks, collaborating with other agents to identify those whose past behaviour has been untrustworthy. Trustworthiness is evaluated by combining the local evidence based on direct prior interactions. The authors use Dempster- Shafer evidence theory to model credibility within the framework.

Shin Minho (2012) [124], provides a practical security risk checklist for sensor data quality considering issues such as patients not applying sensors correctly, the device being stolen or compromised, and falsified data due to the compromised device. This paper provides the hardware sensor challenges, however, it does not take into consideration a formalized trust model considering the multi-dimension aspect of trust, which is the focus of our thesis. Han et al. (2014) [53], present a detailed survey on various trust models for wireless sensor networks (WSN), to detect unexpected node behaviours. This work analyzes various applications of trust models, including secure data aggregation, secure localization, secure node selection, and the calculation of direct and indirect trust. Though a concise mathematical approach is not presented in this survey, it helps in understanding trust computation using the direct and indirect approach, and it also sheds light on trust enhancement considering factors such as energy and bandwidth, which are also factors of consideration for a WMD. Jayasinghe et al. (2017) [65], suggest a hybrid trust framework for evaluating data trust and entity trust with attributes from the knowledge and experience domain

based on social inclusion. The weighted sum of social and non-social factors is considered for trust metric quantification.

We also investigated literature to study frameworks considering uncertainty. Bayesian networks have been employed for inference and prediction in diverse applications. These applications include cyber-physical systems (CPS) [142], and mobile ad hoc networks (MANET) [160], medical imaging [40, 60], and industrial applications [98, 87, 85]. Bao et al. (2013) [15], propose a design to evaluate a scalable and adaptive trust management protocol for dynamic IoT environments. They use Bayesian inference for the reputation system, where each node calculates the trust using Bayesian estimation over historical observations. An iterative approach is taken to aggregate new direct observations with past information.

Zouridaki et al. (2005) [160], propose a trust establishment scheme for MANET, which aims to improve the reliability of packet forwarding over multi-hop routes in the presence of potentially malicious nodes. Using a Bayesian framework, each node assigns a "trustworthiness" value to each of its neighbour nodes based on direct observations of packet forwarding behaviour, thus forming an opinion about each other for trust establishment. Wang (2020) [142], demonstrate Bayesian inference for the trustworthiness of CPS with a mix of subjective and objective statistics probability.

Karakostas (2016) [66], propose an architecture that employs a Bayesian event prediction model for the dynamic IoT environment with uncertainty to predict flight delays using the historical event data generated by the IoT cloud. In [13, 3], the authors propose Bayesian inference for online social networks to limit the bias on the edges of the nodes. Maruyama et al. (2021) [85], propose a Bayesian framework to quantify the epistemic uncertainty associated with the closure model in relevant aircraft configurations with computational fluid dynamics. In these works, the framework is tailored towards a specific use case with a limited amount of data to quantify sources of uncertainty using a fully Bayesian approach.

Wang and Vassileva (2003) [143], demonstrate a BN-based trust model for a file sharing peer-to-peer application based on historical files (e.g. music) to select file providers that match peers' preferences. Flouri et al. (2019) [40], use a data-driven approach (with motion data) for Bayesian

parameter estimation to incorporate motion correction in the placenta during gestation. Holden et al. [60], propose a Bayesian inference in medical imaging run with priors created from knowledge based on the publicly available historical datasets and encoded by a neural network. Meng et al. (2022) [87], propose a data-driven BN model integrating a mix of historical and collected data for structure and parameter learning of the BN, to prioritize the risk-influencing factors in an oil well. In these works, historical data is used to estimate the priors of the BN. However, domain-specific historical data may be limited and not always readily available.

Table 2.1: Summary of research work on trust quantification methods used to evaluate trust in IoT-based systems in diverse applications including healthcare, MANET, CPSs, and WSN

Work	**Method**	**System**
Yan et al. [153]	Adaptive trust control model with fuzzy logic criteria	Trust management for a component-based system influencing quality attributes
Lee et al. [76]	Fuzzy-based algorithm for knowledge trust metric	Trust evaluation of social entities of IoT services
Nitti et al. [101]	Multi-trust with static weighted sum	Trust management in social IoT influencing quality of service
Chen et al. [29]	Probabilistic model	Trust evaluated from direct interaction and social recommendation system
Hamadi et al. [7]	Probability method to evaluate trust	Trust calculated as aggregated user ratings
Saied et al. [116]	Dynamic weighted sum	Trust management in multi-service event-driven environment

Continued on next page

Table 2.1 continued from previous page

Work	Method	System
Govindrajan et al. [48]; Sahoo et al. [115]	Weighted average approach	Trust modeled for a communication network.
Yu et al. [155]	Dempster Shafer evidence theory	Trust modeled as belief about information reliability
Jayasinghe et al. [65]	Weighted sum of social and non-social trust metrics	Data quality trust evaluation framework for IoT
Bao et al. [15]	Bayesian inference to aggregate trust	Reputation of the system is calculated by considering trust of each node using Bayesian estimation over historical observation
Zouridaki et al. [160]	Bayesian network	Trust establishment in MANET utilizes an opinion-based scheme, observing packet forwarding behaviour
Yan Wang [142]	Bayesian probabilistic trust model	Ability- benevolence and integrity based trust metric from direct and indirect observations for CPS
Karakostas et al. [66]	Bayesian inference model	Model to predict flight delays using the historical data

Continued on next page

Table 2.1 continued from previous page

Work	Method	System
Maruyama et al. [85]	Bayesian network	Framework to quantify the epistemic uncertainty for a specific use case (limited data)
Wang and Vassileva [143]	Bayesian-based trust model	Model to study peer music preferences based on historical files
Flouri et al.[40]	Data-driven approach for Bayesian network	Motion correction in the placenta
Holden et al. [60]	Bayesian inference in medical imaging	Priors created from knowledge based on publicly available historical data and encoded by a neural network
Meng et al. [87]	Data-driven Bayesian network	Historical data to prioritize risk-influencing factors in an oil well

2.4.1 Research Gaps in Trust Quantification Models

The scholarly contributions discussed above and listed in Table 2.1, towards quantifying trust in specific domains, particularly within communication and social networks, reveal significant research gaps in the context of trust in WMD. First, though WMDs are inherently socially oriented by relationships, we take the view that a valid trust model must consider factors related to the different layers (data sensing, collection, and communication) of the WMD in addition to the social, security, and privacy aspects. From our literature review, we have identified a lack of detailed information on the formal quantitative approaches used for trust computation and aggregation, specifically focusing on the sensing layer of WMDs.

Secondly, current research work lacks consideration to quantify trust in WMD using a data-driven probabilistic Bayesian approach. Probabilistic models, particularly PGMs, offer an effective approach for quantifying metrics that involve inherent uncertainties, such as trust. Given the subjective and uncertain nature of trust, employing PGMs, such as BN, enables the integration of uncertainty into the quantification process. Bayesian networks have been applied across other domains to address the challenges of trust quantification, as highlighted in the literature [142, 66, 85]. However, in our knowledge there is no current research that uses BN to quantify the subjective nature of trust in WMD in a multidimensional context. Also, the selection of prior in BN is critical, since the parameter estimation is heavily weighted by the prior distribution. However, the parameter estimations in the literature are mostly based on simulated data or a mix of expert knowledge and historical datasets, which may not be a true representation of the priors. There may not be a correct way to translate subjective prior beliefs into a mathematically formulated prior for inferencing. Analytic priors do not usually accurately describe the probability distribution [60, 85]. Treating priors as a regularization term dictated by the variational bound for estimating posteriors may not guarantee convergence [16]. Gaussian (or other standard distributions) with assumed parameters can lead to over-confidence resulting in over-fitting. Additionally, when experts provide the distribution for the prior it may require an exponential number of feedback based on the complexity of the BN structure, which can be costly [142, 160]. The gaps are summarized as follows:

1. In the literature trust is predominantly evaluated in social contexts, such as recommendations, reputation, or communication networks. A multidimensional approach considering factors of different layers of WMD lacks consideration.

2. Trust evaluation models are either complex or have poor weighted sum approximations with arbitrary weights based on subjectivity. These models are mostly based on simulation methods. Also, they have little consideration of the various challenges, limitations, and uncertain (noisy) factors of the different layers in WMD.

3. In models using probabilistic approaches (e.g., BNs), the selection of prior is either not clear or assumed to be uniform.

In this thesis, we address these gaps by presenting a data-driven BN to quantify trust in Chapter 5.

2.5 Summary

In this chapter, an introduction to WMDs was presented in Section 2.1. We then presented the literature review on definitions and perspectives of trust from different viewpoints (Section 2.2), factors influencing the trust of WMD (Section 2.3), and formal methods and approaches for trust quantification (Section 2.4). Based on the review, three important challenges were identified that became the focus of the thesis:

1. Currently, the factors influencing trust in WMDs are largely derived from regulatory standards and literature on the specifications related to device design, development, and implementation. However, trust is a subjective concept which is also shaped by the specific goals, needs, and intentions of those interacting with or using the system. The influence of the multidimensional construct of trust on WMD users' concerns and behaviours has not been investigated. This gap is the focus of our Chapter 3.

2. Bayesian network has been used in inference and prediction in applications such as medical imaging [40, 60], industrial applications [98, 87, 85], CPSs [142], MANETs [160], IoT [66], and social networks [13, 3]. However, a compact BN structure to quantify trust in WMDs (considering the hardware and software challenges) has not been presented. This gap is the focus of Chapter 4.

3. Bayesian networks presented in the literature are mostly based on simulated data or a mix of expert knowledge and historical datasets, which may not be a true representation of the priors. A data-driven approach for the estimation of priors and parameters using a BN to quantify trust in a WMD has not been investigated. This gap is the focus of Chapter 5.

In the next chapter, we will explore the multidimensional construct of trust, to identify the factors that influence user selection and adoption of WMDs.

Chapter 3

Trust Concerns and Behaviour of WMD Users

In this chapter, we report on an empirical study we have conducted to understand trust concerns, attitudes, and behaviour of WMD users. The primary objective of this study is to empirically validate factors that are likely to be important in evaluating the trustworthiness of WMDs from a users' perspective which can serve as foundational components for establishing trust specifications. This objective will be achieved by pursuing two goals. The first goal is to evaluate the user perspectives and awareness of factors that influence trust in WMDs. The factors considered are sourced from existing literature on remote patient health monitoring, cybersecurity, and human factors. Our second goal is to assess the impact of these and other factors on user trust, adoption, and use of WMDs.

We present the research model and the rationale for the development of the hypotheses in Section 3.1. Section 3.2 describes the research methodology. The results of the study are reported in Section 3.3. Section 3.4 reports the evaluation of the hypotheses and discusses the implications of this study, and Section 3.5 summarizes the study.

3.1 Research Model and Hypotheses Development

This section formulates the research questions and illustrates the conceptual model for the empirical research. A number of hypotheses regarding users trust concerns are developed followed by the additional hypotheses developed to understand WMD users trust perception and behaviour.

3.1.1 Research Questions

To address the gaps in the literature described in Section 2.3, this study seeks to provide empirical evidence to answer the following research questions:

RQ1: What is the user's awareness of the literature-based trust factors of WMDs?

RQ2: To what relative extent do these and other factors impact the trust concern and influence the adoption of WMDs?

3.1.2 Model Design

To design our research model, we selected three groups of literature with the common denominator that aim to model and/or measure subjective concerns and behaviour of users (e.g. privacy and trust) when encountering new technologies. The first group includes literature related to the functional product (FP) theory [51, 20], which refers to sustainable business models that focus on providing full life cycle functionality and performance through integrated hardware, software, and service support systems. We relate the FP theory to our work to study the user trust concerns in WMD because of the similarity between the two in addressing factors related to product characteristics. Literature [120, 20] shows that FP theory is extensively used in product development and performance optimization. As WMDs encompass hardware, software, and services, we use the FP theory in our investigation to examine how product characteristics influence user trust and intention to use WMDs. We identified seven device-related factors concerning functional and operational attributes, such as device accuracy, reliability, validation, security, privacy, technical support, and ease of use of the device, to assess their impact on user trust and use. We grouped

them as device-related determinants (DRD).

The second group of literature includes Xu et al. (2008) [151] and Smith et al. (2011) [126], where six dimensions influencing an individual's privacy concerns are introduced: (1) privacy experiences, (2) privacy awareness, (3) personality differences, (4) sensitivity of service or information involved, (5) trust in the service, and (6) privacy regulations governing data sharing practices. The focus group in these studies was general internet users, and they used the general measure of the Internet Users Information and Privacy Concerns (IUIPC) [83]. We relate their work with our focus group of WMD users to identify dimensions influencing trust concerns, based on the literature regarding privacy concerns of general internet users [126], and surveys on privacy concerns in personal health records (PHR) platforms [118]. We apply the methodology used to investigate user privacy concerns in PHR [118] to our study in examining user trust concerns in WMDs, as privacy and trust share similarities in subjectivity and are user-centric concerns. PHR is an electronic platform individuals use to maintain and manage their health information in a private, secure, and accessible manner. Samavi et al. (2014) [118] conducted an empirical study to determine which aspects of the underlying factors considered for general internet users [126, 152] were consistent with the privacy concerns of PHR users. Based on our literature review for trust, we used the user privacy factors for PHR [118] and identified trust factors such as user prior experience and user demographic differences for our study with WMDs. We grouped them as user-related determinants (URD).

The third group includes the extended Unified Theory of Acceptance and Use of Technology model (UTAUT2) [130], which is used to determine user acceptance and usage patterns of any innovative and complex technology. The model describes five constructs to determine the behavioural intention of consumers in adopting complex technology: (1) performance expectancy, (2) effort expectancy, (3) social influence, (4) facilitating conditions, and (5) consumer habit [31]. Literature shows that the UTAUT2 model is used to study users' behaviour towards wearable health technology, considering social factors such as interaction and consumer emotional aspects influenced by symbolic or emotional value or human belief [130, 141]. We relate the UTAUT2 model to our work to study social interacting factors, such as the recommendation of a WMD by

others, as an external determinant (ED), and the users' emotional aspect and human beliefs (such as user-specific belief in trusting the device), as other determinants (OD).

Digital divide, factors characterized by access to technology, internet, and digital literacy also significantly impact the trust and adoption of WMDs, especially among general population groups. These factors can lead to disparities in how different communities engage with WMDs, influencing the perceived reliability, usability, and overall trustworthiness of these technologies. Additionally, social determinants of health, cultural, and environmental factors also influence trust in WMDs. Understanding these broader determinants will also help in the development of more inclusive models that better capture the diverse experiences and needs of WMD users [112, 140]. While the inclusion of these factors is vital, they were not added within the scope of our current work. Our study prioritized determinants directly associated with device functionality, user experience, and social interactions, based on existing literature models. However, societal and environmental aspects, present valuable directions for future research.

Figure 4.1 shows our research model, which conceptualizes the relationships between different antecedents of WMD trust concerns (TC) and their influence on users' decision-making to trust and adopt a WMD. In this research model, we aim to understand the underlying factors that affect the TC in WMDs. These factors are shown in the model as incoming arrows to TC. The TC is in the center of this model. We treat TC as a dependent variable. To answer our first research question (RQ1), we investigated the WMD users' awareness of the different trust factors identified from regulatory standards and literature. Factors were selected based on reports for wearable devices published by regulatory bodies such as FDA [138, 139], HC [54, 92], and the EU [133], and current research on trust concerns relating to hardware [30, 6, 80], software [125, 27, 63], and communication [8, 103]. To answer the second research question (RQ2), we investigated the relative influence of different factors in the literature and other user-specific factors on the user's concerns about the trust, adoption, and use of a WMD [94].

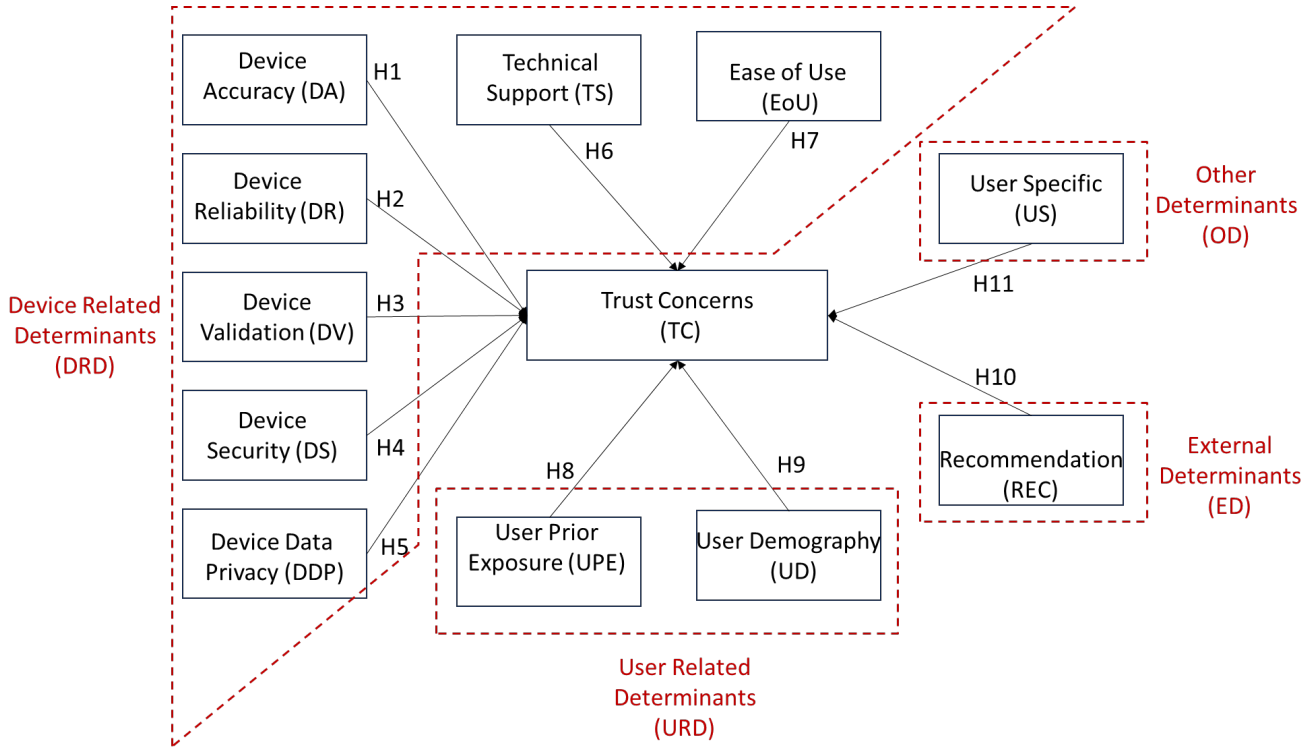

Figure 3.1: Proposed research model incorporating eleven hypotheses to study the relationship among different factors of WMD user trust concerns (TC), user awareness, and their implications on WMD adoption

3.1.3 Factors Moderating Trust Concerns

After an extensive investigation of the existing literature reporting factors moderating user TC in WMDs and previous surveys that address trust issues in WMDs, the following factors are selected as constructs for further study as potential drivers of the concerns about trust of WMD users and the development of hypotheses.

Device Accuracy (DA) refers to the degree to which an individual believes in the accuracy of the WMD sensors for making measurements of their vital bio-signals. Research suggests that trust is influenced by user awareness of the precision and accuracy of the measurements made by a device [6, 150, 124].

H1: Device accuracy is positively correlated with an individual's behavioural intention to trust and use a WMD.

Device Reliability (DR) refers to the degree to which an individual can depend on consistent and stable performance, communication, and operation as intended without frequent malfunctions or errors [80, 8, 103, 12, 146].

H2: Device reliability is positively correlated with an individual's behavioural intention to trust and use a WMD.

Device Validation (DV) refers to the degree to which an individual trusts in the WMD based on clinical validation and regulatory standard approvals. Previous research shows that device validation leads users to trust the effectiveness and safety of the device [30, 139, 54, 133, 62, 41].

H3: Device validation is positively correlated with an individual's behavioural intention to trust and use a WMD.

Device Security (DS) refers to the degree to which an individual believes in the security measures (such as encryption, authentication, and access control) that protect the WMD and its associated data from unauthorized access, attacks, or breaches. Device security supplements the factors noted above in helping users trust the safe and secure operation of the device [125, 7, 63, 89].

H4: Device security is positively correlated with an individual's behavioural intention to trust and use a WMD.

Device Data Privacy (DDP) refers to the degree to which an individual trusts a WMD based on whether the manufacturer uses privacy measures and provides a transparent outline of how the user's personal data is collected and used [6, 27].

H5: Device data privacy is positively associated with an individual's behavioural intention to trust and use a WMD.

Technical Support (TS) refers to the degree to which an individual believes that responsive and reliable support is available from the manufacturer to resolve issues with the device [80, 130].

H6: The amount of technical support is positively associated with an individual's behavioural

intention to trust and use a WMD.

Ease of Use (EoU) refers to the degree to which user-friendly interfaces help the user to intuitively and efficiently interact with the devices with a minimal learning curve [132, 6].
H7: Ease of use is positively associated with an individual's behavioural intention to trust and use a WMD.

User Prior Experience (UPE) refers to the degree to which an individual's trust in a WMD is influenced by previous experience with the device [37, 132].
H8: People have different user experiences (positive or negative), partly stemming from the different roles that people have concerning the WMD (e.g., as the patient who uses the device or the doctor who uses the device data in making clinical decisions), which will affect their trust and intention to use a WMD. For example, a WMD user who has had a positive experience with continuously monitoring glucose levels over an extended period of time will tend to trust and adopt the WMD.

User Demography (UD) refers to the degree to which an individual's trust in a WMD is influenced by demographic factors such as age, gender, education, and culture [6, 103, 62].
H9: People with different user demographics have varied views on WMDs which will impact their trustworthiness and their intention to use a WMD. For example, older (and more technologically challenged groups of people) may have poorly calibrated trust in WMDs (either trusting too little or too much because of ignorance about how the technology works).

Recommendations (REC) refers to the degree to which an individual's trust in a WMD is influenced by external endorsement or advice from healthcare professionals or acquaintances who have experience with similar devices [6, 77].
H10: Recommendation is positively correlated with an individual's trust and use of a WMD.

User Specific (US) refers to the factors that are important from the user's perspective (which may not be the factors identified in the literature) to trust and adopt a WMD. Like the Subject Factor in an analysis of variance, the US can be considered as a trust-related factor at the individual level or as part of the error term in looking at aggregate behaviour across a group of individuals. This construct mainly focuses on the user's internal belief system [6, 103, 25].
H11: User-specific factor is positively associated with an individual's behavioural intention to adopt and use the WMD.

In Fig. 4.1, the factors listed above are shown as arrows linking the determinants to the user's TC. Trust concerns is the dependent variable affected by the eleven different factors, each acting as an independent variable. We carried out the survey reported below to measure TC with the trust-inducing factors so that the hypothesized relationships could be empirically investigated. Our goal is to interpret the results in terms of a model [135] to quantify trust in WMD.

3.2 Research Methodology

The research was conducted as an online survey. In this section we first report on survey item development followed by explaining how the survey was administered to collect data.

3.2.1 Survey Item Development

We evaluate WMD users'TC and behaviour using a self-reported scale and multiple-answer questions. We developed the questionnaire after a review of existing surveys relevant to the topic. Boenisch et al. (2021) [18], studied machine learning practitioners' awareness of and practices related to security and privacy using an online survey to identify influencing factors. Samavi et al. (2014) [118], also sought to understand privacy through an empirical study on privacy concerns, attitudes, and behaviours held and exhibited by PHR users. In a similar study, Montague et al. (2009) [94], assessed how patients and care providers make decisions about the trustworthiness of mutually used medical technology in an obstetric work system. Grosse et al. (2022) [50], analyzed

the answers to a survey of industrial practitioners about attack occurrence and concern.

In this research, we avoid designing yet another new scale to measure individual's concerns on trust, but instead we aim to study which aspects of the underlying factors described in the literature are consistent with the user's trust in the WMD context. We believe that to have a safe and trustworthy experience, users need to be aware of the different trust-impacting factors addressed in the literature and regulatory bodies. The questions in our survey are designed to assess participants' awareness of and perspective on multiple aspects of trust and subsequent behaviour. The questionnaire is structured to consist of eleven constructs as described above, each corresponding to one of the hypotheses and manifested by one or more questions.

3.2.2 Questionnaire Design

To the best of our knowledge, there is no established scale for measuring trust in WMD for the determinants we have identified in the "Factors Moderating Trust Concerns" section (3.1.3). In order to develop our scales and frame our questions for each construct, we examined several existing works on the concerns of trust and mistrust of WMDs. As shown in Fig. 4.1, we designed the scale for 11 hypotheses (DA, DR, DV, DS, DDP, TS, EoU, UPE, UD, REC, and US) which are grouped into four determinants (DRD, URD, ED, and OD). In this section, we describe the development of questions for each hypothesis.

Device Accuracy: We studied the literature for sensor accuracy and quality concerns in WMDs. Adjekum et al. (2018) [6] and Xie et al. (2018) [150] report accuracy as a notable concern of WMD users and its major impact on health promotion. Shin (2012) [124] reports sensor accuracy as an important factor in developing a secure health monitoring system with wearable sensors. Based on the above literature, we developed three questions (Q10 – Q12) to study the user perspective of device accuracy in a WMD. Q10 is intended to assess the influence of accuracy on users' trust in using a WMD, and Q11 is intended to assess users' awareness of WMD sensor accuracy. Q12 is developed as a cross-validation question for Q11.

Device Reliability: We investigated the literature concerning stable and consistent communication and operation in WMDs. Liu et al. (2023) [80] report that users have a preference for reliable technology for WMDs. Sarawi et al. (2017) [8] present reliability as among the potential risks that users foresee while using WMDs. Pal et al. (2019) [103] and Sarawi et al. (2017) [8] present the trade-off between high-speed communication, high power consumption, and memory storage volume for resource-constrained WMDs. Austen (2015) [12] and Wen et al. (2017) [146] report how the increase in the growth of WMDs may result in issues with network communications and stable operation. Drawing from the literature mentioned above, we formulate five questions (Q13 – Q17) to investigate users' perspectives on reliability in a WMD. Q13 is intended to assess the influence of reliability on users' trust and use. Q14 is intended to assess users' awareness of reliability issues such as connectivity, battery life, and memory storage, and Q15 - Q17 are developed as cross-validation questions for Q14 to understand user's awareness of the three reliability aspects (connectivity, battery life, and memory storage).

Device Validation: We studied the literature for concerns about clinical validation and regulatory standard approvals for WMDs. Chib et al. (2023) [30] and Izmailova et al. (2018) [62] report the validation of WMDs as an important issue due to the increasing availability of wearable technologies and solutions. The authors recommend a set of guidelines to design specific protocols and regulatory frameworks for the validation of wearables, which is also supported by regulatory bodies [139, 54, 133]. Fusca et al. (2018) [41] show that clinical-grade wearables provide a valid measurement of health parameters compared to devices that are not clinically validated. Izmailova et al. also report the need for methodological and logistical considerations for implementation in clinical trials, including key elements of analytical and clinical validation in the specific context of use, to facilitate trust and adoption of WMDs. Considering the increasing growth of wearables and their relative impacts as described above, we develop three questions (Q18 – Q20) to understand how users perceive a WMD's clinical validation and regulatory protocols. Q18 is intended to assess the influence of clinical validation and regulatory standards on users' trust. Q20 is intended to assess users' awareness of device validation, and Q19 is developed as a cross-validation question

for Q20 to understand users' confidence in their knowledge of validation methods used by WMD manufacturers.

Device Security: We also investigated relevant literature for security concerns such as unauthorized access to the device, cybersecurity attacks, and potential data breaches in WMDs and the impact of the measures on users. Sicari et al. (2015) [125] and Mills et al. (2016) [89] present security issues such as data encryption, authentication, and access control within the IoT network as the main challenges in trusting WMDs. Jaigirdar et al. (2019) [63] emphasize physicians' concerns about the trustworthiness of the medical data using WMDs and show the impact of security requirements in each layer of the WMD. Sarawi et al. (2017) [8] present the trade-off between network communication protocols (encryption and authentication mechanisms) and power consumption for resource-constrained WMDs. Given that security is presented as a challenging factor in trusting WMDs, we develop three questions (Q21 – Q23) to examine user perspectives regarding security. Q21 is intended to assess the influence of security factors on users' trust and use of a WMD. Q23 is intended to assess users' awareness of device security. Q22 is used to cross-validate (Q23) participants' awareness of security breaches.

Device Data Privacy: We studied the literature for privacy measures such as compliance with privacy standards for data collection, usage, and storage transparency. Adjekum et al. (2018) [6] present the considerations about privacy and data protection, highlighting the ethical challenges that directly influence the trustworthiness of WMDs. Chen et al. (2016) [27] present the challenges of storing user's sensitive data in the cloud and the need to implement privacy measures for a trustworthy system. Based on the literature discussed, we designed three questions (Q24 – Q26) to understand user's perspectives on privacy in a WMD. Q24 is intended to assess the influence of data privacy measures on users' trust and use of a WMD. Q25 is intended to assess users' awareness of privacy practices used in WMDs, and Q26 is developed as a cross-validation question for Q25 to understand users' confidence in their knowledge of privacy methods used for preserving sensitive user data.

Technical Support: We studied the literature for the impact of responsive and reliable support on the trust of WMDs. Liu et al. (2023) [80] report how elderly and stroke patients were willing to trust and use if they received technical support. Talukder et al. (2020) [130] present the use of facilitating conditions, such as technical support, to improve trust in WMD. We developed three questions (Q27 – Q29) to study if the availability of good technical support impacted users in trusting and using a WMD. Q27 is intended to assess the influence of responsive technical support on users' trust and use. Q28 – Q29 are used to assess users' awareness of support in trusting WMD.

Ease of Use: In studying the literature on the impact of usability in trusting WMDs, we found Adjekum et al. (2018) [6] report ease of use as the enabler of trust in their scoping review. Thapa et al. (2023) [132] describe the propensity for systems to require minimal effort for use as an influencing trust factor. We developed two questions (Q30 – Q31) to assess the influence of ease of use on users' trust.

User Prior Experience: We investigated the literature on the influence of prior user experiences on trusting and using WMDs. Arkenberg et al. (2021) [37] report in their study how users with prior experiences in WMDs are proponents of trusting and adopting the device for self-health monitoring and tracking. Thapa et al. (2023) [132] report that experience theory is prevalent in finding users' perceptions of trusting WMDs. Samavi et al. (2014) [118] proposes that an individual's experience with privacy breaches may influence their privacy concern when using PHR technologies. We hypothesized that users' prior experiences (positive or negative) will directly impact their trust in WMD. We developed four questions (Q1 – Q4) to study the impact of user experiences in trusting a WMD. Q1 is intended to assess for what application they had used a WMD (e.g., self-monitoring, or monitoring patients). Q2 asks which device they had experience using (e.g., fitness trackers, continuous glucose monitors, etc.), Q3 asks for how long they had experience using WMD, and Q4 asks if their experiences were positive or negative. With these four questions, we expect to explore how a user's experience can impact their trust in using a WMD.

User Demography: We were also interested in studying the influence of users' demographic factors such as age, gender, ethnicity, and education on trusting and using WMDs. Pal et al. (2019) [103] have shown how individual characteristics such as age and gender can impact the trust and adoption of WMDs. Izmailova et al. (2018) [62] show that the influence on trust is population-based and dependent on factors such as age, gender, and other demographic factors. Adjekum et al. (2018) [6] show that demographic factors influence trust in WMDs. We developed five questions to assess the influence of demographic factors in trusting WMDs (Q5 – Q9). Q5 is a gender question, Q6 gathers information of ethnic origin, Q7 asks the participant's age group, Q8 asks about their highest level of education, and Q9 asks about their professional background. With participant data on these five questions, we can study if demographic factors influence trust and the use of a WMD.

Recommendation: We studied the literature on the influence of social factors, such as external recommendations on trusting and using WMDs. Lee et al. (2020) [77] and Adjekum et al. (2018) [6] report that an individual's social characteristics are influenced by others. For example, if an influencer recommends the use of a WMD, people are more likely to trust and use the device. We developed four questions (Q32 – Q35) to assess the user's perspectives on external recommendations. Q32 asks if the participants had received recommendations from anyone. Q32 – Q33 are follow-up questions asking who recommended them to use a WMD and to what extent the recommendations influenced their behaviour. Q35 asks if they would recommend a WMD to anyone else.

User Specific: We investigated the literature for factors related to human belief systems and trust in WMD. Chang et al. (2020) [25] report that the user's internal belief system that a technical infrastructure exists to support the WMD is an enabler to trust and use the device. Adjekum et al. (2018) [6] report that perceived behavioural control is shown to be related to appropriate sets of salient behavioural, normative, and control beliefs about the behaviour, but the exact nature of these relations is still uncertain. Pal et al. (2019) [103] report that personal innovativeness is

a moderator in predicting trust and, thereby, the adoption of WMDs. The authors propose that it may be worthwhile to study personal innovativeness because, with any new technology, there can be a variety of uncertainties associated with it. We developed two questions (Q36 and Q37) for this construct to assess if there are any behavioural differences in terms of innovativeness as a personal trait or other internal human belief systems impacting trust and, consequently, the use of WMDs.

The questionnaire's contents were subsequently refined and validated with the scales introduced in the After Scenario Questionnaire [79] , the user privacy concern on the PHR survey questionnaire [118], and the mobile phone usability questionnaire [114]. We then formulated our questions for each of the constructs to assess trust in WMDs. The questionnaire was reviewed by experts, including five graduate students and four faculty professors (Medicine, Health Science, Biomedical, and Electrical and Computer Engineering programs). The questionnaire eventually had 37 questions, consisting of a mix of different question types, including (a) Five-point Likert scale questions to understand the awareness and influence of trust factors from the user's perspective, (1- Not at all influential, 2- Somewhat un-influential, 3-Neutral, 4- Somewhat influential, 5-Very influential), and, (b) Multiple- answer questions to capture the impact of different factors on users level of trust in WMD. We also included one open-ended question to capture additional factors influencing users' belief in trust in WMDs. Table 3.1 summarizes the constructs and the corresponding questions. The cross-validation questions are shown in bold in Table 3.1. The complete questionnaire is provided in Appendix A.

Table 3.1: Hypotheses, determinants of trust in WMDs, and corresponding survey items. (Boldface question numbers are cross-validation questions)

Hypothesis	Survey Items	References
Device-Related Determinants		
H1 - Device Accuracy (DA)	Q10: To what extent does the accuracy of the wearable medical device influence your trust in using the device? **Q11:** Have you ever cross-checked the readings of a wearable medical device with another device to verify and be aware of its accuracy? **Q12:** How do you typically verify the accuracy of a wearable (e.g., compare the wearable measurements with other devices)?	[6, 150, 124]
H2 - Device Reliability (DR)	Q13: To what extent does the reliability of the wearable medical device influence your trust in using the device? **Q14:** Have you been aware of or ever experienced any issues with the reliability of a wearable (e.g., reliable operation in terms of battery life, memory storage, and connectivity)?	[80, 8, 103, 12, 146]

Continued on next page

Table 3.1 continued from previous page

Hypothesis	Survey Items	References
	Q15: Which technical challenges related to connectivity issues have you encountered?	
	Q16: Has battery consumption been a problem in your experience with wearables?	
	Q17: How often have you lost data from your wearable device due to memory overload?	
H3 - Device Validation (DV)		
	Q18: To what extent does knowing that your wearable medical device has been clinically validated influence your trust in using it?	[30, 62, 139, 54, 133, 41]
	Q19: How confident are you that the wearables available in the market undergo sufficient clinical validation to ensure the device measurement effectiveness and safety?	
	Q20: Are you aware of the validation methods used by the manufacturer of your wearable medical device?	
H4 - Device Security (DS)		
	Q21: Does the presence of robust security features influence your trust in using your wearable medical device?	[125, 89, 63, 8]

Continued on next page

Table 3.1 continued from previous page

Hypothesis	Survey Items	References
	Q22: What factors are important to evaluate security in wearables?	
	Q23: Are you aware of security breaches in wearable medical devices?	
H5 - Device Data Privacy (DDP)		
	Q24: To what extent does the awareness or assurance of strong privacy practices influence your trust in using wearable medical devices?	[6, 27]
	Q25: Are you aware of privacy practices implemented by wearable medical device developers to protect user data?	
	Q26: How confident are you in the ability of wearable medical device manufacturers to protect the privacy of the data collected?	
H6 - Technical Support (TS)		
	Q27: To what extent does the availability of good technical support influence your willingness to adopt and use wearable medical devices?	[80, 130]
	Q28: Have you ever contacted technical support for assistance with your wearable medical device? If yes, please share your experience.	

Continued on next page

Table 3.1 continued from previous page

Hypothesis	Survey Items	References
	Q29: Do you think that responsive customer support by the manufacturer is an important factor in trusting a wearable medical device?	
H7 - Ease of Use (EoU)		
	Q30: Do you consider an intuitive user interface as an influential factor in trusting and adopting wearables?	[132, 6]
	Q31: Does your confidence in using a new wearable medical device depend on the user interface?	
User-Related Determinants		
H8 - User Prior Experience (UPE)		
	Q1: What did you use wearable medical devices for?	[37, 132, 118]
	Q2: What kinds of wearable medical devices have you used in the past?	
	Q3: How long have you used or worked with wearable medical devices?	
	Q4: How has your experience been in using and trusting wearable medical devices?	
H9 - User Demography (UD)		

Continued on next page

Table 3.1 continued from previous page

Hypothesis	Survey Items	References
	Q5: How do you self-identify in terms of gender?	[103, 62, 6]
	Q6: What is your ethnic or cultural origin(s)?	
	Q7: What is your age?	
	Q8: What is your highest level of education?	
	Q9: What is your professional background or area of expertise?	
External Determinants		
H10 - Recommendation (REC)		
	Q32: Have you ever received recommendations from healthcare professionals (e.g., doctors, nurses) or acquaintances to use wearables?	[77, 62, 6]
	Q33: If the answer to the above question is yes, please specify the type of healthcare professional(s) or acquaintances who recommended wearables to you.	
	Q34: To what extent were the recommendations influential in your decision to trust and use wearable medical devices?	
	Q35: Would you recommend using wearables for your family or your patients?	
Other Determinants		

Continued on next page

Table 3.1 continued from previous page

Hypothesis	Survey Items	References
H11 - User Specific (US)	[25, 103]	
	Q36: According to you, what are the hindrance(s) in trusting and adopting wearables?	
	Q37: According to you, what aspects (factors) of wearable medical devices help gain trust in adoption by users and medical professionals?	

3.2.3 Survey Administration

We made the survey available to a community of digital health advocates, mainly healthcare professionals, gym trainers, faculty, students, academic researchers, and the authors' colleagues. It was disseminated in the different departments and research groups of McMaster University, Toronto Metropolitan University, and Mohawk College in Ontario, Canada. We disseminated the survey in two cohorts. For the first cohort, we had 120 participants, of which only 83 were usable for various reasons, including some failing to meet the criteria and others providing incomplete responses to most questions. Prompted by the first cohort's completion rate, we modified the survey in terms of language and ease of understanding (the questions were the same; only the language for some technical questions was simplified). For the second cohort, we had 129 participants, of which 104 were usable. The total number of usable responses for the survey from the two cohorts was 187. In order to have a more effective reach, we conducted the survey online (Lime Survey) and distributed the URL via e-mail through a community mailing list. The survey could be taken in any online environment without any restriction on the platform or device. As an incentive for participation, we offered participants a chance to enter a draw to win one PayPal gift card valued at $20. The survey was available for four weeks for the first cohort and five months for the second cohort. The survey was unsupervised. Participants could exit the survey at any time, even after responding to

some questions, and they could respond to survey items out of sequence. If participants did not respond to more than half of the survey items, their data was not used in subsequent analyses. For the incomplete surveys that were used in analyses, the missing values were estimated using the regression imputation method. This survey is part of a study that has been reviewed and cleared by the McMaster Research Ethics Board (protocol number 5788), Toronto Metropolitan University Research Ethics Board (protocol number 2022-072-1), and Mohawk College Research Ethics Board (protocol number 23-014).

3.3 Results and Analysis

This section presents the findings derived from the analysis of the survey data collected. First, we describe the characteristics of the sample. Then, we describe the statistical methods to test our set of hypotheses.

3.3.1 Sample Characteristics

We exported the survey data from the Lime Survey online questionnaire, and data preprocessing was implemented in Python. There was a total of 249 participants in our study. The participants had to meet two criteria to participate in the survey. The first criterion required participants to be over 18 years old, while the second criterion specified a minimum of two weeks' experience using a WMD. However, not all participants met the survey requirements. A total of 187 out of 249 participants (75.10%) satisfied the criteria and completed the survey. Table 3.2 summarizes the sample characteristics. In terms of age, about 55% of participants were less than 34 years old. 36 out of 187 (19.25%) were in the range of 35-44 years, 19 (10.16%) were in the range of 45-54 years, 6 (3.2%) were in the range of 55-64 years, and 22 (11.76%) did not disclose their age. Participants were well distributed in terms of gender, where 74 out of 187 (39.57%) participants identified themselves as female, 73 (39.03%) as male, 9 identified as others (4.81%), and 31 (16.5%) declined to reveal their gender status. In terms of education, 73 out of 187 participants (39.03%) completed their Bachelor's, 67 (35.82%) completed their master's and post-graduation,

21 (11.22%) completed their Doctorate or professional degrees, 17 (9.09%) completed their high school / post-secondary diploma, and 9 (4.81%) did not disclose their education. In terms of ethnic origin, 62 out of 187 participants (33.13%) belonged to Asia and Europe (including origins from Bangladesh, Pakistan, Philippines, India, Nepal, Sri Lanka, China, Maldives, Japan, Korea, Vietnam, Greece, Caucasia, Hungary, Germany, and Britain), 38 (20.3%) identified themselves as American / Canadian (including origins from Canada, America, Caribbean, Dominican Republic, and Ecuador), 35 (18.76%) were Africans (including origins from Nigeria, Rwanda, Sudan, Tanzania, Zambia and Uganda), 19 (10.1%) identified themselves as Middle Eastern (including origins from Egypt, Iran, Iraq, Syria, Saudi Arabia, Tunisia, Jordan, Morocco, and the United Arab Emirates), and 33 (17.64%) did not answer the question.

The distribution of the sample data was not normal in terms of age, education, and ethnicity. The sample represents WMD users who are highly educated, are from younger age groups and belong to Asian, European or Canadian ethnicity. Checking assumptions of normality is reported in Appendix B.1.

In order to assess the reliability and internal consistency of the questionnaire, we employed Cronbach's Alpha. This measure is crucial for understanding whether the different questions consistently lead to similar results, thus confirming that they measure the underlying attribute being assessed. Cronbach's Alpha coefficients were calculated for scale questions associated with constructs of device-related and external determinants. The specific questions analyzed included Q10, Q11, Q13, Q14, Q16, Q17, Q18, Q19, Q21, Q23, Q24, Q26, Q27, Q29, Q30, Q31, Q32, Q34, and Q35. We obtained an overall alpha value of 0.78, which shows an acceptable level of reliability. The results of the reliability check are reported in Appendix B.2.

To increase the validity of our survey, we included questions to cross-validate other questions, as shown in Table 3.1 (question numbers in bold). When the results of two cross-validating questions are highly correlated, we conclude the validity of the measurement. The results of the cross-validation questions are described in Section 3.4 (Discussion) and reported in detail in Appendix B.3.

Table 3.2: Sample characteristics of the participants (N = 187)

Variable	**Categories**	**Percent (%)**
Age Group		
	18-24	19.78%
	25-34	35.82%
	35-44	19.25%
	45-54	10.16%
	55-64	3.20%
	Above 65	0.00%
	Did not disclose	11.76%
Gender		
	Female	39.57%
	Male	39.03%
	Other	4.81%
	Did not disclose	16.59%
Education		
	High School/Diploma	9.09%
	Bachelors	39.03%
	Masters	35.82%
	Doctorate/Prof. degrees	11.22%
	Did not disclose	4.81%
Ethnicity		
	Asian & European	33.13%
	American & Canadian	20.30%
	African	18.76%
	Middle-Eastern	10.10%
	Did not disclose	17.64%

3.3.2 Device-Related Determinant

To investigate the awareness of the WMD users of the various trust factors identified in the literature, we posed targeted questions specific to each device-related factor (as described in Table 3.1). Participant awareness was measured based on responses to a 5-point Likert scale, considering them "aware" if they responded "sometimes," "often," or "always." Responses indicating "never" or "seldom" were not included in the awareness assessment. The summary of the participants' responses for all the constructs of the device-related determinant is presented in Table 3.3. The findings indicate that participants were generally highly aware of the diverse trust factors, with a

slightly lower awareness of clinical device validation and technical support factors.

Table 3.3: WMD users' awareness of the different trust factors identified from literature

Factor	**Awareness (%)**
DA	63.63%
DR	70.05%
DV	46.52%
DS	75.40%
DDP	77.01%
TS	48.66%
EoU	83.42%

To investigate the influence of the factors on WMD users' trust and their intentional behaviour in adopting a device we tested hypotheses H1 – H7. We first studied the simple correlation between the independent variables (predictors) – DA, DR, DV, DS, DDP, TS, and EoU with the dependent variable TC as reported in the Spearman correlation matrix (Table 3.4). The results reveal strong positive correlations between DA (0.657), DR (0.673), EoU (0.797) and TC. We see moderate positive correlations between the DS (0.440), DDP (0.361), and TC. Conversely, we see insignificant correlations with TC for DV (-0.076), and TS (0.124).

Table 3.4: Spearman correlation results between WMD users' trust concerns and device-related factors

	DA	**DR**	**DV**	**DS**	**DDP**	**TS**	**EoU**	**TC**
DA	1.000	0.330	0.030	0.303	0.254	0.110	0.232	0.657
DR	0.330	1.000	-0.024	0.323	0.149	0.093	0.373	0.673
DV	0.030	-0.024	1.000	-0.019	0.175	0.093	-0.064	-0.076
DS	0.303	0.323	-0.019	1.000	0.155	0.146	0.395	0.440
DDP	0.254	0.149	0.175	0.155	1.000	0.112	0.025	0.361
TS	0.110	0.093	0.093	0.146	0.112	1.000	0.208	0.124
EoU	0.232	0.373	-0.064	0.395	0.025	0.208	1.000	0.797
TC	0.657	0.673	-0.076	0.440	0.361	0.124	0.797	1.000

Then, we tested H1 – H7 with hierarchical regression analysis to assess each factor's degree of influence on users' decisions to adopt and use a WMD. We also performed factor analysis to test

Table 3.5: Hierarchical regression results of device-related factors for WMD users' trust concerns

Predictors	**Model**	R^2	**Adj.** R^2	**F-statistic**	**Prob. (F-stat)**
Baseline	-	0.000	0.000	-	-
DA	Block 1	0.247	0.243	60.480	5.12e-13
DR	Block 2	0.406	0.400	62.66	1.87e-21
DV	Block 3	0.429	0.420	45.58	5.11e-22
DS	Block 4	0.429	0.416	34.00	3.75e-21
DDP	Block 5	0.433	0.417	27.51	1.24e-20
TS	Block 6	0.435	0.416	22.97	5.16e-20
EoU	Block 7	0.647	0.633	46.570	4.79e-37

the validity and analyze dimensionality reduction by identifying underlying factors that explain the correlations among observed variables.

Table 3.5 reports the results of the hierarchical regression to assess the influence of multiple blocks of predictors (DA, DR, DV, DS, DDP, TS, EoU) on TC. The results reveal a notable enhancement in model fit with the incorporation of the initial blocks of predictors, as evidenced by substantial increases in R^2, adjusted R^2, and significant F-statistics, particularly for DA (R^2 = 0.247, adjusted R^2 = 0.243, F= 60.480) and DR (R^2 = 0.406, adjusted R^2 = 0.400, F= 62.66). However, the subsequent addition of blocks, including DV, DS, DDP, and TS, results in only marginal improvements in model fit, with minimal increase in adjusted R^2 values (from 0.429 to 0.435), while a decrease in F-statistics value (from 45.58 to 22.97), thereby showing less contribution of the predictors in blocks 3 through 6, to the overall model fit compared to the previous blocks. The final inclusion of EoU as the last block leads to a notable boost in both R^2 and adjusted R^2 values, signifying its critical role in supporting the model.

Tables 3.6 (a) and (b) report the results of factor analysis to explore the underlying patterns among the device-related factors. Table 3.6(a) reveals a multidimensional structure, characterized by various underlying factors. Factor 1 shows strong loadings for variables DA, DR, and EoU, moderate loadings for DS and DDP, and weak loading for DV and TS, indicating their contributions to TC. Factor 2 demonstrates weak loadings across all variables except for DV, which has a more substantial negative loading. Factors 3, 4 and 5 exhibit weaker loadings and contribute less to

Table 3.6: Factor analysis results to assess the impact of device-related factors on WMD users' trust concerns

(a) Factor loadings

Variable	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
DA	0.513	0.012	-0.167	0.236	-0.086
DR	0.651	-0.200	-0.188	-0.148	0.129
DV	0.051	-0.535	0.0169	-0.008	0.145
DS	0.429	0.070	0.064	-0.374	-0.102
DDP	0.370	0.404	-0.154	0.085	-0.070
TS	0.172	0.145	0.212	0.038	-0.017
EoU	0.537	-0.242	0.193	0.171	-0.053

(b) The variance of the factors

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
SS Loadings	1.404	0.575	0.299	0.256	0.063
Proportion Var	0.200	0.082	0.042	0.036	0.009
Cumulative Var	0.200	0.282	0.325	0.362	0.371

overall variance. Table 3.6(b) provides a summary of the variance explained by each factor: the sum of Squares (SS Loadings), the proportion of total variance (Proportion Var), and the cumulative proportion of total variance (Cumulative Var). Factor 1 accounts for the highest variance, with an SS Loading of approximately 1.404 units, indicating that Factor 1 captures a substantial portion of the variability present in the data.

From the data collected and results reported in Tables 3.4, 3.5, and 3.6, H1, H2, and H7 were conclusive, while H3, H4, H5, and H6 were inconclusive.

3.3.3 User-Related Determinant

We conducted multiple Kruskal-Wallis tests to evaluate hypotheses H8 and H9, examining whether significant differences in user-related determinants (user personal experience and user demography) were linked to TC and use of a WMD. According to the results reported in Table 3.7, there were significant differences for Q1 (purpose of use), Q2 (type of device used), and Q3 (length of duration of experience with WMDs) for the UPE construct, thus accepting our hypothesis H8 (although for

Q4, the p-value was > 0.05 with our data collected, we thought the formation of this question was confusing; therefore, we removed this question from our analysis). For the UD construct, there were not any statistically meaningful differences for Q5 (gender), Q6 (ethnicity), Q7 (age), and Q8 (education) with the data collected, thus rejecting our hypothesis H9.

In terms of what the WMDs were used for (Q1), 129 out of the 187 (68.98%) had experience using WMDs to monitor their health or their family's health, 43 (22.99%) had experience using WMDs for their patients (12 out of the 43 also used it for self-health monitoring), 26 (13.90%) used it for research and study (15 out of the 26 also used it for self-health monitoring). In terms of what kinds of devices were used (Q2), we got mixed responses as some participants had experience using multiple devices, 145 out of the 187 (77.54%) had experience with motion, fall detection, accelerometer and fitness devices, 54 (28.87%) had experience using pulse oximeters and respiratory monitors, 29 (15.50%) had experience with remote continuous glucose monitors, 12 (6.40%) had experience using blood pressure monitor, 10 (5.34%) had experience using cardiac monitors, and 50 (26.73%) had experiences with other devices such as temperature monitoring, rehabilitation devices, sleep tracking, and hearing aids. In terms of length of experience using the WMD (Q3), out of the 187 participants who completed the survey, 68 (36.36%) had the experience of more than two weeks but less than six months, 44 (23.52%) had experience of more than six months but less than one year, 43 (22.99%) had the experience of more than one year but less than three, 17 (9.09%) had the experience of more than three years but less than five, 12 (6.41%) had the experience of using a WMD for more than five years, and 3 (1.60%) did not respond to the question.

3.3.4 External Determinant

We tested our hypothesis H10 to study the impact of recommendations on WMD users' trust and use of a WMD. We used a linear regression model with recommendation (REC) as the independent variable and TC as the dependent variable for the external determinant. The overall results in Table 3.8 ($R^2 = 0.188$; F = 21.28; Prob (F)= 4.85e-09) reveal that the coefficient's standard error (SE) is relatively large at 0.776, indicating considerable uncertainty in the estimate. The t-statistic for REC is 1.487, with a p-value of 0.139. Since p-value > 0.05, we failed to reject the

Table 3.7: Kruskal-Wallis test results for the relation of user-related determinants with WMD users' trust concerns

Construct	Question	H-statistic	p-value
UPE			
	Q1 (Purpose of Use)	6.911	**0.031**
	Q2 (Type of Device Used)	15.745	**0.046**
	Q3 (Length of Experience)	10.070	**0.039**
	Q4 (Quality of Experience)	7.593	0.107
UD			
	Q5 (Gender)	7.091	0.069
	Q6 (Ethnicity)	9.303	0.079
	Q7 (Age)	5.980	0.308
	Q8 (Education)	13.988	0.082

Table 3.8: Linear regression results for the relation of external determinants with WMD users' trust concerns

Variable	Coef.	SE	t	p>t	95% CI Lower	95% CI Upper
REC	0.1078	0.776	1.487	0.139	-0.035	0.251
R^2 = 0.188, F = 21.28, Prob (F) = 4.85e-09						

null hypothesis; therefore, our model was inconclusive in testing H10 with the data collected.

3.3.5 Other Determinant

We studied the impact of the user-specific factor on trust and adoption of WMDs for our hypothesis H11. In order to evaluate the factors, we asked the participants two questions: (1) Q36 - According to you, what are the hindrances to trusting and adopting wearables for remote patient monitoring? (multiple answer question) (2) Q37 - According to you, what aspects (factors) of wearable medical devices will help gain the trust of users? (open-ended question). The responses to our survey for (Q36) showed that 64.17% reported buy-in from healthcare professionals as one of the major hindrances in the adoption of WMDs, followed by the learning curve (47.59%), buy-in from patient users (34.22%), and language barriers (12.29%). The responses for (Q37) revealed some other factors users are concerned about in the adoption of a WMD, such as cost (67.85%) and lack of medical liability standards and rules (38.98%). Our results indicated that user-specific

Table 3.9: Summary of test results for hypotheses (H1-H11) to assess their influence on WMD users' trust concerns

Hypothesis	**Influence on Trust Concerns**	**Conclusion**
H1	Accuracy of the WMD sensors	Conclusive
H2	Reliability of the WMD	Conclusive
H3	Clinical validation of WMDs	Inconclusive
H4	Security methods used in WMDs	Inconclusive
H5	Privacy of the data collected by the WMDs	Inconclusive
H6	Technical support provided by the WMD manufacturer	Inconclusive
H7	Ease of use of the WMDs	Conclusive
H8	User experience with WMDs (purpose, type, and duration)	Conclusive
H9	User demography (age, gender, ethnicity, and education)	Inconclusive
H10	Recommendations from social contacts to use WMDs	Inconclusive
H11	Other user concerns for WMDs	Conclusive

factors positively influence users' trust and behaviour, thus accepting our hypothesis. Table 3.9 summarizes the accepted and rejected hypotheses (H1 - H11) based on the analysis conducted.

3.4 Discussions and Implication

The objectives of this study, as expressed in our research questions, were twofold. First, we were interested in understanding our target group's awareness of underlying trust-influencing factors for WMDs, as shown in the literature (RQ1). Our results in Table 3.3 indicate that participants generally exhibited substantial awareness of the various trust factors from the literature.

Second, we were interested in understanding the relative importance of the investigated trust-influencing factors on users' decisions to adopt WMDs (RQ2). We analyzed eleven trust-influencing factors identified from existing literature clustered into four determinant groups (device-related, user-related, external, and other determinants). The hierarchical regression model developed and reported in Table 3.5 suggests that WMD users are most concerned with device reading accuracy, consistency, and reliability, as well as how easily and intuitively they can interact with and operate WMDs to achieve their goals. We also see similar strong positive loadings in our factor analysis, as reported in Table 3.6. The participants, in their answers to the open-ended question, "Have you been aware of, or ever experienced any issues with the reliability of a wearable?" (Q14) indicated,

"I think the heart rate accuracy is challenging when flexing forearms", "connectivity issues", "raw data makes no sense to lay users", and "loss of data". The study findings and the open-ended responses suggest that participants are primarily concerned with the reliability and accuracy of the WMDs. These factors are mainly consistent with the factors influencing trust in digital technology and sensor-based medical devices, as reported in the literature [30, 125, 8].

We also cross-validated the results associated with device-related factors. For DA, although responses to Q11 and Q12 were expected to be correlated, there was a discrepancy in the answers of participants who answered these questions - only 11 participants responded they always trusted and never cross-verified a WMD (Q11), whereas 57 participants in Q12 did not cross-verify, indicating that the participants did not know the significant methods of cross-verification. For DR, we analyzed the impact of a subset of reliability factors on issues such as connectivity problems, battery concerns, and memory overload issues, using responses to Q15–17. Our results demonstrate that participants were mainly concerned with signal loss due to connectivity issues (53%) and moderately concerned about recharging the batteries of WMDs (38%). Participants were minimally aware of nor concerned with data loss due to memory overload, as has been reported by Peake et al. [104]. Another factor that was found to influence trust and adoption of WMDs was their ease of wear and use, which has been described previously as a factor influencing the adoption of new and emerging technologies [48]. This result is also validated by Q36, wherein learning curves are reported as the main concern associated with adopting WMDs. Users appear to appreciate a more simplified, easy-to-use technology for monitoring health with wearable devices. For example, in their comments, participants indicated, "I would rather just have a device which is comfortable and easy to use". The original questions, their cross-validation questions, and the analysis of their responses are presented in Appendix B.3.

In terms of the user-related determinants, participants in the 18–34 age group were more active WMD users for personal health-monitoring purposes compared to the other participant age groups. This finding aligns with existing knowledge that new and emerging digital technologies are well-embraced by the younger age groups [73]. However, in terms of external determinants such as recommendations from others, a prior assertion that endorsement is an important factor

influencing trust and use of WMDs [118] could not be validated from the results of our survey since there is no significant relationship between the predictor factor of recommendation and the independent factor TC, as indicated by the low R^2 and non-significant F-statistic value by the multiple regression (Table 3.8). For the other determinants, some new factors, such as cost (67.85%) and lack of medical liability standards and rules (38.98%), were highlighted as trust influencing factors.

3.4.1 Implications

The hierarchical regression results indicate that WMD users are generally aware of trust-influencing factors previously acknowledged and described in the literature [30, 125, 8]. However, the results demonstrated stronger correlations to TC for some factors, such as accuracy [150], reliability [80], and ease of use [132], over others described in the literature, such as security [63] and privacy [28]. Our results aligned with those of other researchers [149, 144, 110], demonstrate that most participants are willing to adopt and use WMDs despite data security and privacy risks. Patients who have chronic illnesses willingly use the WMD because they are more concerned about physical health [144]. The younger generation seems to care more about their current physical and mental conditions and does not perceive security and privacy as significant potential risks (though they can be harmful in the future if not carefully assessed and understood) [4, 33]. The implication of the result reported in this thesis, particularly for critical trust factors among WMD users, is that there is a need to provide greater awareness of the security and privacy risks of WMD [132]. With the predicted growth rate of the WMD market [37], users need to be aware of the different aspects of WMDs, such as hardware, software, cloud computing, and storage. With the exponential growth in this market for diverse applications, regulatory bodies are yet to fully catch up to have a clear and transparent standard for the different aspects of WMDs to ensure security and protect privacy [132, 122, 90]. As such, it is recommended that alternative decision-making strategies be implemented to improve the WMD user's awareness of important trust factors such as security and privacy by entrusting the task of developing standards and certifications to non-government bodies such as the International Organization for Standardization (ISO) or Institute of Electrical

and Electronics Engineers (IEEE).

Another interesting finding of our study was the uncertainty surrounding medical liability when using WMDs. There is a lack of legal standards and guidelines for using WMD for remote patient monitoring because of the nascency of the field [113]. Moreover, medical liability issues related to remote patient monitoring using WMDs have yet to be fully addressed by Canadian and international courts [137]. The absence of definitive case law and guidance from professional standards and guidelines makes it difficult to predict legal standards, especially for medical professionals [127]. We recommend that guidelines be produced by engineering research groups in collaboration with associations of medical professionals from specific clinical specialties such as cardiology, radiology and psychiatry for clinical applications using WMD. These guidelines will help physicians understand their responsibility and duty to instruct patients appropriately on using WMD technology. By equipping physicians with documentation and guidelines, they can execute the setup efficiently and provide adequate directions to patients for self-management of health using WMDs effectively, thereby minimizing associated medical liability risks.

A further significant finding from our study was the concern of buy-in from healthcare professionals. A majority of healthcare professionals indicated that the steep learning curve is a challenge when adopting WMDs. Our results showed that 34% of those who found it challenging to use WMDs were medical professionals. Additionally, physicians consider the relationship with their patients crucial to the treatment process. Although it is recognized that autonomous systems incorporating WMDs provide autonomy for the patients, there is a fear of adverse effects on these relationships, such as the "depersonalization" of the treatment compared to conventional face-to-face sessions [102]. There is a need to redeem trust amongst physicians, especially regarding technology. An informative shift in mindset is needed to change these dynamics. We are left with a question: "Can medical professionals' lack of engagement and resistance halt innovation?" As such, we recommend working with physicians and healthcare professionals, educating organizational partners, and creating physician champions to shift their perspectives and overcome barriers in learning to adopt WMDs for remote patient monitoring and tackle the challenge of

Table 3.10: Summary of findings and recommendations of the empirical study on WMD user's trust concerns

Findings	**Recommendations**
1) Participants using WMDs for chronic illness physical and mental fitness care more about their current conditions and do not perceive security and privacy as potential risks due to a lack of awareness [4, 33].	We recommend that alternative decision-making strategies be implemented to improve the WMD user's awareness of factors such as security and privacy by entrusting the task of developing standards and certifications to non-government bodies such as ISO or IEEE.
2) Medical liability uncertainty is a challenge in using WMDs because of the lack of legal standards [113].	We recommend that guidelines be produced by engineering research groups in collaboration with associations of medical professionals from specific clinical specialties such as cardiology, radiology, and psychiatry for clinical applications using WMD to help physicians and patients understand their responsibility, thereby minimizing associated medical liability risks.
3) Physician buy-in is a challenge due to the steep learning curve and "depersonalization" compared to conventional face-to-face sessions [102].	We recommend training physicians to surmount technical barriers and fostering physician champions who understand the advantages of using WMDs for continuous remote patient monitoring, thus overcoming the challenge of physician buy-in.

healthcare professional/physician buy-in. Ontario Health Team (OHT), in their FY2022/23 Funding Guidelines for Remote Care Management & Surgical Transitions for OHTs [56] highlighted the importance of "champions" for physician buy-in and a mature communication plan to ensure the adoption of digital remote patient systems. A journal in medical economics [86] also brings forward this challenge and suggests identifying clinical and operational "champions" as the main point of contact to promote education and training. Table 3.10 concisely summarizes our findings and recommendations.

3.4.2 Study Limitations

The primary limitation of this study is the non-normal distribution of the participant's demographic characteristics. As previously discussed, the distributions of participants' age groups,

education levels, and years of experience with WMDs were skewed. The majority of the participants were young adults and educated (having completed a Bachelor's degree), in addition to having experience using WMDs for more than one year. Furthermore, the population was underrepresented due to the specific groups that predominantly utilize wearable medical devices. Thus, there are concerns about the fairness and equity of using and recommending wearable technology for health in the general population [22, 159]. This limitation has two essential yet different implications in terms of understanding the user's concerns on trust and intentions to adopt WMD. First, we do not see that demographically skewed data has threatened our conclusions. Several studies have shown that young and educated people are more open to using and adopting emerging technology, including autonomous medical systems with wearables [73]. Hence, most users will be in this demographic group. Therefore, it is less likely that we will lose any aspect of TC for our study. Second, although our results support insight from the group of people who predominantly use WMDs, further study with a larger sample size representing the general population can improve the results of the study and enhance the generalizability of the results and implications on the trust of people who distrust or decide not to use WMDs. We believe expanding the scope of the study would help the WMDs to be designed and developed to include the general population as well.

3.5 Summary

In this chapter, we reported the results of an empirical study on TC and the behaviour of WMD users. We developed a research model based on our extensive literature review and tested several hypotheses. An interesting observation was made concerning the inconsistency between user perceptions of trust and behaviour, where we found that the self-reported perception of trust factors in the literature was significantly different from the actual factors influencing the adoption of WMD. We found that factors of greater importance in the literature, such as security and privacy [94], did not significantly influence the behaviour of users in adopting WMD. Our results showed that about 45% of participants are willing to use and adopt WMDs despite security and privacy risks. Another important finding was the resistance from medical and healthcare professionals to buy

into the use and adoption of WMDs. We conclude with three important implications from our study reported in this thesis: First, there is a need to provide greater awareness of critical factors such as security and privacy to WMD users. Thus, we recommend developing and implementing alternative decision-making strategies to improve WMD users' awareness of critical factors. Second, the uncertainty surrounding medical liability needs to be addressed. Therefore, we recommend that engineering research groups, in collaboration with associations of medical professionals from specific clinical specialties, develop guidelines to help physicians and patients understand their responsibilities, thereby minimizing associated medical liability risks. Third, there is a need to redeem trust regarding technology among medical professionals and physicians. As such, we recommend educating organizational partners, creating physician champions, and assisting them in shifting their perspective to use WMDs safely. This study can serve as a basis for providing trust factors from a user's perspective for designing a machine-learning trust model to quantify trust in WMDs, thereby providing a way towards bringing the study from theory to practice.

In this chapter, we reported the factors that WMD users are concerned about, from our empirical study, thus serving as a foundation to develop the model to quantify trust and measure trustworthiness. The way in which these factors (singly or in combination) lead to trust in WMDs has yet to be determined. In the next chapter, we will explore how these factor(s) influence the quantification of trust in WMDs.

Chapter 4

Trust Quantification Model

In this chapter, we present a formal method for quantifying the complex, multidimensional, and uncertain nature of trust, considering the trust factors of WMD users. We utilize a probabilistic graph method, employing BN to succinctly represent the factors (as random variables) and their interconnections, and present a detailed Bayesian structure for a WMD. We set the inter-node relationships, using the NFR techniques, by refining a high-level goal of trustworthiness to lower-level goals that can be objectively implemented as measurable factors. The structured relationships encoded within a BN are leveraged to derive statistical inferences about each node utilizing conditional probabilities and Gaussian prior distributions. We compute the probability of trust for the entire system or its individual constituents and evaluate a relative measure of trustworthiness under identical test conditions. We show how the BN can be used to make inferences on a factor based on new evidence.

This chapter is organized as follows. In Section 4.1, we expand on the motivating scenario (introduced in Chapter 1) to illustrate our proposed model. The probabilistic graph theory and the preliminaries of BN are described in Section 4.2. The development of the proposed trust BN structure and trust quantification are described in Sections 4.3 and 4.4. The experimental evaluation is presented in Section 4.5, followed by an analysis of results in Section 4.6. A summary of this chapter is provided in Section 4.7. The author brings to the reader's attention that the contents of this chapter are published in Thomas et al., 2021 [134].

4.1 Motivating Scenario

In the motivating scenario introduced in Chapter 1, we explore the challenges faced by Alex, an athlete who participates in various demanding indoor and outdoor athletics events globally, often on the same day. To maintain peak performance, Alex adheres to a rigorous training and exercise regimen, which is both physically and biologically demanding. Regular monitoring of his health parameters is essential to keep track of his well-being. However, frequent visits to a physician to regularly monitor health are not feasible due to their extensive travel and event schedule. Therefore, a WMD becomes the most practical solution for Alex.

Due to Alex's frequent national and international travel, it is crucial that the WMD is reliable across different geographic locations. Since Alex performs in potentially crowded venues such as stadiums or arenas, the reliability of network connectivity is a major concern for his trainer [80]. Alex's physician is concerned about how well the WMD's sensors perform under varying weather conditions in different places [124]. Furthermore, Alex has significant concerns regarding the privacy of his health data, which includes sensitive personal information [6], while, Alex's family is worried about the safety of wearing the device, given the potential risks and reported incidents of device misuse that have led to dangerous and even fatal outcomes [36]. Adding to the complexity of this decision-making process, situations may arise where two devices offer comparable functionalities. This amplifies the challenge as it becomes essential to determine which device is more reliable, thus complicating the decision criteria further and requiring a strong methodology to effectively compare and quantify their trustworthiness. The important question here is how can different stakeholders with different perspectives of trust select the best WMD for Alex.

Figure 4.1 illustrates this motivating scenario and outlines the different layers of the WMD (as described in Section 2.1). Stakeholders have varying and subjective perceptions of what constitutes trustworthiness in a WMD: Alex may prioritize privacy or security, his trainer may emphasize network reliability, and his physician may focus on sensor accuracy. Each of these aspects highlights a different dimension of device trustworthiness, reflecting the unique priorities of each stakeholder.

To navigate these diverse TC effectively, we propose a trust framework using a probabilistic

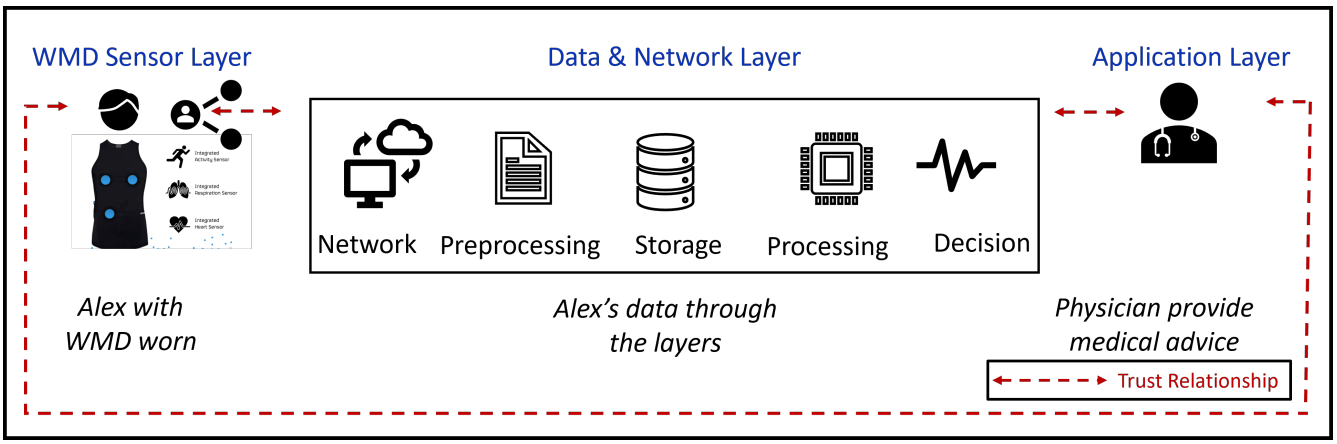

Figure 4.1: Motivating scenario showing the trust relationship between different stakeholders (Alex and medical advisor) in the different layers (sensor, data and network, and application) of the WMD

model. This framework is designed to provide a comprehensive quantification of trust by considering all pertinent factors and offering specific inferences on individual or a combination of selected factors, based on available evidence. This method is particularly useful in complex scenarios such as Alex's, where it systematically assesses and identifies the most trustworthy WMD option based on the stakeholders preferences.

4.2 Probabilistic Graph Models and Bayesian Networks

In order to quantify the subjective nature of trust, we need to have a model, which considers the uncertainty. Probabilistic graph models are used for representing stochasticity using random variables. The PGMs are usually represented with random variables (usually called vertices or nodes) drawn as circles or ovals and relations as lines or arrows. They provide a framework for compactly encoding probability distributions on random variables that interact with each other and capture conditional independence relationships between interacting random variables [71]. Probabilistic graph models can be classified based on the type of graph used to represent the dependence relations as undirected graphs representing symmetric relations (e.g., Markov random fields), and directed graphs in which the direction between relations are important (e.g., Influence diagrams) [128]. The PGMs represented by directed graphs can further be classified as cyclic and

acyclic. Directed cyclic graphs are circuit wherein the final vertex coincides with the initial one. The directed acyclic graphs (DAGs) however, do not form a circuit [128]. An example of a DAG is the BN.

Bayesian networks are DAGs that represents a set of random variables and their conditional dependencies in a compact way using *nodes* and *edges*. The nodes in the BN are probabilistic random variables and the directional edges (arrows) between them represent relationships [71]. There are two types of nodes in the BN: non-descendant (nodes that have no parents) and descendant (nodes that have one or more parents).

Formally, a BN is represented as $G = (X, E)$, where $X = \{X_1, ..., X_n\}$ is a finite set of n discrete random variables or nodes, and E is a finite set of directed edges. Following the terminologies in [71], the non-descendant nodes are denoted as $(X_{iNonDescendants})$ and the descendant nodes are (X_i). The edges of our trust graph model are either $X_i \rightarrow X_j$ or $X_j \rightarrow X_i$, (but not both as G is a DAG) for a pair of nodes X_i, X_j in X. Let Pa_{X_i} denote the parents of a node X_i in the BN structure G.

A BN is specified by 1) the *structure* (encoding a set of conditional independence relations between the variables), and 2) the *parameters* (a set of local conditional probabilities for each variable given its parents in the graph) [128]. Bayesian network structures and parameters can either be predetermined (fixed) or learned from data, as depicted in Fig. 4.2. Figure 4.2(a) illustrates a configuration in which both structure and parameters are fixed based on domain knowledge. In Fig. 4.2(b), the structure is fixed by the developer, while parameters are learned from data. Figure 4.2(c) shows the instance where the structure is learned from data, while parameters are fixed by the developer. Lastly, Fig. 4.2(d) represents a configuration where both structure and parameters are learned from data. In this thesis, we present a BN for quantifying trust in a WMD, by pre-defining the structure and learning the parameters.

Bayesian networks have been used to quantify trust in various fields, including medical imaging [60, 40], industrial applications [66, 87], CPSs [142], and IoT [65]. We use BN to quantify trust in WMD. Using the BN, the stochastic nature of trust is viewed in terms of probabilities as subjective degrees of belief by a set of random variables in the domain. We scope each random

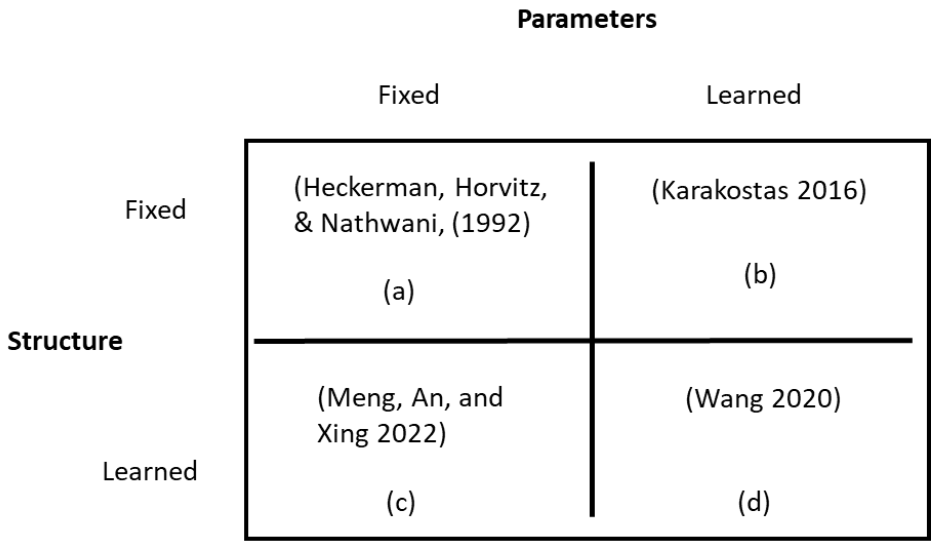

Figure 4.2: The two primary components of a Bayesian network- the structure and the parameters, can each be either learned from data or predefined (fixed)

variable in the network by the trust factors that are identified from the literature and validated by our empirical study.

Quantifying trust can be a challenge, even after scoping trustworthiness to specific determinants (such as robustness, security, and privacy) as these determinants are not granular enough to be directly measured. Therefore, we need to decompose each determinant in the hope of reaching a level at which the relevant quality can be directly measured.

4.2.1 Non-Functional Requirement Techniques for Refinement

Requirements engineering, a crucial subarea of software engineering, provides structured frameworks to refine functional and non-functional requirements. While functional requirements specify the actions a system must perform (such as continuous monitoring of heart rate), NFRs articulate user expectations of system performance, including trust, security, and privacy [156, 32].

Non-functional requirements such as trust (or security) are hard to measure. To address this, requirements engineering uses a technique called *goal refinement* [156, 32]. For example, a high-level goal like "improve system security" can be refined into specific, measurable sub-goals (also

called tasks in i* notation [156]) such as "implement two-factor authentication" or "encrypt all data at rest." If the subgoals (tasks) are still not directly measurable, they can be further refined until they become actionable and measurable. This process ensures that every NFR is translated into actionable tasks that can guide software development. We adopt this technique to refine the high level goal of trust to measurable factors of the device. For example, consider the NFR framework applied to our motivating scenario. The top goal requiring that the WMD is trustworthy can be refined to two sub-goals: reliability of the sensors to provide correct readings and robustness of the WMD to perform efficiently in different conditions (e.g., indoors or outdoors). The top goal is refined to sub-goals based on the expert knowledge, literature review, or stakeholder's survey until we reach a level where the responsibilities can be measured or assigned to a human or software agent. In this thesis, we apply NFR techniques to refine the factors, specifically focusing on eliciting user requirements to define the essential properties of WMDs.

4.2.2 Hierarchical Bayesian Network Structure

We apply the NFR techniques and present a hierarchical layering approach to reach to measurable and quantifiable quality of trust. Below, we demonstrate how this refinement can be performed for WMD, with a simple BN example with seven nodes. For example, in Fig. 4.3, trust (X_7) cannot be measured directly and also it is not directly related to measurable observations such as the heart rate sensor validation (X_1) or the quality of heart rate signal (X_2). However, we know the relationship between the reliability of the device (X_5), the measured observations in X_1 and X_2, and the relationship between reliability (X_5) and trust (X_7).

We present a hierarchical BN structure such that the nodes in the first layer ($X_1 - X_4$), shown in grey in Fig. 4.3, capture the trust factors that can be measured. For example, X_1 and X_2 represent the probability of the trust factors associated with heart rate sensor validation and heart rate quality. The nodes X_3 and X_4 represent the probability of the energy power and memory storage of the WMD. The remainder of the network structure represents the probability of subjective belief on the state of each refined attribute such as reliability (X_5), robustness (X_6), and trust (X_7). Nodes X_1 to X_4 are non-descendant and X_5 to X_7 descendant nodes, such that,

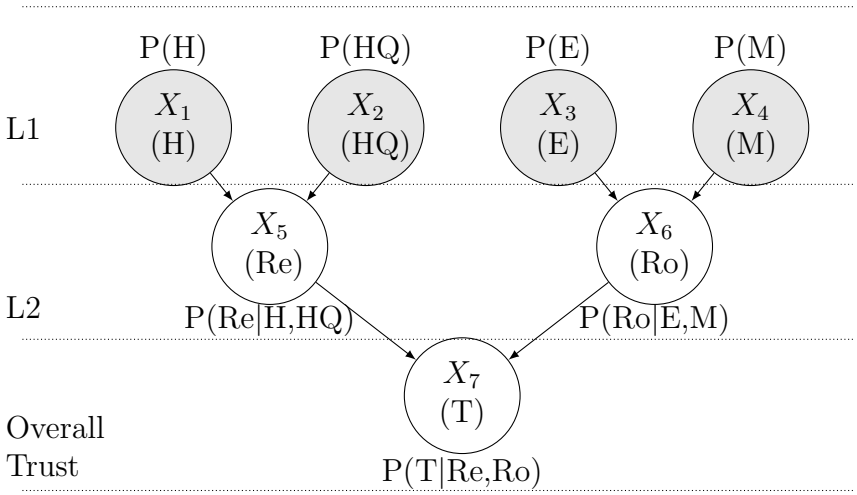

Figure 4.3: Bayesian network to quantify trust (T) in a wearable medical device having one sensor, analyzing patient's heart (H) and heart rate quality (HQ) rate and the device's energy (E) and memory (M) in Level 1 (L1); reliability (Re) and robustness (Ro) in Level 2 (L2). The priors (P) for $X_1 - X_4$ (gray) are arbitrary values based on a Gaussian distribution. The arbitrary values are propagated through the network to estimate the parameters for $X_5 - X_7$

$Pa_{X_5} = \{X_1, X_2\}$, $Pa_{X_6} = \{X_3, X_4\}$, and $Pa_{X_7} = \{X_5, X_6\}$. The nodes have binary levels of data - 0 (low) and 1 (high). The trustworthiness of a WMD from a stakeholder's perspective is represented in terms of the probability of trust to be in a specific state (e.g., low or high).

With its simple structure, a BN can capture many random phenomena in the presence of multiple interrelated aspects that relate to a specific reasoning task. For example, we might be interested to know the probability that a patient has flu given several interrelated pieces of evidence, including the season and symptoms where the same symptoms might also indicate another diagnosis, such as hay fever ([71]-Ch.1). In our approach, we define the trustworthiness of a WMD in a similar way and in terms of the probability of the *trust* node in the BN (X_7) to be in a specific state (e.g., low, or high). In the next section, we show the construction of the BN structure for a WMD to quantify trust.

4.3 Proposed BN Structure to Quantify Trust

The first step in the development of a BN framework to quantify trust in a WMD is to develop the BN structure, which is the focus of this chapter.

Bayesian Network Structure

A BN structure is a PGM that represents the relationship between the random variables (nodes) with directional arrows (edges) through a DAG.

4.3.1 Development of the BN Structure

In order to show the development of the BN structure to quantify trust in a WMD, we first define the scope based on the factors identified from our literature study and validated by our empirical study. We identified three potential determinants related to the device for our scenario: robustness, security, and privacy. To scope trust quantification, we state that the WMD is trustworthy if it is robust (i.e. performs what it is expected to do accurately), secure (i.e. communication between any part of the system and the system interface with the outside world is secure), and preserve privacy (i.e. any and all personal data stays private as the state of the system changes).

Inspired by [142, 160], we present the development of the trust BN for our pilot BN structure [134]. We apply a four-level hierarchical layering approach to reach the measurable and quantifiable quality using NFR techniques. We call the first level of decomposition *determinants*, as the main determinants impacting trust. The determinants are further decomposed to *indicators*, they indicate the factor to be considered for the measurement of a determinant. The indicators also may not be quantifiable, in this case we further refine indicators to *components* of each indicator. Below, we demonstrate the refinement performed for three determinants that we identified for our WMD trust BN structure.

Robustness

Robustness is a firm belief in the competence of the entity to act dependably and perform as expected within the acceptable time frame [49]. Shin and Sarawi et al. [8, 124] provide the important properties required in the sensor layer for resource-constrained sensors. We identify two main properties, among others, that contribute to the robust functioning of a sensor: *performance* and *reliability*. Performance measures the sensor behaviour [7, 44] and is important for WMD where a patient's physiological signals need to be continuously monitored. Reliability measures

the behaviour of the network in the IoT sensor layer [103, 121].

Each of these indicators is affected by various factors of the resource-constrained WMD. For example, performance can be represented in terms of *sensor accuracy* and *energy consumption* [124]. Maitra et al. [82] present the importance of accurately collecting the required data from the sensors for optimal use of energy. Therefore, sensor accuracy and energy consumption are not only suitable candidates to refine the performance indicator but are also quantifiable. Similarly, the reliability indicator can be refined to quantifiable components such as: *Received Signal Strength Indication (RSSI)* and *latency*. Latency is the time between when data is sent from a connected device to when it is received. The time delay between the sent and received packet can be estimated by ping or latency test program measurement [74]. For example, in our use case, alerting the trainer or physician about critical physical changes or events is important, and hence latency plays an important role. RSSI describes the strength of the signal, ensuring the data can be delivered to the receiver considering the environmental factors on a particular radio and can be predicted by an RSSI algorithm [64, 5].
We use a similar multilevel decomposition approach for security and privacy determinants.

Security

The system is considered secure if the sensors are accessed and controlled only by authorized users [125, 93]. Security can be further refined to *safeguards*, which ensure secure sensor operation, and *standards*, which ensure compliance to security protocols. For the safeguards indicator, we consider two components: *confidentiality* and *integrity* [27]. Confidentiality is to ensure the sensor data is accessible only by authorized parties. Integrity is to ensure the sensor data is not changed by unauthorized parties. The *standards* indicator can be further refined to *compliance* and *use safe*. Compliance is to ensure that standards from regulatory bodies, such as the FDA are built into the system to make sure medical sensors are safe and secure to wear [122]. The use safe component is to ensure that the sensors are easy, safe and protective to use.

Privacy

Privacy ensures that the data collected by the sensors are protected from being exposed [63, 129]. Privacy can be refined to *data reuse*, referring to the safe reuse of data and *data protection*, referring to protecting the personal data collected [36]. The data reuse indicator ensures that the collected data is used only for the purpose it is collected [24]. Data protection of a system can be ensured if the data collected by the WMD is used in a confidential and private manner [36]. The data reuse indicator can be refined to *data storage* and *data usage*. The data storage would ensure that the data is stored and archived safely. The data usage component is to ensure that the data collected by the sensors are used only for specific and related analysis or research [63]. Data protection indicators can be further refined to *transparency* and *anonymization*. Transparency is the visibility of the flow of data in the system for the primary stakeholders to gain trust in the system [125]. Anonymization ensures that when the data is used for training, the model follows the Health Insurance Portability and Accountability Act (HIPAA) standard for protecting sensitive patient data [129].

We should emphasize that the refinement of trust, its determinants and indicators we introduced so far is not an exhaustive list of all possible refinements for the WMD. In fact, such an exhaustive list might be unattainable due to the domain dependency of trust definition and limitation of abstraction. However, the approach is significant as it demonstrates that when such a refinement exists we have a succinct representation of the trust network [121, 44].

The development of our BN structure is inherently explainable and interpretable due to its transparent and systematic refinement approach. By decomposing trust in a hierarchical manner into distinct, measurable components, our model provides clear pathways for understanding how each factor influences overall trust in WMDs. This hierarchical structure allows stakeholders to trace the impact of individual variables, such as robustness or security, on the overall trust score. The non-descendant nodes of the BN explicitly links real-world factors, making the model practically interpretable for decision-makers. Confidence in the model's output is enhanced, as each inference can be traced back to specific, well-defined factors, providing an intuitive understanding of trust in WMDs.

Table 4.1: Trust factors identified from the literature and their corresponding domain values © 2021 IEEE

Variable	**Value**	**Description**
Energy, Accuracy, RSSI, Latency, Confidentiality	$\{0, 1\}$	representing low or high levels of the component factors [82, 9, 8, 124, 64, 5]
Integrity, Compliance, Use Safe, Transparent, Anonymize, Data storage, Data usage	$\{0, 1\}$	representing no or yes; ensuring if these component factors are implemented [27, 122]
Performance, Reliability, Safeguards, Standards, Data Protection, Data Reuse	$\{0, 1\}$	representing low or high values of the indicators [7, 44, 103, 125]
Robustness, Security, Privacy	$\{0, 1\}$	representing if the determinants are weak or strong [49, 125, 63, 129]
Trust	$\{0, 1\}$	representing if the system trust is low or high [93, 7]

For our exemplary trust elements, we summarize the variables involved in Table 4.1. For simplicity, we considered all variables to be in two states: 1 (high) and 0 (low). The states indicate the level of contribution of the variable towards computing the trust. Fig. 4.4, shows the BN (we described above) for quantifying trust in WMD (assuming for simplicity that the WMD has one sensor). In a real system, we may have multiple sensors or constituents and each one will have a similar hierarchical relationship with each other. In Fig. 4.4, the nodes $X_1, ..., X_{12}$ are the nodes that are not descendants (have no parents) and are shaded in grey, and the nodes $X_{13}, ..., X_{22}$ are descendant nodes that have one or more parents and maybe parents to one or more nodes. The nodes $X_1, ..., X_n \in X$, represent probabilities such that, X_1, X_2, X_3, X_4 represent energy, accuracy, RSSI, latency; X_5, X_6, X_7, X_8 represent confidentiality, integrity, compliance, use safe; $X_9, X_{10}, X_{11}, X_{12}$ represent data storage, data usage, transparent, anonymize; X_{13}, X_{14} represent performance, reliability; X_{15}, X_{16} represent safeguards, standards; X_{17}, X_{18} represent data reuse, data protection; X_{19}, X_{20}, X_{21} represent robustness, security, privacy, and X_{22} represents trust. Each variable has a domain of values as given in Table 4.1, for example, we consider the variable X_1 to have a binary value, $X_1 = \{low, high\}$ representing if the local energy consumed by the medical sensor is low or high.

We assume the following conditional independence for each variable in X:

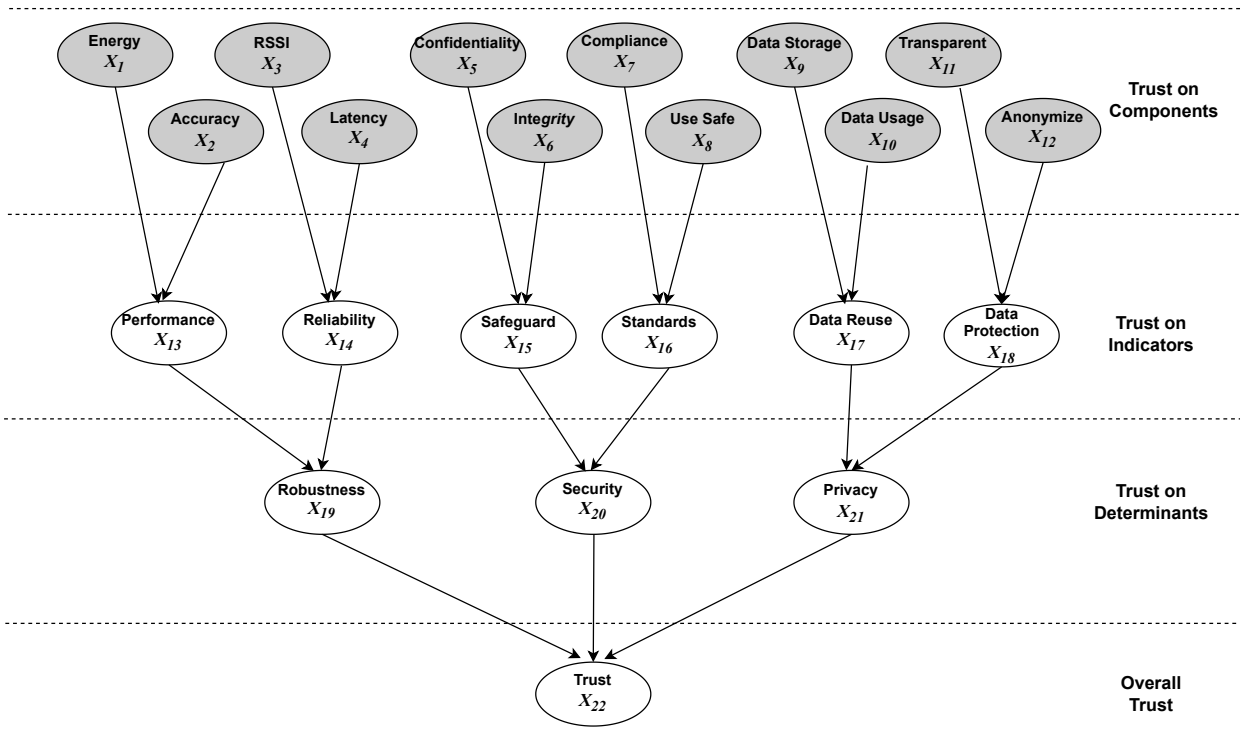

Figure 4.4: Probabilistic graph Bayesian network to represent the probability of trustworthiness for one sensor or component of the system, given the state of trust of other events. Ovals: the random variables of the trust components, indicators, determinants, and overall trust and arrows represent the relationship between them © 2021 IEEE

$(X_i \perp X_{iNonDescendants} | Pa^G_{X_i})$, signifying that each node X_i is conditionally independent of its non-descendant nodes given its parents. For example, referring to Fig. 4.4, the node X_1 (energy) and X_2 (accuracy) are conditionally independent of each other, but each contributes to the probability of the node X_{13} (performance), given the direct relationship of X_1 with X_{13} and X_2 with X_{13}.

The connection from X_1 given the probability of energy is low $P(X_1^L)$ to X_{13} for the probability of high performance $P(X_{13}^H)$ is represented by the directed edge and can be shown by the conditional probability $P(X_{13}^H|X_1^L)$. Multiple conditional probabilities can be computed by the joint probability distribution of the node. For example, the probability that performance is high, given that energy consumption and sensor accuracy are low can be computed by $P(X_{13}^H|X_1^L, X_2^L)P(X_1^L)P(X_2^L)$.

We are interested in answering specific queries about trust such as:

1) What is the probability that trust is high when robustness is low, security is high, and privacy is high - $P(X_{22}^H|X_{19}^L, X_{20}^H, X_{21}^H)$? or,

2) What is the probability that robustness is high when performance is high and reliability is low

- $P(X_{19}^H|X_{13}^H, X_{12}^L)$?

The probability of direct measurements in an ideal world is fixed, however, in a real case, there are uncertainties. To incorporate the uncertainty, we need an approach that computes the probability of an event of a descendant node, for a range of different probabilities. Hence, we need an iterative process that provides an updated probability for direct measurements. We generate random probability distribution functions (PDFs) to mimic the update of the probabilities of the sensor measurements, considering uncertainties.

4.4 Quantifying Trust

In order to effectively quantify trust within a Bayesian framework, it is essential to define both the structure and parameters of the model. Parameter learning necessitates knowledge of the prior probabilities for all nodes. The priors for the nodes can be assumed to have a probability distribution (such as a Gaussian distribution) or can be obtained from the data. This data acquisition process and subsequent parameter learning are the primary focus of Chapter 5.

In this chapter, we proceed under the assumption that we have the prior for all nodes. This assumption facilitates the demonstration of computing the conditional probability and finding the probability of the trust node or any other node for a given state for the WMD network structure. The BN structure (such as in Fig. 4.4) encodes the probability density of the random variables with their conditional dependencies in the form of a DAG. Each node within the BN is associated with a CPD, which articulates the distribution of the node's values contingent upon each possible joint value assignment of its parent nodes. Specifically, for any descendant node X_i in X, the CPD is given by Eq. (4.4.1) [71],

$$P(X_i|X_1, ..., X_{i-1}) = P(X_i|Pa_{X_i}^G), \tag{4.4.1}$$

Here, we assume, all of X_i's parents are in the set $\{X_1, ..., X_{i-1}\}$, and none of X_i's descendants can possibly be in the set. Nodes without parents are described by marginal probability distributions, reflecting their inherent uncertainty independent of other variables.

Using the Bayes rule, the posterior probability of a node X_i in the graph model can be defined in terms of the probability of its parent node and the prior probability as given in Eq. (4.4.2) [71],

$$P(X_i|Pa^G_{X_i}) = \frac{P(Pa^G_{X_i}|X_i)P(X_i)}{P(Pa^G_{X_i})}, \tag{4.4.2}$$

where $P(Pa^G_{X_i})$ is the marginal probability of the parents of X_i.

In general, the posterior probability of any intermediary node given a state can be found by the Bayes rule based on the prior and the CPDs. For each random variable i, the prior probability, $P(X_i)$, captures the aleatory uncertainty of the node.

Bayesian parameter estimation leverages the structured relationships encoded within a BN to derive posterior probability for each node, utilizing both the conditional structures provided by CPDs and the prior distributions. Through this comprehensive approach, the methodology addresses both the dependencies and uncertainties inherent in complex probabilistic models [71]. This dual reliance on CPDs and prior knowledge facilitates a dynamic updating mechanism, which is central to Bayesian inference and critical for applications requiring adaptive probabilistic reasoning.

4.4.1 Query Based Learning

In BNs, query-based learning leverages the PGM structure to efficiently compute complex queries involving multiple variables and evidence (as detailed in Section 4.3.1). This is achieved by decomposing the joint probability distribution into a product of conditional probabilities, facilitating streamlined inferencing.

The essence of query-based learning in BN lies in the calculation of joint probabilities across multiple nodes. This involves computing various conditional probabilities related to the nodes in question. These probabilities are systematically combined to form the joint distribution of the network. The general form of this distribution is represented as follows [71]:

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i|Pa^G_{X_i}), \tag{4.4.3}$$

where $Pa_{X_i}^G$ denotes the parents of node X_i in the graph G.

For example, the probability that the system is robust when performance is low, reliability, confidentiality, latency, signal strength, sensor accuracy are high, and the energy consumption is low, can be written as Eq. (4.4.4),

$$P(X_{19}^H, X_{13}^L, X_{14}^H, X_5^H, X_4^H, X_3^H, X_2^H, X_1^L) = P(X_{19}^H|X_{13}^L, X_{14}^H) \\ P(X_{13}^L|X_1^L, X_2^H)P(X_{14}^H|X_3^H, X_4^H, X_5^H) \\ P(X_1^L)P(X_2^H)P(X_3^H)P(X_4^H)P(X_5^H), \quad (4.4.4)$$

If we assume the discrete probabilities for the variables, such that, $P(X_{19}^H|X_{13}^L, X_{14}^H) = 0.08$; $P(X_{13}^L|X_1^H, X_2^H) = 0.8$; $P(X_{14}^H|X_3^H, X_4^H, X_5^H) = 0.6$; $P(X_1^L) = 0.5$; $P(X_2^H) = 0.7$; $P(X_3^H) = 0.4$; $P(X_4^H) = 0.7$; $P(X_5^H) = 0.6$, then $P(X_{19}^H, X_{13}^L, X_{14}^H, X_5^H, X_4^H, X_3^H, X_2^H, X_1^L) = 0.002$

We can use the same process for any state in the joint probability space. By leveraging these relationships and the modular design of BN, query-based learning not only mitigates computational demands but also makes the probabilistic manipulations for complex problems tractable. Note, that we have considered fixed distributions to demonstrate our pilot WMD Bayesian trust network. However, in the next chapter, we show the parameter estimation of the BN using a data-driven approach.

4.5 Experimental Evaluation

The goal of this experiment is to evaluate the applicability and scalability of our trust model. We evaluate the applicability of our approach, by implementing our trust structure and calculating the posterior probability by propagating the CPDs and joint probability, to check the soundness of the system. We evaluate scalability, by computing the posterior probability for the increasing number of nodes and check the increase in the computation time. For both experiments, since we do not have real data we use simulation methods to generate data for the probability distribution of the random variables. The code and dataset for the experiments are publicly available on GitHub[1].

[1]https://github.com/tailabTMU/TrustQ_MD

4.5.1 Evaluation Metrics

We consider the following metrics to evaluate the trust:

1. **Expectation**: The expectation ($\mathbb{E}$) is the perceived probability of trust [142] and is defined in terms of the mean of the probability density function given by Eq. (4.5.1),

$$\mathbb{E} = \int_{i=0}^{n} x.f(x)dx, \tag{4.5.1}$$

   where $f(x)$ represents the posterior probability density function for X.

2. **Area Under the Curve (AUC)**: AUC can be interpreted by approximating the integral by the total area and can be integrated numerically as given by Eq. (4.5.2) [81],

$$AUC = \int_{X} f(x)_X dx, \tag{4.5.2}$$

The expected mean of a node assigns a real or absolute score to each event in a given state. $\mathbb{E}$ of all the nodes in a structure follows linearity and can be given as the sum of the expected mean of each descendant nodes [142]. $\mathbb{E}$ helps us to compare the absolute score for different cases with a gold standard (baseline) case as reference to provide a *relative measure*.

The AUC for a given interval is the probability of a event in that interval. The larger the area, the more likely is the probability of an event. AUC for the posterior probability distribution of an event, for different cases can be computed and compared to validate the performances.

4.5.2 Experimental Setup

To evaluate overall trust under different conditions, we consider three cases, generating the PDFs of the non-descendant nodes with different ranges of standard deviation (SD).

Case 1: The PDFs of all the nodes are generated to produce wider and more spread distributions with SD ranging from 100-300 leading to poor distribution. This is our *weak* case.

Case 2: The PDFs are a mix of poor and strong distribution with components $X_1 - X_6$ having a

poor distribution and $X_7 - X_{12}$ having strong distribution. This is our *average* case.

Case 3: The PDFs of all the nodes are stronger and sharper with SD ranging from 1-3 are generated. This is our *strong* case.

We used the Python software on an Intel Core i7, 2.6GHz CPU. We generated datasets with the normal (Gaussian) distributions for each non-descendant node X_1 to X_{12} to mimic the prior PDF. We performed 50 repetitions for each case. In order to update the PDFs of the non-descendant nodes, we used the random generator to generate new means and standard deviations in every iteration. The low and high states of the nodes are distinguished by boundary conditions: $0 \leq \mathbb{E}(X) \leq 50$ for low state, and $50 < \mathbb{E}(X) \leq 100$ for high state.

We used simulation methods to compute the trust probability distribution with normal (Gaussian) distribution. We used normal distribution mainly because many natural phenomena follow a normal distribution, and also for simplicity, since a normal prior with a normal likelihood function, gives a normal posterior distribution. The normal distribution parameters that we have considered are the mean (μ), the standard deviation (σ), and the variance (σ^2). The Bayesian approach framework with a multivariate posterior distribution, can be represented as $f(\mu, \sigma^2|x)$ where x is any value the random variable X can take, and $X \sim N(\mu, \sigma^2)$, where N is the normal distribution [81].

4.5.3 Algorithm

The steps taken to compute the probability of any descendant node in our BN trust structure are described in Algorithm 1. The goal of the algorithm is to find the probability of any descendant node to be in a specific state of either "high" or "low", given the conditional and prior probability of the respective nodes.

The inputs to the algorithm are: $G(X, E)$ the graph structure of the system where X are the nodes of the system and the relationships between nodes are given by E; k is the number of iterations; $P(X_i)$, and $P(X_{iNonDescendants})$ are the probability of the descendant and the non-descendant nodes, X_d is the desired node of which we want to find the probability; $state_{X_d}$ is the state of the node (high or low state). The prior probabilities of the non-descendant nodes are updated every iteration with a random Gaussian distribution, to mimic new sensor measurements.

Algorithm 1: Compute the probability of any descendant node to be in a specific state given the conditional probability of the respective parents © 2021 IEEE

Input: $G(X, E)$, $P(X_i : X_i \text{ is descendant})$, $P(X_{iNonDescendants})$, X_d, $state_{X_d}$, k
Output: $\mathbb{E}_T$, AUC_T

$j \leftarrow 1$, $\mathbb{E} = 0$, $AUC = 0$, $Total_{\mathbb{E}} = 0$, $Total_{AUC} = 0$
while $j \leq k$ **do**
  $P(X_{iNonDescendants}) \leftarrow$ Update prior probability
  Compute posterior probability $P(X_i|Pa^G_{X_i})$ Eq. (4.4.2)
  Query Inference $P(X_d = state_{X_d}|Pa^G_{X_d})$ Eq. (4.4.3)
  Compute $\mathbb{E}$ Eq. (4.5.1)
  Compute AUC Eq. (4.5.2)
  $Total_{\mathbb{E}} = Total_{\mathbb{E}} + \mathbb{E}$
  $Total_{AUC} = Total_{AUC} + AUC$
end
return $\mathbb{E}_T = Total_{\mathbb{E}}/k$, $AUC_T = Total_{AUC}/k$

The algorithm computes the posterior probability of all nodes in the BN using the Bayes rule in Eq. (4.4.2). The probability of the desired descendant node given the evidence for a specific state, $P(X_d = state_{X_d}|Pa^G_{X_d})$, is then computed by Eq. (4.4.3). The program returns the expected mean of the posterior probability (score) $\mathbb{E}_{\mathbb{T}}$ computed using Eq. (4.5.1) and the average of the total area under the curve AUC_T computed using Eq. (4.5.2).

Our technique for inference (sampling) from multinomials is based on the forward sampling method, where the variables are sampled in topological order. We start by sampling the variables without parents; then we sample from the next generation by conditioning the CPDs of these variables to values sampled at the first step. We proceed like this until all the variables have been sampled [71]. AUC and expected mean works well assessing the trust for binary observations [96].

4.6 Experimental Results and Analysis

Fig. 4.5 shows the probability of overall trust for the three cases, for 5 iterations. We see from the plots that when the probability of all the components and the CPDs of the determinants are weak, the posterior is spread with weak and slow convergence. However, the distribution becomes stronger, and we get sharper and faster asymptotic convergence for case 3 (strong case).

Table 4.2 shows the log of the AUC and $\mathbb{E}$ for the 3 cases. The decreasing value of AUC from

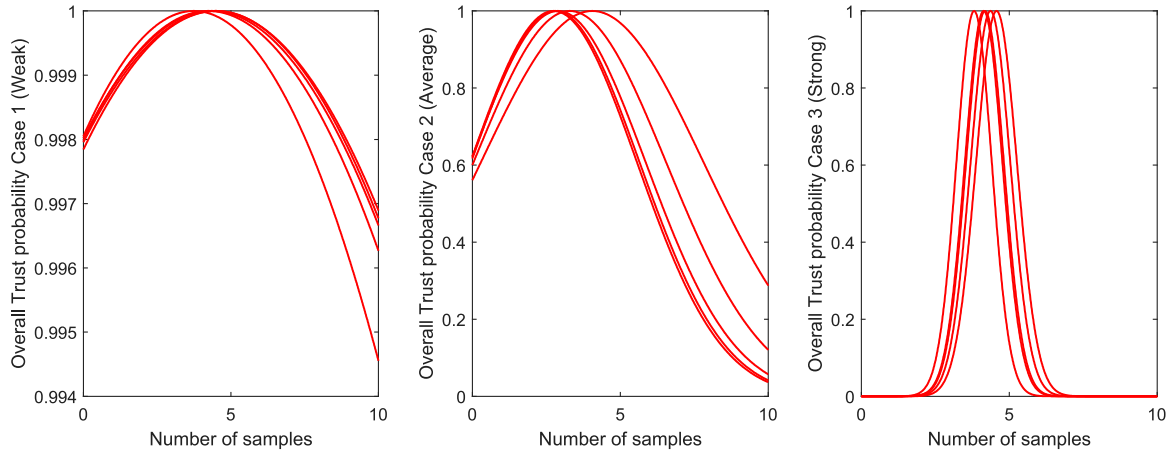

Figure 4.5: Probability of overall trust for Case 1 (weak), Case 2 (average), and Case 3 (strong). The results shown are for 5 iterations © 2021 IEEE

Table 4.2: AUC and Expected mean for Case 1 (weak), Case 2 (average), and Case 3 (strong) © 2021 IEEE

	Cases		
Metric	Case 1	Case 2	Case 3
AUC	-6.887	-19.231	-65.993
$\mathbb{E}$	848.310	124.340	0.060

the weak case (Case 1) to the strong case (Case 3) is the indication of the convergence of the distribution to the expected mean, indicating a higher likelihood of the probability of the event. Case 3 has the strongest convergence and can be considered as a reference to compare with the scores of Cases 1 and 2 inorder to provide a relative measure. The respective expected mean shows the shift in the mean value of the distribution of the three cases.

The model was run with increased nodes to evaluate scalability. We categorize the model with an increasing number of non-descendant and descendant nodes (in multiples of 2), and the number of layers increased by 1 in each category as shown in Table 4.3. The time taken to compute the probability of a descendant node (e.g., overall trust) when each category is run for 10,000 iterations is reported in the last column and demonstrates close to linear increase in time as the number of nodes increases.

From our results, we observe that BNs can be effectively applied for relative measurement of

Table 4.3: Time taken to compute the probability of overall trust, for increased nodes over 10,000 iterations © 2021 IEEE

Non-descendant Nodes	Descendant Nodes	Total Nodes	Layers	Elapsed Time (Secs)
12	3	15	3	1.495
24	6	30	4	2.639
48	12	60	5	5.414
96	24	120	6	11.139
192	48	240	7	21.833
384	96	480	8	43.983

trust for systems with a different configuration, which can be scaled up for higher dimensions. The BN developed for our motivating scenario was assessed based on:

1. Bayesian Structure: We have effectively demonstrated a practical approach to quantifying the multidimensional, abstract, and subjective concept of trust for our motivating scenario for an WMD with one sensor. This approach integrates factors from our empirical studies, and literature into a measurable framework.

2. Query Based Inference: We demonstrated that if we have the priors for all the nodes of the BN, we can effectively provide query-based learning under different conditions, thus helping stakeholders (such as Alex in our motivating scenario) to make informed decisions about trust.

4.6.1 Study Limitations

Our study exhibits some limitations.

1. First, we simulated the priors for all nodes to have a complete dataset using Gaussian distribution based on expert knowledge. However, the Gaussian distributions may not be true representations of the priors for all nodes. Hence, there is a need to have a method by which we can generate more realistic priors for the nodes of the BN.

2. Second, we investigated the model's performance considering a BN with one sensor for simplicity. However, a WMD may have multiple heterogeneous sensors making the system more complex. Hence more complex WMD system should be analyzed.

3. Third, we have assessed our BN with a limited dataset. In order to effectively validate our method we need to assess our approach with datasets for diverse use cases (e.g., normal and noisy conditions).

4. Fourth, a fundamental assumption in our BN structure is the conditional independence of the non-descendant nodes. While this assumption simplifies the model and computational requirements, it may not accurately reflect the true dependencies in complex WMD systems. In reality, some non-descendant nodes may exhibit latent or indirect relationships not captured by this assumption, particularly in cases where non-descendant nodes share common causes or influences. For example, consider "Energy" (X_1) and "Data Storage" (X_9) in Fig. 4.4, although modeled as conditionally independent, in reality, high energy consumption could impact device operation and indirectly affect data storage management, as power constraints may limit the ability to store or process data efficiently. This interaction reflects an underlying dependency that isn't explicitly captured in our current BN framework. To mitigate this kind of interdependent factors, the BN can be refined further to account for such dependencies, by incorporating additional nodes.

4.7 Summary

In this chapter,we presented a formal framework for trust quantification using a BN. We presented the BN structure for a WMD (with a single sensor). Within the framework, we identified various dimensions and composition factors from literature and our empirical study in Chapter 3, to compactly represent the BN, for trust inference in a WMD. To estimate the parameters for the nodes of the BN, we assumed Gaussian distributions for all priors. We then computed the probability of the trust factor (or any other factor), for a given state.

We evaluated our approach for three different cases (weak, average, and strong). Evaluation

metrics such as expected mean and AUC were used to test and compare the model results to demonstrate the applicability of our approach. We also assessed the scalability of the proposed model by increasing the number of nodes.

The formal method introduced in this chapter, used Gaussian priors for each node which might not truly represent the distribution of trust factors. Moving forward, the next chapter will explore a data-driven approach to estimating Bayesian parameters using data collected from WMDs with multiple heterogeneous sensors. This approach aims to provide a more accurate representation of the priors and parameters for all nodes, enhancing trust quantification.

Chapter 5

Data-Driven Approach to Quantify Trust

In this chapter, we continue the work introduced in the previous chapter of quantifying trust in WMD using BN but we shift to a data-driven approach to estimate the parameters of the BN. This approach derives the probability of a trust factor being in a particular state with data collected from the devices, such as sensor quality. It incorporates expert knowledge to define the strength of the relationship between trust factors, integrating this knowledge into the model. We apply propagation rules from requirements engineering to determine the contribution of each trust factor to related intermediate nodes in the network, ultimately calculating the relative score of trustworthiness between different WMDs under identical test conditions. This score offers stakeholders, like Alex and his physician (from our motivating scenario), a measure of trustworthiness to compare and choose between WMDs.

The structure of this chapter is as follows: Section 5.1 introduces the data-driven approach. Section 5.2 details the methodology for our proposed data-driven Bayesian parameter estimation and the construction of a comprehensive BN structure (comprising multiple heterogeneous sensors) to evaluate the data-driven approach experimentally. The experimental setup and evaluation methodology are described in Section 5.3, with the experimental results presented in Section 5.4. The discussions of these results are in Section 5.5. The chapter concludes with a summary in Section 5.6. The author brings to the reader's attention that the contents of this chapter are

published in [135] and [136].

5.1 Introduction to Data-Driven Approach

As described earlier, a BN has two main parts: the structure and the parameters. In Chapter 4, we describe the development of the BN structure. We estimated the parameters and computed the probability of trust (or probability of any other descendant node) assuming a Gaussian distribution for the priors of the nodes in the network. However, Gaussian distribution may not provide accurate representations of the priors for nodes specifically for complex structures like the BNs of the WMD. A data-driven approach uses real data to learn the parameters of the network, improving objectivity and accuracy [71].

There are various approaches for parameter learning of the BN in the literature, the commonly used ones are maximum likelihood estimation (determines parameters by maximizing the likelihood of the observed data under the model) and Bayesian parameter estimation (incorporates prior knowledge and updates beliefs in light of new evidence to provide a probabilistic measure of the parameters).

In this chapter, we describe the Bayesian parameter estimation method, where the parameters of the BN are learned from the data collected from the WMDs. Using real data for parameter learning is a more reliable approach and significantly boosts the precision and reliability of the network by capturing complex, real-world relationships and allowing for dynamic updates and calibration [71, 57]. However, in our trust network, the data is available only for the non-descendant nodes. Therefore, our formalism should include methods for generating data for all intermediary nodes, e.g., generating indirect observations for reliability (X_5) using direct observations of X_1 and X_2 for our example BN in Fig. 4.3. In the following section, we describe the data-driven approach that we have developed for parameter estimation of the BN for a WMD.

5.2 Proposed Data-Driven BN Parameter Estimation

In this section, we describe the second step in the development of the BN framework, which is parameter estimation.

Bayesian Network Parameters

Bayesian network parameters are the set of conditional probability distributions that define the strength of the relationships between the random variables.

5.2.1 BN Parameter Estimation Overview

The Bayesian parameter estimation process uses statistical methods to estimate the CPDs from data, thereby allowing the network to accurately reflect how variables influence each other within the given dataset. To elucidate the development of our data-driven approach, we have refined the example BN with seven nodes, originally introduced in Chapter 4. This revised model incorporates multiple heterogeneous sensors, as depicted in Fig. 5.1(b). Additionally, we provide an example of a prior and conditional probability tables for the binary states of the nodes of this BN, which is illustrated in Fig. 5.1(c). The trustworthiness of a WMD from a stakeholder's perspective (as demonstrated in our motivating scenario) is represented in terms of the probability of trust to be in a specific state (e.g., low or high). Computing this probability requires access to the priors of all the nodes, which appear to be a major challenge [60]. In most cases, we do not have access to priors for all nodes.

For example, in the BN in Fig. 5.1 (b) the BN is structured so that the nodes in the first layer ($X_1 - X_4$) capture trust factors that are directly measurable from multiple sensors and other operational characteristics of the WMD (e.g., quality of the heart rate signals), while the remainder of the network represents the probability of subjective belief on the state of each refined attributes such as reliability (X_5) and robustness (X_6), and cannot be measured directly. As a result, usually, arbitrary values from experts [60] or random values assuming some Gaussian distribution are used as shown in our approach in Chapter 4 [134]. The latter can lead to over-fitting and biased objective quantification. The former requires experts to provide an exponential number of prior and posterior

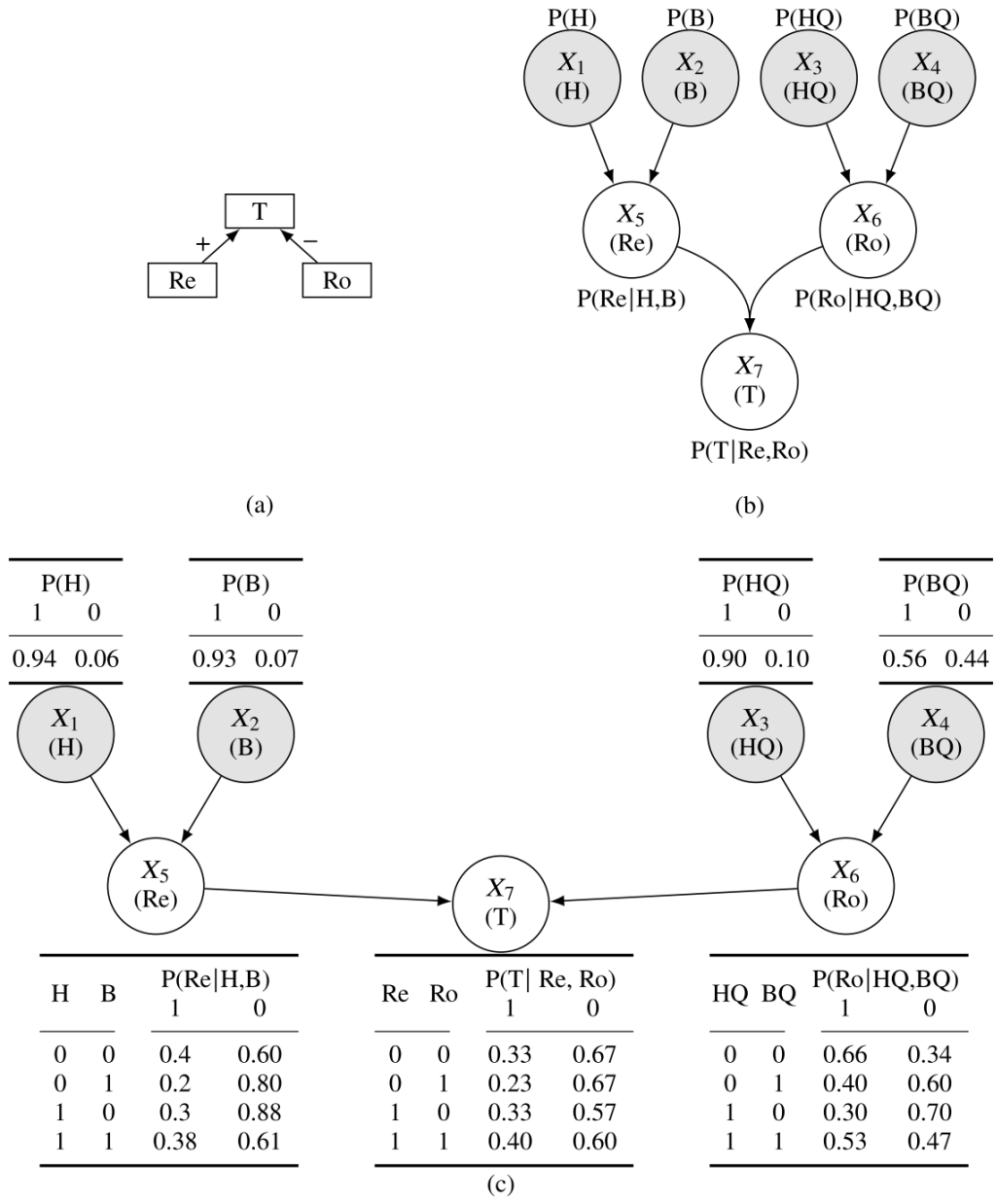

Figure 5.1: Trust Framework (a) Non-functional requirement framework for trust goal with reliability and robustness as sub-goals (b) Bayesian network to quantify trust (T) in WMD by analyzing heart (H) and breathing (B) rate sensor validation, heart (HQ) and breathing (BQ) rate quality, reliability (Re), and robustness (Ro), and (c) Bayesian network to quantify trust (T) in WMD with prior and conditional probability for each node © 2024 by the Society for Experimental Biology and Medicine

probabilities, which is impractical and may lead to expensive, biased, or contradictory estimates.

To bridge this gap, we propose a data-driven parameter estimation method for BN to quantify trust, that employs a hybrid approach, utilizing both measured and generated data. As shown in

Fig. 5.1 (c), our main goal is to compute the probability of X_7. We extract prior probabilities of the measurable attributes ($X_1 - X_4$) directly from the data collected from the WMD sensors. For example, in Fig. 5.1 (b), the probability that node X_1 is high (1) and low (0) (both the states of the nodes) is calculated from the data collected by the heart rate sensor of the WMD.

We use the requirements engineering NFR framework to express the relationships between the nodes [78]. The requirements engineering framework provides engineers with a mechanism to explicitly define their subjective views on the relationship between the factors when choosing a WMD. We capture the subjectivity of stakeholders' perspectives by asking the experts to provide the impact of each measurable attribute on its immediate intermediary node. For example, how X_1 and X_2 impacts reliability (X_5). Using this framework we identify the strength (impact) of the relationships between sub-goals (e.g., the relationship between reliability and trust) based on expert knowledge. The impact of the sub-goals can be expressed qualitatively and quantitatively with contribution links of varying strengths. For example, consider the structure in Fig. 5.1 (a), the sub-goals reliability and robustness have positive (+1) and negative (-1) impacts on trust. The expert-defined strength between sub-goals is then quantified and aggregated using propagation rules [78]. In this way, instead of expecting experts to estimate numerous probabilities, their subjective views are captured by a minimal number of qualitative labels.

Considering our example BN in Fig. 5.1(b) the nodes represent the probabilities of the NFR goal (trust) and sub-goals, which we refer to as factors. Factors are measurable or non-measurable. Measurable factors are observations directly measured from the system (e.g., sensors). Non-measurable factors are subjective aspects of the system (e.g., robustness). The BN is structured so that the nodes in the first layer capture directly measurable trust factors. For example, in Fig. 5.1(b), X_1 and X_2 represent the probability of the measurable trust factors heart and breathing rate sensor validation. The probability X_1 and X_2 is high when the WMD (e.g., Astroskin) sensors measurement agrees with a validating device (e.g., Apple Watch). The nodes X_3 and X_4 represent the probability of the heart and breathing rate sensor quality. The probability that X_3 and X_4 are high is calculated by comparing the measured observations to the manufacturer's specifications. The remainder of the network represents the probability of the subjective belief on the state of

each refined non-measurable trust factor, such as reliability (X_5) and robustness (X_6). Then, the trustworthiness (X_7) of a WMD is represented in terms of the probability of trust to be in a specific discrete state (e.g., low or high).

Following the terminologies in [71], the parameters in this thesis are represented as $\boldsymbol{\theta}$, which is a set of CPDs for the descendant nodes of the BN. For example in Fig. 5.1, $\boldsymbol{\theta}$ is the CPDs of nodes X_5, X_6, and X_7, and θ_i represents the CPD of a particular node i.

5.2.2 Granular Discretization of Nodes

The first step of our data-oriented parameter estimation approach is to compute the discrete probability distributions for the measurable trust factors in the BN. We discretize the continuous observations from the WMD to mutually discrete levels (e.g., discretize heart rate quality as low or high), which we refer to as granular discretization. Let y be a constant integer that represents the number of qualitative levels of an observed or inferred evidence in the BN, i.e., values of every node in the BN are always mapped to a fixed set of discrete values, $k = 1, \ldots, y$. For every i representing a non-descendant node, let $S_i^q = S_i^1, \ldots, S_i^m$ represent the raw samples of m number of observations, where $q = 1, \ldots, m$ is the number of non-discretized observed values of S_i^q. Then the equivalent discretized values of S will be computed as:

$$D^q_{i(non-descendant)} = \begin{cases} 1 & \text{if } l_{i,1} \leq S_i^q < u_{i,1} \\ 2 & \text{if } l_{i,2} \leq S_i^q < u_{i,2} \\ \vdots & \\ y & \text{if } l_{i,y} \leq S_i^q < u_{i,y} \\ 0 & \text{otherwise,} \end{cases} \tag{5.2.1}$$

where D_i^q is the discrete values with operational range values and $l_{(i,k)}$ and $u_{(i,k)}$ are the lower and upper thresholds for each discrete level of S_i^q, respectively. Lower and upper thresholds are defined as:

$$l_{i,k} = \min(O_i) + \Delta H \cdot k \quad \text{and} \quad u_{i,k} = l_{i,k} + \Delta H, \tag{5.2.2}$$

where $O_{(i)} = \{\min, \max\}$ is the set of the minimum and maximum operating values with user-defined thresholds based on the manufacturer's datasheet or a validation device, $\min(O_i)$ is the minimum threshold value, $\max(O_i)$ is the maximum threshold value, and $\Delta H = (\max(O_i) - \min(O_i))/y$ is the step size between the levels. For example, in Fig. 5.1(b), if the raw heart rate validation X_1 is $S_1^q = \{20, 65, 114, 1\}$, the discrete sample observation for $y = 5$ is $D_1 = \{1, 3, 5, 1\}$ with $l_{(1,1)} = 1$ and $u_{(1,1)} = 23.6$; $l_{(1,2)} = 23.6$ and $u_{(1,2)} = 46.2$; $l_{(1,3)} = 46.2$ and $u_{(1,3)} = 68.8$; $l_{(1,4)} = 68.8$ and $u_{(1,4)} = 91.4$; $l_{(1,5)} = 91.4$ and $u_{(1,5)} = 114$; and 0 otherwise.

5.2.3 Data Generation of Descendant Nodes

After the measured observations of non-descendant nodes are discretized, we generate data for each immediate descendant node (non-measurable trust factor) that ultimately contributes to the trust goal. As in requirements engineering, domain experts qualitatively describe the impact of the relationships between related NFRs, in our approach, we incorporate expert knowledge on factor impacts into the edges of our BN. For example, impacts are described as strongly positive (++), positive (+), neutral (O), negative (-), and strongly negative (- -) [78]. We then replace the qualitative contributions with numerical weight coefficients. The weight coefficient for the parent of node i are defined as:

$$W^q_{\mathrm{Pa}(i),j} = \begin{cases} 2^p & \text{positive impact} \\ -(2^p) & \text{negative impact} \\ 1 & \text{neutral impact,} \end{cases} \tag{5.2.3}$$

where $p = 1, \ldots, P$ is the strength of the impact where 1 is lowest and P is the strongest impact. In Fig. 5.1b, the impact of heart and breathing rate on reliability is qualitatively given as positive (++, $p = 2$) and neutral (O, $p = 0$). The impact is then quantified with coefficients (Eq. 5.2.3), $W^q_{\mathrm{Pa}(5),1} = 2^2 = 4$ and $W^q_{\mathrm{Pa}(5),2} = 2^0 = 1$, respectively.

Propagation rules are then used to aggregate the contributions of the sub-goals in the requirements engineering framework to the overall goal. We use a propagation rule to aggregate data for the descendant nodes based on discretized values of the measured observations of the non-descendant nodes. To propagate the impacts, we aggregate the weighted elements of the immediate parent nodes to generate the data for the descendant node as follows:

$$S^q_{i(descendant)} = \sum_{j=1}^{T_i} (W^q_{\mathrm{Pa}(i),j} \cdot D^q_{\mathrm{Pa}(i),j}), \quad \forall q = 1, \ldots, m, \tag{5.2.4}$$

where T_i are the number of parents of descendant node i. We can generalize this approach of aggregation to any discrete level in Eq. 5.2.3. For example, if $D_2 = \{2, 4, 3, 1\}$, then multiplying the weight coefficients $W^q_{\mathrm{Pa}(5),1} = 4$ and $W^q_{\mathrm{Pa}(5),2} = 1$, we aggregate to get $S_5 = W^q_{Pa(5),1} \cdot D^q_{5,1} + W^q_{\mathrm{Pa}(5),2} \cdot D^q_{5,2} = 4 \times \{1, 3, 5, 1\} + 1 \times \{2, 4, 3, 1\} = \{6, 16, 23, 5\}$. We compute $S^q_{i(descendant)}$ for each descendant node using Eq. 5.2.4. To get $D^q_{i(descendant)}$, the data $S^q_{i(descendant)}$ is discretized using Eq. 5.2.1. In our example, S_5 is discretized following Eq. 5.2.1 to $y = 5$ levels, and we get $D_5 = \{1, 4, 5, 1\}$ with $l_{5,1} = 5$ and $u_{5,1} = 8.6; l_{5,2} = 8.6$ and $u_{5,2} = 12.2; l_{5,3} = 12.2$ and $u_{5,3} = 15.8; l_{5,4} = 15.8$ and $u_{5,4} = 19.4; l_{5,5} = 19.4$ and $u_{5,5} = 23; 0$ otherwise.

An advantage of incorporating impact weights into our BN is that we can study the impact of different opinions on trust quantification. For example, in Fig. 5.1 (b), another domain expert gives the degree of strength for heart and breathing rate validated values as very positive (++, $p = 2$), so $W^q_{\mathrm{Pa}(5),1} = W^q_{\mathrm{Pa}(5),2} = 4$. Using Eq. 5.2.4, $S_{5(\mathrm{alternate})} = W^q_{\mathrm{Pa}(5),1} \cdot D^q_{5,1} + W^q_{\mathrm{Pa}(5),2} \cdot D^q_{5,2} = 4 \times \{1, 3, 5, 1\} + 4 \times \{2, 4, 3, 1\} = \{12, 20, 36, 4\}$. Note the difference between $S_5 = \{6, 14, 24, 1\}$ and $S_{5(\mathrm{alternate})} = \{12, 28, 32, 8\}$ due to alternate weight impacts, which would impact the computed overall trust score.

Now that we have data (observed and generated) for all nodes of our BN, we compute prior and posterior probabilities for all nodes. The prior probability for node X_i in our BN with multinomial dataset D_i that takes y discrete levels is computed as:

$$P(X^k_i) = \sum_{q=0}^{m} \left(\frac{D^q_i = k}{m} \right), \quad \forall k = 1, \ldots, y, \tag{5.2.5}$$

where $P(X_i^k)$ is the short form for $P(X_i = k)$. For example, the prior probabilities of nodes X_1 and X_2 (heart and breathing rate sensor validation), and X_5 (reliability) are $P(D_1^1) = 0.5; P(D_1^3) = P(D_1^5) = 0.25; P(D_2^1) = P(D_2^2) = P(D_2^3) = (D_2^3) = 0.25; P(D_5^1) = 0.5$; and $P(D_5^4) = P(D_5^5) = 0.25$.

Next, we estimate the parameter θ_i for each descendant node. Note that we now have discrete datasets for all nodes, i.e., $D_i[1], ..., D_i[m]$ for $i = 1, \ldots, n$. The structure of the BN allows us to reduce the parameter estimation to a set of unrelated (disjoint) problems. Let $P(D_i^k[1], ..., D_i^k[m])$ represent the joint probability of instances of node i to have value k. The joint distribution for the dataset of D_i^k and θ_i is then given as:

$$\begin{aligned} P(D_i^k[1], ..., D_i^k[m], \theta_i) &= P(D_i^k[1], ..., D_i^k[m]|\theta_i)P(\theta_i), \\ &= P(\theta_i)\prod_{q=1}^{m} P(D_i^k[m]|\theta_i) \quad \forall i = descendant, \end{aligned} \tag{5.2.6}$$

where $\prod_{q=1}^{m} P(D_i^k[m]|\theta_i)$ is the likelihood function $L(\theta_i : D_i^k)$ of the parameters, and $P(\theta_i)$ is the prior probability of θ_i in concern [71]. Then the posterior probability of θ_i given the instances of D_i^k is computed as:

$$P(\theta_i|D_i^k[1], ..., D_i^k[m]) = \frac{P(D_i^k[1], ..., D_i^k[m]|\theta_i)P(\theta_i)}{P(D_i^k[1], ..., D_i^k[m])}, \tag{5.2.7}$$

where $P(D_i^k[1], ..., D_i^k[m])$ is the marginal probability [71].

Following these equations, for our example (Fig. 5.1), we compute θ_5 for the descendant node X_5 (reliability) as $P(\theta_5|D_5^1[1], ..., D_5^1[5])$. Similarly, we can find the Bayesian parameters for descendant node X_6 (robustness) represented by $P(\theta_6|D_6^1[1], ..., D_6^1[5])$. We can then make inferences of trust (X_7), probability to be in a specific discrete state. For example, we compute the probability that trust is high (X_7^1), given reliability is high (X_5^1), and robustness is low (X_6^0) as $P(X_7^1|X_5^1, X_6^0)$.

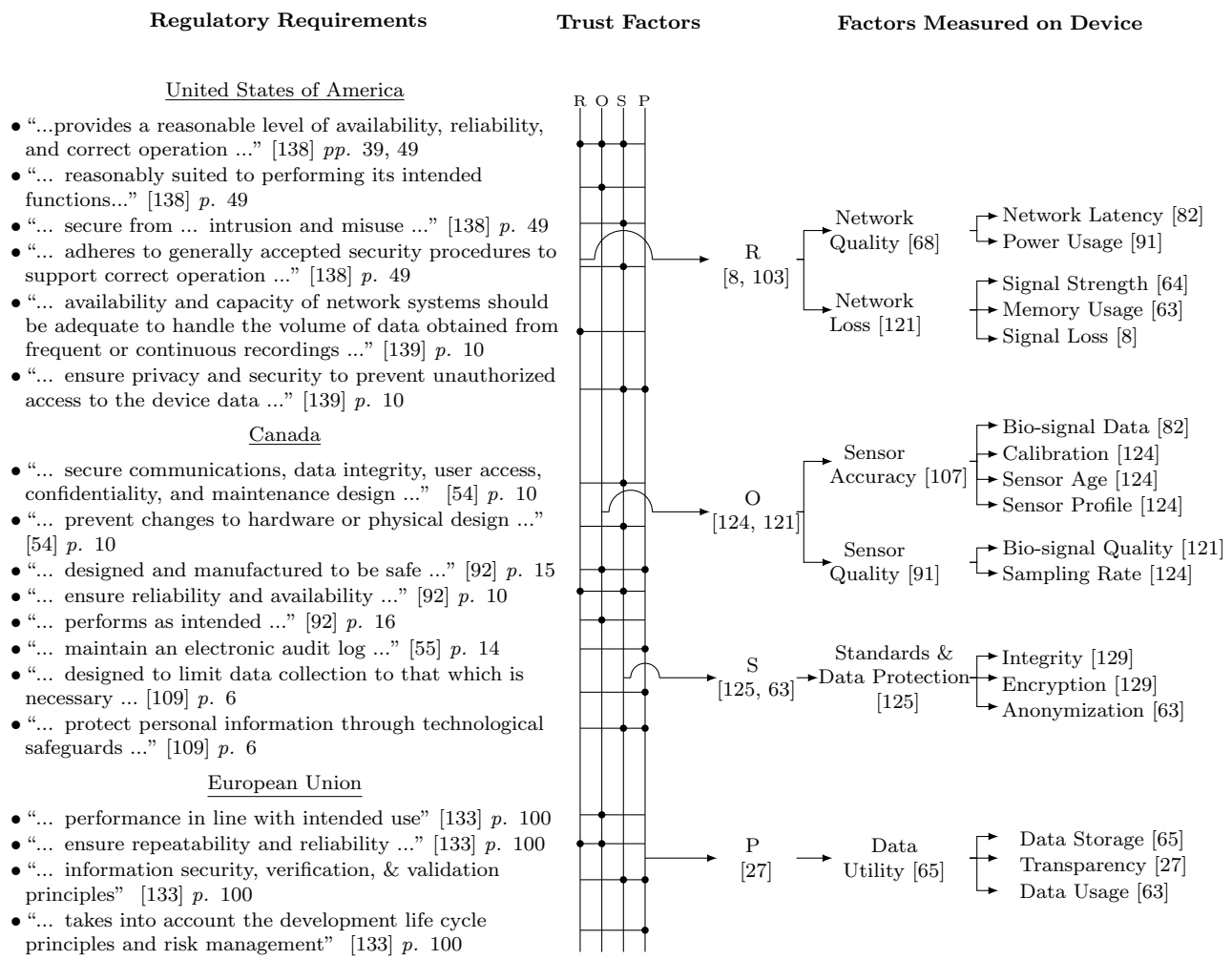

Figure 5.2: Mapping USA, Canada, and European Union regulatory requirements to trust factors from the literature that can be directly measured on the devices. Dots denote mapping between regulatory requirements and reliability (R), operations (O), security (S), and privacy (P) trust factors © 2024 by the Society for Experimental Biology and Medicine

5.2.4 BN Structure for Data-Driven Evaluation

In this section, we describe the construction of a comprehensive BN for a WMD (considering multiple heterogeneous sensors), to experimentally evaluate our data-driven parameter estimation approach. We extract factors from the regulatory standards of the American (USA) [138, 139], Canadian [54, 92, 55], and the EU [133] and map them with the factors in the literature and our empirical study (Chapter 3). We granularize the factors until we reach a level where the factors can be directly measured on a WMD, as shown in Fig. 5.2. The trust determining factors in this

study are not exhaustive due to domain dependency. Nevertheless, our proposed mapping enables stakeholders such as Alex or his family, to define their own trust network for the selection of a suitable WMD.

We identified four main trust-determining factors: reliability, operations, security and privacy. We refined reliability in terms of network quality and loss, to evaluate the efficiency [68, 82] and effectiveness [38, 121] of the network, respectively. For network quality, we measure network latency [82], and power consumption [91]. For network loss, we measure received signal strength [64], memory consumption [63] and the signal loss [8].

We refined operations in terms of sensor accuracy to assess if the sensor behaves according to the manufacturer's specifications [107, 91], and sensor quality to evaluate the recorded signal data quality compared to a baseline [91]. We further refined the sensor accuracy to measurable factors such as bio-signal data from multiple sensors [82], time since the last calibration [124], age of the sensor [124] and sensor profile [124]. For sensor quality, we measure the bio-signal quality data [121] and the sampling rate of the sensors [124].

We refined security in terms of standards and data protection to evaluate compliance with security protocols and confidentiality [125]. Standards and data protection were further refined to evaluate system encryption [129], integrity [129] (unauthorized parties do not change sensor data), and anonymization [63]. We refined privacy in terms of data utility to evaluate if the data is used only by authorized users and for its approved purpose [65]. Data utility was further refined to safe data storage [65], transparency [27] (data flow is visible to stakeholders) and authorized data usage [63].

Figures 5.3 and 5.4 present an instantiation of a subset of the trust factors mapped in Fig. 5.2. For our proof-of-concept BNs, we considered two main trust-determining factors: operations and reliability. The refined measurable factors forming the non-descendant nodes are the heart (X_1) and breathing (X_2) rate sensor validation; heart rate (X_3) and breathing rate quality (X_4); memory consumption and signal loss (X_5 and X_6) and power consumption and latency (X_7 and X_8). The instantiated trust factors forming the descendant nodes for the second layer of our hierarchical BN are sensor accuracy (X_9) and quality (X_{10}), network loss (X_{11}), and network quality (X_{12}).

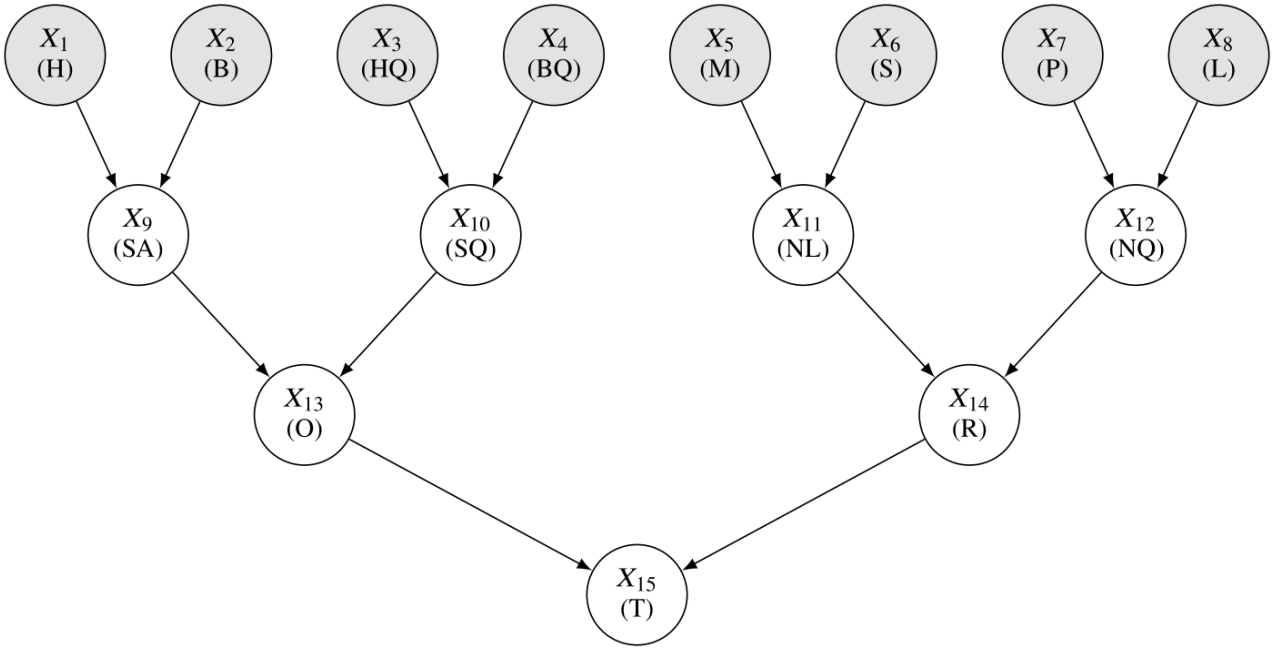

Figure 5.3: Homogeneous Bayesian network (BN3) to quantify trust (T) in WMD analyzing heart (H) and breathing (B) rate, heart (HQ) and breathing rate quality (BQ), signal loss (S), memory (M) and power usage (P), latency (L), sensor accuracy (SA) and quality (SQ), network loss (NL) and quality (NQ), operations (O), and reliability (R) © 2024 by the Society for Experimental Biology and Medicine

The descendant nodes of the second layer contribute to the trust-determining factors of operations (X_{13}) and reliability (X_{14}) in the third layer of the BN, which then determine the overall trust (X_{15}) of a WMD. Data for the descendant nodes are generated based on the impacts between the parent nodes defined by the experts (Eqs. 5.2.3 and 5.2.4).

Undoubtedly, this structure is subjective both in terms of structure and also in terms of the qualitative levels that we define to indicate the relative trustworthiness of a WMD. However, in this thesis, we show that if these subjective aspects are given by the domain experts, then the parameters of the entire structure can be estimated, i.e. the conditional probability of all descendant nodes, including the trustworthiness captured in the ultimate node of the BN. In this way, instead of domain experts subjectively guessing the trustworthiness of a device, they use observed data from the device's behaviour, which are informative but not fully indicative of the

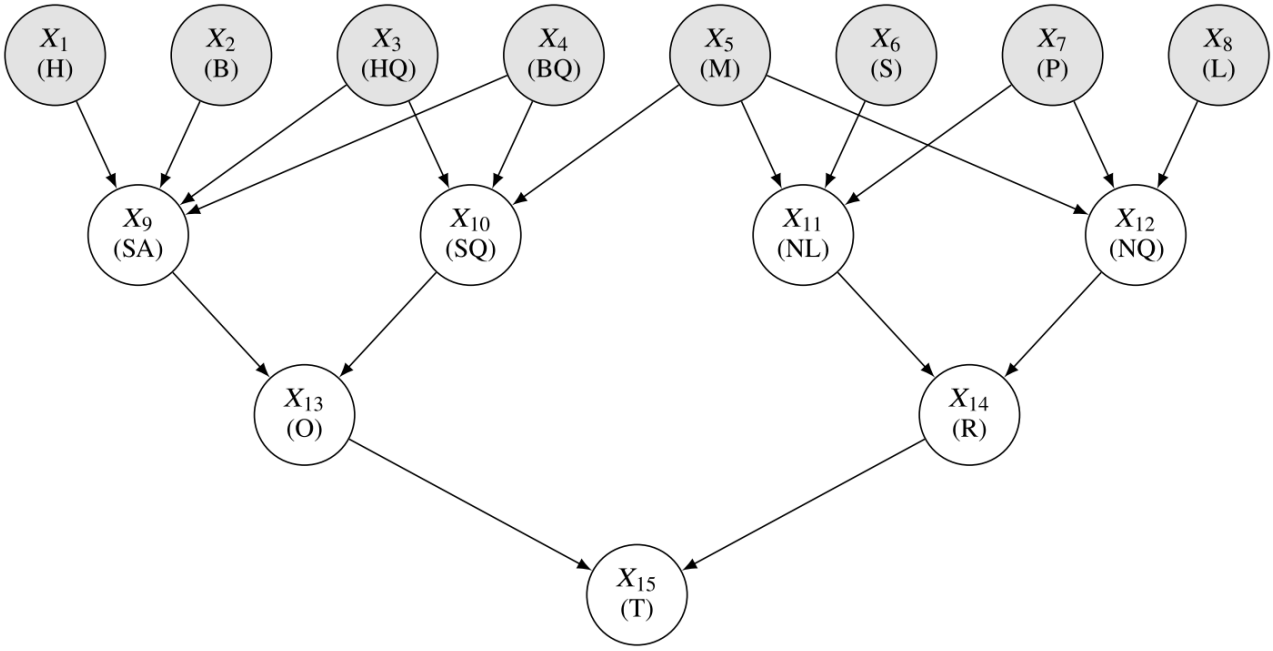

Figure 5.4: Heterogenous Bayesian network (BN4) to quantify trust (T) in WMD analyzing heart (H) and breathing (B) rate, heart (HQ) and breathing rate quality (BQ), signal loss (S), memory (M) and power usage (P), latency (L), sensor accuracy (SA) and quality (SQ), network loss (NL) and quality (NQ), operations (O), and reliability (R) © 2024 by the Society for Experimental Biology and Medicine

device's trustworthiness, to quantify and measure trust.

5.3 Experimental Evaluation

In this section, we evaluate our data-driven approach to estimate Bayesian parameters in terms of *learnability* and *generalizability*. For learnability, we determine how similar the relative trust scores are for the two (or more) WMDs under the same test conditions. We hypothesize that the trust score (denoted as X_{15}) will be similar across devices in the same usage conditions (e.g., walking in the evening). We also compare the BN quantified scores for individual nodes (e.g., reliability) between the WMDs. Additionally, we conducted a comparative analysis of proof-of-concept BNs to investigate the effect of different model configurations (impact weights, propagation rules, and homogeneous and heterogeneous nodes) on computing the trust scores.

For generalizability, we assess how reducing the number of data samples from 10,000,000 to 100 by a magnitude of ten affects the trust score. We also evaluate how the trust score is affected by increasing the noise level from 0% to 30% in 10% increments. The experimental analysis was conducted on a 64-bit Windows 10 laptop equipped with a 2.5 GHz Intel Core i9 CPU.

In addition to the real data collected from WMD, we also used synthetic data in this research to simulate a range of operational conditions that WMDs might encounter, allowing for comprehensive testing and validation beyond what is feasible with limited real datasets. This approach facilitates the exploration of device performance and trust assessment under controlled yet diverse scenarios, including varying levels of noise and environmental factors. By comparing the trust scores from synthetic data with those derived from real data, the research can validate the robustness and sensitivity of the BNs, ensuring they are capable of reliable predictions in real-world applications.

5.3.1 Test Environment

We outline the setup and conditions under which the data-driven parameter estimation and structural analysis for the BNs were conducted in terms of the specific WMDs utilized for data collection, the diverse use cases including the various levels of noisy conditions, and a comprehensive overview of the four BNs (BN1, BN2, BN3, and BN4) used for testing the data.

Wearable Medical Devices

We used three WMDs for data collection and validation:

1. Zephyr BioHarness 3.0 [157]
2. Astroskin Vital Signs Monitoring System [59]
3. Apple Watch Series 7 [11]

The Zephyr and Astroskin were used as test WMDs to evaluate the trust, while the Apple Watch was used as our validation device.

Use Cases

We evaluated the WMDs in two use cases:

1. Use Case 1 (UC1) - involved a set of predefined activities (sitting, standing, and walking at a speed of 4 MPH) for 30 minutes each, both indoors and outdoors, performed by three participants: one female (aged 45-55 years) and two males (aged 20-30 years). An *indoor condition* refers to activities performed in a room of 50 m $\times$ 50 m containing two or fewer people with wireless devices (e.g., phones, smartwatches, and laptops), excluding the participant. An *outdoor condition* refers to activities performed in a park containing at least ten people with wireless devices, excluding the participant. The outdoor activity times were in the morning (6-8 AM) and evening (6-8 PM) when the park had 10-12 and 40-50 people with wireless devices, respectively.

2. Use Case 2 (UC2) - involved daily activities (e.g., sleeping, sitting, working, and walking) performed by a female participant (aged 45-55 years) for two 24-hour periods under hybrid (indoor and outdoor) conditions. All participants wore the Astroskin on the right side of the lower abdomen, Zephyr on the left side of the upper abdomen, and Apple Watch on the left wrist simultaneously.

Bayesian Structure

We compared the learnability and generalizability performances for four BNs (BN1-BN4) for trust quantification. The configurations of BN1-BN4 are summarized in Table 5.1. BN1–BN3 have the same homogeneous BN structure as shown in Fig. 5.3 and use the same propagation rule as given in Eq. 5.2.4. In BN1, all edges have a uniform weight equal to one, denoted as $W^q_{\mathrm{Pa}(i),j} = 1$, and binary discretization levels. BN2 has uniform weights, but the nodes are granularized to six discrete levels. We use six discretized levels indicating the probability distribution of the nodes with levels: $y = \{\text{"invalid"}, \text{"very low"}, \text{"low"}, \text{"average"}, \text{"high"}, \text{"very high"}\} = \{0, 1, 2, 3, 4, 5\}$ based on the user-defined thresholds (as specified in Eq. 5.2.3 with $p = 5$). BN3 has granular discretization ($y = 6$), but weights are defined by expert knowledge. We generated Alternative A, B, and C (see Table 5.2) to investigate the impact of different weights on the BN performance.

In BN1–BN3, we use homogeneous BN structures as shown in Fig. 5.3, where we assume each node has two parents, which is not always the case. For example, network loss (X_{11}) is impacted by memory consumption (X_5), signal loss (X_6), and power consumption (X_7). To investigate the impact of interrelated parent nodes, we developed a heterogeneous BN4 with different impact weights (alternatives A, B and C) as shown in Fig. 5.4 with discretization levels $y = 6$ and the same propagation rule as in Eq. 5.2.4.

The BNs were developed in a Python environment using the bnlearn package (to construct the BNs efficiently and learn the parameters), matplotlib and graphviz (for visualization), and pandas (for data handling). Network parameters were generated based on datasets under different conditions to assess the performance and validate its generalizability and robustness in varying operational scenarios.

Table 5.1: Features of the four Bayesian network configurations (BN1, BN2, BN3, and BN4) © 2024 by the Society for Experimental Biology and Medicine

Feature	**BN1**	**BN2**	**BN3**	**BN4**
Discretization Levels (y)	2	6	6	6
Impact Weight ($W^q_{\mathrm{Pa}(i),j}$)	Uniform (1)	Uniform (1)	Alt. A to C (Table 5.2)	Alt. A to C (Table 5.2)
BN Nodes	Homogeneous	Homogeneous	Homogeneous	Heterogeneous

5.3.2 Real Data Collection Using WMD

We used the Zephyr BioHarness 3.0 device [157] and Astroskin Vital Signs Monitoring System [59] to quantify trust in Use Cases 1 and 2. We collected measured observations $\{S^1_1, \ldots, S^m_8\}$ for the non-descendant nodes $\{X_1, \ldots, X_8\}$ of our BN from each WMD for Use Cases 1 and 2. The raw samples of the measured observations were discretized with user-defined thresholds from the datasheet to $y = 6$ discrete levels using Eqs. 5.2.1 and 5.2.2. Datasets D^q_1 and D^q_2 (corresponding to X_1 and X_2) were obtained by validating the heart and breathing rate sensor observations with the Apple Watch [11]. For D^q_3 and D^q_4 (corresponding to X_3 and X_4), we compared the heart and breathing rate quality level with a predetermined noise threshold ($l_{i,k}$) and ($u_{i,k}$) [157, 58]. For D^q_5 (corresponding to X_5), the memory consumed during use was compared with user-defined

Table 5.2: Alternative (Alt) qualitative and quantitative impacts for the homogeneous (BN3) and heterogeneous (BN4) networks for sensor accuracy (SA), sensor quality (SQ), network loss (NL), network quality (NQ), operations (O), reliability (R), and trust (T) nodes

Node	Impact Weights	BN3			BN4		
		Alt A	Alt B	Alt C	Alt A	Alt B	Alt C
SA	$W^q_{Pa_{(9),1}}$	++ (+4)	+ (+2)	++ (+4)	++ (+4)	+ (+2)	++ (+4)
	$W^q_{Pa_{(9),2}}$	+ (+2)	+ (+2)	++ (+4)	+ (+2)	+ (+2)	++ (+4)
	$W^q_{Pa_{(9),3}}$	N/A	N/A	N/A	++ (+4)	+ (+2)	O (+1)
	$W^q_{Pa_{(9),4}}$	N/A	N/A	N/A	- (-2)	+ (+2)	O (+1)
SQ	$W^q_{Pa_{(10),3}}$	+ (+2)	+ (+2)	O (+1)	+ (+2)	+ (+2)	O (+1)
	$W^q_{Pa_{(10),4}}$	- (-2)	+ (+2)	O (+1)	- (-2)	+ (+2)	O (+1)
	$W^q_{Pa_{(10),5}}$	N/A	N/A	N/A	O (+1)	– (-4)	- (-2)
NL	$W^q_{Pa_{(11),5}}$	++ (+4)	+ (+2)	- (-2)	++ (+4)	+ (+2)	- (-2)
	$W^q_{Pa_{(11),6}}$	++ (+4)	- (-2)	- (-2)	++ (+4)	- (-2)	- (-2)
	$W^q_{Pa_{(11),8}}$	N/A	N/A	N/A	O (+1)	- (-2)	– (-4)
NQ	$W^q_{Pa_{(12),5}}$	N/A	N/A	N/A	O (+1)	– (-4)	- (-2)
	$W^q_{Pa_{(12),7}}$	O (+1)	++ (+4)	O (+1)	O (+1)	O (+1)	O (+1)
	$W^q_{Pa_{(12),8}}$	++ (+4)	++ (+4)	+ (+2)	++ (+4)	++ (+4)	+ (+2)
O	$W^q_{Pa_{(13),9}}$	+ (+2)	++ (+4)	O (+1)	+ (+2)	++ (+4)	O (+1)
	$W^q_{Pa_{(13),10}}$	+ (+2)	+ (+2)	++ (+4)	+ (+2)	+ (+2)	++ (+4)
R	$W^q_{Pa_{(14),11}}$	- (-2)	+ (+2)	O (+1)	- (-2)	+ (+2)	O (+1)
	$W^q_{Pa_{(14),12}}$	+ (+2)	++ (+4)	- (-2)	+ (+2)	++ (+4)	- (-2)
T	$W^q_{Pa_{(15),13}}$	+ (+2)	+ (+2)	++ (+4)	+ (+2)	+ (+2)	++ (+4)
	$W^q_{Pa_{(15),14}}$	+ (+2)	++ (+4)	+ (+2)	+ (+2)	++ (+4)	+ (+2)

thresholds. For D^q_6 (corresponding to X_6, signal loss), we calculated the amount of null or missing data during use [158, 72]. For D^q_7 (corresponding to X_7, power usage), we compared the battery level with a predetermined threshold [39]. For D^q_8 (corresponding to X_8, latency), we compared the delay with the user threshold for the application.

Sixteen datasets were collected for this study. In Use Case 1, we collected a dataset for each of the three participants wearing WMDs in indoor and outdoor conditions (six datasets each). In Use Case 2, we collected two datasets for one participant wearing WMDs in hybrid conditions over

two 24-hour periods (four datasets). The local research ethics committee approved the study and all subjects consented.

5.3.3 Synthetic Data Generation

In order to train and validate our data-driven model effectively, a larger collection of real WMD datasets is required. However, there is a scarcity in obtaining appropriate datasets. This scarcity is multifaceted, rooted in privacy concerns, regulatory restrictions, and the logistical challenges of collecting and annotating vast amounts of data.

The generation of synthetic datasets engineered to replicate the intricate patterns and correlations present in real data helps circumvent the challenge and serves as a proxy for algorithm development and testing [69]. This method enables a sustainable, scalable approach to data procurement for healthcare applications, particularly in the realm of wearable devices. The application of synthetic data, in our work, is pivotal to leveraging and fine-tuning the node relationships, ensuring the reliability of the BN under diverse conditions. Moreover, synthetic data significantly enhances the robustness of our model by expanding the variety of data scenarios available for training. This expansion allows for thorough testing and validation of device performance across rare and diverse conditions, including hypothetical scenarios that are difficult to replicate in real life. Consequently, this leads to improvements in the accuracy and reliability of the Bayesian Network (BN) framework [26, 152].

In the context of complex datasets such as those from WMDs, selecting appropriate synthetic data generation methods is essential. Different techniques, such as Variational Autoencoders (VAE), differential privacy, and Generative Adversarial Networks (GAN) methods, used to generate synthetic data for complex systems and relationships were studied. The VAEs offer a probabilistic approach to data generation, allowing for the creation of new data points within the distribution of existing data; however, it lacks precision for complex datasets [69]. The differential privacy method, on the other hand, offers strong privacy guarantees but may compromise data utility [1]. GANs are known to generate high-fidelity data [46]. However, they require careful tuning and can be computationally intensive.

In this thesis, we generate synthetic data using GANs. The process of generating synthetic data using GANs involves several critical parameters that are meticulously tuned to ensure high-quality data. These include the architecture of the generator and discriminator networks, the choice of activation functions, learning rates, and the optimization algorithm. Proper tuning of these parameters is essential to balance the capabilities of the generator and discriminator, thereby ensuring the diversity and statistical properties of the generated dataset [46]. The architecture of the generator and discriminator networks, including the depth and width of the layers, significantly influences the GAN's ability to learn and replicate the data distribution. The choice of activation function, such as the Rectified Linear Unit (ReLU), impacts the network's learning dynamics by introducing non-linearity and preventing the problem of vanishing gradient.

We implemented the GAN for synthetic data generation within the Python environment, leveraging the deep learning library TensorFlow. We used real WMD data collected under different conditions (indoor and outdoor) as our input dataset. The GAN is then trained to learn the patterns or regularities in the given input data in such a way that the model can be used to generate or output new synthetic data that plausibly is drawn from the original (input) dataset. We integrated the ReLU activation functions within the GAN architecture to integrate the non-linearity of the data, allowing the model to learn more complex patterns.

From an engineering standpoint, the model parameters were fine-tuned for an accurate representation of the original input dataset (using optimizers), expanded sample size (using dense layers, epochs, and batch size), and categorical data resembling our input dataset (using activation function). Additionally, we also generated data under various levels of noisy conditions to test and validate our BN model.

Noisy Conditions

We computed the trust score under noise levels from 0% to 30% with synthetic datasets for the two WMDs. We made a comparative study between the two devices (Astroskin and Zephyr) for Use Case 1 (outdoor), with real and synthetic data with varying noise levels for BN1. The noise represents missing values from signal loss due to connectivity problems, which is a common concern

of WMD users [25, 63, 147]. The BN had a uniform weight equal to one for all edges, denoted as $W^q_{\mathrm{Pa}(i),j} = 1$, and binary discretization levels.

5.4 Experimental Results

In this section, we present the findings from two distinct sets of experiments designed to test and validate four BNs (BN1, BN2, BN3, and BN4) across two use cases (UC1 and UC2) for two WMDs, Astroskin and Zephyr. The first set of experiments utilizes real data to assess the trustworthiness of these devices under controlled conditions. To extend our analysis and test the validity and applicability of our Bayesian approach, we replicate these tests using a synthetic dataset. The code and datasets for the experiments are publicly available on GitHub[1].

5.4.1 Analysis of Model with Real Sensor Data

Table 5.3: Inference scores (average ± standard deviation) of trust and reliability (R) for BN1 using real data from Astroskin (A) and Zephyr (Z) for Use Case 1 (indoor and outdoor) and Use Case 2 (Hybrid) for n=3

Inference	**WMD**	**Indoor (n=3)**	**Outdoor (n=3)**	**Hybrid (n=2)**
Trust $P(X^1_{15} \mid X^1_{13}, X^1_{14})$	A	0.960 ± 0.004	0.690 ± 0.012	0.670 ± 0.063
	Z	0.810 ± 0.012	0.561 ± 0.016	0.780 ± 0.007
Trust $P(X^1_{15} \mid X^0_{13}, X^0_{14})$	A	0.500 ± 0.009	0.540 ± 0.018	0.331 ± 0.008
	Z	0.437 ± 0.004	0.356 ± 0.004	0.464 ± 0.060
R $P(X^1_{14} \mid X^1_{11}, X^1_{12})$	A	0.860 ± 0.017	0.886 ± 0.004	0.647 ± 0.007
	Z	0.802 ± 0.090	0.705 ± 0.098	0.889 ± 0.009

Tables 5.3 to 5.6 show the inference scores for BN1 to BN4, respectively. The average trust score for Astroskin was higher than Zephyr in Use Case 1 but lower in Use Case 2. Our results show Astroskin was more trustworthy during short-duration activities (Use Case 1). Conversely, Zephyr was more trustworthy during long-duration activities (Use Case 2).

Table 5.3 shows the inference scores for BN1. The probabilities for the overall trust descendant node X_{15} for low and high states for the parents' operations (X_{13}) and reliability (X_{14}) are shown.

[1] https://github.com/tailabTMU/TrustQ_MD

Table 5.4: Inference scores (average ± standard deviation) of trust and reliability (R) for BN2 using real data from Astroskin (A) and Zephyr (Z) for Use Case 1 (indoor and outdoor) and Use Case 2 (Hybrid) for n=3

Inference	**WMD**	**Indoor (n=3)**	**Outdoor (n=3)**	**Hybrid (n=2)**
		Trust		
$P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$	A	0.370 ± 0.030	0	0
	Z	0.320 ± 0.040	0	0
$P(X_{15}^5 \mid X_{13}^4, X_{14}^4)$	A	0.752 ± 0.040	0.530 ± 0.016	0.650 ± 0.012
	Z	0.729 ± 0.140	0.660 ± 0.029	0.760 ± 0.020
$P(X_{15}^5 \mid X_{13}^3, X_{14}^3)$	A	0.786 ± 0.012	0.830 ± 0.016	0.890 ± 0.016
	Z	0.723 ± 0.020	0.860 ± 0.029	0.870 ± 0.029
$P(X_{15}^5 \mid X_{13}^2, X_{14}^2)$	A	0.323 ± 0.040	0.282 ± 0.070	0.062 ± 0.070
	Z	0.282 ± 0.100	0.231 ± 0.004	0.041 ± 0.004
$P(X_{15}^5 \mid X_{13}^1, X_{14}^1)$	A	0.003 ± 0.080	0	0
	Z	0	0	0
$P(X_{15}^5 \mid X_{13}^0, X_{14}^0)$	A	0	0	0.050 ± 0.008
	Z	0	0	0
		Reliability (R)		
$P(X_{14}^5 \mid X_{11}^3, X_{12}^3)$	A	0.276 ± 0.117	0.547 ± 0.007	0.477 ± 0.007
	Z	0.254 ± 0.010	0.649 ± 0.090	0.541 ± 0.090

The probability that trust is high (X_{15}^1), given the parents operations and reliability are high (X_{13}^1 and X_{14}^1) is $P(X_{15}^1 \mid X_{13}^1, X_{14}^1) \approx 0.9$ and ≈ 0.7 for both WMDs in Use Cases 1 and 2, respectively. The trust score $P(X_{15}^1 \mid X_{13}^0, X_{14}^0) > 0.3$ over all test cases, which was higher than hypothesized (≈ 0).

Table 5.4 shows the average inference score for BN2 for all impact levels of the parent nodes. With six granularized levels, the probability space for our BN is 2,016 as compared to 448 for binary granularization. The trust scores for $P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$ were ≈ 0.35 for both WMDs, despite our expectation that the probability of the devices should be large (≈ 0.9) given the high levels of the parents. The trust scores for $q = 3, 4$ (average and high) of the parents were ≈ 0.7 to 0.8, and for $q \leq 2$ (invalid, very low, and low) were ≈ 0 to 0.3.

Table 5.5 shows the average inference score for BN3 Alternatives A-C. The trust scores for BN3 (user-defined impacts) were higher than the scores of BN2 (uniform impact) for Use Cases 1 and 2. The average trust score for Alternative B was higher than Alternatives A and C for all WMDs

Table 5.5: Inference scores (average ± standard deviation) of trust and reliability (R) for BN3 Alternatives A-C using real data from Astroskin (A) and Zephyr (Z) for Use Case 1 (indoor and outdoor) and Use Case 2 (Hybrid) for n=3

Use Case	**WMD**	**Alt A**	**Alt B**	**Alt C**
		Trust $P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$		
Use Case 1 (In)	A	0.810 ± 0.016	0.853 ± 0.002	0.856 ± 0.004
	Z	0.650 ± 0.014	0.831 ± 0.012	0.482 ± 0.020
Use Case 1 (Out)	A	0.390 ± 0.020	0.340 ± 0.004	0.135 ± 0.002
	Z	0.476 ± 0.004	0.416 ± 0.012	0.168 ± 0.012
Use Case 2	A	0.305 ± 0.040	0.288 ± 0.020	0.260 ± 0.020
	Z	0.430 ± 0.007	0.445 ± 0.021	0.430 ± 0.007
		Trust $P(X_{15}^5 \mid X_{13}^3, X_{14}^3)$		
Use Case 1 (In)	A	0.790 ± 0.018	0.798 ± 0.005	0.730 ± 0.040
	Z	0.830 ± 0.009	0.731 ± 0.020	0.795 ± 0.009
Use Case 1 (Out)	A	0.750 ± 0.040	0.860 ± 0.014	0.790 ± 0.003
	Z	0.707 ± 0.005	0.824 ± 0.012	0.800 ± 0.034
Use Case 2	A	0.635 ± 0.030	0.695 ± 0.007	0.535 ± 0.020
	Z	0.790 ± 0.049	0.810 ± 0.014	0.831 ± 0.028
		Trust $P(X_{15}^5 \mid X_{13}^0, X_{14}^0)$		
Use Case 1 (In)	A	0	0.020 ± 0.004	0
	Z	0	0	0
Use Case 1 (Out)	A	0.001 ± 0.002	0	0
	Z	0.007 ± 0.020	0	0
Use Case 2	A	0.021 ± 0.026	0.075 ± 0.007	0.025 ± 0.028
	Z	0.060 ± 0.003	0.018 ± 0.002	0.160 ± 0.197
		Reliability (R) $P(X_{14}^5 \mid X_{11}^3, X_{12}^3)$		
Use Case 1 (In)	A	0.526 ± 0.119	0.605 ± 0.007	0.605 ± 0.117
	Z	0.816 ± 0.001	0.723 ± 0.005	0.880 ± 0.016
Use Case 1 (Out)	A	0.526 ± 0.119	0.713 ± 0.117	0.117 ± 0.127
	Z	0.516 ± 0.001	0.580 ± 0.005	0.816 ± 0.016
Use Case 2	A	0.426 ± 0.119	0.410 ± 0.012	0.596 ± 0.110
	Z	0.496 ± 0.011	0.480 ± 0.015	0.722 ± 0.006

and Use Cases.

Table 5.6 shows the result of inference scores for BN4 with heterogeneous parents. We hypothesized that heterogeneous nodes would change the trust computation. However, our results show

Table 5.6: Inference scores (average ± standard deviation) of trust and reliability (R) for BN4 Alternatives (Alt) A - C using real data from Astroskin (A) and Zephyr (Z) for Use Case 1 (indoor and outdoor) and Use Case 2 (Hybrid) for n=3

Use Case	**WMD**	**Alt A**	**Alt B**	**Alt C**
		Trust $P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$		
Use Case 1 (In)	A	0.400 ± 0.004	0.420 ± 0.016	0.416 ± 0.002
	Z	0.420 ± 0.016	0.380 ± 0.020	0.309 ± 0.012
Use Case 1 (Out)	A	0.370 ± 0.020	0.300 ± 0.004	0.380 ± 0.018
	Z	0.300 ± 0.020	0.330 ± 0.004	0.320 ± 0.020
Use Case 2	A	0.315 ± 0.006	0.380 ± 0.004	0.360 ± 0.020
	Z	0.390 ± 0.014	0.300 ± 0.007	0.310 ± 0.040
		Trust $P(X_{15}^5 \mid X_{13}^3, X_{14}^3)$		
Use Case 1 (In)	A	0.770 ± 0.032	0.800 ± 0.100	0.740 ± 0.036
	Z	0.860 ± 0.028	0.940 ± 0.020	0.900 ± 0.018
Use Case 1 (Out)	A	0.790 ± 0.033	0.770 ± 0.004	0.789 ± 0.010
	Z	0.710 ± 0.050	0.816 ± 0.050	0.780 ± 0.060
Use Case 2	A	0.615 ± 0.020	0.670 ± 0.020	0.670 ± 0.040
	Z	0.810 ± 0.014	0.820 ± 0.007	0.810 ± 0.021
		Trust $P(X_{15}^5 \mid X_{13}^0, X_{14}^0)$		
Use Case 1 (In)	A	0	0	0
	Z	0.002 ± 0.002	0	0.030 ± 0.030
Use Case 1 (Out)	A	0	0	0
	Z	0.015 ± 0.002	0.021 ± 0.020	0
Use Case 2	A	0	0	0
	Z	0	0	0
		Reliability (R) $P(X_{14}^5 \mid X_{11}^3, X_{12}^3)$		
Use Case 1 (In)	A	0.426 ± 0.119	0.470 ± 0.007	0.412 ± 0.117
	Z	0.516 ± 0.001	0.477 ± 0.005	0.476 ± 0.016
Use Case 1 (Out)	A	0.466 ± 0.119	0.518 ± 0.117	0.562 ± 0.127
	Z	0.411 ± 0.001	0.589 ± 0.005	0.506 ± 0.017
Use Case 2	A	0.333 ± 0.119	0.340 ± 0.012	0.322 ± 0.110
	Z	0.396 ± 0.011	0.380 ± 0.015	0.399 ± 0.006

that the effect of heterogeneous compared to homogeneous nodes was negligible for both WMDs in Use Cases 1 and 2.

In addition to computing the trust scores, we can compute the probability of any descendant

Table 5.7: Trust scores $P(X_{15}^5|X_{13}^4, X_{14}^4)$ for a sample dataset with different sample sizes for BN1 - BN2, and BN3 - BN4 (Alt B) for Use Case 1 (Indoors) with Astroskin (A) and Zephyr (Z)

Samples	**BN1**		**BN2**		**BN3 (Alt B)**		**BN4 (Alt B)**	
	A	**Z**	**A**	**Z**	**A**	**Z**	**A**	**Z**
10,000,000	0.889	0.873	0.854	0.821	0.760	0.724	0.858	0.830
1,000,000	0.889	0.873	0.854	0.821	0.760	0.724	0.858	0.830
100,000	0.860	0.867	0.820	0.810	0.740	0.723	0.833	0.800
10,000	0.850	0.861	0.780	0.770	0.718	0.700	0.800	0.780
1,000	0.850	0.834	0.780	0.755	0.717	0.690	0.790	0.768
100	0.590	0.550	0.449	0.443	0.420	0.410	0.500	0.590
10	0.520	0.500	0.455	0.400	0.400	0.380	0.490	0.460

nodes for BN1 to BN4, given the parents' discretized level. We present the computed score for the probability of reliability $P(X_{14}^1 \mid X_{11}^1, X_{12}^1)$ and $P(X_{14}^5 \mid X_{11}^3, X_{12}^3)$ in Table 5.3 and Tables 5.4 - 5.6 respectively. The reliability scores for Astroskin were higher than Zephyr in Use Case 1 but not in Use Case 2. The reliability performance scores were similar to the trust score for both WMDs in Use Cases 1 and 2.

Table 5.7 shows our results to assess generalizability. The results show that the trust score slightly dropped when we reduced the sample size from 10,000,000 to 100. The trust score dropped from ≈ 0.8 (1,000 samples) to ≈ 0.5 (ten samples). We observed the same performance when we decreased the sample sizes for BN1 to BN4 for both WMDs.

5.4.2 Analysis of Model with Synthetic Data

Our objective in generating synthetic data was two-fold. Firstly, we sought to validate our BN by utilizing an extensive and representative dataset. This allowed us to ensure that our model accurately reflected real-world scenarios and was robust enough to handle diverse data inputs. Secondly, we aimed to validate our model under various levels of noisy conditions by incorporating datasets with differing degrees of noise. This approach enabled us to assess the model's performance and reliability across a spectrum of potential noisy factors, thus enhancing its applicability and effectiveness in practical settings.

We generated synthetic data under four levels of noise: no noise (0%), low (10%), medium (20%), and high (30%). The synthetically generated data was then used in our four BNs (BN1 - BN4), to validate our model. For a comparative study, we selected a subset of our BNs: BN1, BN2, BN3 (Alt. A), and BN4 (Alt. A). We compared the probability of the trust node (X_{15}), for varied conditions of the parents being high/ low, for indoor and outdoor activities for both WMDs (Astroskin and Zephyr). Tables 5.8 to 5.11 shows the trust inference score for BN1 – BN4 with synthetic data.

Table 5.8: Inference scores of trust for BN1 using synthetic data for Astroskin (A) and Zephyr (Z) for Use Case 1 indoor and outdoor (UC1 - In and UC1 - Out) for n=3. The readings of the real data used for the generation of the synthetic data (for the two conditions) are added for reference

			Inference Score				
			Real	**Synthetic Data**			
Use Case	**WMD**	**Probability**	**Data**	**No Noise**	**10% Noise**	**20% Noise**	**30% Noise**
UC1-In	A	$P(X_{15}^1 \mid X_{13}^1, X_{14}^1)$	0.960 ± 0.003	0.980 ± 0.011	0.955 ± 0.033	0.910 ± 0.033	0.789 ± 0.053
		$P(X_{15}^1 \mid X_{13}^0, X_{14}^0)$	0	0	0	0	0.011 ± 0.033
	Z	$P(X_{15}^1 \mid X_{13}^1, X_{14}^1)$	0.880 ± 0.014	0.910 ± 0.003	0.881 ± 0.011	0.805 ± 0.017	0.642 ± 0.114
		$P(X_{15}^1 \mid X_{13}^0, X_{14}^0)$	0	0	0	0	0.009 ± 0.017
UC1-Out	A	$P(X_{15}^1 \mid X_{13}^1, X_{14}^1)$	0.764 ± 0.014	0.740 ± 0.003	0.693 ± 0.014	0.670 ± 0.063	0.535 ± 0.167
		$P(X_{15}^1 \mid X_{13}^0, X_{14}^0)$	0.022 ± 0.070	0	0	0.008 ± 0.015	0.017 ± 0.089
	Z	$P(X_{15}^1 \mid X_{13}^1, X_{14}^1)$	0.714 ± 0.011	0.700 ± 0.037	0.690 ± 0.011	0.639 ± 0.011	0.505 ± 0.091
		$P(X_{15}^1 \mid X_{13}^0, X_{14}^0)$	0	0	0	0.006 ± 0.010	0.054 ± 0.012

Table 5.8 shows the inference scores for BN1 for Use Case1 (indoor and outdoor) with synthetic data for Astroskin and Zephyr under various noise conditions revealing notable differences in trustworthiness between indoor and outdoor environments and between real and synthetic data. In the indoor settings, for $P(X_{15}^1 \mid X_{13}^1, X_{14}^1)$, Astroskin demonstrates high trust with a real data

Table 5.9: Inference scores of trust for BN2 using synthetic data for Astroskin (A) and Zephyr (Z) for Use Case 1 indoor and outdoor (UC1 - In and UC1 - Out) for n=3. The readings of the real data used for the generation of the synthetic data (for the two conditions) are added for reference

			Inference Score				
			Real	**Synthetic Data**			
Use Case	**WMD**	**Probability**	**Data**	**No Noise**	**10% Noise**	**20% Noise**	**30% Noise**
UC1-In	A	$P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$	0.378 ± 0.033	0.366 ± 0.016	0.318 ± 0.033	0.270 ± 0.057	0.218 ± 0.047
		$P(X_{15}^1 \mid X_{13}^0, X_{14}^0)$	0	0.021 ± 0.003	0.016 ± 0.023	0.033 ± 0.068	0.071 ± 0.007
	Z	$P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$	0.330 ± 0.044	0.297 ± 0.015	0.260 ± 0.019	0.200 ± 0.011	0.196 ± 0.036
		$P(X_{15}^5 \mid X_{13}^0, X_{14}^0)$	0	0	0.019 ± 0.010	0.028 ± 0.017	0.029 ± 0.016
UC1-Out	A	$P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$	0.544 ± 0.017	0.540 ± 0.023	0.503 ± 0.014	0.489 ± 0.077	0.389 ± 0.100
		$P(X_{15}^5 \mid X_{13}^0, X_{14}^0)$	0.022 ± 0.070	0	0.024 ± 0.018	0.073 ± 0.011	0.102 ± 0.090
	Z	$P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$	0.444 ± 0.004	0.468 ± 0.017	0.437 ± 0.011	0.419 ± 0.027	0.385 ± 0.031
		$P(X_{15}^5 \mid X_{13}^0, X_{14}^0)$	0	0	0.010 ± 0.027	0.026 ± 0.015	0.034 ± 0.022

score of 0.960, which slightly decreases in synthetic scenarios, especially as noise increases. Zephyr starts with a 0.880 score and experiences a similar decline under noisy conditions. Outdoors, both devices exhibit lower trust scores than indoors, with significant drops at 30% noise—Astroskin to 0.535 and Zephyr to 0.505. This indicates that environmental factors significantly impact device performance and trust assessments.

Table 5.9 shows the inference scores for BN2 with synthetic data for Astroskin and Zephyr in both indoor and outdoor settings (UC1) under varying noise conditions reveals a trend of decreasing trust scores as noise levels increase. For $P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$, in the indoor environment, Astroskin starts with a real data score of 0.378, which decreases to 0.218 at 30% noise, while Zephyr begins at 0.330 and declines to 0.196 under high noise scenarios. Outdoors, Astroskin's scores drop from 0.544 to 0.389, and Zephyr from 0.444 to 0.385 at 30% noise.

Table 5.10 shows the inference scores for BN3 with synthetic data. The analysis of BN3 for

Table 5.10: Inference scores of trust for BN3 Alt. A using synthetic data for Astroskin (A) and Zephyr (Z) for Use Case 1 indoor and outdoor (UC1 - In and UC1 - Out) for n=3. The readings of the real data used for the generation of the synthetic data (for the two conditions) are added for reference

			Inference Score				
			Real	**Synthetic Data**			
Use Case	**WMD**	**Probability**	**Data**	**No Noise**	**10% Noise**	**20% Noise**	**30% Noise**
UC1-In	A	$P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$	0.810 ± 0.016	0.820 ± 0.079	0.779 ± 0.056	0.753 ± 0.016	0.626 ± 0.004
		$P(X_{15}^1 \mid X_{13}^0, X_{14}^0)$	0	0.020 ± 0.004	0.036 ± 0.039	0.050 ± 0.013	0.091 ± 0.077
	Z	$P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$	0.650 ± 0.014	0.645 ± 0.011	0.623 ± 0.023	0.601 ± 0.201	0.545 ± 0.117
		$P(X_{15}^5 \mid X_{13}^0, X_{14}^0)$	0	0	0.010 ± 0.014	0.068 ± 0.055	0.079 ± 0.017
UC1-Out	A	$P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$	0.390 ± 0.003	0.400 ± 0.033	0.373 ± 0.074	0.315 ± 0.147	0.289 ± 0.109
		$P(X_{15}^5 \mid X_{13}^0, X_{14}^0)$	0.022 ± 0.070	0	0.045 ± 0.017	0.075 ± 0.011	0.110 ± 0.090
	Z	$P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$	0.476 ± 0.004	0.468 ± 0.017	0.417 ± 0.009	0.399 ± 0.070	0.315 ± 0.033
		$P(X_{15}^5 \mid X_{13}^0, X_{14}^0)$	0.040 ± 0.003	0	0.019 ± 0.022	0.029 ± 0.015	0.014 ± 0.038

Astroskin and Zephyr under various noise conditions reveals a decrease in trust scores as noise increases. For $P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$, indoors, Astroskin starts with a real data score of 0.810, reducing to 0.626 at 30% noise, while Zephyr begins at 0.650 and drops to 0.545 under similar conditions. Outdoors, the initial trust scores are lower, with Astroskin dropping from 0.390 to 0.289, and Zephyr from 0.476 to 0.315 as noise reaches 30%. The results indicate a strong sensitivity to noise, affecting the reliability of trust assessments and emphasizing the need to consider environmental factors in the quantification of trust for WMDs.

Table 5.11 shows the inference scores for BN4 with synthetic data. The evaluation of BN4 for Astroskin and Zephyr under different noise conditions illustrates their trustworthiness in various environments. For $P(X_{15}^5 \mid X_{13}^5, X_{14}^5)$, indoors, Astroskin starts with a real data inference score of 0.678, decreasing to 0.508 at 30% noise, while Zephyr's score reduces from 0.590 to 0.487 under

Table 5.11: Inference scores of trust for BN4 Alt. A using synthetic data for Astroskin (A) and Zephyr (Z) for Use Case 1 indoor and outdoor (UC1 - In and UC1 - Out) for n=3. The readings of real data used for the generation of the synthetic data (for the two conditions) are added for reference

			Inference Score				
			Real	**Synthetic Data**			
Use Case	**WMD**	**Metric**	**Data**	**No Noise**	**10% Noise**	**20% Noise**	**30% Noise**
UC1-In	A	$P(X_{15}^5 \mid X_{13}^5 X_{14}^5)$	0.678 ± 0.033	0.606 ± 0.016	0.578 ± 0.061	0.534 ± 0.033	0.508 ± 0.013
		$P(X_{15}^5 \mid X_{13}^0 X_{14}^0)$	0	0	0	0.043 ± 0.013	0.061 ± 0.031
	Z	$P(X_{15}^5 \mid X_{13}^5 X_{14}^5)$	0.590 ± 0.044	0.597 ± 0.015	0.560 ± 0.019	0.579 ± 0.011	0.487 ± 0.036
		$P(X_{15}^5 \mid X_{13}^0 X_{14}^0)$	0	0	0.010 ± 0.010	0.032 ± 0.017	0.059 ± 0.010
UC1-Out	A	$P(X_{15}^5 \mid X_{13}^5 X_{14}^5)$	0.604 ± 0.014	0.540 ± 0.023	0.503 ± 0.014	0.489 ± 0.077	0.389 ± 0.107
		$P(X_{15}^5 \mid X_{13}^0 X_{14}^0)$	0.002 ± 0.070	0	0	0.017 ± 0.055	0.010 ± 0.090
	Z	$P(X_{15}^5 \mid X_{13}^5 X_{14}^5)$	0.514 ± 0.011	0.500 ± 0.007	0.500 ± 0.011	0.509 ± 0.031	0.465 ± 0.091
		$P(X_{15}^5 \mid X_{13}^0 X_{14}^0)$	0	0	0.017 ± 0.017	0.036 ± 0.027	0.054 ± 0.012

similar conditions, showing their vulnerability to noise. Outdoors, both devices show an initial high trust score which significantly drops as noise increases, with Astroskin falling to 0.389 and Zephyr to 0.465 at 30% noise.

Table 5.12 presents a comprehensive summary of the percentage errors in inference scores for Astroskin and Zephyr, under (indoor and outdoor) conditions for UC1, across BN1 to BN4. The summary table indicates variations in performance under different noise conditions for both indoor and outdoor settings of UC1. Initial results show minimal errors in no noise scenarios, but significant errors emerge as noise increases, particularly in BN2 where errors reach up to 42.32% indoors for Astroskin. BN1 maintains lower error rates, suggesting greater stability. Zephyr exhibits heightened sensitivity to noise, with substantial errors, peaking at 40.60% in BN2 indoors.

Table 5.12: Comprehensive summary of percentage errors for inference scores of two wearable medical devices, Astroskin (A) and Zephyr (Z), for Use Case 1 (UC1), indoor and outdoor, across four Bayesian networks (BN1, BN2, BN3 (Alt. A), and BN4 (Alt. A)) using synthetic data compared with the real score for $P(X_{15}^5|X_{13}^5X_{14}^5)$

Use Case	Device	Data Type	BN1	BN2	BN3	BN4
Indoor	Astroskin	No Noise	2.08%	3.17%	1.23%	10.61%
		30% Noise	17.81%	42.32%	22.71%	25.07%
	Zephyr	No Noise	3.41%	10.00%	0.76%	1.18%
		30% Noise	27.04%	40.60%	16.15%	17.45%
Outdoor	Astroskin	No Noise	3.14%	0.75%	4.35%	10.59%
		30% Noise	29.97%	28.49%	25.81%	35.59%
	Zephyr	No Noise	1.96%	6.36%	1.68%	2.72%
		30% Noise	29.29%	13.28%	33.82%	9.53%

5.5 Discussions

Our BN provides a relative measure of trustworthiness between WMDs. We present a BN structure for trust quantification identified from our mapping of regulatory requirements, trust factors from the literature, and measurements from WMD. Using a BN structure enables stakeholders to define the trustworthiness factors used to assess WMD explicitly. Our data-driven approach can then estimate the parameters for any BN structure using the measured observations from the WMD instead of defining the parameters based on Gaussian distributions or expert knowledge. Our proposed BN provides a mechanism for the stakeholders in our motivating scenario to select a trustworthy WMD. The relative trust score helps Alex choose the WMD based on trustworthiness factors such as operations. Alex's doctor and the trainer also decide which WMD can be used for monitoring Alex's health during indoor and outdoor activities based on the WMD's sensor accuracy and reliability.

The results for BN1 for real and synthetic data, as depicted in Table 5.3 and 5.8 respectively, show that Astroskin and Zephyr exhibit high trust in indoor settings with real data, but experience a decrease in synthetic scenarios, particularly as noise increases. This reduction in trust scores in noisy environments highlights the sensitivity of the BN to environmental disturbances. The disparity between real and synthetic data outcomes emphasizes the necessity for robust validation

techniques that can accurately reflect real-world conditions. Also, the binary discretization used in real data analysis captures a broad range of values, leading to a higher probability for low discrete states, which demonstrates the need for more granular discretization levels to enhance the precision and applicability of the model. For example, if power consumption threshold limits are $60 \leq S_i^q \leq 100$, then the power consumption of 61% and 99% would both be discretized to high with binary discretization and the remaining range of values would be low. The wide range resulted in a higher probability for the low discrete state. Using more granularized levels ($y = 6$), the trust score for low, average, and high levels performed as expected.

For BN2, the real and synthetic data, as shown in Tables 5.4 and 5.9 respectively, shows a decline in trust scores with increasing noise levels across both real and synthetic data indicates the model's vulnerability to environmental noise. The analysis of BN3, for real and synthetic data as per Tables 5.5 and 5.10 respectively, illustrates that noise substantially impacts trust assessments, with both devices showing reduced trustworthiness under higher noise levels in both indoor and outdoor settings with the synthetic data. The real data results reflect the successful capturing of weight strengths in trust scores across different WMDs, indicating that BN3 can effectively utilize expert knowledge in modelling.

For BN4, real and synthetic data results detailed in Tables 5.6 and 5.11 respectively, reveal the decline in trust scores from indoor to outdoor settings and the sensitivity to increased noise levels for both Astroskin and Zephyr demonstrate the BN's responsiveness to environmental changes. The negligible effect on average trust scores across use cases and WMDs in real data suggests that while the model handles heterogeneity effectively, the significant drops under noisy conditions in synthetic data highlight its vulnerability.

The comprehensive analysis of all BNs presented in Table 5.12 emphasizes the critical challenges in the quantification of trust under variable environmental conditions. While minimal errors in no-noise scenarios offer a baseline validation of the networks' capabilities, significant discrepancies under high noise conditions, particularly in BN2, reveal the networks' sensitivity to environmental noise. This data suggests BN1's higher stability and Zephyr's elevated noise sensitivity, highlighting the critical challenge of accurately quantifying trust under variable environmental conditions.

The real data analysis reveals challenges related to fine granularization levels, which resulted in few samples per level and influenced the trust scores negatively for BN2- BN4. This outcome suggests that while granularization is good for fine discrete levels, an optimal level of granularization should be identified based on the data size and data quality.

The learned parameters for our BNs with real and synthetic data resulted in similar trust scores for both WMDs for different use cases as shown in Tables 5.3 to 5.6, and Tables 5.8 to 5.11. Our results indicate that although the measured observations from the two WMDs were unique to the device, the parameter estimation approaches produced similar results under the same conditions, thus demonstrating the learnability of our approach. The overall trust score was consistent across the different use cases for both WMDs. Table 5.7 demonstrates the generalizability of the Bayesian-based parameter estimation with fewer training samples compared to discriminative models [97]. Reducing the sampling had little effect on Bayesian parameter estimation because it is based on discrete distributions, not individual samples, for learning the parameters.

Synthetic data exhibits the highest similarity to real data when generated in a no-noise environment, potentially surpassing real data in capturing underlying patterns. As noise levels increase (10%, 20%, 30%), the synthetic data diverges more from the real data, causing deterioration in the performance of trust as seen in the percentage errors, under noise.

The comparative analysis of synthetic data generated by GANs, underscores the significant impact of noise on data quality, alongside the foundational influence of real data characteristics sourced from Astroskin and Zephyr devices. Variations in error rates across different noise conditions indicate that the quality of the initial real dataset critically affects the performance of GAN-generated synthetic data. The quality, cleanliness, and inherent patterns of real data are pivotal in shaping the initial learning phase of the GAN. Even minor inconsistencies or complex patterns within the real data can challenge the GAN's ability to faithfully replicate these attributes. The effectiveness of GANs in generating realistic synthetic data heavily relies on the integrity and quality of the underlying real data, alongside the model's ability to adapt, and process noise using advanced optimizations algorithms such can be used.

Our study exhibits some limitations: First, the raw samples of the measured observations

for the non-descendant nodes were discretized with user-defined thresholds (Eqs. 5.2.1 and 5.2.2) defined by a small group of domain experts including the authors to demonstrate our proof-of-concept data-driven approach. Next, the generation of prior probabilities for all the descendant nodes is based on the strength of the relationship with the parent nodes are defined by a small group of domain experts. Finally, the credibility of our trust score depends on the trust factors in our BN, which we identified from our limited empirical study and by reviewing and mapping the literature with regulatory standards.

5.6 Summary

In this chapter, we presented a data-driven approach to estimate parameters in the BN for a WMD when subjective and stochastic concepts such as trust need to be quantified. First, we developed a fixed BN structure to quantify trust in a WMD based on our empirical study, regulatory and literature mapping in Chapter 3. Then to compute the relative measure of trust, we used Bayesian parameter estimation. To learn and estimate the parameters, we need access to the prior probabilities of all the nodes in the BN. We demonstrated a method to learn the parameters using a data-driven approach. Inspired by NFR techniques, we incorporated expert knowledge to qualitatively express the strength of the relationships between the nodes (trust factors) in our BN. We then used propagation rules to quantify and aggregate the qualitatively linked trust factors and generate the priors for the network.

We demonstrated our proof-of-concept BNs (BN1-BN4) to quantify trust in WMDs based on the trust factors identified. We learned the BN parameters of our structure using our data-driven approach for two WMDs under identical use cases. To assess the performance of the BN, we performed a comparative analysis with different granularization levels, impact weights, homogeneous and heterogeneous nodes, and noise levels. Our results demonstrate the learnability and generalizability of our data-driven approach.

We also developed a systematic methodology to generate synthetic representative data under various normal and noisy conditions to investigate their effect on trust with our BNs (BN1 -BN4). We validated our model with the larger dataset of synthetic data generated. We compared the

probability of the trust node for the four BNs using both real and synthetic data without noise. The inference scores across the networks were closely aligned, exhibiting an average probability of 0.78, with a standard deviation of 0.03 and minimal variation, underscoring the robustness and consistency of our prototype in trust assessment. Furthermore, we generated synthetic data for different levels of noise to see the trend of results in estimating the probability of the trust nodes. We found that the BN inference score deteriorated as noise levels increased, as expected, thus validating our approach.

In this chapter, we outlined a data-driven methodology for estimating Bayesian parameters using both real and synthetic data from WMDs. This method calculates the probabilities of trust or trust factors being in a specific state, providing stakeholders like Alex with critical data to analyze and select the most trustworthy WMD for their needs. In the next chapter, we present the development of a user-configurable, parameterized prototype to compare the trustworthiness of different WMDs.

Chapter 6

Prototype Development

In this chapter, we present comprehensive steps that were taken in implementing the parameterized prototype that allows WMD users to construct a customizable trust network and effectively compare the trustworthiness of different devices. From initial requirement gathering to final deployment, this exposition ensures a complete understanding of the prototype's functionalities and the technologies deployed.

The structure of this chapter is organized as follows. Section 6.1 describes the prototype. Section 6.2 describes in detail the five phases in the development cycle (requirement elicitation, system analysis, system design, implementation and testing) of our prototype. Results and their discussion are provided in Section 6.3, culminating with a summary in Section 6.4. The code for the prototype is open source and available on GitHub[1].

6.1 Prototype Description

First, we present a detailed description of the system prototype. The system is designed to facilitate the comparison of the trustworthiness among various WMDs using a BN. This system allows stakeholders to evaluate WMDs based on multiple trust determinants dynamically.

The system caters to several key user groups, each interacting with the prototype in ways that reflect their specific needs and expertise. End-users, such as athletes, patients, and general

[1]https://github.com/tailabTMU/TrustQ_MD

consumers, primarily use the system to input their personal preferences and view the trust ratings of different WMDs. These outputs help them understand which devices best align with their individual trust criteria like reliability and privacy. On the other hand users such as medical professionals, including doctors and trainers, utilize the system to delve deeper into the clinical and performance-oriented aspects of WMDs, such as sensor accuracy and data security and view the trust rating. Their interaction enables them to make informed device recommendations.

The prototype operates through a well-defined workflow. First, it captures user specifications for the BN structure through user entry via the command prompt (terminal). This specification includes information such as the total number of nodes of the BN, total number of levels of the BN, relationships between each parent and child node, and strength of each relationship between the nodes (e.g., "++" corresponding to a very strong relationship between parent and child). The qualitative user entry of the relationship is then translated to respective weights as described in Section 5.2 (e.g., qualitative measures such as "++" are given a weight of +4). It then accepts the dataset for all the nodes of the BN structure from the user, in order to learn the parameters of the BN. This dataset, typically related to various trust factors of the BN for WMDs (e.g., real data collected from WMD or synthetic data generated), is uploaded in CSV format.

Next, the software dynamically constructs the BN based on user input, using algorithms that efficiently manage nodes, relationships, and strengths. The parameters are learned employing the data-driven approach, with the dataset for all nodes and the strength of relationships. Then the system computes the conditional and marginal probabilities for all nodes for every state using BN algorithms. The prototype also facilitates query inferencing and computes the probability of trust for a certain state of the node given the evidence based on the network structure.

The results of the trust BN structure for WMDs are then presented through a visualization graph, which aids in the straightforward interpretation of complex relationships between the different nodes in the network. The conditional probabilities and queries are represented in a tabular format. The results of the system are adapted to various user needs, from simple trust assessments to detailed trust factor analyses.

This prototype - "the trust assessment prototype" - is an advanced tool designed to compare

the trustworthiness of different WMDs and enhance the decision-making process regarding the use of WMDs using probabilistic BNs. By accommodating a broad spectrum of users and allowing detailed customization and interaction, the system ensures that different stakeholders can effectively assess device trustworthiness based on factors influencing trust as inputted by individual users.

6.1.1 Objective and Scope of the system

The main objective is to create a parameterized prototype to allow the user to compare the trust of different WMDs. The scope of this prototype includes:

1. Designing a flexible, user-configurable trust BN with dynamic input for customization (e.g. user inputs such as the number of nodes in the BN, number of levels in the BN, relationships between the nodes, and the strength of the relationships).

2. Generating the CPDs and probabilistic query inferencing for a certain state of the node given the evidence based on the BN (e.g., from Fig. 5.1 (b), what is the probability that trust is high, given reliability is low and robustness is high?)

3. Providing a graphical visualization of the customized BN generated as per the user input with clear annotations of nodes and strengths, and presenting the probabilities in a tabular format to interpret trust determinants effectively.

6.2 Development Phases of the Prototype

The development of our WMD trust assessment prototype adheres to a structured five-phase software development life cycle (SDLC) adapted from Ghezzi et al. [43], to ensure a rigorous and thorough approach from concept to deployment (as shown in Fig. 6.1). This approach is critical to crafting a robust system that meets both the explicit and implicit needs of our stakeholders.

1. Requirement Elicitation: This foundational phase involves collecting requirements from

Figure 6.1: Steps taken for the development of our prototype

stakeholders, including end-users such as athletes and medical professionals, to capture comprehensive requirements such as functional and non-functional. Here, we clarify what stakeholders expect the system to do, guiding the development of clear specifications that will drive all subsequent design and implementation efforts.

2. System Analysis: Building on the requirements gathered, this phase analyzes them to establish a complete set of system specifications. This involves detailed scrutiny of the system's necessary capabilities and potential constraints, ensuring that the specifications are both feasible and aligned with user needs. It sets a definitive groundwork for developing the system architecture.

3. System Design: Transitioning from analysis to design, this phase transforms the specifications into a detailed system blueprint. It includes the creation of logical and physical designs, detailing the system's architecture, defining major components, planning the database structure, and specifying software interfaces. These specifications ensure the system is built on a solid foundation that meets all outlined requirements.

4. Implementation: With a detailed design in place, this phase focuses on the actual construction of the prototype according to the specifications. Develop code and integrate various components based on the designs, adhering closely to the specifications to ensure the system functions as intended.

5. Testing: The final verification phase before deployment, testing involves rigorous evaluation of the prototype to ensure it meets all the specified requirements and functions correctly under various conditions. This phase includes unit testing, integration testing, and system testing, which are essential for identifying and fixing any discrepancies between the developed system and the specifications.

6.2.1 Requirement Elicitation

The first step in the development of our prototype is the requirement elicitation phase. Elicitation of requirements is one of the initial phases of the software development lifecycle. This critical phase involves systematically *gathering and documenting both functional requirement and NFRs* from stakeholders, end-users, and others involved with the system. Functional requirements for our BN based prototype to compare trust in WMDs specify the precise operations the system must execute, such as dynamic BN structuring, probabilistic calculations, and data processing tasks essential for assessing trust factors effectively. These requirements are framed in terms of direct actions the system must facilitate, ensuring operational precision and relevance to user tasks.

Non-functional requirements in contrast, focus on defining the operational quality and constraints of the system, addressing essential attributes such as performance, usability, and scalability. These aspects are crucial not only for ensuring the prototype's functionality but also for maintaining its efficiency and effectiveness in diverse real-world applications, particularly in healthcare settings where data system reliability are paramount.

The overarching goal of the requirement elicitation for this prototype is to meticulously capture and articulate all the needs of the stakeholder(s) to align the capabilities of the system with their objectives. This detailed understanding is pivotal for designing a system that accurately reflects stakeholder expectations and optimally serves their needs. Moreover, early identification of potential risks through this process aids in preemptively addressing possible technical challenges and stakeholder conflicts, ensuring smoother project progression and a more robust system architecture.

Functional Requirements

The functional requirements identified for the prototype are listed below:

1. User Interface: The prototype should offer an intuitive and user-friendly interface that enables users to effortlessly input and modify the BN structure. This encompasses entering factors influencing trust, defining relationships between nodes, and the strength of these relationships. The interface should not only facilitate seamless navigation and interaction

but also incorporate robust validation checks for user input types, ensuring that all data entries adhere to expected formats. It should promptly provide clear error messages for any invalid inputs, enhancing usability and preventing potential errors in the BN configuration. This feature ensures that users can engage with the system effectively, without technical hindrances, and contribute to accurate trust assessments.

2. Structure: The prototype should be able to create a BN with an adjustable structure based on the user's input in terms of the factors influencing users (e.g., nodes), the relationships between the nodes (e.g., edges), and the strength (influence) of the relationships (e.g., highly influential).

3. Parameter Estimation: The prototype should learn the CPDs of the BN structure created.

4. Data File: The prototype should be able to learn the parameters of the BN from real or synthetic WMD data provided by the user.

5. Visualization: The prototype should be able to provide a visualization of the BN structure, showing nodes and edges with annotations of the strength (influence).

6. Query-based inferencing: The prototype should be able to provide query-based inferencing for the different nodes of the BN based on the evidence and user request. For example, providing queries such as (Fig. 5.1b, in our example BN): 1) What is the probability that trust is high when reliability is low, but robustness high; $P(X_7^H|{X_5}^L, {X_6}^H)$? or, 2) What is the probability that reliability is high when heart rate sensor validation is high and breathing rate sensor validation is low; $P(X_3^H|{X_1}^H, {X_2}^L)$?

Non-functional Requirements

The NFRs identified for the prototype are listed below:

1. Scalability: The prototype must demonstrate the capacity to efficiently manage an increase in workload by effectively scaling its resources, including Graphics Processing Unit (GPU), Random Access Memory (RAM), and processing power, to accommodate a higher number of nodes and hierarchical levels.

2. Performance: The prototype should be able to efficiently compute the probabilities of different nodes even with the maximum number of nodes and levels, for different parent-child combinations.

Risk Management

In the development of our prototype for evaluating WMD trustworthiness using a BN, risk identification and management play pivotal roles in safeguarding against potential pitfalls that could jeopardize the project's objectives.

1. One significant risk is the efficient handling of the complex BNs, especially as the number of nodes and interdependencies increase. This is crucial for accurately generating BNs for WMDs with multiple heterogeneous sensors and relationships. To mitigate this, scalable resources for the BN modelling and parameter estimation that can dynamically adjust to the load need to be implemented.

2. Another technical risk involves the accuracy and integrity of data used to train the BN. Inaccurate or biased data could lead to unreliable trust assessments, which would compromise the utility of the system in real-world applications. To manage this risk, data validation and cleansing protocols should be established, prior to the integration of the system to ensure high data quality.

3. Lastly, given the sensitive nature of data involved in WMD assessments, privacy concerns such as disclosure of personal information can be a risk. To counter these risks, the personally identifiable information in the data collected through the WMDs, should be completely anonymized, ensuring compliance with the health data protection standards (such as HIPAA). Through these measures, the project aims not only to mitigate risks effectively but also to enhance the overall robustness and reliability of the system architecture.

Use Case Diagram

When developing sophisticated systems such as a prototype for evaluating the trustworthiness of WMDs, it is crucial to articulate and understand user interactions clearly. Use case diagrams play an instrumental role in the early stages of software development, particularly during the requirement elicitation phase. By visually representing the interactions between users and the system, these diagrams provide a clear and structured overview of user requirements. They highlight the functionalities the system must deliver from the perspective of the user, making them essential for defining and clarifying what the system is expected to achieve. Use case diagrams help bridge the gap between user expectations and system development, ensuring that all stakeholders have a common understanding of the objectives and capabilities of the system.

We present usage scenarios that illustrate the system's functionality. A usage scenario is a specific sequence of actions that illustrates how interactions are carried out within a system. It provides a narrative description of a user's workflow or the system's process flow and helps developers and stakeholders understand the context in which the system operates and the requirements necessary to support user tasks. Usage scenarios are instrumental during the requirements-gathering phase as they outline the real-world use of the system in various contexts, making them pivotal for functional testing and user experience design.

The usage scenarios are distilled into use cases that abstract these interactions into structured, goal-oriented sequences that define the required functionalities of the system. This relationship supports an iterative design process where scenarios inform the creation and refinement of use cases, and together, they guide testing and validation efforts. By integrating both, we ensure the system is both functionally sound and effectively meets user needs, thereby aligning the final product closely with business objectives and user expectations.

We present six use case scenarios for our WMD trust assessment prototype. Each scenario is presented with actors (defined by their roles like athletes or medical professionals), and use cases (specific actions taken by the actors within the system like entering preferences or analyzing WMD clinical aspects). We also present the graphical representation of the use cases and their interactions with actors within a system by a use case diagram. The use case diagram for our trust

assessment prototype provides a clear, high-level overview of the system's interactions across different user types, such as athletes (who use WMDs to monitor health parameters during rigorous training), general users (who use WMDs for self-health monitoring), medical professionals (who recommend WMDs to their patients or community), WMD evaluation specialist (who evaluates WMDs from different manufacturers), and regulatory compliance officer (who ensures that the WMDs are complaint with regulatory standards). Each actor is associated with specific functionalities within the system: End-users are primarily involved in entering their preferences, viewing trust ratings, and making selections of a device, reflecting the system's ability to cater to personal trust criteria. Medical professionals gain in-depth analysis, utilizing the system to evaluate the WMD considering clinical aspects and making recommendations based on these analyses. The use case diagram for our use cases is presented in Fig. 6.2.

1. **Usage Scenario 1: Trust Assessment of a WMD by an Athlete**

   An athlete wants to choose a WMD that is trustworthy for their specific training needs. They are particularly interested in devices that are renowned for their reliability and accuracy. They use the WMD trust assessment system to compare the trustworthiness of different devices and choose a suitable one.

   - Actors: Athlete (End-user)
   - Use Cases:
     (a) Select Device Criteria: Athlete selects the WMD parameters of interest.
     (b) Submit Preferences: Athlete submits their selected preferences.
     (c) View BN: View the BN constructed based on the parameters selected.
     (d) Receive Measures: The system generates the trustworthiness measures related to selected criteria.
     (e) View Device Measures: Athlete views related measures on trustworthiness to choose the right WMD for their use.
     (f) Compare Devices: The athlete compares two or more devices based on their trustworthiness metrics derived from the BNs to select the best device.

2. **Usage Scenario 2: Evaluating the Clinical Trustworthiness of a WMD by a Medical Professional**

   A doctor needs to recommend a WMD to a patient for continuous health monitoring. The doctor assesses devices based on clinical accuracy and data security.

   - Actors: Medical Professional (Doctor)
   - Use Cases:
     (a) Select Clinical WMD Parameters: Doctor selects clinical parameters of the WMDs.
     (b) View BN: View the BN constructed based on the parameters selected.
     (c) Review Device Measures: Analyze the detailed performance of WMDs for the clinical parameters selected.
     (d) Make Recommendations: Enter device recommendations on trust for patients based on the parameter performance.

3. **Usage Scenario 3: Comparative Analysis of WMD Trustworthiness by an Organization**

   An organization involved in the evaluation and assessment of WMDs wants to compare the trustworthiness of multiple devices simultaneously. They construct individual BNs for each device to model and analyze various trust factors such as reliability, data privacy, and user feedback.

   - Actors: WMD Evaluation Specialist
   - Use Cases:
     (a) Input Device Data: WMD evaluation specialist inputs the specifications and performance data of different WMDs into the system.
     (b) Construct BN for Devices: The system constructs separate BNs for each device based on the provided data.
     (c) View BN: View the BN constructed based on the parameters selected.

(d) Compare Devices: The WMD evaluation specialist queries the system to compare two or more devices based on their trustworthiness metrics derived from the BNs.

(e) Review Comparative Results: The specialist reviews the comparative analysis to identify the most trustworthy devices and prepares a report for technical evaluation purposes.

4. **Usage Scenario 4: Consumer Decision Support for WMD Purchase**

A consumer shopping for WMDs uses an online platform integrated with our prototype to make an informed purchase decision based on trustworthiness assessments.

- Actors: Consumer (End-user)
- Use Cases:

(a) Access Platform: Consumer logs into the online platform integrated with the BN-based trust assessment tool.

(b) Input Preferences: Consumer inputs their health conditions and specific needs for a WMD.

(c) View Device Options: The platform presents various WMD options available on the market.

(d) Evaluate Trustworthiness: The system generates trustworthiness scores for each WMD based on the consumer's input and displays them.

(e) Compare Device and Make Purchase Decision: The consumer compares the different WMDs and selects a WMD based on the trustworthiness scores and additional product information.

5. **Usage Scenario 5: Regulatory Compliance Monitoring**

Regulatory agencies use the prototype to monitor and ensure compliance with standards for trustworthiness in the manufacture and performance of WMD.

- Actors: Regulatory Compliance Officer
- Use Cases:

(a) Set Compliance Criteria: Define specific trustworthiness criteria that WMDs must meet according to health regulations.

(b) Monitor WMDs: Use the prototype to assess compliance of various WMDs with the set criteria.

(c) Generate Compliance Reports: System generates reports detailing each device's compliance status.

(d) Take Regulatory Actions: Based on reports, take necessary actions such as certifications, warnings, or recalls.

6. **Usage Scenario 6: WMD Trustworthiness Training for Healthcare Providers**
Healthcare providers use the prototype to understand various trust factors and how they affect the suitability of WMDs for different patient needs.

- Actors: Healthcare Provider (Doctor, Nurse)
- Use Cases:

(a) Training Session Login: Healthcare providers log into a training module integrated with the BN prototype.

(b) Learn Trust Factors: Interactive sessions that explain different trust factors modelled in the BN.

(c) Simulate Scenarios: Run simulations with hypothetical patient scenarios to see how different WMDs perform.

(d) Assess Learning: Providers take assessments to ensure they understand how to evaluate WMD trustworthiness.

6.2.2 System Analysis

System analysis serves as the second critical phase in the development of our prototype, focusing on *transforming the gathered requirements into comprehensive system specifications*. This phase delves deep into analyzing both functional requirements and NFRs to ensure they are clearly

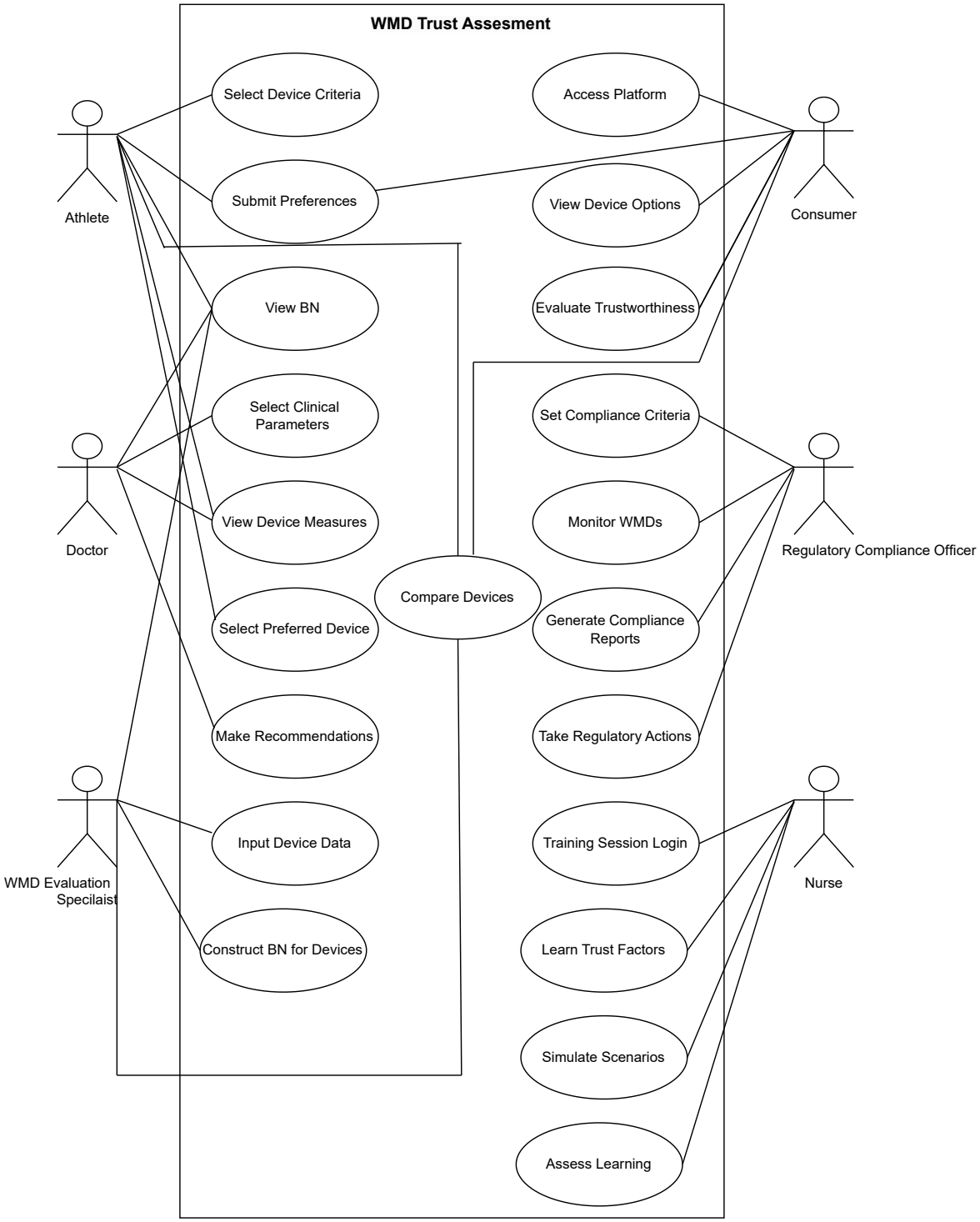

Figure 6.2: The use case diagram for our use cases depicting the interactions between the different actor(s) and the system in assessing the trust measure of a WMD

defined, feasible, and closely aligned with user needs. It lays a robust foundation for the system's architecture, establishing clear specifications that guide subsequent development stages.

The system analysis phase is pivotal in the development of our WMD trust assessment prototype, where the transition from conceptual requirements to concrete system specifications occurs. This phase critically examines both requirements - functional and non-functional, to ensure that they are detailed, technically feasible, and aligned with the stakeholders' needs. It also sets a solid foundation for the system's architecture by establishing explicit specifications that guide all subsequent development activities.

1. **Technical Feasibility and Requirements Verification:**

   - *Analysis of Complex Structure:* Assess the system demands of maintaining and dynamically updating BNs, especially those that may include up to 100 nodes with intricate multilayer interdependencies.
   - *Data Integrity:* Ensure the accuracy and reliability of the data used to train the BN by addressing potential biases and inaccuracies, and by implementing robust data anonymization techniques to protect user privacy.

2. **System Requirements Specification:**

   - *Functional Requirements:* Specifications include dynamic creation and modification of the BN structure based on user inputs, along with the capability to perform detailed probabilistic inferencing to compute and display trust probabilities.
   - *Non-functional Requirements:* Emphasis on scalability, ensuring the system can efficiently manage an increasing number of nodes and hierarchical levels while maintaining optimal performance in terms of processing speed, memory utilization, and power consumption.

3. **Modeling and Simulation:**

   - *Probabilistic Model Development:* Develop a hierarchical PGM using Bayesian techniques tailored to specific user trust factors of WMDs.

- *Simulation:* Employ both real and synthetic data to simulate the BN's performance under diverse operational scenarios, validating the system's effectiveness and readiness before full-scale deployment.

4. **Interface Requirements:**

   - *Input Terminal Design:* Construct a user interface that supports comprehensive input capabilities, allowing users to specify detailed information such as node numbers, hierarchical layers, and node interrelations. This interface should also facilitate dynamic visualization of BNs to illustrate how modifications affect trust probabilities.

5. **Data Analysis and Integration:**

   - *Data Management:* Design data handling protocols to integrate seamlessly with current healthcare ecosystems using WMDs, focusing on maintaining compliance with health data protection standards.
   - *Output Management:* Systematically format and present results to support clear, actionable insights for decision-making by end-users and medical professionals alike.

The system analysis phase ensures that every aspect of the prototype, from data input to the final trust assessment, is rigorously defined, forming the basis for all future development phases.

Sequence Diagram

Based on the identified specifications, we present the sequence diagram for the athlete use case for our prototype. The sequence diagram illustrates user interactions over time, details the operation sequence of processes, and clarifies system behaviour across various scenarios. This diagram depicts the order of the interactions from one object to another and can detail the conditions under which several objects collaborate to accomplish a task or a use case. It enhance the structural insights provided by use case diagrams by detailing the temporal and operational interactions within the system. For example, for our athlete use case the sequence diagram will illustrate the interaction process as athletes assess the trustworthiness of the WMDs. It starts with athletes entering their

preferences, followed by the system dynamically constructing a BN, calculating trust metrics, and displaying results. Athletes can adjust inputs and request updated assessments, providing a clear visualization of the system's operations. This diagram is essential for understanding the sequential interactions and supports system development, testing, and troubleshooting. The detailed interaction flow is explained below and the sequence diagram is shown in Fig. 6.3.

Sequence Diagram - Interaction Flow for Athlete's Assessment of WMD Trustworthiness

For the use case of an athlete (user) assessing the trustworthiness of WMDs using the trust assessment prototype, we present the sequence diagram that can effectively illustrate the detailed interaction flow. The breakdown of steps of the sequence diagram shown in Fig. 6.3 is given below:

Step 1: Initialize System

- Step 1.1: The athlete accesses the system through a user interface.
- Step 1.2: System displays initial greeting and request for input.

Step 2: User Interface

- Step 2.1: Athlete enters personal preferences related to trust factors (such as device accuracy, comfort, or data privacy).
- Step 2.2: Athlete enters levels of abstraction and the total number of nodes.
- Step 2.3: Athlete enters the relationship and the strength between the nodes.
- Step 2.4: Athlete enters the dataset file (CSV) for all the nodes.
- Step 2.5: System validates the entered data for each entry and confirms acceptance or requests re-entry of data.

Step 3: System Core

- Step 3.1: System captures all the user inputted data.

Step 3.2: System pre-processes and performs statistical analyses on the dataset to meet the basic requirements of the BN configuration.

Step 3.3: System stores the updated data in the data file.

Step 4: Bayesian Network

Step 4.1: Based on the athlete's inputs, the system dynamically constructs the BN structure.

Step 4.2: System generates a visualization of the BN graph.

Step 4.3: System calculates the necessary parameters (conditional and marginal probabilities) for the BN nodes based on the dataset, relationship, and strengths of the relationship provided by the athlete for trust assessment.

Step 4.4: System presents the parameters in a tabular format.

Step 5: User Interaction with Results

Step 5.1: The athlete queries the system to derive measures of a certain trust factor for a certain state (low, medium or high).

Step 5.2: System presents the probabilistic inferencing as query results.

Step 5.3: Athlete reviews the output, to compare with and gain insights into various factors of the WMD.

Step 5.4: Athlete specifies additional criteria or modifies existing parameters.

Step 5.5: Athlete reviews the output, to compare various WMD.

Step 5.6: System recalculates and updates the output based on the new queries.

Step 5.7: System provides options to start a new assessment cycle.

Step 6: End Session

Sub-step 6.1: Athlete decides to end the session.

Sub-step 6.2: System securely logs out the user and closes the session.

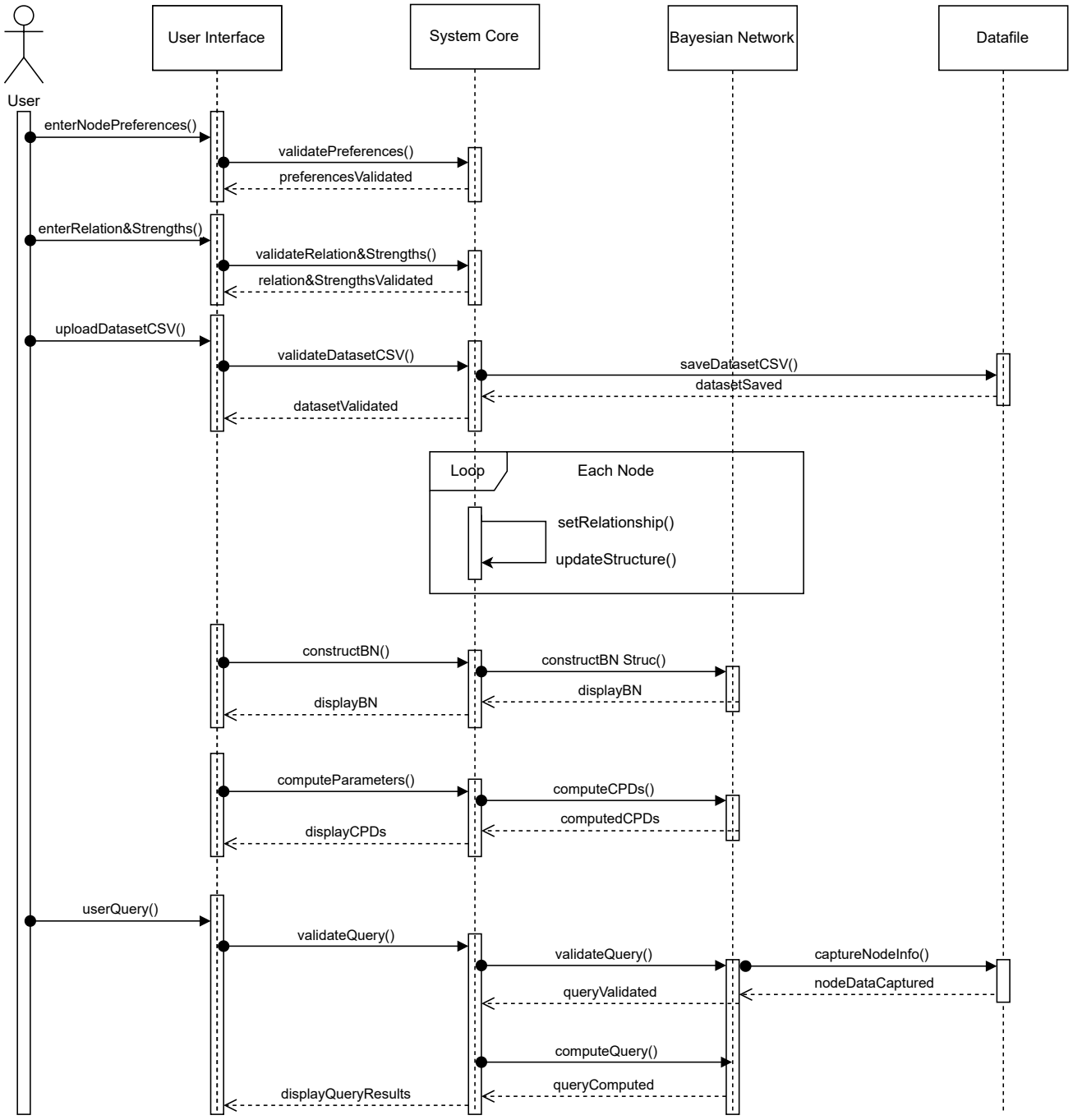

Figure 6.3: A sequence diagram for the use case of an athlete using the trust assessment protocol to evaluate the trustworthiness of a WMD. This diagram depicts the interactions between the actor (athlete) and the system in evaluating the trust of a WMD

6.2.3 System Design

The third phase is the system design. System design is pivotal in the development of our prototype, as it meticulously transforms detailed requirements both functional and non-functional,

and risk management into a robust software architecture that forms the backbone for subsequent development and deployment phases. It encompasses the detailed definition of 1) system architecture (layered design), and 2) detailed design (components including modules, interfaces, and data interactions, ensuring these elements collectively meet the precise scientific requirements and performance standards necessary for effective trust evaluation) [21]. We describe the two main parts here:

1. **Architectural Design:** The prototype's system design serves as a strategic blueprint, outlining a comprehensive architecture that defines how each component of the BN interacts within the system and with external entities (users).

   A layered architecture of n levels (n = 4) is selected for our prototype due to its ability to manage complexity and improve *modularity* to facilitate the comparison of trustworthiness between various WMDs [21]. This design pattern allows each layer to operate independently while interacting seamlessly with adjacent layers, making the system easier to manage, adapt, and scale. Given the prototype's need to handle various user inputs, process complex data, and generate interpretable outputs, a layered architecture ensures that each component can be developed and maintained with a clear focus on its specific responsibilities without interference from unrelated system aspects. Common functionalities like logging, error handling, and data management can be centralized in the infrastructure layer, promoting *reuse* across the system. Also, it increases *flexibility* as new user interface technologies or data storage solutions can be integrated with minimal impact on the business logic and presentation layers. The layered architecture of our prototype comprises of four key layers:

   (a) Presentation Layer: Manages the user interface and interaction, allowing users to input data and view results. This layer is essential for providing a clear and intuitive interface for athletes, doctors, and other end-users to input their trust determinants and view the trustworthiness outputs of various WMDs.

   (b) Business Layer: Primarily handles the logic and rules associated with the BN calculations, orchestrating the dynamic construction and modification of the network based

on user inputs. This layer is crucial for processing trust determinants and generating trust probabilities, ensuring that the analytical outputs are robust and reflective of the complex relationships and dependencies inherent in the WMD trust BN. This layer also ensures scalability with enhanced algorithms.

(c) Persistence Layer: Responsible for managing the storage and retrieval of data through CSV files, ensuring data integrity and consistency across sessions. This layer encapsulates all file-handling operations, providing a reliable and straightforward mechanism for persisting data, which facilitates easy data manipulation and scalability without the complexities of a traditional database system.

(d) Datafile Layer: Functions for data management that handles operations with CSV files, serving as the system's primary method for data storage and retrieval. This layer abstracts the complexities, maintains privacy and provides data access, and processing of the CSV files.

2. **Detailed Design:** During the system design phase, the prototype is decomposed into manageable, well-defined modules, each responsible for specific functionalities within the system. The architecture [47] of the prototype is designed with modularity at its core, incorporating a series of distinct yet interrelated components. Each module is meticulously engineered to fulfill specific functions within the system, enabling seamless integration and scalable performance tailored to the demands of trust analysis in WMDs as described below [43]:

(a) Input Management Module: Handles user inputs for defining the structure and parameters of the BN.

- Functionality: Captures user specifications for the number of nodes, levels, and types of relationships.
- Interfaces: User inputs are taken from the command prompt (terminal) where users can specify and modify the BN structure. The dataset for all the nodes of the BN with the readings collected from is taken as a CSV file from the user.
- Processing: Validates input data for type and logical conditions. Processes input

data to fit the BN's requirements, including weights, normalization, and handling missing values.

- Integration: Ensure all data matches the user-specified configurations.
- Customization:

  i) Dynamic Structure Definition: Allows users to dynamically define and adjust the BNs structure, including adding or removing nodes and edges based on evolving analysis requirements and relationships in the BN with user input command. This flexibility supports experimental and iterative approaches to understanding trust determinants in WMDs.

  ii) Parameter Specification: Users can provide the real or synthetic dataset for the nodes of the BN to learn from.

  iii) Advanced Validation Features: Implements advanced validation rules when gathering inputs from the user; the code validates both the type (ensuring the data type such as integer or string) and conditions (like range limits for nodes and levels) for the inputs. This helps prevent errors downstream by ensuring all inputs are in the expected format and within logical boundaries. In addition, checks are performed to ensure that the parent nodes entered by the user exist and are correctly positioned in relation to the child node. This prevents cycle creation and structural inconsistencies for the DAG structure in the BN.

  iv) Data Ingestion: Load data from a CSV file and validate that the dataset contains all required columns (nodes) as specified by the user when defining the network structure.

  v) Data Mapping: Ensure that the data matches the user-specified configurations for the network, matching the headers of the data columns with the corresponding nodes in the BN.

(b) BN Configuration Engine: Dynamically constructs and updates the BN structure and parameters based on user inputs.

- Functionality: Constructs the BN based on user-defined specifications. Allows for

dynamic modifications to the structure.

- Algorithm: Utilizes algorithms in specialized libraries in Python such as BN modelling (e.g., bnlearn) and probabilistic graphical models (e.g., pgmpy) for graph construction, ensuring efficient management of nodes and edges.
- Data Handling: Stores and retrieves configuration settings provided by the user are managed by dictionaries.
- Customization:

  i) The computation of CPDs in the code is based on the user-specified structure and strengths.

  ii) Data-Driven CPD Calculation: When CPDs are generated, all possible combinations of parent states are considered. This uses a Cartesian product to enumerate these combinations, ensuring the CPD covers all possible states. The CPDs for the nodes with parents and those without parents are considered differently. For the nodes without parents, the marginal probability is calculated by taking the frequency of each node's states within the dataset. For the nodes with parents, the CPDs are computed with the given states of its parent nodes.

  iii) Application of Weights: The manually specified weights by the user are applied to influence the probabilities in the CPD tables. This allows the network to reflect stronger or weaker influences between nodes based on real-world knowledge or data insights. For example, if a parent-child relationship is marked with a higher strength, the influence of that parent on the child's distribution is scaled by a factor (for example, relationships with strengths '++' are scaled by a factor of 4).

  vi) Normalization: After applying weights, the probabilities are normalized to ensure that the total probability for each state combination sums to 1. This is essential for maintaining valid probability distributions.

(c) Inference Engine: Executes queries on the BN to compute trust probability based on the current state of the network.

- Functionality: Performs probabilistic inferencing to calculate trust probability.

- Techniques: Uses Bayesian inference algorithms in the "pgmpy" library to update beliefs in the network based on new evidence or query inputs.
- Customization:

  i) Query Inferencing: Allows for calculating the probabilities of certain node states (low, medium, and high) given evidence about other nodes.

(d) Visualization Module: Provides a graphical representation of the BN and the results of queries to facilitate easier interpretation of trust relationships.

- Functionality: Generates intuitive visualizations of the BN and query results.
- Tools: Uses the NetworkX library for graphical representations.
- Customization:

  i) Node and Edge Representation: It draws nodes and edges, where nodes represent variables and edges denote direct dependencies between nodes of the BN.

  ii) Weight Annotations: The graph visually represents the weights of relationships. This is done by annotating the edges with the strength of the relationship (like "++", and "+"), based on the user-defined weights.

  iii) Aesthetic Settings: Nodes are displayed with specific sizes, colours, and labels, making the graph visually appealing and easier to interpret.

(e) Output Management Module: Manages the presentation of results and any data exports.

- Functionality: Formats and displays the results of BN queries in a tabular format.
- Tools: Uses the Tabular CPD library for graphical representations.
- Customization:

  i) Query Inferencing Result Filtering and Sorting: Allows users to make query inferencing based on their determinants of interest. This customization helps users focus on the most relevant information for their specific analysis needs.

  ii) Conditional Probabilities: Presents the CPDs for all states (e.g., low: 0, average: 1, and high: 2) for each node in a tabular form to facilitate visibility.

Component Diagrams

To accurately illustrate the architecture of our trust assessment prototype, we will construct a component diagram that delineates each layer of the system: presentation, business, persistence, and database. In a component diagram, a 'component' represents a modular part of a system defined by its interfaces and encapsulation, essential for system integration and functionality. This diagram will include components within each layer, specifying their interactions and dependencies, which highlight how individual software components integrate to form a coherent and functional system architecture. Below is a detailed explanation of the different components' of our prototype followed by the component diagram in Fig. 6.4:

1. **Components in the Presentation Layer:**

    - User Interface: It manages user interaction. It collects the user inputs for generating the customized BN structure for the trust assessing prototype and displays outputs in tabular and graphical format. Essential for ensuring a user-friendly experience and facilitating user engagement by presenting trust metrics and other relevant data.

2. **Components in the Business Layer:**

    - Input Management: It captures and validates user inputs regarding the structure and parameters of the BN.
    - BN Configuration Engine: Dynamically adjusts the BN based on user inputs, handles the logic for modifying network structures, and applies user-defined relationship strengths.
    - Inference Engine: Performs probabilistic computations to derive trust metrics from the BN.

3. **Components in the Persistence Layer:**

    - Data Handler: It manages the reading and writing operations for CSV files, ensuring data integrity and facilitating data reusability across sessions.

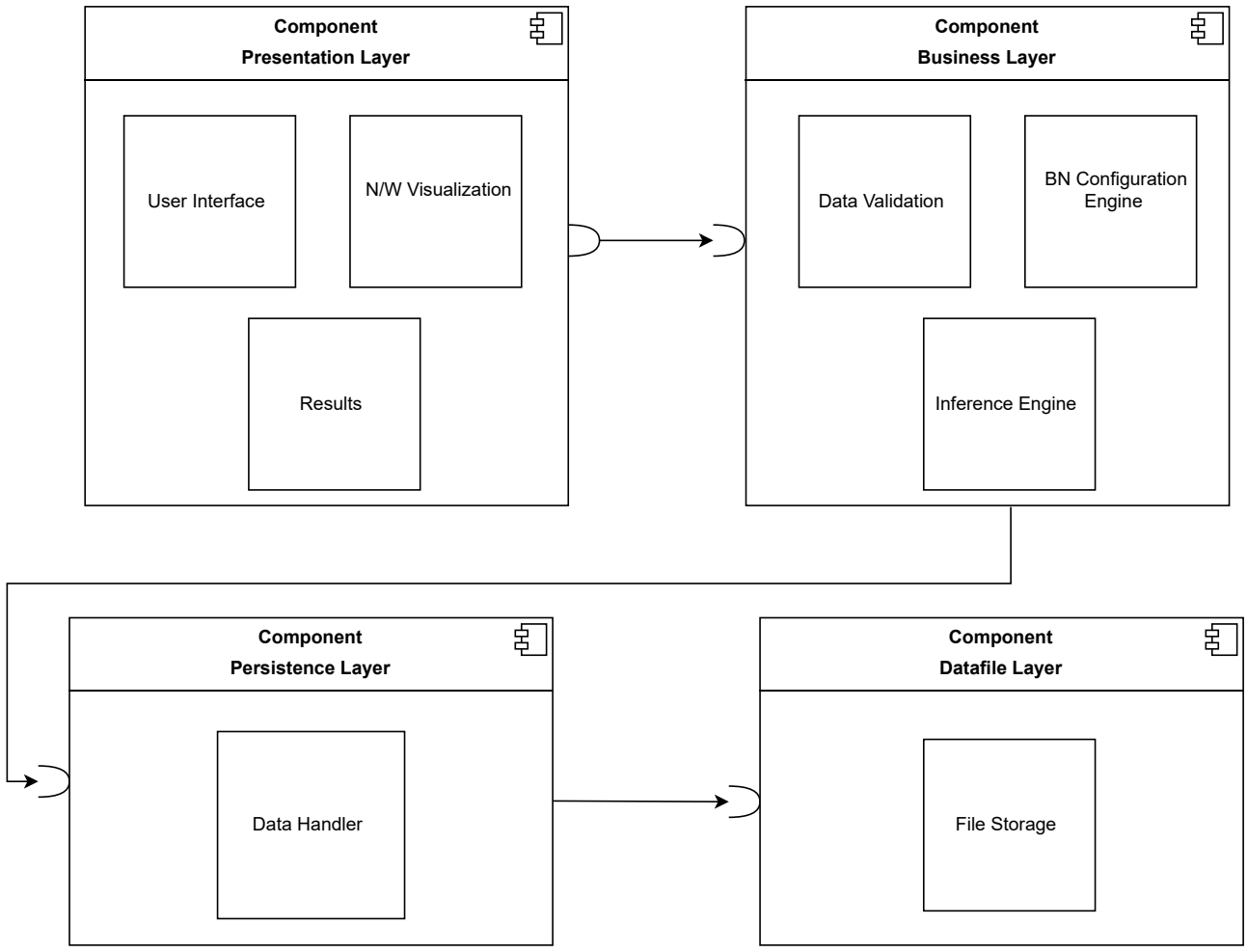

Figure 6.4: Component diagram for the trust assessment prototype for wearable medical devices based on Bayesian network.

4. **Components in the Datafile Layer:**

   - File-based Storage System: Utilizes CSV files for storing and retrieving data, acting as the primary data storage method without the complexity of a traditional database.

Deployment Diagrams

In the system design phase of our prototype development, we also incorporate deployment diagrams to encapsulate the physical layout and interaction of software components with hardware resources. A deployment diagram illustrates the physical aspects of an information system by

detailing how software artifacts are deployed on hardware components, which is crucial for understanding the environment in which a system operates. These diagrams are essential for visualizing the distribution of components across a network of nodes, each represented as a physical device like a computer, server, or sensor. For our prototype, which utilizes BNs to evaluate the trustworthiness of WMDs, the deployment diagram delineates how the software components, such as the BN Configuration Engine and Data Processing Unit, are allocated across different hardware units. It also shows the communication pathways and deployment relationships between these components, highlighting how data and control flow throughout the system. This visualization aids in ensuring that the prototype's software is correctly deployed to meet performance, scalability, and reliability requirements, thereby enhancing the overall system architecture.

Deployment diagrams are highly beneficial for systems like our prototype, where hardware components are often activated by external stimuli such as sensors. These diagrams are crucial for clarifying how the software, specifically our BNs, interacts with hardware elements within the system. In our case, the deployment diagram details how software components manage and utilize data files (artifacts) to evaluate the trustworthiness of WMDs. The diagram effectively illustrates the distribution of these software components across the hardware setup, enhancing our understanding of the system's operational dynamics. The key elements of the deployment diagram for our prototype are detailed as follows:

1. Nodes: Nodes represent physical entities where the software is deployed and are typically shown as 3D boxes in the diagrams. They include:

    - Servers (e.g., web servers, application servers)
    - Client devices (e.g., user computers, mobile devices)

2. Artifacts: Artifacts are tangible elements or files from the development process, such as executables, libraries, and configuration files. They are deployed on nodes and represented by smaller rectangles within the node boxes.

3. Associations: Associations depict the communication pathways between nodes, involving various protocols and networks. They are shown using lines or arrows, indicating data flow

and control flow between nodes.

4. Components: The components like the Input Management Module, BN Configuration Engine, etc., detail how system functions are spread across the hardware.

Currently, our prototype runs locally on a PC so we only have one node (a PC) in our deployment. All components are within this single node, as shown in Fig. 6.5. However, when we deploy it in the cloud in the future, our deployment diagram will change as we will have separate nodes that will include:

1. A *application server* node which will have the graphical user interface (GUI) and the components of business and persistence layers

2. A *user's PC* node where end-users will interact with the system to enter various inputs and datasets.

3. A *data file* node which will be represented as a separate node if data is handled externally or through cloud services (as of now the PC holds the data).

6.2.4 Requirement - Feature Matrix

We present the Requirement - Feature matrix to show how both the requirements functional and non-functional, for the prototype, identified in the requirement elicitation phase, are mapped with the customized features described in the detailed design section. This matrix systematically maps out how each functional requirement—ranging from user interface design to complex query-based inferencing—is addressed by corresponding system features. The columns of the matrix are represented by the key requirements of the system, such as the user interface, structure modification, parameter learning, visualization capabilities, and query-based inferencing capabilities are mapped with key features offered by the prototype. The checkmarks indicate which features fulfill the particular requirements.

This matrix is crucial to ensure transparency and alignment between the development team and stakeholders, facilitating clear communication about what the system is expected to perform.

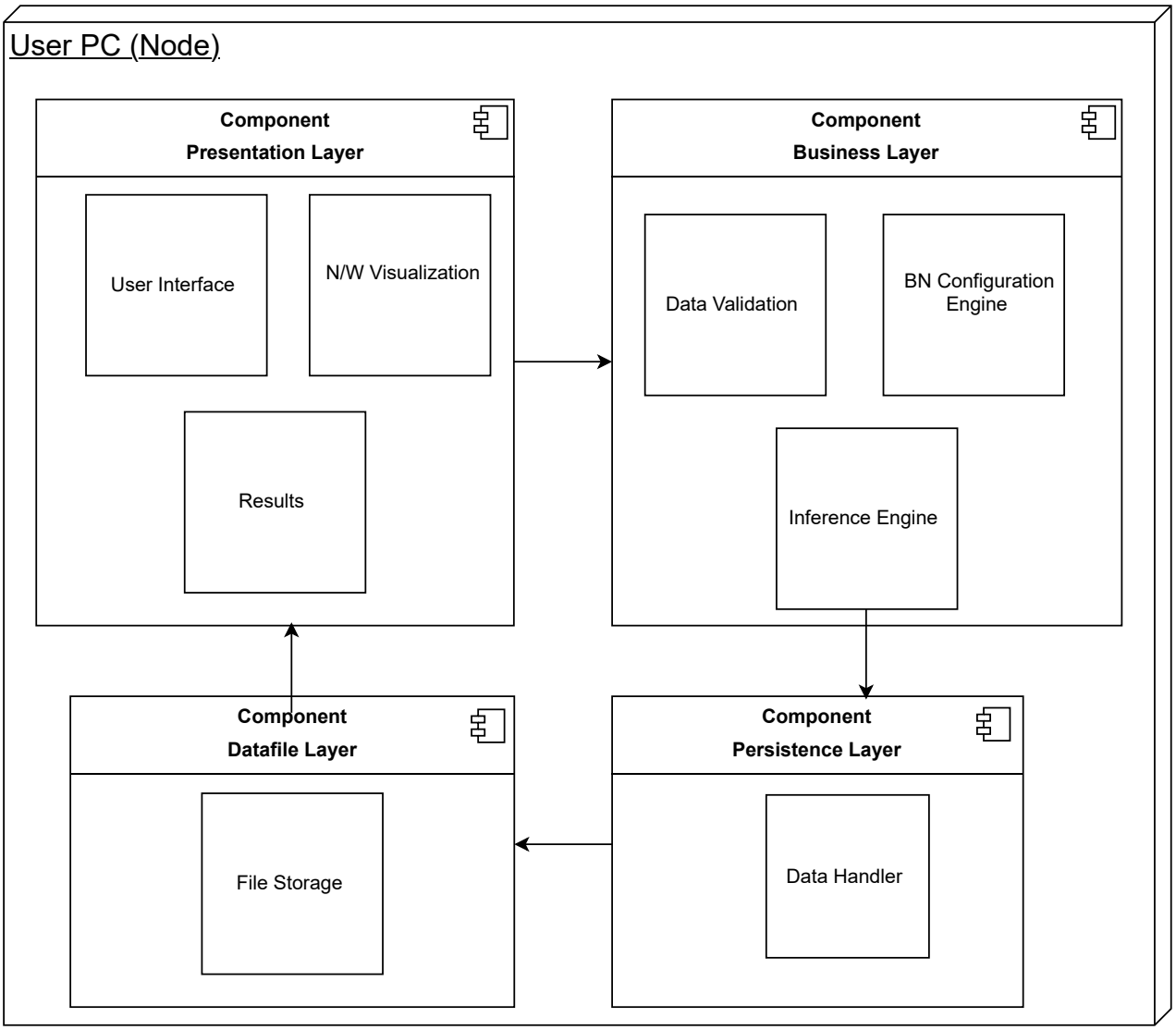

Figure 6.5: A deployment diagram representing the allocation of components to different nodes. A PC can access the application through the local file that provides information from a datafile

It also aids in verifying that every critical requirement has a corresponding feature implemented in the system, providing a foundation for thorough testing and validation phases. Table 6.1 shows the mapping for our prototype.

6.2.5 Implementation

In this section, we describe the development environment, the tools used, and the specific techniques employed to code the prototype. The implementation phase focuses on adhering to the design specifications and ensuring code quality and maintainability. Our prototype mainly has the

Table 6.1: Mapping of the stakeholder's requirements identified in the requirement elicitation phase with the features offered as described in the system design (detailed design) phase by the WMD trust assessment prototype

Requirements / Features	User Interface	Dynamic BN Structure	Parameter Learning	Visualization	Query-Based Inferencing
User - friendly Interface	✓			✓	
Customized User-defined BNs	✓	✓			
User - Data Validation	✓				
Real/Synthetic Data Handling	✓	✓	✓		✓
Data Integrity (Bias)		✓	✓		
Protect PII	✓	✓	✓	✓	✓
Data-Driven BN	✓	✓	✓	✓	✓
Graphs with Annotations				✓	
Ability to find trust impact by any BN factor			✓		✓
Handle large- complex BN	✓	✓	✓	✓	✓
Scalability		✓	✓	✓	✓

following parts:

A. Hardware

1. Wearable Medical Devices:

- *Device:* Selecting the different devices to be tested.
- *Factors:* Identifying the specific factors of each device.
- *Data Collection:* Collecting real data from each device for different use cases.

B. Software

1. Welcome and User Guidance:

- *Feature:* Initial greeting and user instruction.
- *Technique:* Simple print statements provide essential guidance.
- *Tool/Library Used:* Python's built-in `print()` function.

2. User Input Validation:

- *Feature:* Robust system for validating user inputs.
- *Technique:* Repeated prompts and error handling ensure valid user data.

- *Tool/Library Used:* Python's standard input and error handling mechanisms (`input()`, `try-except`).

3. Relationship Configuration:

- *Feature:* Allows users to define relationships and interaction strengths between nodes.
- *Technique:* Collects user inputs for parent-child relationships and strength assignments.
- *Tool/Library Used:* Python dictionaries for storing relationships and strengths, ensuring quick access and manipulation.

4. Data Handling:

- *Feature:* Manages data loading and ensures format compatibility with the network.
- *Technique:* Performs data loading and preprocessing.
- *Tool/Library Used:* `pandas` for reading CSV files and handling data frames efficiently.

5. CPD Computation:

- *Feature:* Computes conditional probability distributions for BN nodes.
- *Technique:* Applies mathematical operations to compute probabilities based on parent states and user-defined strengths.
- *Tool/Library Used:* `numpy` and `bnlearn` for numerical operations and `itertools` for generating parent state combinations.

6. Network Construction:

- *Feature:* Constructs the BN with nodes, edges, and CPDs.
- *Technique:* Integrates nodes and edges based on user configurations and attaches CPDs.
- *Tool/Library Used:* `pgmpy`, specifically its `BayesianNetwork` class for network modeling and `TabularCPD` for probability distributions.

7. Visualization:

- *Feature:* Graphical representation of the BN.
- *Technique:* Visualizes nodes and edges with annotations for relationship strengths.
- *Tool/Library Used:* `NetworkX` for graph operations and `matplotlib.pyplot` for rendering visual representations.

8. Query Processing and Inferencing:

- *Feature:* Allows for probabilistic queries and inferencing within the network.
- *Technique:* Employs probabilistic inference algorithms to compute node probabilities based on given evidence.
- *Tool/Library Used:* `pgmpy`'s `VariableElimination` for efficient computational inferencing.

9. Error Handling:

- *Feature:* System-wide error management to enhance tool robustness.
- *Technique:* Implements extensive exception handling.

The flowchart for our prototype is shown in Fig. 6.6. The flowchart outlines the procedural framework for assessing the trustworthiness of WMDs through a BN. The process begins with a welcome note to the user, followed by a decision point where the user confirms whether to test a device. If the user opts not to proceed, the process is stopped.

Upon choosing to test a device, the user is prompted to input specifications related to the device, which are immediately validated for format correctness. If the input is invalid, the process loops back, allowing the user to reenter the specifications until they are correct.

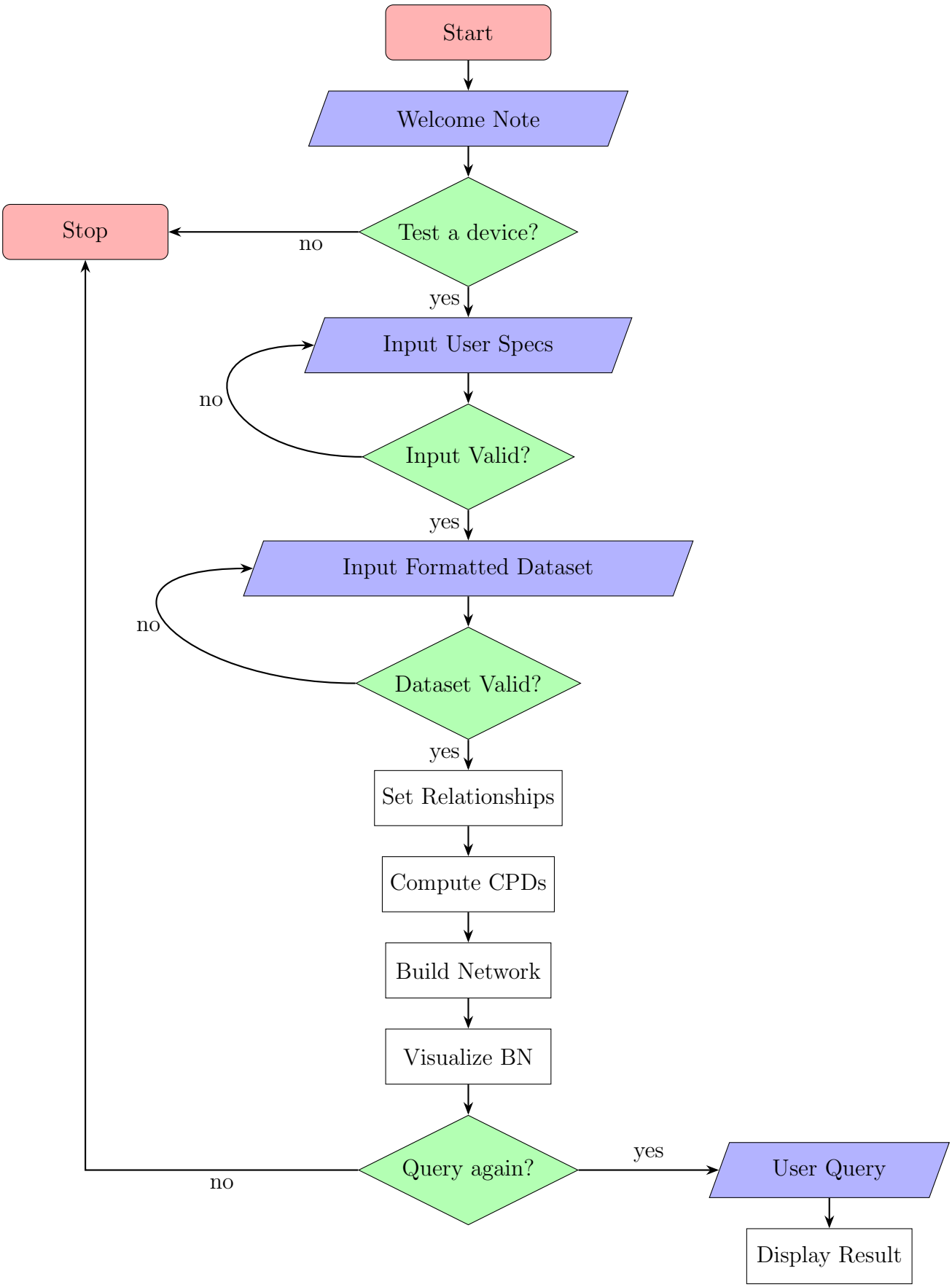

Figure 6.6: Flowchart representing the process flow of the prototype to measure trust of WMD using Bayesian network

Once the user input is validated, the next step involves loading and validating the formatted dataset. If the dataset is found to be invalid, similar to user input, the process loops back to allow for corrections.

With valid inputs and datasets, the system proceeds to set the relationships necessary for constructing the BN. Following this setup, the system computes the CPDs and constructs the network. Subsequently, the system visualizes the BN, providing a graphical representation for better user comprehension. After visualization, the system enters a query phase where the user can interactively query the BN multiple times. Each query is processed and the results are displayed. This interaction continues until the user decides not to initiate further queries, at which point the system stops.

This flowchart provides a clear, step-by-step representation of the operational sequence, ensuring that the user's interaction with the system is structured and efficient, leading to accurate trust assessments of medical devices.

6.2.6 Testing

This section provides a detailed overview of the comprehensive testing strategy implemented for the trust assessment prototype for WMDs using the BN. The testing plan described herein is designed to rigorously evaluate the system's performance across various scenarios, including edge cases and stress conditions. By systematically addressing each phase of testing—from input validation and error handling to performance under maximum operational loads—this documentation aims to affirm the robustness of the tool. Detailed test cases, expected outcomes, and results are presented to give a clear pathway for replication and to document the outcomes effectively. This thorough approach ensures that the tool not only adheres to the specified requirements but also delivers a seamless and effective user experience in constructing and analyzing BNs.

Test Plan

The following steps outline the test plan for the trust assessment prototype. These guidelines are meant to ensure comprehensive testing of the tool under various conditions and inputs.

- Ensure that the input data files are in the correct CSV format with appropriate headers matching the expected node names.
- Confirm that the system has Python installed with all necessary libraries: pandas, pgmpy, networkx, numpy, and matplotlib.
- Ensure that input values, especially for node numbers and levels, stay within the defined ranges to avoid system crashes or unexpected behaviour.

Test Cases

Testing is a crucial component in the software development life cycle of our prototype, aimed at assessing the trustworthiness of WMDs. This section outlines a series of structured test cases designed to rigorously evaluate the system's capabilities and ensure its reliability and accuracy under various operational conditions. Each test case is created to challenge different aspects of the system, from basic functionality and error handling to performance under complex scenarios. These tests as shown in Table 6.2.

Table 6.2: Test cases for prototype configuration

Test Case ID	**Description**	**Input Data**	**Expected Outcome**
TC1	Validate system with minimum nodes	3 nodes, 1 levels, all interconnected	Successful network creation and visualization
TC2	Validate system with maximum nodes	100 nodes, 3 levels, all interconnected	Successful network creation and visualization
TC3	Test with non-existent CSV file	Invalid file path	Error message and safe failure
TC4	Boundary test for node levels	101 nodes, 3 levels	Error message about node limit
TC5	Test with missing data in CSV	Valid CSV missing some node data	Handle missing data gracefully, with warnings
TC6	Check response to negative weights	Nodes with negative relationship weights	Proper handling of negative weights in computations

6.3 Results and Discussions

The evaluation of the prototype to assess trust in WMDs through our comprehensive testing plan has yielded insightful results. This section discusses the results of various test cases that were designed to assess different aspects of the system's functionality, robustness, and user interaction capabilities.

6.3.1 User Interface and Data Handling

The initial phase of testing focused on user interaction and data handling capabilities, crucial for ensuring the system's practicality in real-world settings. The interface was tested for user-friendliness and the ability to guide users through complex configurations without prior technical knowledge.

- **User Input Validation:** Inputs through the system were rigorously tested to validate the handling of diverse data formats and user errors. The tool successfully rejected invalid inputs with informative error messages, effectively preventing potential data processing errors. as shown in Fig. 6.7 (example screenshots of the validation process).
- **Data Loading and Validation:** The system demonstrated robust capabilities in loading and validating structured datasets. Tests specifically designed to address common data issues, such as missing values or incorrect data types, were handled correctly, ensuring that the system could proceed with accurate network computations even under suboptimal data conditions as shown in Fig. 6.7.

6.3.2 Dynamic Bayesian Network Construction and Visualization

Following data validation, the focus shifted to the core functionalities of network construction and BN visualization, which are critical to the practical application of the tool in assessing the trustworthiness of WMD.

```
 Welcome to the Trust Measuring App
Please ensure your data file is in CSV format with appropriate headers for each node.
Enter the total number of nodes in the system (min 3, max 100): 200
Error: The number of the total nodes must be between 3 and 100.
Enter the total number of nodes in the system (min 3, max 100): 7
Enter the total number of levels in the system (min 2, max 50): 51
Error: The number of the total levels must be between 2 and 50.
Enter the total number of levels in the system: 3
Enter the number of nodes in Level 1: 4
Enter the number of nodes in Level 2: 2
Enter the number of nodes in Level 3: 1
Entering parent nodes for node X5:
Enter the parent nodes of X5, separated by commas (e.g., X1,X2): X1 -X2
Error: Invalid parent nodes. Ensure they are lower level and correctly formatted.
Enter the parent nodes of X5, separated by commas (e.g., X1,X2): X1, X2
Enter the relationship strength (choose ++, +, 0, -, --): *
Error: Invalid strength. Choose one of the following: ++, +, 0, -, --.
Enter the relationship strength (choose ++, +, 0, -, --): +
Enter the relationship strength (choose ++, +, 0, -, --): ++
Entering parent nodes for node X6:
Enter the parent nodes, separated by commas (e.g., X1, X2): X3, X4
Enter the relationship strength (choose ++, +, 0, -, --): -
Enter the relationship strength (choose ++, +, 0, -, --): --
Entering parent nodes for node X7:
Enter the parent nodes, separated by commas (e.g., X1,X2): X5, X6
Enter the relationship strength (choose ++, +, 0, -, --): 0
Enter the relationship strength (choose ++, +, 0, -, --): 0
```

Figure 6.7: Example of user input validation process. Error message displayed in red. User input is displayed in green

- **Logical Checks in Constructing the BN:** The system checked the rules of the Bayesian DAG and returned error if the criteria for acyclic digraph was not met as shown in Fig. 6.8 (example screenshots).

- **Bayesian Network Construction & Visualization:** The system was able to construct BN dynamically according to user-defined specifications. We tested the various conditions in the test case Table 6.3. The BN was developed for a minimum of 7 nodes and 3 levels to test the prototype. Even under the stress test with hundred nodes and 12 levels (TC2), the prototype efficiently managed the complexity, showcasing its scalability and performance. Fig. 6.9 shows the visualization of a complex network structure for a case of 45 nodes and 8 levels.

```
Do you want to perform query-based inferencing (yes/no): yes
Enter the node for query (or type 'exit' to stop): X3
Enter evidence in the form Node1=Value1,Node2=Value2,...: X1=88, X2=1
Error in processing query 88: Use Valid states. Please try again.
Enter the node for the query (or type 'exit' to stop): X1=0, X1
Enter evidence in the form Node1=Value1,Node2=Value2,...: X1=2, X2=1
Error in processing query: The node X1=0 is not in the digraph. Please try again.
```

Figure 6.8: Example of user-directed acyclic graph logical validation. Error message displayed in red. User input is displayed in green

- **Visualization and Interactivity:** The visualization module proved highly effective, providing clear and interpretable graphical representations of the BN. This feature greatly helped the users for its ability to illustrate the influence of various nodes and their interdependencies within the network as shown in Fig. 6.9.

6.3.3 Parameter Estimation and Query Handling

The system's ability to handle queries and present results was tested to assess its analytical capabilities and the effectiveness of its output in supporting decision-making processes.

- **Query-Based Inferencing:** Test cases involving complex queries demonstrated the system's capacity to execute inferencing tasks efficiently. The results were presented in a user-friendly manner, allowing users to gain meaningful insights into the trustworthiness aspects of various WMDs based on the computed probabilities. Illustrative results from query tests are presented in Table 6.3.

The results for each test case in Table 6.2 are presented in Appendix C. In general, the testing phase not only validated the functionality of the prototype but also highlighted its potential to enhance the decision-making process with respect to the use of WMDs. The results have substantiated the prototype's readiness for deployment in a real-world setting, offering a robust platform for healthcare professionals and end-users to assess and compare the trustworthiness of WMDs.

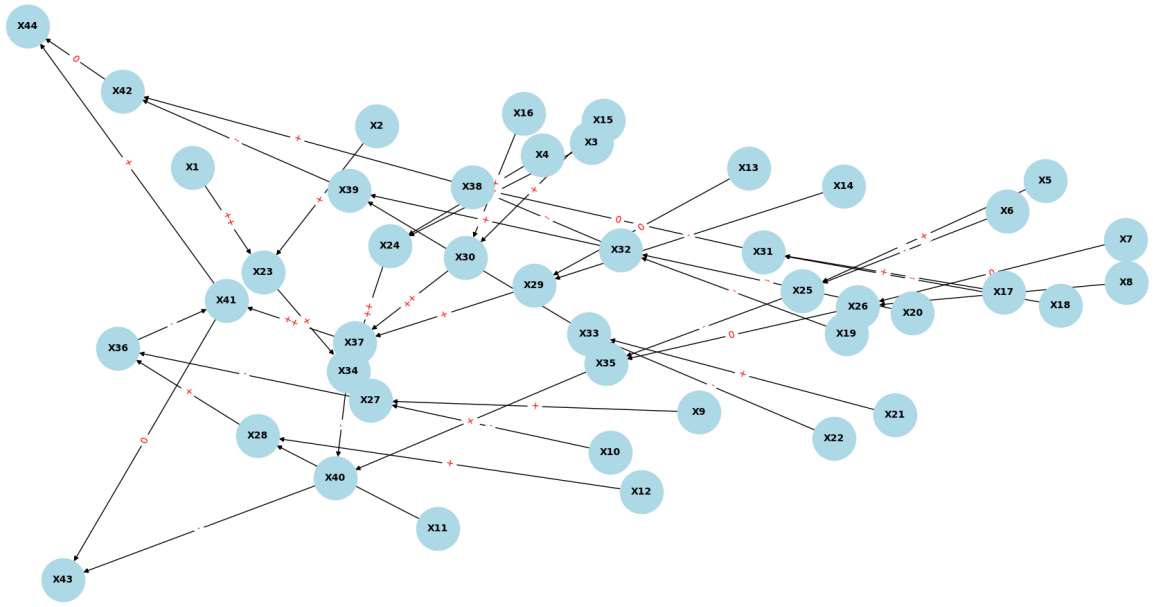

Figure 6.9: Bayesian network graph with forty-five nodes and eight levels of hierarchy

6.4 Summary

This chapter outlines the development process, testing, and validation of a BN prototype designed to evaluate the trustworthiness of WMDs. Starting with the requirement elicitation phase, we identified and documented both the requirements- functional and non-functional, crucial for the prototype's success. Functional requirements focused on the dynamic structuring, parameter estimation and query inferencing of the BN, while NFRs emphasized scalability, performance, and robust error handling to ensure the system's reliability and efficiency in real-world applications.

Table 6.3: Examples of query inferencing for nodes in the trust Bayesian network for evaluating the probability of a certain node in a state given its parents

Total Number of Nodes	BN Query Inference	Value
7	$P(X_7^2 \mid X_5^1, X_6^0)$	0.333
19	$P(X_{19}^2 \mid X_{16}^1, X_{17}^0, X_{18}^0)$	0.217
19	$P(X_{16}^2 \mid X_1^2, X_2^2, X_3^0, X_4^1, X_5^2)$	0.713
36	$P(X_{36}^2 \mid X_{34}^1, X_{35}^2)$	0.816
36	$P(X_{33}^2 \mid X_{27}^1, X_{28}^1, X_{29}^1)$	0.567
99	$P(X_{75}^2 \mid X_{19}^1, X_{20}^2, X_{21}^0, X_{22}^2)$	0.660
99	$P(X_{99}^2 \mid X_{95}^1, X_{96}^2, X_{97}^2, X_{98}^2)$	0.789

The second phase was the system analysis phase, which effectively transforms the initial conceptual requirements into detailed actionable system specifications. This critical phase tackles both requirements functional and non-functional to ensure they are well defined, technically feasible, and align with stakeholder expectations, thereby laying a solid foundation for the system's architecture and providing clear guidelines in terms of technical feasibility, computational complexities, data integrity, and efficient scalability with increasing nodes and levels. This stage also involves modelling and simulation of a PGM tailored to WMDs with interface requirements and validating this model with real and synthetic data to ensure the system's readiness and effectiveness.

The system analysis phase was followed by the system design phase, which transformed these specifications into a comprehensive architecture, illustrating a clear, modular approach with distinct layers to facilitate maintenance and scalability. Each component's role was clearly defined, supporting the system's overall functionality from data input to the visualization of the BN. Subsequent to the design phase, the implementation involved setting up the environment, programming the necessary functionalities, and integrating different components of the system. This phase was critical in bringing the conceptual model to life, ensuring that each element functioned as intended.

Testing was extensive, encompassing various scenarios to verify the system's robustness and responsiveness. We employ a series of test cases, from simple validations like input format checks to complex operational tests such as handling maximum nodes and levels. The results from these tests confirmed that the prototype met all specified requirements and handled edge cases effectively, showcasing its capability to perform reliably under different conditions.

The discussions provided insights into the potential applications of the prototype in real-world scenarios, emphasizing its utility in providing accurate trust assessments of WMDs. In future work, prototype functionalities can be improved with its adaptability to new types of wearable devices and emerging technologies. The prototype can be extended to a Graphical User Interface (GUI) for an intuitive and more engaged user experience. Overall, this chapter not only demonstrated the successful implementation and testing of the prototype, but also highlighted its potential impact on the decision-making process regarding the use of WMDs, ultimately contributing to enhanced device selection based on trustworthiness.

Chapter 7

Conclusion and Future Work

In this chapter, we summarize the thesis contributions and discuss future directions for research.

7.1 Summary of Contributions

We started this research with the primary goal of quantifying trust in WMDs. To accomplish this, we established two key objectives: firstly, to empirically identify and validate the factors that influence the trustworthiness of WMDs; and secondly, to develop a model that effectively quantifies trust, taking into account the inherent complexity and uncertainty of trust. The process of developing a solution to meet these objectives began by validating the underlying assumptions and gaining an understanding of the concerns of trust of WMD users. Thus, we narrowed our scope and defined an interim research objective to understand the TC and behaviours of WMD users in a health application domain.

7.2 Empirical Study

Our empirical investigation (Chapter 3) revealed a discrepancy between the perceived importance of trust factors by users and the actual determinants influencing the adoption of WMDs. Notably, our study identified that the factors deemed crucial in the literature, such as security and privacy, did not have a significant impact on user behaviour concerning WMD adoption. Approximately

45% of the participants indicated a willingness to utilize WMDs despite potential security and privacy risks. Additionally, we observed significant reluctance among medical and healthcare professionals regarding the endorsement and usage of WMDs.

This research yields several critical implications: Firstly, there is an evident need to enhance the awareness of essential trust factors like security and privacy among WMD users. We advocate for the development and implementation of alternative decision-making strategies that can elevate user awareness regarding these pivotal factors. Secondly, the prevalent uncertainty surrounding medical liability with regard to WMDs must be addressed. We recommend that engineering research groups collaborate with professional medical associations to formulate guidelines that clarify the responsibilities of both physicians and patients, thereby reducing the risks associated with medical liabilities. Thirdly, efforts should be made to address physician buy-in among medical professionals. This involves educating organizational partners, cultivating physician champions, and supporting them in adopting WMDs safely and effectively.

These findings provide a foundational framework for identifying user-specific trust factors, which is instrumental in designing a model to quantify trust in WMDs. This approach bridges the gap from theoretical research to practical application, setting a pathway for future developments in trust quantification within digital health technologies.

7.3 Trust Quantification Model

The second research objective focused on developing a model to quantify trust. In this thesis, we introduced a Bayesian-based trust quantification model to quantify trust in WMDs in Chapter 4. Bayesian networks are directed acyclic probabilistic graphs, represented by *nodes* and *edges*, which compactly depict a set of random variables and their conditional dependencies. In a BN, nodes represent random variables, while directed edges illustrate relationships between these variables, distinguishing between non-descendant nodes (nodes without parents) and descendant nodes (nodes with one or more parents). The BN is defined by its *structure* and *parameters*. In this thesis, we presented a framework using BN to quantify the uncertain and subjective nature of trust in WMDs.

The development of the proposed BN framework involved a structured two-step process. First, a BN structure is constructed, comprising of nodes, representing trust factors and edges depicting their interconnections. These factors were identified by our empirical findings and existing standards in literature and regulatory frameworks. Non-functional requirement techniques are employed to structure inter-node relationships. These NFRs help in refining broad trust goals into specific, measurable objectives that can be effectively assessed. We developed a hierarchical BN with layers decomposed into components (measurable factors), indicators and determinants (abstract and subjective aspects of trust). A simplified BN structure was developed for a single-sensor WMD as a preliminary study.

The second phase of the two-step process was to estimate the parameters of the BN. Probability density functions for the priors of all nodes were created using a Gaussian distribution based on expert domain knowledge to estimate parameters. This Bayesian-based trust network enabled query-based inferences about different states of the nodes based on existing evidence (e.g., estimating the probability of the trust node being in a high state when reliability is low). The trust BN was evaluated across various conditions. Evaluation metrics such as expectation mean and area under the curve were employed to assess and demonstrate the applicability of our model. Additionally, the scalability of the proposed model was tested by increasing the number of nodes.

The developed BN effectively demonstrated a viable method for quantifying the multidimensional, and abstract concept of trust in a WMD, incorporating diverse domain-related factors. The results confirmed that BNs could be effectively utilized to measure trust in systems under various configurations, and these systems could be expanded to accommodate higher dimensions. Furthermore, with established priors for all BN nodes, we could provide effective query-based learning under different conditions, aiding stakeholders like Alex in our motivational scenario to make informed trust decisions.

7.4 Data-driven Approach to Quantify Trust

The initial BN model introduced in our preliminary study in Chapter 4 was constructed with a focus on a single-sensor WMD, utilizing simulated Gaussian priors. Recognizing that WMDs

typically comprise of multiple complex and heterogeneous sensors with potentially non-uniform distributions, we extended our model considering multiple sensors and learned the BN parameters using real data collected from the WMDs, in Chapter 5.

In this extended framework, we adopted a data-driven approach to generate priors and estimate parameters within the BN, to accurately quantify trust. This method leverages actual data from WMDs to establish base probabilities for all nodes, facilitating the generation of priors. Subsequently, propagation rules are applied to define and quantify relationships based on the strengths of the impacts of the relationship, allowing for the calculation of CPDs for the descendant nodes of the BN. This setup enables precise computation of the probability of trust (or any other specific node) based on available evidence to evaluate the relative trustworthiness across different WMDs under uniform testing conditions.

Our methodology was validated through the deployment of four distinct BN configurations (BN1-BN4), aimed at quantifying trust in WMDs based on identified trust factors. The parameters for our BN framework were learned using our data-driven strategy across two WMDs under identical scenarios. We conducted a comparative analysis to evaluate the BN's performance, varying granularization levels, impact weights, and node heterogeneity. The outcomes affirmed the learnability and generalizability of our data-driven method. Furthermore, the BNs were tested with a larger set of data. We systematically created synthetic data under various normal and noisy conditions to explore their impact on trust assessments using our BNs (BN1-BN4). Synthetic data, created with different noise intensities, enabled the simulation of various operational conditions and offered a strong foundation for model validation. This ensured that our models maintained accuracy and dependability across varied and possibly challenging environments. Utilizing both real and synthetic data essentially helped in verifying the performance of our BNs in real-world scenarios.

An additional finding of our thesis is that if manufacturers of WMDs consider releasing sample datasets along with the datasheets of their respective devices, the practice would enable researchers, developers, and end-users to benchmark the performance and trust metrics of their devices with the models discussed in this thesis. By providing sample datasets, manufacturers can facilitate more

transparent and comparative evaluations, helping users to better understand the product in terms of important factors (such as reliability, accuracy, and overall trustworthiness) of their devices. Such an initiative would not only enhance user confidence but also encourage more evidence-based decision-making in the selection of WMDs.

7.5 Parameterized Prototype for Trust Quantification

In order to show the proof-of-concept of our model, we developed a prototype. We followed the software development life cycle steps to systematically develop the prototype. We documented the systematic development of our parameterized prototype through various phases, beginning with a detailed description of the prototype, progressing through its rigorous testing, and culminating in the evaluation of trustworthiness for WMDs, in Chapter 6. This comprehensive approach ensures a thorough understanding and validation of the prototype's capabilities in assessing the trustworthiness of WMDs in real-world scenarios.

Initially, we presented a detailed description of the prototype, that established the foundation for the first phase—requirement elicitation. We provided various use cases and meticulously documented the functional requirements and the NFRs essential for the prototype's efficacy. Functional requirements centred on the dynamic structuring of the BN, whereas NFRs highlighted scalability, performance, and robust error handling to maintain the system's reliability and efficiency in practical settings.

Next, the system analysis phase transitioned the initial conceptual requirements into comprehensive, actionable specifications. We addressed both requirements functional and non-functional, ensuring technical feasibility, proper alignment with stakeholder expectations, and robust system architecture. This phase also encompassed the evaluation of computational complexities and assurance of data integrity. Additionally, it focused on scalable system design as node numbers increase.

Following this, the system design phase transitioned these requirements into a detailed architectural blueprint, showcasing a clear, modular strategy with distinct layers to enhance maintenance and scalability. Each component was carefully defined to support the system's comprehensive

functionality from data input through to the visualization of the BN. The implementation phase was pivotal in actualizing the BN prototype tailored to WMDs, incorporating interface requirements, environment setup, programming of necessary functionalities, and integration of various system components to ensure seamless operation and interaction. Further, the implementation was tailored to incorporate various interface requirements.

Finally, extensive testing covered a range of scenarios to ensure the system's robustness and responsiveness. A variety of test cases were conducted, ranging from basic input format checks to complex operational tests designed to manage the maximum number of nodes and levels. The outcomes of these tests confirmed that the prototype fulfilled all specified requirements and effectively managed edge cases, demonstrating its reliability across various conditions. We successfully established a replicable framework that organizations across healthcare can adopt and adapt.

7.6 Future Work

In this section, recommendations for future work are discussed in four areas: 1) improve the outcomes of the empirical study, 2) enhance the structure learning, 3) enhance the BN parameter learning, and 4) GUI based user intuitive, and interactive system.

Firstly, we would like to address the limitation of the empirical study reported in Chapter 3, of a relatively small number and skewed participants. Due to this limitation, some of our hypotheses could not be analyzed effectively and hence were inconclusive. Our study also focused only on WMD users. The findings from our study underscore a significant issue with demographic representation among users of WMDs, revealing a potential bias in data that could adversely affect the fairness and utility of health recommendations derived from these technologies. The under representation noted may lead to data insights that do not accurately represent the diverse needs and health outcomes of the general population, potentially exacerbating existing inequities if WMDs are predominantly designed or calibrated based on data from specific groups who use these devices more frequently [70, 23].

The lack of diverse user data can result in several critical challenges. Models developed with data from a homogeneous group may not perform accurately across other demographics, leading to

erroneous assessments. This situation risks widening the gap in health disparities among different socioeconomic and racial groups by making WMDs less trustworthy and accessible or less accurate for some populations. Additionally, research and conclusions drawn from such biased datasets might lack generalizability, limiting the effectiveness of health interventions based on these findings. To mitigate the limitations in this area of the research we propose an alternate method of data collection and strategy for the empirical study as future work:

1. The inclusion of broader societal factors, such as the digital divide and social determinants of health, presents an essential direction for future research. Our current study prioritized determinants related to device functionality, user experience, and social interactions; however, expanding the model to incorporate digital access disparities and socio-environmental influences would provide a more comprehensive evaluation of trust in WMDs. Addressing these determinants could significantly improve the inclusivity and equity of WMD adoption, enhancing our understanding of trust across diverse populations and reducing biases in health technology deployment [112, 140].

2. It is essential to adopt inclusive data collection strategies that encourage the use of WMDs across a broader range of population groups including the group who distrust and do not use WMDs. The survey language can be refined to be more accessible for the general population. Engaging with communities that are underrepresented in WMD usage to co-develop solutions tailored to their needs and continuously monitoring the impact of WMD trustworthiness on different population groups will further enhance fairness and effectiveness, ultimately ensuring that WMDs benefit the entire spectrum of society [22, 159].

3. Furthermore, due to the limitation of using the survey as the study's instrument, we relied on self-reporting measures to understand the WMD users' awareness of trust and behaviour in adopting WMD. To maintain the accuracy of the self-reported behaviour, we asked indirect questions and/or questions about familiar environments from which participants could more accurately report their behaviour. However, the accuracy could be improved if the study is re-designed to include group interviews (with focus groups) or a lab experiment (with direct observation of participants' trust behaviour when using technical devices such as WMD) [61].

The group interview or a lab experiment has its own challenges in terms of logistics and controlled lab settings. However, it also has advantages compared to the survey instrument, including decreasing inaccurate self-reporting behaviour, as well as capturing aspects of users' behaviour that can only be determined by the manipulation of independent variables [70] through direct measurements. Measuring users' perspectives and behavioural aspects when the parameters controlling a WMD change are interesting future research directions that can be pursued through direct observations.

Secondly, we would like to address the limitation in terms of BN structure generation discussed in Chapter 4. The nodes in the trust BN are subjective in terms of structure and also the qualitative levels that define the relationships between the nodes in the structure to indicate the relative trustworthiness of a WMD. To address the subjective nature of the node structure and qualitative relationships in the trust BN proposed for WMDs, several strategic improvements are recommended:

1. Adopting automated structure learning can significantly reduce reliance on expert-defined node relationships. By employing algorithms that derive the most probable network structure from data, such as constraint-based or score-based methods, the model can autonomously identify optimal configurations, thus minimizing subjective bias [71, 123].

2. Another important step is to expand the diversity of the expert panel involved in defining the nodes and their interrelationships is crucial. This panel should include specialists from multiple disciplines such as biomedical engineering, data science, and ethics, and representatives from end-user groups including patients and healthcare providers. This broadened perspective ensures a more comprehensive and unbiased approach to model construction [140].

The third limitation we would like to address is related to the learning of the parameters, as discussed in Chapter 5. In the proposed data-driven method for quantifying trust using BN, the parameter learning are based on the data collected by the WMDs for the first layer (non-descendant nodes) of our BN. The data for all the remaining nodes are generated from the data collected from the non-descendant nodes based on the relationships between the nodes. The relationships between

the nodes are based on the qualitative levels of influence between the nodes (the relationship of the parent with the child) as defined by the domain expert, which is subjective. To mitigate the limitation due to the subjective nature of parameter estimation in our trust BN, the following strategy can be considered:

1. Incorporating data from multiple sources. By expanding the dataset to include user feedback, clinical outcomes, and varied device performance metrics, the model can capture a more diverse and representative understanding of the factors influencing trust. This approach not only balances the data inputs but also enhances the overall objectivity of the parameter estimation process [105].

Lastly, we would like to address the user interface limitation of our prototype as discussed in Chapter 6. The prototype is developed for interactions via the terminal (command prompt), which may not be the most interactive or user-friendly method, particularly for users without technical expertise. This approach can result in a steep learning curve and may deter user engagement. To enhance usability and accessibility, the following improvement is recommended:

1. Develop and integrate a GUI for the BN-based trust prototype. This interface should be designed to facilitate intuitive interactions, allowing users to visually manipulate the BN structure, adjust parameters, and view results in real time. The GUI should aim to reduce complexity and make the process of trust assessment more accessible to all users, regardless of their technical background. This strategy aligns with established usability principles that emphasize the importance of intuitive graphical interfaces in enhancing user experience and engagement [99, 75].

In summary, future work can be considered in four areas: Firstly, the empirical study's limitation due to a small sample size is acknowledged, which may have led to inconclusive results for some hypotheses. To mitigate this, future studies should employ inclusive data collection strategies to encompass a broader demographic, including those who typically do not use WMDs. This approach would ensure the data accurately reflects the diversity of the general population and aids in eliminating bias in health recommendations derived from WMDs. The second area of focus would be enhancing the BN for trust quantification in WMDs. The current model's reliance

on expert-defined structures and relationships can introduce subjectivity, affecting the trustworthiness assessments. Future improvements could include adopting automated structure learning techniques to reduce reliance on expert input and incorporating a diverse panel of experts to ensure a balanced and comprehensive model construction. Thirdly, expanding the parameter learning by integrating data from multiple sources would enhance the model's objectivity and representatives. Lastly, an enhanced user interactive interface is recommended by developing a GUI that simplifies user interaction with the prototype, supporting intuitive operations, and allowing users to easily modify the network and visualize results. These strategic enhancements aim to refine the effectiveness of the prototype and expand its applicability to diverse user groups, ensuring equitable and efficient trust assessments in WMDs.

Appendix A

Survey Questionnaire

In this appendix, we introduce the questionnaire developed for our survey. The questionnaire's design aims to explore which factors, as identified in existing literature, align with users' trust in the context of WMDs. It is crucial for users to understand various trust-influencing factors recognized by the literature and regulatory authorities to ensure a safe and reliable experience. Our survey questions are crafted to gauge participants' awareness and viewpoints on several trust-related aspects and their subsequent behaviours. The questionnaire is organized around eleven constructs, each linked to a specific hypothesis and represented by one or more questions.

Survey Questionnaire

Survey on Trust in Wearables

This survey is administered by Mini Thomas, Department of Computing and Software Engineering, McMaster University. The purpose of the survey is to understand the users' concerns, attitudes, and behaviours regarding trusting and adopting wearable medical devices (WMD).

A wearable medical device is defined as a device that is autonomous, non-invasive and that performs a specific medical function such as monitoring or support over a prolonged period of time. The term wearable implies that the device is worn and either supported by the human body or clothing.

To learn more about the survey and the researcher's study, particularly in terms of any associated risks or harms associated with the survey, how confidentiality and anonymity will be handled, withdrawal procedures, how to obtain information about the survey's results, how to find helpful resources should the survey make you uncomfortable or upset, please read the accompanying letter of information:
https://docs.google.com/document/d/1625rMiZ3QK8rYsUFQ-mckhRAO4DU2aYCV_dxsZtT0sM/edit?usp=sharing

This survey should take approximately 10 minutes to complete. People filling out this survey must be 18 years of age or older.

This survey is part of a study that has been reviewed and cleared by the Mohawk College Research Ethics Board (MCREB), McMaster Research Ethics Board (MREB), and Toronto Metropolitan University Research Ethics Board (TMU REB). The MCREB protocol number associated with this survey is 23-014, the MREB protocol number is 5788, and the TMU REB protocol number is 2022-072-1.

You are free to complete this survey or not. If you have any concerns or questions about your rights as a participant or about the way the study is being conducted, please contact:

Mohawk College Research Ethics Board, E-mail: reb.coordinator@mohawkcollege.ca
or
McMaster Research Ethics Office, Telephone: (905) 525-9140 ext. 2314, E-mail: mreb@mcmaster.ca
or
Toronto Metropolitan University Research Ethics Board E-mail: rebchair@torontomu.ca

CRITERIA

This survey is about trust in wearable medical devices. A wearable medical device is an electronic device that can be worn as an accessory or embedded in clothing that can monitor health data such as temperature, blood pressure, blood sugar, heart rate, motion, oxygen saturation, pulse, etc. for a prolonged period of time. For example, using a wearable glucose meter to monitor sugar level or an accelerometer to monitor activity.

To participate in this survey, you must meet the following criteria:

1) Be over 18 years of age, and,

2) Have at least two weeks or more experience using wearable medical devices (such as continuous glucose monitors, ECG, heart rate, sleep tracking, activity tracking, oxygen saturation, wearable insulin pumps, pulse oximeters, respiratory rate, body temperature, wearable hearing aids, sleep trackers or other similar devices) for health monitoring of self or other members (such as family/ friends/ patients or working on a research project).

Please choose only one of the following:

☐ Yes, I meet the criteria

☐ No, I do not meet the criteria

CONSENT

- I have read the information presented in the information letter about a study being conducted by Mini Thomas.
- I have had the opportunity to ask questions about my involvement in this study and to receive the additional details I requested.
- I understand that if I agree to participate in this study, I may withdraw from the study at any time until I submit my responses, but once my responses have been submitted, they cannot be withdrawn due to the anonymous nature of the study.
- I have been given a copy of this form.

Having read the above, I understand that by clicking the "Yes" button below, I agree to take part in this study under the terms and conditions outlined in the accompanied letter of information. Please choose one of the following:

☐ Yes, I agree to participate in the study.

☐ No, I do not wish to participate in the study.

USER PRIOR EXPERIENCE QUESTIONS

1) What did you use wearable medical devices for? Select all that apply.
   - ☐ To monitor health parameters for self, family, or friend
   - ☐ To monitor the health parameters of patients in hospitals, clinics, or homecare
   - ☐ In research, projects, course studying/ teaching

   Other (please specify:)

2) What kinds of wearable medical devices have you used in the past? Select all that apply.
   - ☐ Activity and fitness trackers (e.g., Fitbit Charge 4)
   - ☐ Cardiac monitors (e.g., Apple Watch Series 4 and later with ECG app)
   - ☐ Continuous glucose monitors (e.g., Libre, Dexcom G6)
   - ☐ Blood pressure monitors (e.g., Omron Heart Guide)
   - ☐ Respiratory monitors (e.g., Empatica E4)
   - ☐ Temperature monitors (e.g., Apple Watch)
   - ☐ Sleep trackers (e.g., Oura Ring)
   - ☐ Pulse oximeters (SPO2) (e.g., Masimo)
   - ☐ Hearing aids (e.g., Phonak)

   Other (please specify:)

3) How long have you used or worked with wearable medical devices?
   - ☐ More than 2 weeks but less than 6 months
   - ☐ More than 6 months but less than 1 year
   - ☐ More than 1 year but less than 3 years
   - ☐ More than 3 years but less than 5 years
   - ☐ More than 5 years

   Other (please specify:)

4) How has your experience been in using and trusting wearable medical devices?

☐Very Difficult
☐Difficult
☐Manageable
☐Convenient
☐Very convenient

USER DEMOGRAPHY QUESTIONS

5) How do you self-identify in terms of gender?

6) What is your ethnic or cultural origin(s)?

Ethnic groups have a common identity, heritage, ancestry, or historical past. Examples include Canadian, Chinese, East Indian, English, Italian, Filipino, Scottish, Irish, Portuguese, German, Polish, Dutch, French, Jamaican, Pakistani, Iranian, Sri Lankan, Korean, Ukrainian, Lebanese, Guyanese, Somali, Colombian, and Jewish. (Ethnic or cultural origins should not be confused with citizenship, nationality, language, or place of birth.

7) What is your age?

☐ 18-24
☐ 25-34
☐ 35-44
☐ 45-54
☐ 55-64
☐ 65 or older

8) What is your highest level of education?

☐ High School
☐ Post-Secondary Diploma
☐ Associate Degree
☐ Bachelor's Degree
☐ Master's degree
☐ Doctoral Degree
☐ Professional Degree

Other (please specify:)

9) What is your professional background or area of expertise?

☐ Healthcare / Medical (including Medical Doctors, Nurses, Paramedics, Dentists, Pharmacists, Physiotherapists, Dieticians, and Personal Care Workers

☐ Engineering / Technology / IT

☐ Marketing / Advertising

☐ Finance / Banking/Business

☐ Education

☐ Manufacturing

☐ Arts/ Entertainment

Other (please specify:)

DEVICE ACCURACY QUESTIONS

Device accuracy measures an individual's belief in the correct and accurate measurements of their health parameters using the wearable medical device.

10) To what extent does the **accuracy** of the wearable medical device influence your trust in using the device?

☐ Not at all influential (I always trust the device)

☐ Slightly

☐ Neutral

☐ Moderately

☐ Very influential

11) Have you ever cross-checked the readings of a wearable medical device with another device to verify and be aware of its accuracy?

☐ Never

☐ Seldom

☐ Sometimes

☐ Often

☐ Always

12) If the answer to the above question is yes, how do you typically verify (e.g., compare the wearable measurements with other devices) the accuracy of a wearable? Select all that apply.

☐ I always trust the wearable and do not verify

☐ Cross-referencing with professional medical devices (compare the wearable measurements with other devices such as Smartwatch)

☐ Relying on personal experience and consistency of readings over time

☐ Trusting the device manufacturer's claims

☐ Seeking advice from healthcare professionals

Other (please specify:)

DEVICE RELIABILITY QUESTIONS

Device reliability measures an individual's belief in the device for seamless operation and connectivity for the measurement of their health parameters (e.g., heart rate)

13) To what extent does the **reliability** of the wearable medical device influence your trust in using the device?

☐ Not at all (I always trust the device)

☐ Slightly

☐ Neutral

☐ Moderately

☐ Highly

14) Have you been aware of or ever experienced any issues with the reliability of a wearable? (e.g., reliable operation in terms of battery life, memory storage, and connectivity)

☐ Never

☐ Seldom

☐ Sometimes

☐ Often

☐ Always

Other (please specify:)

15) Which technical challenges related to connectivity issues have you encountered?

- ☐ Nothing in particular
- ☐ Poor Bluetooth/Wi-Fi connectivity
- ☐ Difficulty pairing the device
- ☐ Slow data transfer
- ☐ Compatibility issues with other devices

16) Has battery consumption been a problem in your experience with wearables?

- ☐ No at all, Battery lasted for more than 24 hours or even more
- ☐ Seldom, Battery lasted for about 20-24 hours
- ☐ Sometimes, the Battery needed to be charged twice a day
- ☐ Often, the Battery needed to be charged every 6-8 hours or even more often
- ☐ Yes always, Battery needed to be charged every 6-8 hours or even more often

17) How often have you lost data from your wearable device due to memory overload?

- ☐ Never
- ☐ Seldom
- ☐ Sometimes
- ☐ Often
- ☐ Always

DEVICE CRITERION VALIDATION QUESTIONS

Device validation measures an individual's belief that the wearable has undergone rigorous clinical validation (e.g., the performance of a wearable medical device designed to monitor heart rate to detect abnormal heart rhythms is validated by comparing with gold standards such as electrocardiogram (ECG) and testing, providing evidence of its effectiveness and safety.

18) To what extent does knowing that your wearable medical device has been **clinically validated** influence your trust in using it?

- ☐ Not at all influential (I always trust the device)
- ☐ Slightly
- ☐ Neutral
- ☐ Moderately
- ☐ Very influential

19) How confident are you that the wearables available in the market undergo sufficient clinical validation to ensure the device measurement effectiveness and safety?
    - ☐ Not at all confident
    - ☐ Slightly
    - ☐ Neutral
    - ☐ Somewhat
    - ☐ Very confident
20) Are you aware of the validation methods used by the manufacturer of your wearable medical device?
    - ☐ Not at all aware
    - ☐ Slightly
    - ☐ Neutral
    - ☐ Somewhat
    - ☐ Extremely aware

DEVICE SECURITY QUESTIONS

Device security measures an individual's belief that the device uses security measures such as encryption, authentication, and access control.

Definitions:

- ***Encryption*** in wearables refers to the process of encoding sensitive data or information transmitted by the device to protect it from unauthorized access or interception.
- ***Authentication*** in wearables refers to the process of verifying and confirming the identity of users or systems before granting access to sensitive data or functionalities. It ensures that only authorized individuals or entities are allowed to interact with the device and access its data.
- ***Access control*** in wearables refers to the process of managing and regulating the permissions and privileges granted to users or systems regarding the device's functionalities, data, and resources. It involves mechanisms and policies that determine who can access what information, perform specific actions, or modify settings within the device's ecosystem.

21) Does the presence of robust **security** features influence your trust in using your wearable medical device?
    - ☐ Not at all influential (I always trust the device)
    - ☐ Slightly
    - ☐ Neutral
    - ☐ Moderately
    - ☐ Very influential

22) According to you, what factors are important to evaluate security in wearables? Select all that apply.

- ☐ Authorized access control
- ☐ Multi-factor authentication
- ☐ Device compliance with security compliance and protocols
- ☐ Encryption
- ☐ None of the above

Other (please specify:)

23) Are you aware of security breaches in wearable medical devices?

- ☐ Not at all aware
- ☐ Slightly
- ☐ Neutral
- ☐ Somewhat
- ☐ Extremely aware

DEVICE DATA PRIVACY QUESTIONS

Device privacy measures an individual's belief in the compliance of privacy standards that provide a transparent outline of how the user data is collected, used, and shared.

24) To what extent does the awareness or assurance of strong **privacy** practices influence your trust in using wearable medical devices?

- ☐ Not at all influential (I always trust the device)
- ☐ Slightly
- ☐ Neutral
- ☐ Moderately
- ☐ Very influential

25) Are you aware of privacy practices implemented by wearable medical device developers to protect user data?

- ☐ No
- ☐ Yes

26) How confident are you in the ability of wearable medical device manufacturers to protect the privacy of the data collected?

- ☐ Not at all confident
- ☐ Slightly
- ☐ Neutral
- ☐ Somewhat
- ☐ Very confident

TECHNICAL SUPPORT QUESTIONS

Technical support measures an individual's belief in the responsive and reliable support available from the manufacturer to resolve issues with the wearables.

27) To what extent does the availability of good **technical support** influence your willingness to adopt and use wearable medical devices?

- ☐ Not at all influential (I always trust the device)
- ☐ Slightly
- ☐ Neutral
- ☐ Moderately
- ☐ Very influential

28) Have you ever contacted technical support for assistance with your wearable medical device? If yes, please share your experience.

- ☐ Yes
- ☐ No

29) Do you think that responsive customer support by the manufacturer is an important factor in trusting a wearable medical device?

- ☐ Not at all important
- ☐ Slightly
- ☐ Neutral
- ☐ Somewhat
- ☐ Very important

EASE OF USE QUESTIONS

Ease of use questions measure how comfortably and efficiently the user-friendly interfaces of the wearables allow the users to interact with the device.

30) Do you consider an **intuitive user interface** as an influential factor in trusting and adopting wearables?

- ☐ Not influential at all
- ☐ Slightly
- ☐ Neutral
- ☐ Somewhat
- ☐ Very influential

31) Does your confidence in using a new wearable medical device depend on the user interface?

- ☐ Not at all important
- ☐ Slightly
- ☐ Neutral
- ☐ Somewhat
- ☐ Strongly

RECOMMENDATION QUESTIONS

Recommendation questions measure an individual's belief in wearables based on endorsement or advice from healthcare professionals or acquaintances.

32) Have you ever received recommendations from healthcare professionals (e.g., doctors, nurses) or acquaintances to use wearables?

- ☐ Never
- ☐ Seldom
- ☐ Sometimes
- ☐ Often
- ☐ Always

33) If the answer to the above question is yes, please specify the type of healthcare professional(s) or acquaintances who recommended wearables to you. Select all that apply

- ☐ Healthcare professional
- ☐ Gym Trainers
- ☐ Friend/ Family

Other (please specify:)

34) To what extent were the recommendations influential in your decision to trust and use wearable medical devices?

- ☐ Not influential at all
- ☐ Slightly
- ☐ Neutral
- ☐ Somewhat
- ☐ Very influential

35) Would you recommend using wearables for your family or your patients?

- ☐ Definitely not
- ☐ Probably not
- ☐ Unsure
- ☐ Probably not
- ☐ Definitely yes

USER SPECIFIC QUESTIONS

User-specific questions measure the factors that are important to an individual to trust and adopt a wearable medical device (it may be the top current factors as given in the literature or it may be something entirely new /different that an individual is concerned about).

36) According to you, what are the hindrance(s) in trusting and adopting wearables for remote patient monitoring. Select all that apply.

- ☐ Buy-in from health professionals
- ☐ Buy-in from patient users
- ☐ Learning curve
- ☐ Language barriers

Other (please specify:)

37) According to you, what aspects (factors) of wearable medical devices help gain trust in adoption by users and medical professionals? List the top three factors

THANK YOU!

We appreciate your time to fill out this survey. We would be happy to share with you the survey result link. We are also grateful for your time in filling out this survey. In gratitude for answering the survey, we will offer participants a chance to enter a draw to win 1 PayPal card valued at $20. If you are interested in receiving the result information and entering the draw, please click Yes below

- ☐ Yes, I am happy to receive the survey results and enter a draw for a gift prize
- ☐ No, I am not interested in the survey results and enter a draw

Appendix B

Survey Analysis

B.1 Checking Assumptions of Normality

Normality testing is essential to verify the distribution of the data and ensure that the statistical methods used in the analysis yield reliable and valid results. In contexts where the data significantly deviates from a normal distribution, non-parametric methods might be necessary to accurately analyze the data. For our study, the normality of the distribution was tested for key demographic variables such as age, education, and ethnicity using the Shapiro-Wilk test as shown in Table B.1. The results indicate a significant deviation from normality, as the p-value is much less than 0.05.

Table B.1: Normality test results for the user demography construct

Characteristic	**Score**	***p*-value**
Age	0.780	4.851e-09
Education	0.918	1.485e-05
Ethnicity	0.890	0.132e-08

The skewness values derived from our analysis indicated that the sample data were not normally distributed with respect to age, education, and ethnicity and were skewed toward participants who are younger, highly educated and likely of Asian, European, and Canadian ethnicity.

B.2 Reliability of Survey Questionnaire

To evaluate the reliability and internal consistency of our questionnaire, we used Cronbach's Alpha, a statistical measure commonly used in social sciences research. Cronbach's Alpha assesses the degree to which a set of items or questions in a questionnaire are related to one another, thereby indicating their overall internal consistency. This metric is vital as it helps determine whether different items on a test consistently measure the same underlying attribute. A higher value of Cronbach's Alpha suggests that the items are more closely related, which in turn signifies that they are likely effective at assessing the intended construct.

For our analysis, we computed the Cronbach Alpha coefficients for questions related to specific constructs concerning device-related and external determinants that influence user trust and behaviour as shown in Table B.2.

The Cronbach's Alpha value obtained from our dataset was 0.78. This value indicates an acceptable level of reliability, confirming that the questionnaire is a reliable tool for measuring the specified constructs within our study.

Table B.2: Survey questions and Cronbach coefficient to test reliability of the questionnaire

Question	**Cronbach's Alpha**
Q10, Q11, Q13, Q14, Q16, Q17, Q18, Q19, Q21, Q23, Q24, Q26, Q27, Q29, Q30, Q31, Q32, Q34 and Q35	0.78

B.3 Validity of Survey Questionnaire

To enhance the validity of our survey, we incorporated a set of cross-validation questions that are specifically designed to confirm the accuracy of responses to other related questions. The approach involves analyzing the correlation between paired questions that are intended to measure similar constructs. A high correlation between these pairs indicates a strong validity of the measurements obtained. The findings from these cross-validation efforts, including the specific correlations observed and their implications for the overall survey validity, are discussed in this Section. A

comprehensive breakdown of the cross-validation results is provided in Table B.3, which provides a deeper insight into the methodological robustness of the survey.

Table B.3: Survey construct, original questions, related cross-validation questions, and responses to test validity of questionnaire

Construct	**Question**	**Cross-validation Question**	**Response**
Device Accuracy	Q11: Have you ever cross-checked the readings of a wearable medical device with another device to verify and be aware of its accuracy?	Q12: If the answer to the above question is yes, how do you typically verify (e.g., compare the wearable measurements with other devices) the accuracy of a wearable?	There were discrepancies in the responses to the paired questions- only 11 participants consistently trusted WMDs without cross-verification, while 57 admitted to not doing so, due to a lack of knowledge about cross-verification methods.

Continued on next page

Table B.3 continued from previous page

Construct	Question	Cross-validation Question	Response
Device Reliability	Q14: Have you been aware of or ever experienced any issues with the reliability of a wearable (e.g., reliable operation in terms of battery life, memory storage, and connectivity)?	Q15: Which technical challenges related to connectivity issues have you encountered? Q16: Has battery consumption been a problem in your experience with wearable? Q17: How often have you lost data from your wearable device due to memory overload?	There was a significant correlation in the responses to the paired questions in terms of reliability concerns among users, including connectivity issues such as signal loss (53%), battery recharge concerns (38%), and low awareness of memory overload.
Device Validation	Q20: Are you aware of the validation methods used by the manufacturer of your wearable medical device?	Q19: How confident are you that the wearable available in the market undergo sufficient clinical validation to ensure the device measurement effectiveness and safety?	Responses were coherent showing low awareness and influence from the clinical device validation methods and studies.

Continued on next page

Table B.3 continued from previous page

Construct	Question	Cross-validation Question	Response
Device Security	Q23: Are you aware of security breaches in wearable medical devices?	Q22: According to you, what factors are important to evaluate security in wearable?	Responses consistently showed that participants had a good idea of different security breaches and standards.
Device Data Privacy	Q25: Are you aware of privacy practices implemented by wearable medical device developers to protect user data?	Q26: How confident are you in the ability of wearable medical device manufacturers to protect the privacy of the data collected?	Responses consistently showed participants were aware of privacy issues approximately 60%) though it did not influence many of the users in their decision to adopt WMD.

Appendix C

Test Case Results for the Trust Assessment Prototype

In this appendix, we present the figures for each test case discussed in Section 6.2.6.

Table C.1: Test cases for Bayesian network configuration prototype

Test Case ID	**Description**	**Expected Outcome**	**Result Verified**
TC1	Validate system with minimum nodes	Successful network creation and visualization	Verified. See Fig. C.1
TC2	Validate system with maximum nodes	Successful network creation and visualization	Verified. See Fig. C.2
TC3	Test with non-existent CSV file	Error message and safe failure	Verified. See Fig. C.3
TC4	Boundary test for node numbers	Error message about node limit	Verified. See Fig. C.4
TC5	Test with missing data in CSV	Handle missing data gracefully, with warnings	Verified. See Fig. C.5
TC6	Check response to negative weights	Proper handling of negative weights in computations	Verified. See Fig. C.6

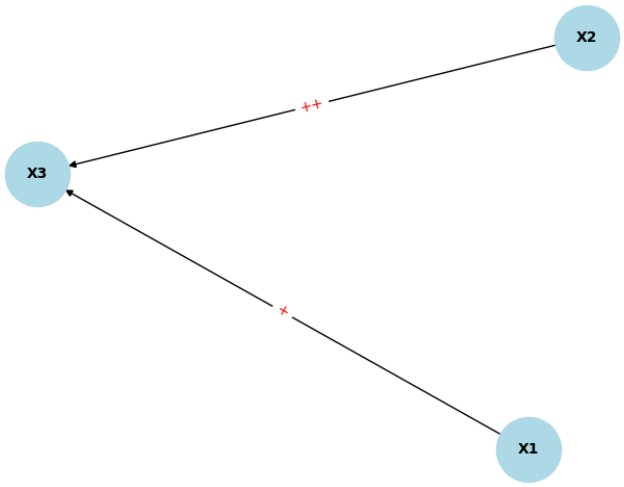

Figure C.1: Test Case 1- Successful Bayesian network creation and visualization with minimum nodes

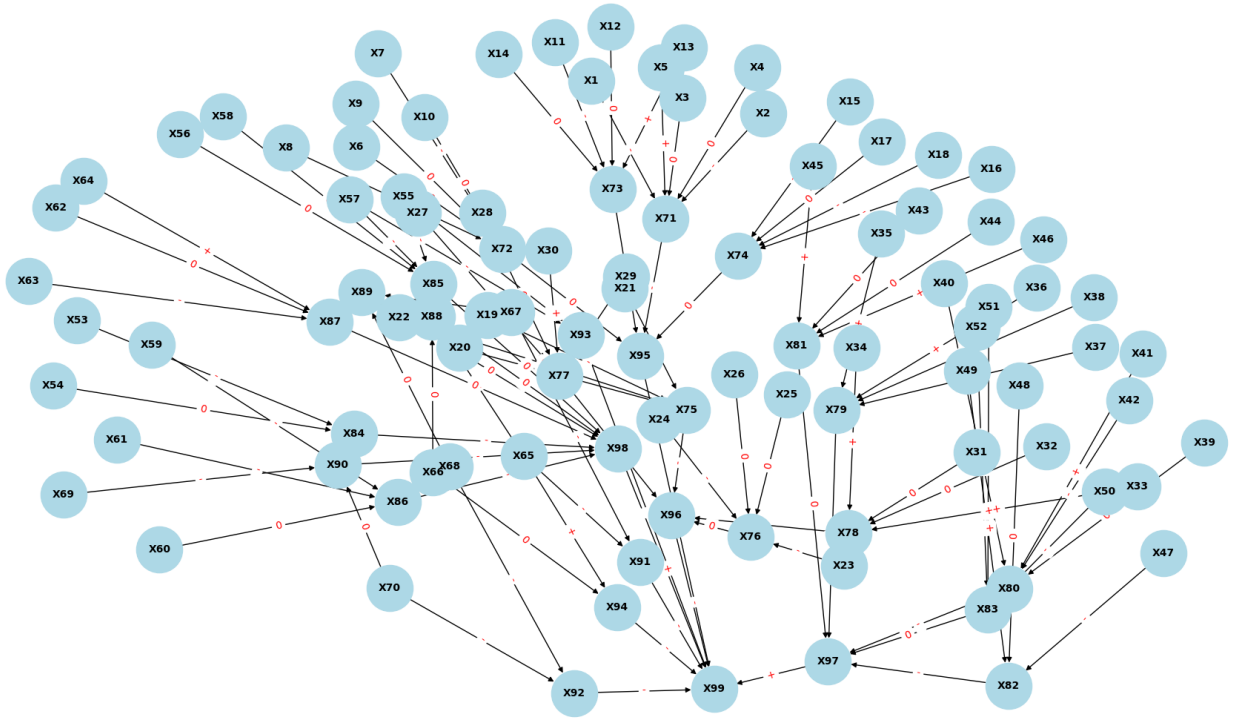

Figure C.2: Test Case 2- Successful Bayesian network creation and visualization with maximum nodes

```
Enter the parent nodes for node X3, separated by commas (e.g., X1,X2) or leave blank if
none: X1,X2
Enter the relationship strength for X1 (choose ++,+,0,-,00) (or type 'exit' to stop): 0
Enter the relationship strength for X2 (choose ++,+,0,-,00) (or type 'exit' to stop): 0
Enter the path to the CSV file containing data for all nodes: syntheticdata1
Error: File not found. Please enter a valid file path with a valid extension.
Enter the path to the CSV file containing data for all nodes: syntheticdata1.csv
```

Figure C.3: Test Case 3- Test with non-existent CSV file. Error message displayed in red. User input is displayed in green

```
Enter the total number of nodes in the system (min 3, max 100): 101
Error: The total number of nodes must be between 3 and 100.
Enter the total number of nodes in the system (min 3, max 100): 99
Enter the total number of levels in the system (min 2, max 50): 51
Error: The total number of levels must be between 2 and 50.
Enter the total number of levels in the system (min 2, max 50): 49
```

Figure C.4: Test Case 4- Boundary test for node numbers and levels. The error message is displayed in red. User input is displayed in green

```
Enter the path to the CSV file containing data for all nodes: syntheticdata1.csv
Error: Node X2 missing in file. Enter the correct file name with all nodes.
Enter the path to the CSV file containing data for all nodes: syntheticdata2.csv
Value Error: the sum or integral of conditional probabilities for node X2 is not equal to 1.
Check the file again.
```

Figure C.5: Test Case 5- Test for missing nodes and respective data in the file. The error message is displayed in red. User input is displayed in green

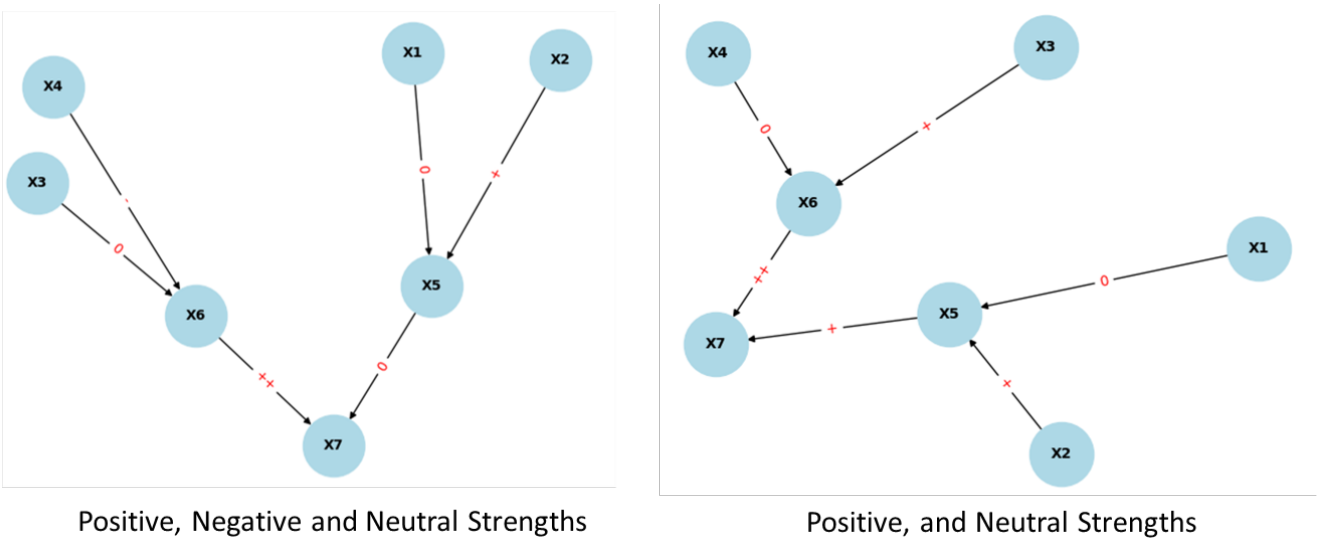

Figure C.6: Test Case 6- Proper handling of positive and negative weights

Bibliography

[1] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. Deep learning with differential privacy. In *Proceedings of the Association for Computing Machinery - Special Interest Group on Security, Audit and Control Conference on Computer and Communications Security (ACM SIGSAC)*, pages 308–318, 2016. DOI: ~10.1145/2976749.2978318.

[2] D. B. Abeywickrama, A. Bennaceur, G. Chance, Y. Demiris, A. Kordoni, M. Levine, L. Moffat, L. Moreau, M. R. Mousavi, B. Nuseibeh, et al. On specifying for trustworthiness. *Communications of the Association for Computing Machinery (ACM)*, 67(1):98–109, 2023. DOI:~10.1145/3624699.

[3] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar. Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236, 2011. DOI:~10.1093/restud/rdr004.

[4] E. Adelowo. College students perception of privacy in smart wearable medical devices. *International Review of Research in Open and Distributed Learning*, 118(1):1–8, 2021.

[5] O. Adewumi, K. Djouani, and A. Kurien. RSSI based indoor and outdoor distance estimation for localization in WSN. In *IEEE International Conference on Industrial Technology (ICIT)*, pages 1534–1539, 2013. DOI:~10.1109/ICIT.2013.6505900.

[6] A. Adjekum, A. Blasimme, and E. Vayena. Elements of trust in digital health systems: scoping review. *Journal of Medical Internet Research*, 20(12):e11254–e11270, 2018. DOI: ~10.2196/11254.

[7] H. Al-Hamadi and R. Chen. Trust-based decision making for health iot systems. *IEEE IoT Journal*, 4(5):1408–1419, 2017. DOI:~10.1109/JIOT.2017.2736446.

[8] S. Al-Sarawi, M. Anbar, K. Alieyan, and M. Alzubaidi. Internet of things (iot) communication protocols. In *IEEE 8th International Conference on Information Technology (ICIT)*, pages 685–690, 2017. DOI:~10.1109/ICITECH.2017.8079928.

[9] L. Alazzawi and A. Elkateeb. Performance evaluation of the wsn routing protocols scalability. *Journal of Computer Networks and Communications*, 08(01):1–8, 2008. DOI:~doi:10.1155/2008/481046.

[10] S. AlMotiri, M. Khan, and M. Alghamdi. Mobile health (m-health) system in the context of iot. In *IEEE 4th International Conference on Future Internet of things and Cloud Workshops (FiCloudW)*, pages 39–42, 2016. DOI:~10.1109/W-FiCloud.2016.24.

[11] Apple.com. Apple watch 7 series, 2021. URL https://support.apple.com/en-us/111909.

[12] K. Austen. The trouble with wearables. *Nature*, 525(7567):22, 2015.

[13] K. Avrachenkov, B. Ribeiro, and J. K. Sreedharan. Bayesian inference of online social network statistics via lightweight random walk crawls. *arXiv preprint arXiv:1510.05407*, 2015. DOI:~10.48550/arXiv.1510.05407.

[14] A. Baldwin-Medsker. Access to care: Using ehealth to limit location-based barriers for patients with cancer. *Clinical Journal of Oncology Nursing*, 24(3):16–23, 2020. DOI:~10.1188/20.CJON.S1.16-23.

[15] F. Bao, R. Chen, and J. Guo. Scalable, adaptive and survivable trust management for community of interest based internet of things systems. In *IEEE 11th International Symposium on Autonomous Decentralized Systems (ISADS)*, pages 1–7, 2013. DOI:~10.1109/ISADS.2013.6513398.

[16] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013. DOI:~10.1109/TPAMI.2013.50.

[17] R. S. Bhatia, C. Chu, A. Pang, M. Tadrous, V. Stamenova, and P. Cram. Virtual care use before and during the covid-19 pandemic: a repeated cross-sectional study. *Canadian Medical Association (CMA) Journal*, 9(1):e107–e114, 2021.

[18] F. Boenisch, V. Battis, N. Buchmann, and M. Poikela. "i never thought about securing my machine learning systems": A study of security and privacy awareness of machine learning practitioners. In *Proceedings of Mensch Und Computer*, pages 520–546. 2021. DOI:~10.1145/3473856.3473869.

[19] K. Bouabida, K. Malas, A. Talbot, M.- Desrosiers, F. Lavoie, B. Lebouché, M. Taguemout, E. Rafie, D. Lessard, and M.-P. Pomey. Remote patient monitoring program for covid-19 patients following hospital discharge: a cross-sectional study. *Frontiers in Digital Health*, 3 (1):721044–721057, 2021. DOI:~10.3389/fdgth.2021.721044.

[20] O. Brännström, B. Elström, and G. Thompson. Functional products create new demands on product development organizations. *Design Management: Process and Information Issues*, 28(1):305–313, 2001.

[21] B. Bruegge and A. H. Dutoit. *Object Oriented Software Engineering: Using UML Patterns and Java*. Prentice Hall, 2010.

[22] S. Canali, V. Schiaffonati, and A. Aliverti. Challenges and recommendations for wearable devices in digital health: Data quality, interoperability, health equity, fairness. *Public Library of Science (PLOS) Digital Health*, 1(10):e0000104–e0000120, 2022. DOI:~10.1371/journal.pdig.0000104.

[23] Y. Cao, R. C. Chen, and A. J. Katz. Why is a small sample size not enough? *The Oncologist*, 1(1):oyae162, 2024. DOI:~10.1093/oncolo/oyae162.

[24] A. Cavoukian, A. Mihailidis, and J. Boger. Sensors and in-home collection of health data: A privacy by design approach. *Information and Privacy Commissioner of Ontario, Technical Report*, pages 1–22, 2010.

[25] C. C. Chang et al. Exploring the usage intentions of wearable medical devices: a demonstration study. *Interactive Journal of Medical Research*, 9(3):e19776–e19788, 2020. DOI: ~10.2196/19776.

[26] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(1):321–357, 2002. DOI:~10.1613/jair.953.

[27] M. Chen, Y. Qian, J. Chen, K. Hwang, S. Mao, and L. Hu. Privacy protection and intrusion avoidance for cloudlet-based medical data sharing. *IEEE Transactions on Cloud Computing*, 1(1):1274–1283, 2016. DOI:~10.1109/TCC.2016.2617382.

[28] R. Chen, F. Bao, M. Chang, and J.-H. Cho. Dynamic trust management for delay tolerant networks and its application to secure routing. *IEEE Transactions on Parallel and Distributed Systems*, 25(5):1200–1210, 2013. DOI:~10.1109/TPDS.2013.116.

[29] R. Chen, F. Bao, and J. Guo. Trust-based service management for social internet of things systems. *IEEE Transactions on Dependable and Secure Computing*, 13(6):684–696, 2015. DOI:~10.1109/TDSC.2015.2420552.

[30] A. Chib, C. Li, and S. Lin. Health wearable tools and health promotion. In *Oxford Research Encyclopedia of Global Public Health*. 2023.

[31] T. H. Chu, C. M. Chao, H. H. Liu, and D. F. Chen. Developing an extended theory of utaut2 model to explore factors influencing taiwanese consumer adoption of intelligent elevators. *SAGE Open*, 12(4):1–16, 2022. DOI:~10.1177/21582440221142209.

[32] L. Chung and J. C. S. do Prado Leite. On non-functional requirements in software engineering. *Conceptual Modeling: Foundations and Applications: Essays in Honor of John Mylopoulos*, pages 363–379, 2009.

[33] L. Cilliers. Wearable devices in healthcare: Privacy and information security issues. *Health Information Management Journal*, 49(23):150–156, 2020. DOI:~10.1177/1833358319851684.

[34] C. A. da Costa, C. F. Pasluosta, B. Eskofier, D. B. da Silva, and R. da Rosa Righi. Internet of health things: Toward intelligent vital signs monitoring in hospital wards. *Artificial Intelligence in Medicine*, 89(1):61–69, 2018. DOI:~10.1016/j.artmed.2018.05.005.

[35] R. Dai, T. Kannampallil, S. Kim, V. Thornton, L. Bierut, and C. Lu. Detecting mental disorders with wearables: A large cohort study. In *ACM/IEEE Proceedings of the 8th Conference on Internet of Things Design and Implementation*, pages 39–51, 2023.

[36] J. Davies and C. Fortuna. *The Internet of Things*. Wiley Online Library, 2020.

[37] Deloitte Team. How the pandemic has stress-tested the crowded digital home, 2021. URL https://deloitte.wsj.com/cmo/how-the-pandemic-has-stress-tested-the-digital-home-01628880858.

[38] J. K. Devine et al. Technical, regulatory, economic, and trust issues preventing successful integration of sensors into the mainstream consumer wearables market. *Sensors*, 22(7):1–7, 2022. DOI:~10.3390/s22072731.

[39] Duracell.com. Duracell battery datsheet, 2024. URL https://docs.rs-online.com/585c/0900766b81249809.pdf.

[40] D. Flouri, D. Owen, R. Aughwane, N. Mufti, M. Sokolska, D. Atkinson, G. Kendall, A. Bainbridge, T. Vercauteren, A. L. David, et al. Improved placental parameter estimation using data-driven bayesian modelling. In *22nd International Conference Medical Image Computing and Computer Assisted Intervention–MICCAI 2019, Part III*, pages 609–616. Springer, 2019.

[41] M. Fusca, F. Negrini, P. Perego, L. Magoni, F. Molteni, and G. Andreoni. Validation of a wearable imu system for gait analysis: Protocol and application to a new system. *Applied Sciences*, 8(7):1167–1182, 2018. DOI:~10.3390/app8071167.

[42] D. Gambetta et al. Can we trust trust. *Trust: Making and Breaking Cooperative Relations*, 13(1):213–237, 2000.

[43] C. Ghezzi, M. Jazayeri, and D. Mandrioli. *Fundamentals of Software Engineering*. Prentice-Hall, Inc., 1991.

[44] V. Gligor and J. Wing. Towards a theory of trust in networks of humans and computers. In *International Workshop on Security Protocols*, pages 223–242, 2011.

[45] K. Goel, R. Sindhgatta, S. Kalra, R. Goel, and P. Mutreja. The effect of machine learning explanations on user trust for automated diagnosis of covid-19. *Computers in Biology and Medicine*, 146(1):105587–105598, 2022. DOI:~10.1016/j.compbiomed.2022.105587.

[46] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Proceedings of the Advances in Neural Information Processing Systems*, pages 1–10, 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

[47] I. Gorton. *Essential Software Architecture*. Springer Science & Business Media, 2006.

[48] K. Govindan and P. Mohapatra. Trust computations and trust dynamics in mobile adhoc networks: A survey. *IEEE Communications Surveys & Tutorials*, 14(2):279–298, 2011. DOI:~10.1109/SURV.2011.042711.00083.

[49] T. Grandison and M. Sloman. A survey of trust in internet applications. *IEEE Communications Surveys Tutorials*, 3(4):2–16, 2000. DOI:~10.1109/COMST.2000.5340804.

[50] K. Grosse, L. Bieringer, T. R. Besold, B. Biggio, and K. Krombholz. "why do so?"-a practical perspective on machine learning security. In *International Conference of Machine Learning: New Frontiers of Adversarial Machine Learning*, pages 1–10, 2022.

[51] Z. Gu and J. Wei. Empirical study on initial trust of wearable devices based on product characteristics. *Journal of Computer Information Systems*, 61(6):520–528, 2021. DOI:~10.1080/08874417.2020.1779150.

[52] S. Gulati, S. Sousa, and D. Lamas. Modelling trust: An empirical assessment. In *Proceedings of Human-Computer Interaction (INTERACT)*, pages 40–61, 2017. DOI:~10.1007/978-3-319-68059-0_3.

[53] G. Han, J. Jiang, L. Shu, J. Niu, and H. C. Chao. Management and applications of trust in wireless sensor networks: A survey. *Journal of Computer and System Sciences*, 80(3): 602–617, 2014. DOI:~10.1016/j.jcss.2013.06.014.

[54] Health Canada. Pre-market requirements for medical device cybersecurity. *Health Canada Guidance Document*, 2019. URL https://www.canada.ca/content/dam/hc-sc/documents/services/drugs-health-products/medical-devices/application-information/guidance-documents/cybersecurity-guidance.pdf.

[55] Health Commissioner Ontario. Personal Health Information Protection Act. *Freedom of Information and Protection of Privacy Act*, 2022. URL https://www.ontario.ca/laws/statute/04p03.

[56] Health Ontario. Guidelines for remote care management and surgical transitions for ohts. *Funding Guidelines*, 2022. URL https://www.ontariohealth.ca/sites/ontariohealth/files/2021-08/FundingCriteria-RemoteCareManagement.pdf.

[57] D. Heckerman. *A Tutorial on Learning with Bayesian Networks*. Springer Netherlands, 1998. DOI:~10.1007/978-94-011-5014-9_11.

[58] Hexoskin Developers. Hexoskin api, 2018. URL https://api.hexoskin.com/docs/resource/datatype/.

[59] Hexoskin Manufacturing. Hexoskin 3.0 user manual, 2018. URL https://hexoskin.com/pages/astroskin-vital-signs-monitoring-platform-for-advanced-research.

[60] M. Holden, M. Pereyra, and K. C. Zygalakis. Bayesian imaging with data-driven priors encoded by neural networks. *Society for Industrial and Applied Mathematics (SIAM) Journal on Imaging Sciences*, 15(2):892–924, 2022. DOI:~10.1137/21M1406313.

[61] J. Horkoff and E. Yu. Interactive analysis of agent-goal models in enterprise modeling. *International Journal of Information System Modeling and Design (IJISMD)*, 1(4):1–23, 2010. DOI:~10.4018/jismd.2010100101.

[62] E. S. Izmailova, J. A. Wagner, and E. D. Perakslis. Wearable devices in clinical trials: hype and hypothesis. *Clinical Pharmacology & Therapeutics*, 104(1):42–52, 2018. DOI: ~10.1002/cpt.966.

[63] F. T. Jaigirdar, C. Rudolph, and C. Bain. Can i trust the data i see? a physician's concern on medical data in iot health architectures. In *Proceedings of the Australasian Computer Science Week Multiconference*, pages 1–10, 2019. DOI:~10.1145/3290688.3290731.

[64] W. S. Jang and W. Healy. Wireless sensor network performance metrics for building applications. *Energy and Buildings*, 42(6):862–868, 2010. DOI:~10.1016/j.autcon.2008.02.001.

[65] U. Jayasinghe, A. Otebolaku, T.-W. Um, and G. M. Lee. Data centric trust evaluation and prediction framework for iot. In *Proceedings of IEEE International Telecommunication Union (ITU) Kaleidoscope: Challenges for a Data-Driven Society*, pages 1–7, 2017. DOI: ~10.23919/ITU-WT.2017.8246999.

[66] B. Karakostas. Event prediction in an iot environment using naïve bayesian models. *Procedia Computer Science*, 83(1):11–17, 2016. DOI:~10.1016/j.procs.2016.04.093.

[67] Y. Khan, A. E. Ostfeld, C. M. Lochner, A. Pierre, and A. C. Arias. Monitoring of vital signs with flexible and wearable medical devices. *Advanced Materials*, 28(22):4373–4395, 2016. DOI:~https://doi.org/10.1002/adma.201504366.

[68] J. Kim. Energy-efficient dynamic packet downloading for medical iot platforms. *IEEE Transactions on Industrial Informatics*, 11(6):1653–1659, 2015. DOI:~10.1109/TII.2015.2434773.

[69] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. DOI:~10.48550/arXiv.1312.6114.

[70] B. A. Kitchenham, S. L. Pfleeger, L. M. Pickard, P. W. Jones, D. C. Hoaglin, K. El Emam, and J. Rosenberg. Preliminary guidelines for empirical research in software engineering. *IEEE Transactions on Software Engineering*, 28(8):721–734, 2002. DOI:~10.1109/TSE.2002.1027796.

[71] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT press, 2009.

[72] L. Kong, M. Xia, X.-Y. Liu, M.-Y. Wu, and X. Liu. Data loss and reconstruction in sensor networks. In *IEEE Proceedings of International Conference on Computer Communications (INFOCOM)*, pages 1654–1662, 2013. DOI:~10.1109/INFCOM.2013.6566962.

[73] M. Kyytsönen, T. Vehko, H. Anttila, and J. Ikonen. Factors associated with use of wearable technology to support activity, well-being, or a healthy lifestyle in the adult population and among older adults. *Public Library of Science (PLOS) Digital Health*, 2(5):e000024–e000055, 2023. DOI:~10.1371/journal.pdig.0000245.

[74] A. Lagerqvist and T. Lakshminarayana. Iot latency and power consumption: Measuring the performance impact of mqtt and coap, 2018. URL https://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-39392.

[75] J. Lazar, J. H. Feng, and H. Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[76] G. Lee and N. Truong. A reputation and knowledge based trust service platform for trustworthy social internet of things. *Innovations in Clouds, Internet and Networks*, pages 1–10, 2016. URL https://researchonline.ljmu.ac.uk/id/eprint/2599/.

[77] S. M. Lee and D. Lee. Healthcare wearable devices: an analysis of key factors for continuous use intention. *Service Business*, 14(4):503–531, 2020. DOI:~10.1007/s11628-020-00428-3.

[78] E. Letier and A. Van Lamsweerde. Reasoning about partial goal satisfaction for requirements and design engineering. In *Proceedings of the 12th ACM: International Symposium on Foundations of Software Engineering (SIGSOFT)*, pages 53–62, 2004. DOI:~10.1145/1029894.1029905.

[79] J. R. Lewis. Psychometric evaluation of the post-study system usability questionnaire: The pssuq. In *Proceedings of the Human Factors Society Annual Meeting*, pages 1259–1260, 1992. DOI:~10.1177/154193129203601617.

[80] J. Y. W. Liu, G. Sorwar, M. S. Rahman, and M. R. Hoque. The role of trust and habit in the adoption of mhealth by older adults in hong kong: a healthcare technology service acceptance (htsa) model. *BioMed Central Geriatrics*, 23(1):73–80, 2023. DOI:~10.1186/s12877-023-03779-4.

[81] S. M. Lynch. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Springer, 2007.

[82] S. Maitra and K. Yelamarthi. Rapidly deployable iot architecture with data security: Implementation and experimental evaluation. *Sensors*, 19(11):2484–2493, 2019. DOI:~10.3390/s19112484.

[83] N. K. Malhotra, S. S. Kim, and J. Agarwal. Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model. *Information Systems Research*, 15(4):336–355, 2004. DOI:~10.1287/isre.1040.0032.

[84] B. Martin, D. C. Tarraf, T. C. Whitmore, J. DeWeese, C. Kenney, J. Schmid, and P. DeLuca. Advancing autonomous systems. *Rand Corporation*, 1(1):9–11, 2019. DOI:~10.7249/RR2751.

[85] D. Maruyama, P. Bekemeyer, S. Görtz, S. Coggon, and S. Sharma. Data-driven bayesian inference of turbulence model closure coefficients incorporating epistemic uncertainty. *Acta Mechanica Sinica*, 37(12):1812–1838, 2021. DOI:~10.1007/s10409-021-01152-5.

[86] Medical Economics Team. Building a successful rpm program, 2022. URL https://www.medicaleconomics.com/view/bootcamp-fall-2022-building-a-successful-rpm-program.

[87] H. Meng, X. An, and J. Xing. A data-driven bayesian network model integrating physical knowledge for prioritization of risk influencing factors. *Process Safety and Environmental Protection*, 160(1):434–449, 2022. DOI:~10.1016/j.psep.2022.02.010.

[88] S. Mierdel and K. Owen. Telehomecare reduces er use and hospitalizations at william osler health system. *Studies in Health Technology and Informatics*, 209(1):102–108, 2015. DOI:~10.3233/978-1-61499-505-0-102.

[89] A. J. Mills, R. T. Watson, L. Pitt, and J. Kietzmann. Wearing safe: Physical and informational security in the age of the wearable device. *Business Horizons*, 59(6):615–622, 2016. DOI:~10.1016/j.bushor.2016.08.003.

[90] M. T. Minen, A. Gopal, G. Sahyoun, E. Stieglitz, and J. Torous. The functionality, evidence, and privacy issues around smartphone apps for the top neuropsychiatric conditions. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 33(1):72–79, 2021. DOI:~10.1176/appi.neuropsych.19120353.

[91] D. K. Ming, S. Sangkaew, H. Q. Chanh, P. T. Nhat, S. Yacoub, P. Georgiou, and A. H. Holmes. Continuous physiological monitoring using wearable technology to inform individual management of infectious diseases, public health and outbreak responses. *International Journal of Infectious Diseases*, 96(1):648–654, 2020. DOI:~10.1016/j.ijid.2020.05.086.

[92] Ministry of Health. Medical Devices Regulations. *Health Canada Document*, 2022. URL https://laws-lois.justice.gc.ca/eng/regulations/sor-98-282/.

[93] V. Mohammadi, A. M. Rahmani, A. M. Darwesh, and A. Sahafi. Trust-based recommendation systems in internet of things: a systematic literature review. *Human-centric Computing and Information Sciences*, 9(1):21–82, 2019. DOI:~10.1186/s13673-019-0183-8.

[94] E. N. Montague, B. M. Kleiner, and W. W. Winchester III. Empirically understanding trust in medical technology. *International Journal of Industrial Ergonomics*, 39(4):628–634, 2009. DOI:~10.1016/j.ergon.2009.01.004.

[95] L. Mui, M. Mohtashemi, and A. Halberstadt. A computational model of trust and reputation. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, pages 2431–2439. IEEE, 2002. DOI:~10.1109/HICSS.2002.994181.

[96] A. H. Murphy and R. L. Winkler. Diagnostic verification of probability forecasts. *International Journal of Forecasting*, 7(4):435–455, 1992. DOI:~10.1016/0169-2070(92)90028-8.

[97] A. Ng and M. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Proceedings of Advances in Neural Information Processing Systems*, pages 1–8, 2001. URL https://proceedings.neurips.cc/paper_files/paper/2001/file/7b7a53e239400a13bd6be6c91c4f6c4e-Paper.pdf.

[98] D. M. Nhat, R. Venkatesan, and F. Khan. Data-driven bayesian network model for early kick detection in industrial drilling process. *Process Safety and Environmental Protection*, 138(1):130–138, 2020. DOI:~10.1016/j.psep.2020.03.017.

[99] J. Nielsen. *Usability engineering*. Morgan Kaufmann, 1994.

[100] N. Niknejad, W. B. Ismail, A. Mardani, H. Liao, and I. Ghani. A comprehensive overview of smart wearables: The state of the art literature, recent advances, and future challenges. *Engineering Applications of Artificial Intelligence*, 90(1):103529–103602, 2020. DOI:~10.1016/j.engappai.2020.103529.

[101] M. Nitti, R. Girau, and L. Atzori. Trustworthiness management in the social internet of things. *IEEE Transactions on Knowledge and Data Engineering*, 26(5):1253–1266, 2013. DOI:~10.1109/TKDE.2013.105.

[102] S. B. Olivencia, K. Zahed, F. Sasangohar, R. Davir, and A. Vedlitz. Integration of remote patient monitoring systems into physicians work in underserved communities: Survey of

healthcare provider perspectives. *arXiv preprint arXiv:2207.01489*, 2022. DOI:~10.48550/arXiv.2207.01489.

[103] D. Pal, S. Funilkul, and B. Papasratorn. Antecedents of trust and the continuance intention in iot-based smart products: The case of consumer wearables. *IEEE Access*, 7(1):184160–184171, 2019. DOI:~10.1109/ACCESS.2019.2960467.

[104] J. M. Peake, G. Kerr, and J. P. Sullivan. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. *Frontiers in Physiology*, 9(1):329783–329799, 2018. DOI:~10.3389/fphys.2018.00743.

[105] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[106] C. Pfleeger, S. L. Pfleeger, and J. Margulies. *Security in Computing*. ProQuest Safari Tech Books Online, 2017.

[107] M. Pobiruchin, J. Suleder, R. Zowalla, et al. Accuracy and adoption of wearable technology used by active citizens: A marathon event field study. *Journal of Medical Internet Research mHealth and uHealth*, 5(2):1–20, 2017. DOI:~10.2196/mhealth.6395.

[108] Privacy Commissioner of Canada. Wearable computing - challenges and opportunities for privacy protection. *Government of Canada Document*, 2014. URL https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2014/wc_201401/.

[109] Privacy Commissioner of Canada. Privacy guidance for manufacturers of internet of things devices. *Canadian Standards Association*, 2020. URL https://priv.gc.ca/en/privacy-topics/technology/gd_iot_man.

[110] G. L. Puplampu, A. P. Fenny, and G. Mensah. *Consumers and consumer behaviour*. Productivity Press, 2019. DOI:~10.4324/9780429400858.

[111] W. Raghupathi and V. Raghupathi. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1):1–10, 2014. DOI:~10.1186/2047-2501-2-3.

[112] A. Ramsetty and C. Adams. Impact of the digital divide in the age of covid-19. *Journal of the American Medical Informatics Association*, 27(7):1147–1148, 2020.

[113] D. Roskams-Edris. The eye inside: Remote biosensing technologies in healthcare and the law. *Dalhousie Journal of Legal Studies*, 27(1):59–65, 2018.

[114] Y. S. Ryu and T. L. Smith-Jackson. Reliability and validity of the mobile phone usability questionnaire (mpuq). *Journal of Usability Studies*, 2(1):39–53, 2006.

[115] R. R. Sahoo, A. R. Sardar, M. Singh, S. Ray, and S. K. Sarkar. A bio inspired and trust based approach for clustering in wsn. *Natural Computing*, 15(3):423–434, 2016. DOI:~10.1007/s11047-015-9491-8.

[116] Y. B. Saied, A. Olivereau, D. Zeghlache, and M. Laurent. Trust management system design for the internet of things: A context-aware and multi-service approach. *Computers & Security*, 39(3):351–365, 2013. DOI:~10.1016/j.cose.2013.09.001.

[117] J. H. Saltzer and M. D. Schroeder. The protection of information in computer systems. *Proceedings of the IEEE*, 63(9):1278–1308, 1975. DOI:~10.1109/PROC.1975.9939.

[118] R. Samavi, M. P. Consens, and M. Chignell. Phr user privacy concerns and behaviours. *Procedia Computer Science*, 37(1):517–524, 2014. DOI:~10.1016/j.procs.2014.08.077.

[119] K. Samhale. The impact of trust in the internet of things for health on user engagement. *Digital Business*, 2(1):100021–100042, 2022. DOI:~10.1016/j.digbus.2022.100021.

[120] M. Sandberg, P. Boart, and T. Larsson. Functional product life-cycle simulation model for cost estimation in conceptual design of jet engine components. *Concurrent Engineering*, 13 (4):331–342, 2005. DOI:~10.1177/1063293X05060136.

[121] A. Sawand, S. Djahel, Z. Zhang, and F. Abdesselam. Multidisciplinary approaches to achieving efficient and trustworthy ehealth monitoring systems. In *Proceedings of the*

IEEE International Conference on Communications in China (ICCC), pages 187–192, 2014. DOI:~10.1109/ICCChina.2014.7008269.

[122] M. Schukat, D. McCaldin, K. Wang, G. Schreier, N. H. Lovell, M. Marschollek, and S. J. Redmond. Unintended consequences of wearable sensor use in healthcare. *Yearbook of Medical Informatics*, 1(1):73–83, 2016. DOI:~10.15265/IY-2016-025.

[123] M. Scutari. Learning bayesian networks with the bnlearn r package. *arXiv preprint arXiv:0908.3817*, 2009. DOI:~10.48550/arXiv.0908.3817.

[124] M. Shin. Secure remote health monitoring with unreliable mobile devices. *Journal of Biomedicine and Biotech.*, 12(1):1–6, 2012. DOI:~10.1155/2012/546021.

[125] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini. Security, privacy and trust in internet of things: The road ahead. *Computer Networks*, 76(1):146–164, 2015. DOI:~10.1016/j.comnet.2014.11.008.

[126] H. J. Smith, T. Dinev, and H. Xu. Information privacy research: an interdisciplinary review. *MIS Quarterly*, 1(1):989–1015, 2011. DOI:~10.2307/41409970.

[127] B. Stanberry. Legal and ethical aspects of telemedicine. *Journal of Telemedicine and Telecare*, 12(4):166–175, 2006. DOI:~10.1258/135763306777488825.

[128] L. E. Sucar. Probabilistic graphical models. *Advances in Computer Vision and Pattern Recognition. London: Springer London.*, 10(978):1–353, 2015. URL https://link.springer.com/content/pdf/10.1007/978-3-030-61943-5.pdf.

[129] W. Sun, Z. Cai, Y. Li, F. Liu, S. Fang, and G. Wang. Security and privacy in the medical internet of things: a review. *Security and Communication Networks*, 1(1):1–10, 2018. DOI:~10.1155/2018/5978636.

[130] M. S. Talukder, G. Sorwar, Y. Bao, J. U. Ahmed, and M. A. S. Palash. Predicting antecedents of wearable healthcare technology acceptance by elderly: A combined sem-neural network

approach. *Technological Forecasting and Social Change*, 150(1):119793–119799, 2020. DOI:~10.1016/j.techfore.2019.119793.

[131] M. L. Taylor, E. E. Thomas, C. L. Snoswell, A. C. Smith, and L. J. Caffery. Does remote patient monitoring reduce acute care use? a systematic review. *British Medical Journal (BMJ) Open*, 11(3):e040232–e040242, 2021. DOI:~10.1136/bmjopen-2020-040232.

[132] S. Thapa, A. Bello, A. Maurushat, and F. Farid. Security risks and user perception towards adopting wearable internet of medical things. *International Journal of Environmental Research and Public Health*, 20(8):5519–5540, 2023. DOI:~10.3390/ijerph20085519.

[133] The European Parliament, and Council of European Union. Medical device regulation. *Official Journal of the European Union*, 2017. URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32020R0561.

[134] M. Thomas, R. Samavi, and T. E. Doyle. Trust quantification for autonomous medical advisory systems. In *18th International Conference on Privacy, Security and Trust (PST)*, pages 1–7, 2021. DOI:~10.1109/PST52912.2021.9647818.

[135] M. Thomas, O. Boursalie, R. Samavi, and T. E. Doyle. Bayesian-based parameter estimation to quantify trust in medical devices. In *International Workshop on Health Intelligence*, pages 95–108, 2023. DOI:~10.1007/978-3-031-36938-4_8.

[136] M. Thomas, O. Boursalie, R. Samavi, and T. E. Doyle. Data-driven approach to quantify trust in medical devices using bayesian networks. *Experimental Biology and Medicine*, 248 (24):2578–2592, 2023. DOI:~10.1177/15353702231215893.

[137] K. W. Tong. Telehealth as a double-edge sword: Lessons from court cases to gain understanding of medico-legal risks. *Medicine and Law*, 38(3):85–100, 2019.

[138] United States Department of Health. Health fda draft for cybersecurity in medical devices. *Food and Drug Administration*, 2022. URL https://www.skadden.com/-/media/files/publications/2022/04/privacy-cybersecurity-update/fn-3-a-cybersecurity-in-medical-devices.pdf.

[139] United States Department of Health and Human Service. Digital health technologies for remote data acquisition in clinical investigations guidance for industry, investigators, and other stakeholders. *Food and Drug Administration*, 2023. URL https://www.fda.gov/media/155022/download.

[140] T. C. Veinot, H. Mitchell, and J. S. Ancker. Good intentions are not enough: how informatics interventions can worsen inequality. *Journal of the American Medical Informatics Association*, 25(8):1080–1088, 2018. DOI:~10.1093/jamia/ocy052.

[141] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis. User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 1(1):425–478, 2003. DOI:~10.2307/30036540.

[142] Y. Wang. Quantifying trust perception to enable design for connectivity in cyber-physical-social systems. In *Emotional Engineering*, pages 85–113. 2020. DOI:~10.1007/978-3-030-38360-2_6.

[143] Y. Wang and J. Vassileva. Bayesian network-based trust model. In *Proceedings OF IEEE International Conference on Web Intelligence (WI 2003)*, pages 372–378, 2003. DOI:~10.1109/WI.2003.1241218.

[144] Y. Wang, L. Lu, R. Zhang, Y. Ma, S. Zhao, and C. Liang. The willingness to continue using wearable devices among the elderly: Sem and fsqca analysis. *BMC Medical Informatics and Decision Making*, 23(1):218–230, 2023. DOI:~10.1186/s12911-023-02336-8.

[145] R. S. Weinstein, A. M. Lopez, B. A. Joseph, K. A. Erps, M. Holcomb, G. P. Barker, and E. A. Krupinski. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. *The American Journal of Medicine*, 127(3):183–187, 2014. DOI:~10.1016/j.amjmed.2013.09.032.

[146] D. Wen, X. Zhang, X. Liu, and J. Lei. Evaluating the consistency of current mainstream wearable devices in health monitoring: a comparison under free-living conditions. *Journal of Medical Internet research*, 19(3):e68–e85, 2017. DOI:~10.2196/jmir.6874.

[147] E. E. Wickel. Reporting the reliability of accelerometer data with and without missing values. *Public Library of Science (PLOS) One*, 9(12):e114402–e114433, 2014. DOI:~10.1371/journal.pone.0114402.

[148] W. Wilkowska and M. Ziefle. Determinants of trust in acceptance of medical assistive technologies. In *Information and Communication Technologies for Ageing Well and e-Health*, pages 45–65, 2019. DOI:~10.1007/978-3-030-15736-4_3.

[149] World Health Organization. The who special initiative for mental health (2019-2023): universal health coverage for mental health. Technical report, 2019. URL https://iris.who.int/handle/10665/310981.

[150] J. Xie, D. Wen, L. Liang, Y. Jia, L. Gao, J. Lei, et al. Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. *JMIR mHealth and uHealth*, 6(4):e9754–e9784, 2018. DOI:~10.2196/mhealth.9754.

[151] H. Xu, T. Dinev, H. J. Smith, and P. Hart. Examining the formation of individual's privacy concerns: Toward an integrative view. In *Proceedings of International Conference on Information Systems*, pages 1–8, 2008. URL https://aisel.aisnet.org/icis2008/6.

[152] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni. Modeling tabular data using conditional gan. In *Proceedings of Advances in Neural Information Processing Systems*, pages 1–8, 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/254ed7d2de3b23ab10936522dd547b78-Paper.pdf.

[153] Z. Yan and C. Prehofer. Autonomic trust management for a component-based software system. *IEEE Transactions on Dependable and Secure Computing*, 8(6):810–823, 2010. DOI:~10.1109/TDSC.2010.47.

[154] Z. Yan, P. Zhang, and A. V. Vasilakos. A survey on trust management for internet of things. *Journal of Network and Computer Applications*, 42(1):120–134, 2014. DOI:~10.1016/j.jnca.2014.01.014.

[155] B. Yu, S. Kallurkar, and R. Flo. A demspter-shafer approach to provenance-aware trust assessment. In *Proceedings of International Symposium on Collaborative Technologies and Systems*, pages 383–390, 2008. DOI:~10.1109/CTS.2008.4543955.

[156] E. Yu, P. Giorgini, N. Maiden, and J. Mylopoulos. *Social Modeling for Requirements Engineering: An Introduction*. MIT Press, 2010.

[157] Zephyr. Zephyr bioharness3 user manual, 2016. URL https://www.zephyranywhere.com/media/download/bioharness-log-data-descriptions-07-apr-2016.pdf.

[158] M. Zhang, A. Raghunathan, and N. K. Jha. Trustworthiness of medical devices and body area networks. *Proceedings of the IEEE*, 102(8):1174–1188, 2014. DOI:~10.1109/JPROC.2014.2322103.

[159] A. Zinzuwadia and J. P. Singh. Wearable devices—addressing bias and inequity. *The Lancet Digital Health*, 4(12):e856–e857, 2022. DOI:~10.1016/S2589-7500(22)00194-7.

[160] C. Zouridaki, B. L. Mark, M. Hejmo, and R. K. Thomas. A quantitative trust establishment framework for reliable data packet delivery in manets. In *Proceedings of the 3rd ACM Workshop on Security of Adhoc and Sensor Networks*, pages 1–10, 2005. DOI:~10.1145/1102219.1102222.