Approaches to Clinical Pathway Protection by means of Artificial Intelligence and Process Mining

Politecnico di Bari

Department of Electrical and Information Engineering
Electrical and Information Engineering Ph.D. Program
SSD: ING-INF/05 - Information Processing Systems

Final Dissertation

Approaches to Clinical Pathway Protection by means of Artificial Intelligence and Process Mining

by

Domenico Lofù

Supervisors
Prof. Tommaso Di Noia
Eng. Felice Vitulano
Co-Supervisor
Prof. Carmelo Ardito

Coordinator of the Ph.D. Program
Prof. Mario Carpentieri

Course n° 35, 01/11/2019 - 31/10/2022

ABSTRACT

The Covid-19 era was characterized by an increase in patients needing constant medical care because of the virus and post-covid symptoms. Patients were increasingly monitored 24 hours a day. Thus, remote monitoring of patients has became an established method of monitoring their vital parameters and keeping track of their activities. Using remote monitoring, healthcare workers can manage patients' clinical activities without physically intervention. Using a dashboard, healthcare professionals can visualize the patient's clinical course and intervene when necessary. Globally and in territorial healthcare, clinical data from patients are an increasingly valuable asset. Collection, analysis, and use of accurate patient information can help consolidate patient information. It is possible to analyze symptoms for a more accurate diagnosis, highlight demographics and geography to aid in disease diagnosis, define scenarios, and plan for the necessary resources. Remote monitoring is once again defined by data. The amount of data collected is closely related to machine learning based analysis of data streams. Having access to real-time data enables convenient dosing of drugs and improves care quality. In this context, machine learning is a paradigm to perform patient monitoring remotely without requiring the physical presence of the healthcare provider. This thesis summarises our efforts to exploit the opportunities offered by machine learning in remote monitoring systems. The model we propose identifies possible data transmission tampering. Our model can also monitor whether patients properly follow the healing path suggested by the doctor. This work is motivated by our research on remote monitoring systems and supervised learning algorithms for clinical data management. It is based on an extensive literature review on remote monitoring systems and a survey of artificial intelligence paradigms and techniques capable of managing clinical data to provide a high level of data quality, which is essential for informed clinical care.

PUBLICATIONS

Some ideas and figures have appeared in previous publications. A complete list of my publications is available in the following (the symbol (*) denotes papers where I contributed as main author).

[1] (*) Carmelo Ardito, Tommaso Di Noia, Eugenio Di Sciascio, Domenico Lofú, Giulio Mallardi, Claudio Pomo, and Felice Vitulano. "Towards a trustworthy patient home-care thanks to an edge-node infrastructure." In: *International Conference on Human-Centred Software Engineering*. Springer International Publishing. 2020, pp. 181–189.

[2] (*) Carmelo Ardito, Tommaso Di Noia, Eugenio Di Sciascio, Domenico Lofù, Andrea Pazienza, and Felice Vitulano. "An Artificial Intelligence Cyberattack Detection System to Improve Threat Reaction in e-Health." In: *ITASEC - Italian Conference on Cybersecurity 2021*. CEUR. 2021.

[3] (*) Carmelo Ardito, Tommaso Di Noia, Eugenio Di Sciascio, Domenico Lofù, Andrea Pazienza, and Felice Vitulano. "User Feedback to Improve the Performance of a Cyberattack Detection Artificial Intelligence System in the e-Health Domain." In: *IFIP Conference on Human-Computer Interaction*. Springer International Publishing. 2021, pp. 295–299.

[4] Carmelo Ardito, Francesco Bellifemine, Tommaso Di Noia, Domenico Lofú, and Giulio Mallardi. "A proposal of case-based approach to clinical pathway modeling support." In: *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*. IEEE. 2020, pp. 1–6.

[5] Carmelo Ardito, Ilaria Bortone, Tommaso Colafiglio, Tommaso Di Noia, Eugenio Di Sciascio, Domenico Lofù, Fedelucio Narducci, Rodolfo Sardone, and Paolo Sorino. "Brain Computer Interface: Deep Learning Approach to Predict Human Emotion Recognition." In: *2022 International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE.

[6] Carmelo Ardito, Tommaso Di Noia, Corrado Fasciano, Domenico Lofù, Nicola Macchiarulo, Giulio Mallardi, Andrea Pazienza, and Felice Vitulano. "Management at the edge of situation awareness during patient telemonitoring." In: *International conference of the italian association for artificial intelligence*. Springer International Publishing. 2020, pp. 372–387.

[7] Carmelo Ardito, Tommaso Di Noia, Corrado Fasciano, Domenico Lofù, Nicola Macchiarulo, Giulio Mallardi, Andrea Pazienza, and Felice Vitulano. "Towards a Situation Awareness for eHealth in Ageing Society." In: *Italian Workshop on Artificial Intelligence for an Ageing Society (AIxAS), 2020. Co-located with AI*IA 2020, The International Conference of the Italian Association for Artificial Intelligence*. Vol. 2804. CEUR. 2020.

[8] Carmelo Ardito, Tommaso Di Noia, Corrado Fasciano, Domenico Lofù, Nicola Macchiarulo, Giulio Mallardi, Andrea Pazienza, and Felice Vitulano. "An Edge Ambient Assisted Living Process for Clinical Pathway." In: *Ambient Assisted Living*. Springer International Publishing, 2022, pp. 363–374.

[9] Carmelo Ardito, Tommaso Di Noia, Domenico Lofú, and Giulio Mallardi. "An adaptive architecture for Healthcare Situation Awareness." In: *6th Italian Conference on ICT for Smart Cities And Communities (I-CiTies 2020)*. 2020.

[10] Simona Aresta et al. "Combining biomechanical features and Machine Learning approaches to identify fencers' levels for training support." In: *SUBMITTED TO: Applied Science.* (2022).

[11] (*) Pietro Boccadoro, Vitanio Daniele, Pietro Di Gennaro, Domenico Lofù, and Pietro Tedeschi. "Water quality prediction on a Sigfox-compliant IoT device: The road ahead of WaterS." In: *Ad Hoc Networks* 126 (2022), p. 102749.

[12] Fabio Castellana, Ilaria Bortone, Tommaso Colafiglio, Tommaso Di Noia, Eugenio Di Sciascio, Domenico Lofù, Fedelucio Narducci, Rodolfo Sardone, and Paolo Sorino. "An Artificial Neural Network Model to Assess Nutritional Factors Associated with Frailty in the Aging Population from Southern Italy." In: *2022 International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE.

[13] Colabuono Consuelo et al. "Approach to sector-specific Cybersecurity Schemes: key elements and Security Problem Definition." In: *2022 International Conference on Multimedia Communications, Services and Security (MCSS)*.

[14] Tommaso Di Noia, Corrado Fasciano, Domenico Lofù, Giulio Mallardi, Graziano Pappadà, Corrado Tatulli, and Felice Vitulano. "INSTAMED: an Integrated Platform for the Advanced Automation of Diagnosis in Precision Medicine." In: *7th Italian Conference on ICT for Smart Cities And Communities (I-CiTies 2021)*. 2021.

[15] (*) Kristine Hovhannisyan, Piotr Bogacki, Consuelo Assunta Colabuono, Domenico Lofù, Maria Vittoria Marabello, and Brady Eugene Maxwell. “Towards a Healthcare Cybersecurity Certification Scheme.” In: *2021 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*. IEEE. 2021, pp. 1–9.

[16] Eufemia Lella, Andrea Pazienza, Domenico Lofù, Roberto Anglani, and Felice Vitulano. “An Ensemble Learning Approach Based on Diffusion Tensor Imaging Measures for Alzheimer’s Disease Classification.” In: *Electronics*. Vol. 10, 249. MDPI, 2021.

[17] (*) Domenico Lofù, Pietro Di Gennaro, Paolo Sorino, Tommaso Di Noia, and Eugenio Di Sciascio. “CPU-side comparison for Key Agreement between Tree Parity Machines and standard Cryptographic Primitives.” In: *TO BE PUBLISHED IN: The 12th IEEE International Conference on Depandable Systems, Services and Technologies (DESSERT)*. IEEE. 2022.

[18] (*) Domenico Lofù, Andrea Pazienza, Carmelo Ardito, Tommaso Di Noia, Eugenio Di Sciascio, and Felice Vitulano. “A Situation Awareness Computational Intelligent Model for Metabolic Syndrome Management.” In: *2022 Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*. IEEE. 2022, pp. 118–124.

[19] (*) Domenico Lofù, Paolo Sorino, Tommaso Colafiglio, Caterina Bonfiglio, Fedelucio Narducci, and Tommaso Di Noia. “MAFUS: a Framework to predict mortality risk in MAFLD subjects.” In: *SUBMITTED TO: Journal of Ambient Intelligence and Humanized Computing* (2022).

[20] Andrea Pazienza, Domenico Lofù, Giampaolo Flace, Marco Salzedo, Pietro Noviello, Eugenio Di Sciascio, and Felice Vitulano. “A Web Crowdsourcing Platform for Territorial Control in Smart Cities.” In: *International Conference on Web Engineering*. Springer International Publishing. 2022, pp. 375–382.

[21] Rossella Tatoli, Luisa Lampignano, Rossella Donghia, Fabio Castellana, Roberta Zupo, Ilaria Bortone, Sara De Nucci, Giuseppe Campanile, Domenico Lofù, Luigi Vimercati, et al. “Dietary Customs and Social Deprivation in an Aging Population From Southern Italy: A Machine Learning Approach.” In: *Frontiers in Nutrition* 9 (2022).

[22] Rossella Tatoli et al. “Dietary Patterns Associated with Diabetes in an Older Population from Southern Italy Using an Unsupervised Learning Approach.” In: *Sensors* 22.6 (2022).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Part I

PRELIMINARIES

That is, what can you expect from this thesis, what I have studied and what I have worked for, what you need to know before going on. Preliminaries is my "*Where are you and where do you want to go? Which are your shoes?*".

1

INTRODUCTION

Remote monitoring of the patient in-home care is an activity involving several domains, including the clinical and healthcare fields. During monitoring, wearable sensors transmit their values electronically. The data from the sensors are fed to artificial intelligence algorithms to perform a remote clinical assessment. Remote monitoring is thus a healthcare delivery model that uses technology to connect patients and caregivers/professionals outside the clinic, doctor's office, or hospital. The use of modern equipment, integrated with monitoring apps, positively impacts patient care and reduces organisational costs. To manage patients' clinical conditions and improve their course of treatment, remote monitoring of patients has become paramount. It eliminates the need to travel, and reduces hospital costs; it also facilitates communication between the healthcare provider and patient. Remote monitoring also facilitates the sharing of patient status information in real-time. Despite the rapid deployment and ease of use of remote monitoring, it is crucial to keep track of any bias that may occur during patient monitoring. Several factors could compromise patient monitoring: environmental, clinical, and technological. Incorrect patient monitoring could also lead to death. In this thesis, we propose two approaches to solve two problems that could compromise the monitoring phase.

The work at hand, whose investigation started in 2019 within a Ph.D. industrial program in the Exprivia S.p.A. Company [1], takes up the goals and the challenges of remote monitoring and explores its applications and implications in the Artificial Intelligence (AI) domain. In particular, we applied AI techniques to solve technological monitoring problems.

In detail, this thesis guides the reader towards three core contributions:

- Introduction and definition of Clinical Pathway (CP) and its representation within the context of remote ICT patient monitoring;
- Presentation of Security Problems (assumptions, organizational security policies, threats) based on five use-cases of Picture Archiving and Communication System (PACS), sub-category systems of Clinical Information Systems;

1 http://www.exprivia.com

- Presentation of two approaches that, by using AI and Process Mining, succeed respectively in reducing the risk of data compromise and ensuring that the patient's CP does not deviate from the physician's directions.

In the next section, we deeply analyze the research questions that guided our work towards its contributions and form the content of this dissertation.

1.1 RESEARCH CONTRIBUTIONS

The concepts that more often will occur in this dissertation are *remote monitoring system*, *security*, *explainability* and *security problem*.

Their connection is exceptionally close, since patient monitoring systems, as mentioned above, must transmit data while ensuring that it is not compromised and would be not used if they do not explain the data that may be compromised.

Despite its usefulness for patients and clinicians, data collection is exposed to threats and often is the source of safety issues for the CP.

Below, we present the research questions that guided this three-year investigation and led to the contributions that outline the central part of this thesis (See The Showcase, Part ii).

1.1.1 *Artificial Intelligence for sequential data analysis in health*

RESEARCH QUESTIONS A

Is it possible to design an ICT system in which the doctor remotely has complete control over the patient's clinical activity? How can the sequencing of clinical actions to be followed by a patient be defined? Are there AI approaches to identify data anomalies? What algorithmic approach can be used to analyse sequential data? How is it possible to explain the anomalous data detected?

Artificial Intelligence has opened the door to new techniques for sequentially analysing data and signals. However, in the early stages of our research, we focused on understanding how we could contribute to our Exprivia company's systems as an added value.

Chapter 2 presents our contribution to taking a step in Ambient Assisted Leaving (AAL) systems' literature that considers the need for sequential analysis of clinical data. We introduce several notions and methods that are useful in the following chapters. In particular, in this

chapter, we define how a CP is generated and how we can use AI to check whether the transmitted data is correct or not. We propose an algorithm for handling data that follows a well-defined chronological order and introduce process mining concepts. Ultimately, we propose an approach to explain how it is also essential, from a security point of view, to explain data using AI.

1.1.2 *Security Functional Requirements for the health domain*

RESEARCH QUESTIONS B
How can a health protection problem be defined? Is it possible to establish a certification scheme for healthcare products? How can security problems be identified to define a baseline for the protection profile? How are security requests identified? What are the safety needs in the health sector? How is it possible to assess the compliance of a healthcare product with safety requirements?

ENISA's pioneering work on the certification of cybersecurity products has also made great strides in healthcare. It has also paid great attention to candidate methods for the certification of health products.

Chapter 3 presents our contribution to defining cybersecurity certification schemes as candidate methods for sector cybersecurity product certification as part of the EU Cybersecurity Certification Framework being prepared by ENISA. Indeed, it is a very recent area of research within the EU landscape.

Our work was undertaken within the H2020 ECHO project[2] and is reported in detail in its deliverables and Papers [48, 80].

The work starts by identifying the sector-specific needs to be addressed for specific critical sectors. As described in the EU Cybersecurity Act, the mandatory Key Elements of a certification scheme have been customized. The sector-specific analysis allowed us to define a Security Problem Definition baseline to quickly draft a Protection Profile of an asset category of the considered sectors.

Security needs have also been identified using the sectoral risk assessment guidelines provided by ENISA for certification purposes.

An inter-sector risk scenario has also been developed to highlight the most critical security needs to mitigate cross-sector security failures.

Finally, Cyber Range technologies are leveraged for the Conformity Assessment activities of a Healthcare product prototype, for which the

2 www.echonetwork.eu

substantial assurance level certification has been simulated to validate our approach.

1.1.3 *Design of a monitoring system*

Research Questions C
Is it possible to define an architecture for remote patient ICT monitoring? What fundamental modules must there be in the architecture? Is it possible to consider a module that verifies that the transmitted data is safe in the architecture? Can the architecture be subject to security problems? What kind of tampering could occur? How does the caregiver interact with the infrastructure?

Although remote patient monitoring has long been considered one of the most promising solutions as it reduces hospitalization costs and still allows remote patient monitoring, it is not without safety concerns. Chapter 4 introduces an anomaly detection system that uses Long Short Term Memory (LSTM) autoencoders to analyze the data detected by the sensors. An application case based on monitoring a patient's ECG is presented. It also introduces a CP monitoring system that allows you to see how far the patient's path differs from that of the doctor. The method is proposed based on a patient suffering from metabolic syndrome. The resulting system shows a high degree of accuracy.

1.2 ORGANIZATION OF THE THESIS

The chapters of this thesis are as self-contained as possible and present the notions of specific problems, architectures, paradigms, data structures, and metrics related with their content. Moreover, each chapter independently surveys the state-of-the-art of the specific problem it deals with, showing the most interesting solutions and their limitations in the literature of that field.

However, in the next chapter, we propose a brief but comprehensive introduction to the most important recurring topics of this dissertation.

Chapter 2 introduces the algorithms used to solve the problems presented below. Chapter 3 defines the security issues and protection profile that a health security solution designer must follow. Next, in chapter 4 we apply the algorithms presented in chapter 2 to identify anomalous data of a transmission between the sensor and the server. We also use conformance-checking algorithms to verify that a patient follows the doctor's recommendations correctly.

All the chapters of Part ii have their own specific research questions and present the design of the proposed approach that we have designed to answer the research questions with the related discussion of the results. However, the three chapters together constitute the single path that guided this three-year work and helped shed light on some of the challenges we wanted to address.

2

INTELLIGENT AMBIENT ASSISTED LIVING (AAL) FOR SAFE PATIENT MONITORING

OUTLINE

Massive collection of data from Internet of Medical Things (IoMT) devices presents Security problems. Monitoring sensitive user data may be exposed to attacks aimed at compromising their structure and integrity.

In this chapter, we analyze the AAL paradigm, used to reduce costs for patient care. In addition, we presents CP as a method for defining the patient's treatment pathway. As a final step, we present two approaches to monitoring patient adherence to the CP: (i) Detecting anomalies in sensor data using AI algorithms and classifying them and, (ii) using Process Mining, we measure how the clinical treatment of the patient differs from CP.

Part of the content of this chapter has been presented in the papers [11, 13, 15, 16].

2.1 AAL FOR PATIENTS' HOME CARE

In healthcare domain, AAL [50] is a paradigm that is increasingly weighing on institutions' budgets. One of the possible approaches to be able to reduce its costs, is to use the AAL paradigm for home care of patients. Ardito *et al.* [13] define AAL as follows:

Definition 1. *"An emerging multi-disciplinary field aiming at providing an ecosystem of different types of sensors, computers, mobile devices, wireless networks and software applications for personal healthcare monitoring and remote care systems".*

The recent COVID-19 emergency has also shown how, to avoid contagion spreading, it is necessary to minimise access to hospital facilities by those chronically ill patients who can be monitored at home. And even COVID patients with mild symptoms can be cared for remotely, without the need to take up hospital places that can be allocated to more severe patients and without the risk of worsening their situation due to

contact with the latter.

The adoption of CPs [38, 150] (next described in Section 2.2), would make AAL implementations much more effective. Indeed, it would allow remote monitoring of patient care and automate the reporting of critical events that deviate from the prescribed care, also thanks to the use of Machine Learning (ML) techniques. Ardito *et al.* [11] defines CP as follows:

Definition 2. *"A set of diagnostic and therapeutic procedures to be performed for a specific patient. It is considered as a process model characterized by two main phases:*

1. *Activities, or sub-processes, managed by the personnel in the health structures;*

2. *Action managed autonomously by the patient and by medical-unsupervised manner. "*

The second action previous described, is processed by an intelligent architecture able to deal with the specific clinical sub-path for the patient at home. It also verifies that a physician or nurse validates the pathway and ensures its compliance with the actual medical indications specified in the Clinical Path.

AAL is becoming more and more successful thanks to the evolution of Internet of Things (IoT) technology and in particular of wearable devices. However, it must be considered that there are limits, mainly due to the latency of the network, that sometimes make the use of such solutions critical in the healthcare.

The security of data being transmitted from sensors to the cloud is another concern, as their transmission could be affected either by technical problems or malicious manipulations, resulting in life-threatening for the patient. Sensitive patient information could also be sniffed.

To strengthen patients' trust, health organizations should follow safety measures based on Health Insurance Portability and Accountability Act (HIPAA) compliance standards [124] that guarantee the proper privacy protection of data.

To address these issues, one of the scenarios described in this thesis and, implemented in a company case study, consists of architecture that couples IoT and Edge-Computing, which also implements an anomaly detection module (described in Section 4.1) able to detect deviations from the patient's CP.

In addition, in Chapter 3 is describes an application example of how it is crucial, including in healthcare, to ensure security compliance standards and adeguate data privacy.

With the advent of IoT, monitoring patients' care and their vital parameters has become easier thanks to wearable devices. Still, there are issues related to performances and security.

Weareable IoT devices are applied in crucial applications: monitoring vital signs, tracking indoor positions, or alerting for some important events (e.g., take a pill) [17]. Besides, with the advent of machine learning, these applications are becoming more and more sophisticated, requiring much computational power. Processes driven by such devices become time-consuming and harvest much computational power, thus also impacting on the battery life.

Low latency to send and receive critical data or high reliability to scale or replace these devices are the most critical constraints in the healthcare context. Standard cloud architectures to manage the network communications cannot be exploited. Indeed, cloud computing is not designed with these goals in mind, thus it doesn't fulfil these requirements [2].

The Multi-access Edge Computing (MEC) approach addresses this issue [160]. MEC is defined as the ability to process and store data at the edge of the network, i.e., in the proximity of the data sources. MEC's the advantage in a smart health environment is multifaceted, as it can provide short response time, decreased energy consumption for battery-operated devices, network bandwidth saving, secure transmission and data privacy [1].

The latter aspect has become predominant in recent times. It is a mandatory constraint in the healthcare application domain: data exchanged between an intelligent device and a central station contains sensible details about patients' personal lives. *Edge computing* is a promising solution to mitigate this issue. It is distributed. Thus sensible data are pre-processed on the edge of the network, and obfuscated sensitive information of the patient are sent to a central server that needs only extracted features to perform related tasks.

Abdellatif *et al.* [2] presented a healthcare system based on the multi-access edge computing paradigm. Such an approach enhances the performance of the system. Moreover, considering the low energy capacity of the devices, the management of the huge amount of data generated by sensors and human and medical factors is optimized. By adopting this architecture, bottlenecks in healthcare systems can be sig-

nificantly reduced thanks to the less amount of data transferred to the cloud. Some computing tasks that can be performed at the edge of the network are also identified. This ensures shorter response times, efficient processing, minimum power consumption and bandwidth saving.

However, these smart devices can also be subject to malfunctions and technical anomalies (intentional or unintentional). It is fundamental to detect in order to avoid serious life-threatening for a patient.

An anomaly detection system is proposed by Huang *et al.* [81]. It is based on the extraction of care-flow records that regularly capture medical behaviors in clinical processes, also identifying the anomalous ones. In order to monitor patient treatment and care behaviors in a variety of clinical settings, this approach requires an high-frequency detection rate of the care-flow records and a specific description of them.

Ahsanul Haque *et al.* [75] present a system for the detection of sensor anomalies in healthcare, able to distinguish real alarms from false alarms. The system is implemented in Java and takes advantage of the SMO regression utility of the WEKA framework. The system was tested on three real medical datasets. The value detected by a sensor is compared with the historical data, in order to detect suspicious variations. The experimental results show a Digital Radiography (DR) of 100 % and a low False-Positive Rate (FPR) for all the datasets.

Unfortunately, none of the solutions was presented previously. The analyzed in these PhD courses (both in industry and academia) fully meet the main requirements and challenges in remote patient monitoring and, consequently, possible anomaly and attack detection systems in healthcare.

Section 4.2 described one of the approaches presented in this thesis aims to provide a comprehensive, autonomous, effective, and *HIPAA*-compliant architecture [124] that is also capable of detecting cyber anomalies and attacks in healthcare. The Section 3.5.2 describe the HIPAA regulamentation, considered in the ECHO EU project[1].

2.2 CLINICAL PATH

Technological innovations have led to the gradual introduction of new methods of diagnosis and treatment, which should not dry up the doctor-patient relationship but rather enrich it. Is there a method to define the patient's treatment pathway and monitor it remotely? This thesis section introduces CP and puts it in context with our research. In the end,

1 https://echonetwork.eu/

an application example is provided.

Identifying some areas of innovation in charge of the Regional Health Services (RHS) in developing digital services for supporting clinical health and care processes is possible. The one that is attracting the most attention is the digital management of CP. CPs are the primary tool for implementing clinical guidelines and evidence-based medicine. Their primary objective is optimizing care and reducing unjustified variations in clinical practice and the health system's costs, thanks to its inter-functional and multidisciplinary nature and its integrated and cooperative coverage of the hospital and extra-hospital settings [37, 149].

CPs are effective at containing clinical complexity in a generalized context where chronic diseases are becoming more prevalent. This due to unhealthy lifestyles and poor eating habits: the elderly and (or) chronic patient typically presents a difficult to read clinical picture and a high diagnostic and therapeutic complexity.

As a result of an aging population and an increase in chronic diseases associated with unhealthy lifestyles and improper eating habits, CPs play a vital role in containing clinical complexity. It is typically difficult-to-read the clinical picture of elderly or chronic patients due to their high diagnostic and therapeutic complexity.

Clinical overload must be mitigated in new ways. Using predictive analysis, clinicians can be guided in dealing with complex pathology cases, from decision support to medical prescriptions. The Figure 2.1 shows an example of a Clinical Path.

Although the development of CPs offers increasingly broad and qualitatively satisfying coverage of clinical cases (both in the diagnostic and therapeutic fields, as well as in the treatment of chronic diseases), their adoption in treatment centers is still slowed down by human, cognitive, organizational, and technological barriers.

The increasing computerization of clinical/health workflows and medical records offers a significant opportunity to reduce these barriers. However, the same computerization of clinical practices has generated new problems - mainly due to a significant technological and informative fragmentation - that generally make progress in disseminating evidence-based medical practice sub-optimal and reduce the contribution of information technology to this progress.

Indeed, Information and Communication Technologies (ICT) is gaining momentum due to the consolidation of technologies and methods for process automation and the availability of important transactional

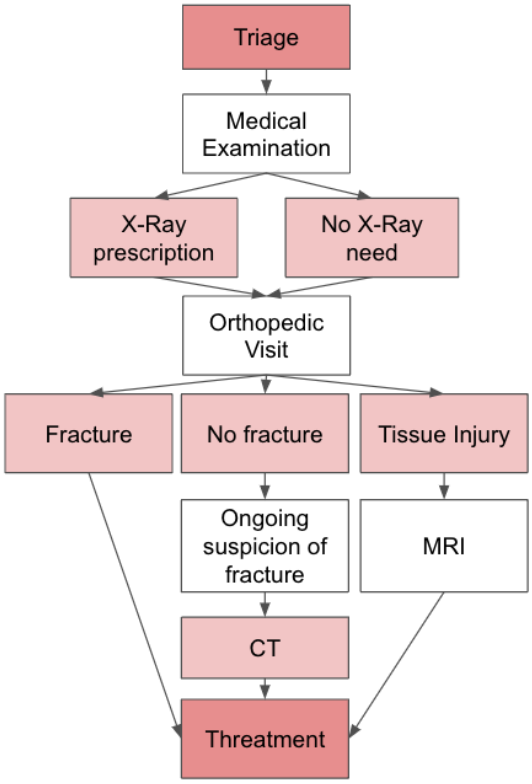

Figure 2.1: Example of Clinical Pathway.

and historical data and tools able to exploit its volume and heterogeneity for creating a prescriptive and predictive value. These techniques, under predictive analytics, are successfully applied in areas such as marketing, customer relations, fraud, or financial risk. Unfortunately, especially in Italy, they are weakly exploited in the clinical-health context, although characterized by the presence of enormous amounts of information, both structured and textual.

The progressive computerization of clinical practice processes in care centers leads to accumulating of an extraordinary amount of data. This process, whose value in terms of support to the medical decision (both for care and secondary uses - such as health administration/government and scientific research), is enormous and, to a large extent, unexpressed.

Another strategic area of the healthcare sector promotes the exploitation of these broad masses of health data. Develop models and solutions for Clinical Governance that boost interoperability between various databases, improving resource management and evidence-based governance.

In the healthcare domain, significant efforts are made to create a level of interoperability and exchange of information between different

systems. Furthermore, the biggest challenge is representing information streams to extract Clinical Pathway information from this scattered dataset. This is illustrated by some contributions that take this approach.

Integrated Clinical Pathway management has been proposed by Li *et al.* [106]. This approach is based on a semiotically inspired system architecture that aims to embed pathway knowledge into treatment processes and existing hospital information systems. To support later analysis, *et al.* [39] proposed a process mining-based approach that enables the extraction of valuable organizational and medical information on past CP executions from the event logs of healthcare information systems.

A contribution to the possibility of managing in a personalized and dynamic way comes from Schlieter *et al.* [148]. They proposed personalized active pathways and a reference architecture for integrating them into existing inter-organizational healthcare information systems. In the previous works, the authors aim to extract information from a series of heterogeneous systems to build a chain of events that, whether in a second phase, will be formalized as a concrete pathway.

Electronic health (e-Health) represents the application of ICT in healthcare. It enables efficient access to services, achieving the highest quality of care processes at low costs. It also makes clinical decisions safer and more appropriate through the structured sharing of clinical information and knowledge between all actors in the healthcare chain so that it can be used and accessed by everyone. The COVID-19 pandemic has necessitated digital healthcare, making it a significant opportunity for patients and doctors to communicate effectively, even at a distance. The security of the information transmitted is made possible by using the latest technologies based on advanced security protocols and data protection mechanisms that make an attack on the systems less likely.

Therefore, the protection of health data is a fundamental task that the engineers of eHealth systems must take into account.

In this direction, two application approaches are described below. Both are aimed at protecting the transmitted clinical data from attacks. In particular, it is assumed that the attack is directed at the patient's CP. There is, therefore, a need to identify when there is an attack. The two approaches are described below:

1. We use an artificial intelligence algorithm to identify anomalies in data transmission. In particular, we use LSTM autoencoders (Sec. 2.3.1.5) to identify data that does not match to the general data representing the patient's clinical pathway.

2. We use a Process Mining algorithm to identify deviation from the CP. In particular, we use Petri Nets (Sec. 4.3.3) to identify whether and how much the patient's clinical treatment deviates from that defined by the physician.

2.3 AI FOR IDENTIFYING ANOMALIES

Anomalies or outliers are rare and unusual patterns in any data stream. We aim to build an AI-powered outlier detection system that uses machine learning techniques for anomaly detection.

Identifying anomalies can be used to solve core business problems, such as fraud detection (i.e., Fintech or Banking) or intrusion detection (i.e., Security Domain).

In the healthcare industry, an anomaly detection is often a valuable tool for enhancing the quality of wearable devices. For instance, in healthcare, an outlier detection system can be used to monitor whether the communication between the devices worn by the patient and the server takes place correctly.

2.3.1 *Anomaly Detection*

Anomaly detection is an unsupervised task [6, 40, 128, 158]. Anomaly Detection is used in serveral fields such as Network Security [29], Smart Hospital [145] and, Remote monitoring and control in eHealth [117].

Anomaly detection, a.k.a. outlier detection or novelty detection, is referred to as the process of detecting data instances that significantly deviate from the majority of data instances. Anomaly detection has been an active research area for several decades, with early exploration dating back as far as the 1960s [72]. Due to the increasing demand and applications in broad domains, such as risk management, compliance, security, financial surveillance, health and medical risk, and AI safety, anomaly detection plays increasingly important roles, highlighted in various communities including data mining, machine learning, computer vision, and statistics.

In recent years, deep learning has shown tremendous capabilities in learning expressive representations of complex data such as high-dimensional data, temporal data, spatial data, and graph data, pushing the boundaries of different learning tasks. Deep learning for anomaly detection, deep anomaly detection for short, aim at learning feature representations or anomaly scores via neural networks for the sake of anomaly detection. A large number of deep anomaly detection methods

have been introduced, demonstrating significantly better performance than conventional anomaly detection on addressing challenging detection problems in a variety of real-world applications.

Definition 3. Given a dataset $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ with $\mathbf{x}_i \in \mathbb{R}^D$, let $\mathcal{Z} \in \mathbb{R}^K (K \ll N)$ be a representation space, then *deep anomaly detection* aims at learning a feature representation mapping function $\phi(\cdot) : \mathcal{X} \mapsto \mathcal{Z}$ or an anomaly score learning function $\tau(\cdot) : X \mapsto \mathbb{R}$ in a way that anomalies can be easily differentiated from the normal data instances in the space yielded by the ϕ or τ function, where both ϕ and τ are a neural network-enabled mapping function with $H \in \mathbb{N}$ hidden layers and their weight matrices $\Theta = \left\{\mathbf{M}^1, \mathbf{M}^2, \ldots, \mathbf{M}^H\right\}$.

In the case of learning the feature mapping $\phi(\cdot)$, an additional step is required to calculate the anomaly score of each data instance in the new representation space, while $\tau(\cdot)$ can directly infer the anomaly scores with raw data inputs. Larger τ outputs indicate greater degree of being anomalous.

2.3.1.1 *Problem Complexities*

Owing to the unique nature, anomaly detection presents distinct problem complexities from the majority of analytical and learning problems and tasks. This section summarizes such intrinsic complexities and unsolved detection challenges in complex anomaly data.

Major Problem Complexities. Unlike those problems and tasks on majority, regular or evident patterns, anomaly detection addresses minority, unpredictable/uncertain and rare events, leading to some unique problem complexities to all (both deep and shallow) detection methods:

- *Unknownness.* Anomalies are associated with many unknowns, e.g., instances with unknown abrupt behaviors, data structures, and distributions. They remain unknown until actually occur, such as novel terrorist attacks, fraud, and network intrusions;
- *Heterogeneous anomaly classes.* Anomalies are irregular, and thus, one class of anomalies may demonstrate completely different abnormal characteristics from another class of anomalies. For example, in video surveillance, the abnormal events robbery, traffic accidents and burglary are visually highly different;

- *Rarity and class imbalance.* Anomalies are typically rare data instances, contrasting to normal instances that often account for an overwhelming proportion of the data. Therefore, it is difficult, if not impossible, to collect a large amount of labeled abnormal instances. This results in the unavailability of large-scale labeled data in most applications. The class imbalance is also due to the fact that misclassification of anomalies is normally much more costly than that of normal instances.

- *Diverse types of anomaly.* Three completely different types of anomaly have been explored. Point anomalies are individual instances that are anomalous w.r.t. the majority of other individual instances, e.g., the abnormal health indicators of a patient. Conditional anomalies, a.k.a. contextual anomalies, also refer to individual anomalous instances but in a specific context, i.e., data instances are anomalous in the specific context, otherwise normal. The contexts can be highly different in real-world applications, e.g., sudden temperature drop/increase in a particular temporal context, or rapid credit card transactions in unusual spatial contexts. Group anomalies, a.k.a. collective anomalies, are a subset of data instances anomalous as a whole w.r.t. the other data instances; the individual members of the collective anomaly may not be anomalies, e.g., exceptionally dense subgraphs formed by fake accounts in social network are anomalies as a collection, but the individual nodes in those subgraphs can be as normal as real accounts.

2.3.1.2 *Main Challenges*

The above complex problem nature leads to a number of detection challenges. Some challenges, such as scalability w.r.t. data size, have been well addressed in recent years, while the following are largely unsolved, to which deep anomaly detection can play some essential roles.

- *CH1:Low anomaly detection recall rate.* Since anomalies are highly rare and heterogeneous, it is difficult to identify all of the anomalies. Many normal instances are wrongly reported as anomalies while true yet sophisticated anomalies are missed. Although a plethora of anomaly detection methods have been introduced over the years, the current state-of-the-art methods, especially unsupervised methods (e.g., References [32, 107]), still

often incur high false positives on real-world datasets [35, 140]. How to reduce false positives and enhance detection recall rates is one of the most important and yet difficult challenges, particularly for the significant expense of failing to spotting anomalies.

- *CH2: Anomaly detection in high-dimensional and/or not-independent data.* Anomalies often exhibit evident abnormal characteristics in a low-dimensional space yet become hidden and unnoticeable in a high-dimensional space. High-dimensional anomaly detection has been a long-standing problem [178]. Performing anomaly detection in a reduced lower-dimensional space spanned by a small subset of original features or newly constructed features is a straightforward solution, e.g., in subspace-based [95, 102] and feature selection-based methods [20]. However, identifying intricate (e.g., high-order, nonlinear and heterogeneous) feature interactions and couplings [36] may be essential in high-dimensional data, but it remains a major challenge for anomaly detection. Further, how to guarantee the new feature space preserved proper information for specific detection methods is critical to downstream accurate anomaly detection, but it is challenging due to the aforementioned unknowns and heterogeneities of anomalies. Also, it is challenging to detect anomalies from instances that may be dependent on each other such as by temporal, spatial, graph-based and other interdependency relationships [5, 8].

- *CH3:Data-efficient learning of normality/abnormality.* Due to the difficulty and cost of collecting large-scale labeled anomaly data, fully supervised anomaly detection is often impractical as it assumes the availability of labeled training data with both normal and anomaly classes. In the last decade, major research efforts have been focused on unsupervised anomaly detection that does not require any labeled training data. However, unsupervised methods do not have any prior knowledge of true anomalies. They rely heavily on their assumption on the distribution of anomalies. However, it is often not difficult to collect labeled normal data and some labeled anomaly data. In practice, it is often suggested to leverage such readily accessible labeled data as much as possible [5]. Thus, utilizing those labeled data to learn expressive representations of normality/abnormality is crucial for accurate anomaly detection. Semi-supervised anomaly detection, which assumes a set of labeled normal training data, is a research direction dedicated to this problem. Another research line is weakly

supervised anomaly detection that assumes we have some labels for anomaly classes yet the class labels are partial/incomplete (i.e., they do not span the entire set of anomaly class), inexact (i.e., coarse-grained labels), or inaccurate (i.e., some given labels can be incorrect). Two major challenges are how to learn expressive normality/abnormality representations with a small amount of labeled anomaly data and how to learn detection models that are generalized to novel anomalies uncovered by the given labeled anomaly data.

- *CH4:Noise-resilient anomaly detection.* Manyweakly/semi-supervised anomaly detection methods assume the labeled training data are clean, which can be vulnerable to noisy instances that are mistakenly labeled as an opposite class label. In such cases, we may use unsupervised methods instead, but this fails to utilize the genuine labeled data. Additionally, there often exists large-scale anomaly-contaminated unlabeled data. Noise-resilient models can leverage those unlabeled data for more accurate detection. Thus, the noise here can be either mislabeled data or unlabeled anomalies. The main challenge is that the amount of noises can differ significantly from datasets and noisy instances may be irregularly distributed in the data space.

- *CH5: Detection of complex anomalies.* Most of existing methods are for point anomalies, which cannot be used for conditional anomaly and group anomaly, since they exhibit completely different behaviors from point anomalies. One main challenge here is to incorporate the concept of conditional/group anomalies into anomaly measures/models. Also, current methods mainly focus on detect anomalies from single data sources, while many applications require the detection of anomalies with multiple heterogeneous data sources, e.g., multidimensional data, graph, image, text, and audio data. One main challenge is that some anomalies can be detected only when considering two or more data sources.

- *CH6: Anomaly explanation.* In many safety-critical domains there may be some major risks if anomaly detection models are directly used as black-box models. For example, the rare data instances reported as anomalies may lead to possible algorithmic bias against the minority groups presented in the data, such as under-represented groups in fraud detection and crime detection systems. An effective approach to mitigate this type of risk is

to have anomaly explanation algorithms that provide straightforward clues about why a specific data instance is identified as anomaly. Human experts can then look into and correct the bias. Providing such explanation can be as important as detection accuracy in some applications. However, most anomaly detection studies focus on detection accuracy only, ignoring the capability of providing explanation of the identified anomalies. To derive anomaly explanation from specific detection methods is still a largely unsolved problem, especially for complex models. Developing inherently interpretable anomaly detection models is also crucial, but it remains a main challenge to well balance the model's interpretability and effectiveness.

Autoencoders (Section 2.3.1.4) are one AI approach for detecting anomalies. Using autoencoders for ehealth task described in this thesi, assumes that a trained autoencoder learn the latent subspace of standard samples. Once trained, it result in a low reconstruction error for traditional models and a high reconstruction error for anomalies.

Keeping track of in-home care parameters in chronological order is essential when analysing health data. Since the sequentiality of the recording is important in this domain, it helps reconstruct the patient's status more accurately.

Application-wise, LSTM autoencoders make the most sense. It compresses the data using an encoder and maintains its original structure via a decoder.

2.3.1.3 *Long Short Term Memory (LSTM)*

This section presents LSTM, particularly important in analysing patients in healthcare.

Learning to store information over extended time intervals via recurrent backpropagation takes a very long time, mostly due to insufficient, decaying error back-flow. LSTM is a recurrent network architecture in conjunction with an appropriate gradient-based learning algorithm. Recurrent Neural Networks (RNNs) are networks in which connections between nodes create cycles. Their great success is due to their ability to find relationships between sequential data sets, such as the words of a sentence or the value of a stock over time. LSTM is designed to overcome these error back-flow problems.

Figure 2.2 represents an example [2] of an internal structure of an LSTM. The structure of a generic recurrent neural network does not dif-

2 https://colah.github.io/posts/2015-08-Understanding-LSTMs/

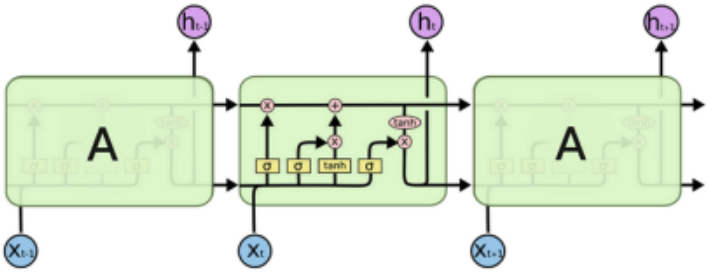

Figure 2.2: Internal structure of an LSTM.

fer significantly from a feed-forward neural network. It can be thought of as multiple copies of the same network, each of which can 'pass a message to its successor. The network's output is influenced not only by weights applied to the input as in a feed-forward network but also by the *hidden* state the message passed, which depends on the context of the previous inputs and outputs. Specifically, the output at step t of the sequence is given by $\tanh([h_{t-1}, x_t])$ where, $[h_{t-1}, x_t]$ indicates the concatenation between the output at the step (t-1)-n and the input at step t. Thus, the same input may produce different outputs depending on the previous inputs in the sequence.

To construct an architecture that allows for constant error flow through special, self-connected units without the disadvantages of the naive approach, we extend the constant error carrousel CEC embodied by the self-connected, linear unit j by introducing additional features. A multiplicative input $gateunit$ is introduced to protect the memory contents stored in j from perturbation by irrelevant inputs. Likewise, a multiplicative $outputgate$ unit is introduced which protects other units from perturbation by currently irrelevant memory contents stored in j.

The resulting, more complex unit is called a $memorycell$ (Figure 2.3). The j-th memory cell is denoted c_j. Each memory cell is built around a central linear unit with a fixed self-connection (the CEC). In addition to net_{c_j}, c_j gets input from a multiplicative unit out_j (called $outputgate$), and from another multiplicative unit in_j (called $inputgate$). in_j's activation at time t is denoted by $y^{in_j}(t)$, out_j's by $y^{\text{out}_j}(t)$. In Equation 2.1 we have:

$$y^{\text{out}_j}(t) = f_{\text{out}_j}\left(\text{net}_{\text{out}_j}(t)\right); y^{in_j}(t) = f_{\text{in}_j}\left(\text{net}_{\text{in}_j}(t)\right) \quad (2.1)$$

where

$$\text{net}_{\text{out}_j}(t) = \sum_u w_{\text{out}_j u} y^u(t-1) \quad (2.2)$$

and

$$n^{net}{}_{in_j}(t) = \sum_u w_{in_j u} y^u(t-1) \tag{2.3}$$

We also have

$$net_{c_j}(t) = \sum_u w_{c_j u} y^u(t-1) \tag{2.4}$$

The summation indices u may stand for input units, gate units, memory cells, or even conventional hidden units if there are any. All these different types of units may convey useful information about the current state of the net. For instance, an input gate (output gate) may use inputs from other memory cells to decide whether to store (access) certain information in its memory cell. There even may be recurrent self-connections like $w_{c_j c_j}$. It is up to the user to define the network topology. An example is represented in Figure 2.3. At time t, c_j 's output $y^{c_j}(t)$ is computed as in Equation 2.5:

$$y^{c_j}(t) = y^{out_j}(t) h\left(s_{c_j}(t)\right) \tag{2.5}$$

where the internal state $s_{c_j}(t)$ is:

$$s_{c_j}(0) = 0, s_{c_j}(t) = s_{c_j}(t-1) + y^{in_j}(t) g\left(\text{net}_{c_j}(t)\right) \text{ for } t > 0 \tag{2.6}$$

The differentiable function g squashes net_{c_j}; the differentiable function h scales memory cell outputs computed from the internal state s_{c_j}.

WHY GATE UNITS? To avoid input weight conflicts, in_j controls the error flow to memory cell c_j 's input connections $w_{c_j i}$. To circumvent c_j 's output weight conflicts, out_j controls the error flow from unit j 's output connections. In other words, the net can use in_j to decide when to keep or override information in memory cell c_j, and out_j to decide when to access memory cell c_j and when to prevent other units from being perturbed by c_j (as shown in Figure 2.3).

Error signals trapped within a memory cell's CEC cannot change - but different error signals flowing into the cell (at different times) via its output gate may get superimposed. The output gate will have to learn which errors to trap in its CEC, by appropriately scaling them. The input gate will have to learn when to release

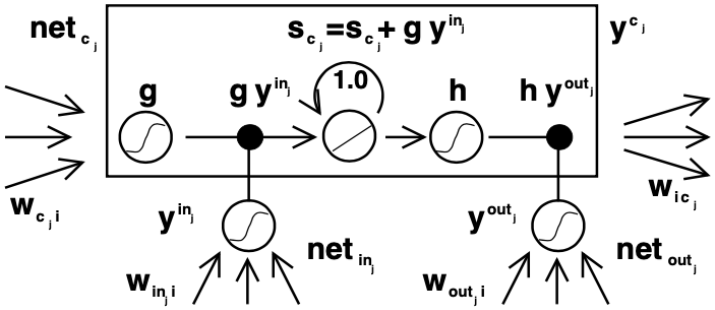

Figure 2.3: Architecture of memory cell c_j (the box) and its gate units in_j, out_j. The self-recurrent connection (with weight 1.0) indicates feedback with a delay of 1 time step. It builds the basis of the constant error carrousel (CEC). The gate units open and close access to CEC.

errors, again by appropriately scaling them. Essentially, the multiplicative gate units open and close access to constant error flow through CEC. Distributed output representations typically do require output gates. Not always are both gate types necessary, though - one may be sufficient.

NETWORK TOPOLOGY. We use networks with one input layer, one hidden layer, and one output layer. The (fully) self-connected hidden layer contains memory cells and corresponding gate units (for convenience, we refer to both memory cells and gate units as being located in the hidden layer). The hidden layer may also contain *conventional* hidden units providing inputs to gate units and memory cells. All units (except for gate units) in all layers have directed connections (serve as inputs) to all units in the layer above (or to all higher layers - Experiments 2a and 2b).

MEMORY CELL BLOCKS. S memory cells sharing the same input gate and the same output gate form a structure called a "memory cell block of size S ". Memory cell blocks facilitate information storage - as with conventional neural nets, it is not so easy to code a distributed input within a single cell. Since each memory cell block has as many gate units as a single memory cell (namely two), the block architecture can be even slightly more efficient (see paragraph "computational complexity"). A memory cell block of size 1 is just a simple memory cell.

COMPUTATIONAL COMPLEXITY. As with Mozer's focused recurrent backpropagation algorithm [78], only the derivatives $\frac{\partial s_{c_j}}{\partial w_{it}}$ need

to be stored and updated. Hence the LSTM algorithm is very efficient, with an excellent update complexity of $O(W)$, where W the number of weights. Hence, LSTM and BPTT for fully recurrent nets have the same update complexity per time step (while RTRL's is much worse). Unlike full BPTT, however, LSTM is local in space and time: there is no need to store activation values observed during sequence processing in a stack with potentially unlimited size.

2.3.1.4 *Autoencoder*

This section details the operation of autoencoders [23]. It is a specific type of a neural network, which is mainly designed to encode the input into a compressed and meaningful representation, and then decode it back such that the reconstructed input is similar as possible to the original one.

Autoencoders have been first introduced in [112] as a neural network that is trained to reconstruct its input. Their main purpose is learning in an unsupervised manner an "informative" representation of the data that can be used for various implications such as clustering. The problem, as formally defined in [21], is to learn the functions $A : \mathbb{R}^n \rightarrow \mathbb{R}^p$ (encoder) and $B : \mathbb{R}^p \rightarrow \mathbb{R}^n$ (decoder) that satisfy the follow:

$$\arg\min_{A,B} E\left[\Delta(\mathbf{x}, B \circ A(\mathbf{x})\right], \tag{2.7}$$

where E is the expectation over the distribution of x, and Δ is the reconstruction loss function, which measures the distance between the output of the decoder and the intput. The latter is usually set to be the ℓ_2-norm. Figure 2.4 provides an illustration of the autoencoder model.

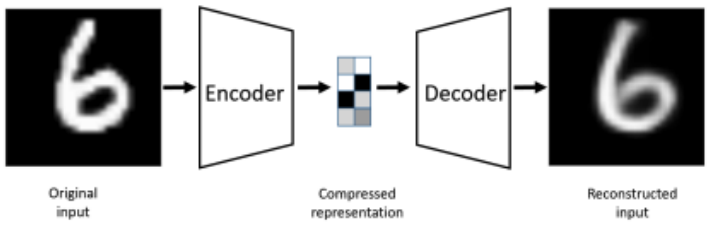

Figure 2.4: An autoencoder example. The input image is encoded to a compressed representation and then decoded.

In the most popular form of autoencoders, A and B are neural networks [141]. In the special case that A and B are linear operations, we get a linear autoencoder [165]. In the case of linear autoencoder

where we also drop the non-linear operations, the autoencoder would achieve the same latent representation as Principal Component Analysis (PCA) [136]. Therefore, an autoencoder is in fact a generalization of PCA, where instead of finding a low dimensional hyperplane in which the data lies, it is able to learn a non-linear manifold. Autoencoders may be trained end-to-end or gradually layer by layer. In the latter case, they are ”stacked” together, which leads to a deeper encoder. The various types of existing autoencoders are described below.

REGULARIZED AUTOENCODERS. Since in training, one may just get the identity operator for A and B, which keeps the achieved representation the same as the input, some additional regularization is required. The most common option is to make the dimension of the representation smaller than the input. This way, a *bottleneck* is imposed. This option also directly serves the goal of getting a low dimensional representation of the data. This representation can be used for purposes such as data compression, feature extraction, etc. Its important to note that even if the *bottleneck* is comprised of only one node, then overfitting is still possible if the capacity of the encoder and the decoder is large enough to encode each sample to an index. In cases where the size of the hidden layer is equal or greater than the size of the input, there is a risk that the encoder will simply learn the identity function. To prevent it without creating a bottleneck (i.e. smaller hidden layer) several options exists for regularization, which we describe hereafter, that would enforce the autoencoder to learn a different representation of the input. An important tradeoff in autoencoders is the bias-variance tradeoff. On the one hand, we want the architecure of the autoencoder to be able to reconstruct the input well (i.e. reduce the reconstruction error). On the other hand, we want the low representation to generalize to a meaningful one. We now turn to describe different methods to tackle such tradeoffs.

SPARSE AUTOENCODERS. One way to deal with this tradeoff is to enforce sparsity on the hidden activations. This can be added on top of the bottleneck enforcement, or instead of it. There are two strategies to enforce the sparsity regularization. They are similar to ordinary regularization, where they are applied on the activations instead of the weights. The first way to do so, is to apply L_1 regularization, which is known to induce sparsity. Thus, the autoencoder optimization objective is described in Equation 2.8:

$$\arg\min_{A,B} E\left[\Delta(\mathbf{x}, B \circ A(\mathbf{x})] + \lambda \sum_i |a_i|\right. \tag{2.8}$$

Where a_1 is the activation at the i-th hidden layer and i iterates over all the hiddens activations. Another way to do so, is to use the KL-divergence, which is a measure of the distance between two probability distributions. Instead of tweaking the *lambda* parameter as in the L_1 regularization, we can assume the activation of each neuron acts as a Bernouli variable with probability p and tweak that probability. At each batch, the actual probability is then measured, and the difference is calculated and applied as a regularization factor. For each neuron j, the calculated empirical probability is $\hat{p}_j = \frac{1}{m}\sum_i a_i(x)$, where i iterates over the samples in the batch. Thus the overall loss function would be represented by the follows Equation:

$$\arg\min_{A,B} E\left[\Delta(\mathbf{x}, B \circ A(\mathbf{x})] + \sum_j KL\left(p \| \hat{p}_j\right)\right. \tag{2.9}$$

Where the regularization term in it aims at matching p to $\hat{p}$.

DENOISING AUTOENCODERS. Denoising autoencoders [168] can be viewed either as a regularization option, or as robust autoencoders which can be used for error correction. In these architectures, the input is disrupted by some noise (e.g., additive white Gaussian noise or erasures using Dropout) and the autoencoder is expected to reconstruct the clean version of the input, as illustrated in Figure 2.5.

Note: $\hat{x}$ is a random variable, whose distribution is given by $C_\sigma(\tilde{x}|x)$. Two common options for C are:

$$C_\sigma(\tilde{x}|x) = N(x, \sigma^2 I), \tag{2.10}$$

and

$$C_p(\tilde{x}|x) = \beta \odot x, \beta \sim Ber(p), \tag{2.11}$$

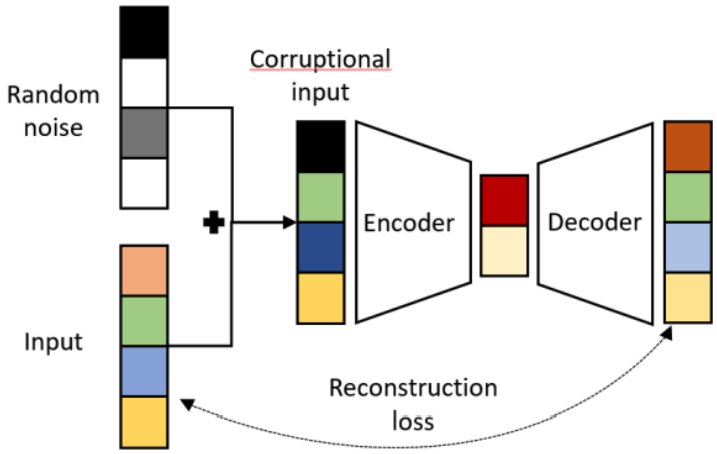

Figure 2.5: Example of denoising autoencoder. The disrupted input image is encoded to a representation and then decoded.

Where $\odot$ detnotes an element-wise (Hadamard) product. In the first option, the variance parameter σ sets the impact of the noise. In the second, the parameter p sets the probability of a value in x not being nullified. A relationship between denoising autoencoders with dropout to analog coding with erasures has been shown in [22].

CONTRACTIVE AUTOENCODERS. In denoising autoencoders, the emphasis is on letting the encoder be resistant to some perturbations of the input. In contractive autoencoders, the emphasis is on making the feature extraction less sensitive to small perturbations, by forcing the encoder to disregard changes in the input that are not important for the reconstruction by the decoder. Thus, a penatly is imposed on the Jacobian of the network. The Jacobian matrix of the hidden layer h consists of the derivative of each node h_j with respect to each value x_i in the input x. Formally: $J_{ji} = \nabla_{x_i} h_j(x_i)$ In contractive autoencoders we try to minimize its L2 norm, such that the overall optimization loss would be:

$$argmin_{A,B} E\lfloor \delta(x, B) \circ A(x) \rfloor + \lambda \parallel J_A(x) \parallel_2^2 \tag{2.12}$$

The reconstruction Loss Function and the Regularization Loss actually pull the result towards opposite directions. By minimizing the squared Jacobian norm, all the latent representations of the input tend to be more similar to each other, and by thus make the reconstruction more difficult, since the differences between

the representations are smaller. The main idea is that variations in the latent representation that are not important for the reconstructions would be diminished by the regularization factor, while important variations would remain because of their impact on the reconstruction error.

VARIATIONAL AUTOENCODERS. A major improvement in the representation capabilities of autoencoders has been achieved by the Variational Autoencoders (VAE) model [98]. Following Variational Bayes (VB) Inference [30], VAE are generative models that attempt to describe data generation through a probabilistic distribution. Specifically, given an observed dataset $X = \{X\}_{i=1}{}^{N}$ of V i.i.d samples, we assume a generative model for each datum x_i conditioned on an unobserved random latent variable z_i where θ are the parameters governing the generative distribution. This generative model is also equivalent to a *probabilistic decoder*. Symmetrically, we assume an approximate posterior distribution over the latent variable z_i given a datum x_i denoted by recognition, which is equivalent a *probabilistic encoder* and governed by the parameters ϕ. Finally, we assume a prior distribution for the latent variables z_i denoted by $p_\theta(z_i)$.

Figure 2.6 depicts the relationship described above. The parameters θ and ϕ are unknown and needs to learned from the data. The observed latent variables z_i can be interpreted as a code given by the recognition model $q_\phi(z|x)$. The marginal log-likelihood is expressed as a sum over the individual data points: $logp_\theta(x_1, x_2, ..., x_N) = \sum_{i=1}{}^{N} logp_\theta(x_i)$ and each point can be rewritten as:

$$logp_\theta(x_i) = D_{KL}q_\phi(z \mid x_i) \parallel p_\theta(z \mid x_i) + L(\theta, \phi; x_i) \quad (2.13)$$

where the first term is the Kullback-Leibler divergence of the approximate recognition model from the true posterior and the second term is called the variational lower bound on the marginal likelihood.

2.3.1.5 *LSTM Autoencoder*

Long Short-Term Memory, or LSTM, networks are recurrent neural networks that can support sequences of input data. Their internal memory enables them to remember or use information across long input

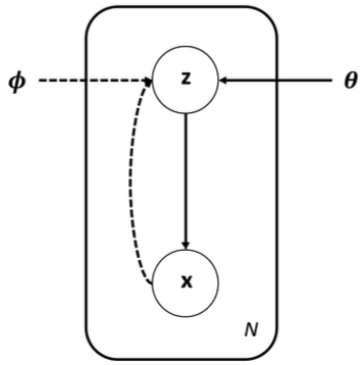

Figure 2.6: A Graphical Representatin of VAE.

sequences as well as to learn complex dynamics within their temporal orderings. LSTM networks can be structured so they can be used to support variable length input sequences as well as predict or output variable length output sequences in an Encoder-Decoder LSTM architecture.

LSTM has been widely used in many fields and achieved great success, such as in music generation, image caption, speech recognition and machine translation. LSTM improves the hidden-layer cell on the basis of RNN. The improvement of cell can make up for the gradient disappearance problem of RNN. LSTM adds some memory units, including forget gate, input gate and output gate. The memory units can further control the data and decide which should be retained and which should be deleted. The dimensionality of the data is reduced. The acquired characteristics are taken as a part of the input data of the prediction network. In this way, the data dimension is not increased too much.

Using an LSTM autoencoder, we present in Section 4.2 an approach to detect whether data transmitted via wearable devices are compromised.

2.4 PROCESS MINING FOR DETECTING ANOMALIES

New frontiers of healthcare delve deeper into eHealth practices, leading to any diagnostic, therapeutic, or social support service provided at home. In fact, home care, when used appropriately, reduces hospitalization and nursing home use without compromising medical outcomes. Furthermore, patients generally prefer to remain in a familiar environment. The medical support of home care services honors that prefer-

ence. Home care includes the use of medical equipment, telemedicine monitoring, and portable diagnostic tools. The technology-intensive services range from the simple recording of vital parameters to the more complex management of the entire therapeutic path.

In this context, healthcare related to remote clinical decision-making can be considered a critical operation. Situation Awareness (SA) [147] deals with this decision process to maintain and understand what is happening in a specific situation and leverage this information to avoid or mitigate eventual risks. Today, the IoT has evolved due to the advanced interconnectivity of hardware devices equipped with sensors and actuators. With the advent of IoT in healthcare, many low-cost devices are used to monitor a patient's health status remotely. IoT has been widely adopted for both in-home and in-hospital care. Therefore, it is possible to achieve a situation-aware IoT smart home/health environment by: (i) provisioning sensed data to enable monitoring of these environments, (ii) detecting situations based on recorded event logs, and (iii) triggering an action based on the recognized situations. Specifically, a situation-aware system in eHealth allows personalizing the therapeutic path for every patient, considering the biological characteristics of the pathology, the aspects of the clinical history, and the living environment.

The complexity and rise of data in the healthcare sector mean that AI will increasingly be applied within the field, potentially transforming many aspects of patient care and administrative processes. Inherently, specific AI sub-fields such as ML, Deep Learning (DL), Natural Language Processing (NLP), and Process Mining enable healthcare stakeholders and medical professionals to identify healthcare needs and solutions faster and more accurately. This using computational models to make informed medical or business decisions quickly. In particular, process mining for healthcare is an appropriate method for extracting information from event logs that are scattered throughout the health system and for defining workflows to be analyzed.

In more recent times, the healthcare industry may be revolutionized by edge-enabled AI, where a series of embedded sensors and IoT devices interconnect to provide diverse intelligent services for the well-being of in-home patients. As Edge AI technology continues to mature, it is increasingly being included in healthcare decision making, as AI is

now a key use case for edge computing and edge is a significant enabler for AI.

2.4.1 *Process Mining*

The increasing adoption of Hospital Information Systems (HISs) and Electronic Health Records (EHRs), together with the recent IoT advancements, are allowing smart homes and hospitals to measure a variety of patient- and process related data. Specifically, process mining has emerged as a suitable approach to analyze, discover, improve and manage real-life and complex processes, by extracting knowledge from event logs. In particular, healthcare processes are renowned as complex, flexible, multidisciplinary and ad-hoc, and, thus, they are difficult to manage and analyze with traditional model-driven techniques [55].

Process mining techniques have been used for various use cases within the healthcare domain. Examples include discovering the actual order of activities in a patient's treatment trajectory [57], evaluating the extent to which clinical guidelines have been followed [174], and predicting patient outcome based on how the treatment is performed [114].

Starting point for process mining is an *event log*. All process mining techniques assume that it is possible to sequentially record events such that each event refers to an *activity* (i.e., a well-defined step in some process) and is related to a particular *case* (i.e., a specific execution of activities in a determined order). *Case traces* are lists of events associated to steps. A *workflow* (or process model) is therefore a formal specification of how an activity sequence can be composed and can end in a valid process. A *process* consists of a suitable combination of different tasks performed by agents. A *task* is a generic piece of work to be executed. In particular, an existing process model can be compared with an event log of the same process. *Conformance Checking* [120, 122, 162] is a specific process mining technique that can be used to check if reality, as recorded in the log, conforms to the model and vice versa. Hence, conformance checking techniques need an event log and a model as input. The output consists of diagnostic information showing differences and commonalities between model and log. In other words, it evaluates how well an event log that records the actual executions matches the model.

The mining process allows automatic discovery of a process model from event logs which provides insight into it and enables various types of model-based analysis. Many diverse model specification techniques have been proposed, e.g. Petri net [133], BPMN [171], UML function diagrams [109]. Petri net is widely used in process mining as a basis for validation. It can be described as a graphical method of the formal definition of logical interaction between components or the flow of activities in a complex system. Petri net is particularly well suited for modelling concurrency and conflict, sequencing, conditional branch and looping, synchronization, limited resource allocation, and mutual exclusion. The log of a series of events will refer as a task. When the task is executed a token is produced at the start state, then when an event is executed if an edge with the same name is enabled (that is to say there is a token on the preceding state) then the token is consumed and one produced at each connected state. The fact that an edge may lead to more than one state allows for parallel execution. If there is no enabled edge matching the event then one which has the same name is chosen at random and the same process is followed, without the consumption.

2.4.2 *Petri Net*

Petri nets are the oldest process modeling language, introduced in 1962 by Carl Adam Petri [134], a helpful abstract model for representing the dynamics of a system characterized by synchronous and concurrent activities (activities that can be performed in parallel). A Petri net is a graph bipartite, oriented and weighted graph composed of places, transitions, and arcs connecting them. Input arcs connect places to transitions, while output arcs connect changes to locations. The structure of the network is static, but tokens can flow through the network; in fact, the state of a Petri net is determined by the distribution of tickets on the places, which indicates the Petri net's marking. Marking refers to the fact that each place contains a non-negative integer number negative integer number of tokens, defines the state of the network, and allows its evolution.

A Petri net is defined by a quadruple $PN = \{P, T, Pre, Post\}$ where P is the finite set of places, T is the finite set of transitions, Pre is a pre-incidence matrix specifying the arcs directed from places to transitions, and $Post$ is a post-incidence matrix specifying the arcs directed from transitions to places. With Petri nets, it is possible to represent and describe a process. Still, it is also possible to follow the evolution of the process, visualising the state the network is in at a particular instant

in a specific instant in a dynamic manner. The state of a Petri net is graphically represented by placing the tokens in places, allowing the status of the operations carried out in the process. The presence of tokens in the areas indicates the availability of the resource in question. The "firing rule defines the dynamic behavior of a Petri net," i.e., a transition is enabled if all places before the transition (pre-set) contain several tokens at least equal to the weight of the arc connecting them to the transition. The firing of a transition causes the removal from each place in the pre-set and the addition of several tokens at areas downstream of the transition (post-set) equal to the weight of the arcs connecting the transition to these places.

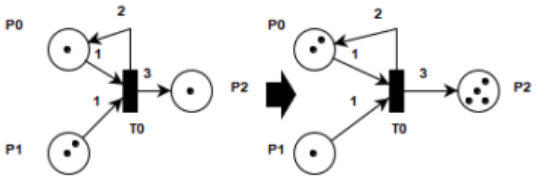

Figure 2.7: Example of shooting a transition.

Fundamental structures of a Petri net, useful for modeling systems:

- **Sequence:** represented as a succession of places and transitions. Two transitions, t_0 and t, are said to be in sequence if t_0 precedes t_1.

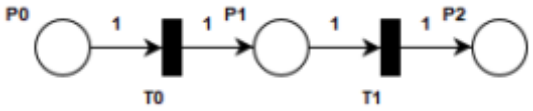

Figure 2.8: Example of transitions in sequence.

- *Parallelism:* one event triggers and enables different events simultaneously, and then it is decided first to trigger.

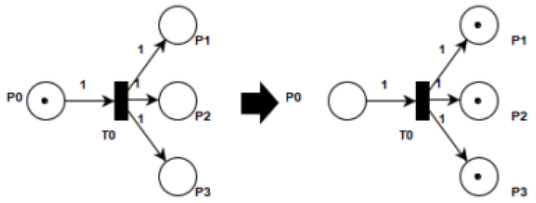

Figure 2.9: Example of parallel transitions.

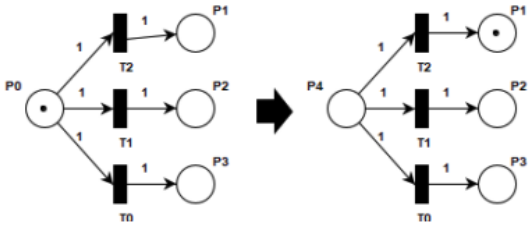

Figure 2.10: Example of a choice situation.

- *Choice:* as opposed to parallelism, where everyone can shoot, with the choice, I can only enable one, and there is a different evolution depending on the event that is triggered.
- *Synchronisation:* transitions without shared input posts, all enabled and followed by output posts that are also input posts for a change common.

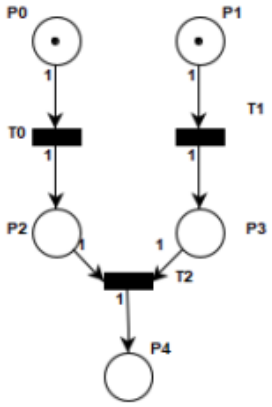

Figure 2.11: Example of synchronous transitions.

Petri Net Properties. The basic properties of Petri nets are the following. They are called behavioural properties as they depend on the structure of the network and the initial marking:

- *Reachability:* indicates the possibility of obtaining a given marking, i.e., a specific state, from another marking. A marking M is said to be reachable from a marking M_0 if at least one sequence of transitions exists, triggering them from M results in M_0;
- *Restrictedness:* a place in a Petri net is said to be k-restricted if in all the markings reached by the network, the number of tokens present in the area never exceeds a set value, k;

- *Reversibility:* a Petri net with initial marking M_0 is said to be reversible if for each marking M reachable from M_0, M_0 is reachable from M; therefore, if from each marking, it is possible to return to the initial marking M_0;
- *Conservative:* a marked Petri net is strictly conservative if, for each reachable marking, the net's number of tokens does not vary;
- *Aliveness:* a transition t is said to be alive if and only if starting from any marking of the graph, I succeed in turning t on.

The using of Petri nets have the following advantages:

- The graphical representationis very compact and concise, allowing an easier understanding of the evolution and functioning of the system;
- Are mathematical models, and this allows the analysis of the network using linear algebra;
- They use a modular representation; in fact, each part of the system can be considered as an independent subsystem, and it is possible to analyse independently of the others;
- Are easy to analyse activities that take place simultaneously.

2.4.3 *Process Mining Techniques*

One of the fundamental aspects of process mining is the emphasis placed on establishing a relationship between the process model and the reality contained in the event log. To describe this type of relationship, the terms play-out, play-in, and replay are detailed:

- *Play-out:* refers to the classic use of process models, i.e., given a Petri net, it is possible to generate behaviour. It can be used for analysis and business process realisation, as it allows business processes to be executed using executable process models. The main idea of simulation is to run a model and collect statistics and confidence intervals repeatedly.
- *Play-in:* is the opposite of play-out, i.e., the example behavior recorded in the logs is taken as input, and the aim is to automatically build a model based on the recorded data. The α-algorithm or other process discovery approaches are examples of play-in techniques.

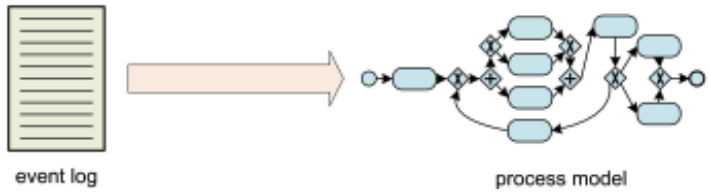

Figure 2.12: Example of Play-out relation.

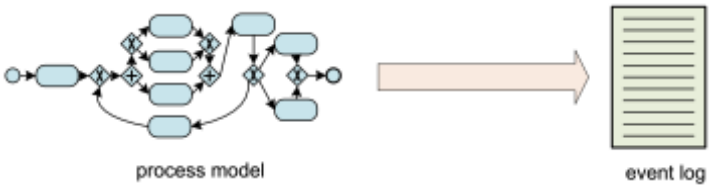

Figure 2.13: Example of Play-in relation.

- *Replay:* uses an event log and a process model as input. The event log is 'replayed' on the process model to check conformity, i.e., discrepancies between the event log and the model. With the log replay, it is possible to see which parts of the model are executed frequently, allowing bottlenecks to be identified. Furthermore, predictive models can be built, i.e., predictions can be made for different states of the model. Replay is not only limited to recorded event data, but it is also possible to replay partial traces of events still in progress. Partial traces of events still in execution to detect deviations during process execution.

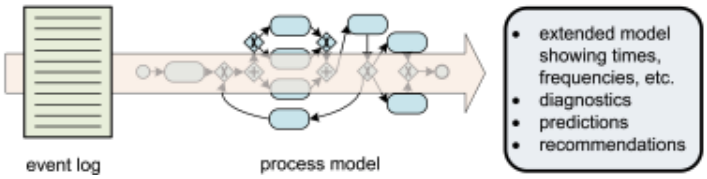

Figure 2.14: Example of Replay relation.

Process mining techniques can be divided into:

- *Process discovery:* takes a log of events as input and produces a model based on the recorded data. In the following, we will analyse some of the most widely used algorithms for process pattern discovery.

- *Conformance checking:* a process model, discovered through process discovery or pre-existing, is compared with the information in an event log. This makes it possible to check whether what is

happening in reality, hence the information contained in the logs, conforms to the model and vice versa.

- *Enhancement:* the idea is to extend or improve an existing process model by using process-related information stored in the logs. Unlike conformance checking, which measures how well a model is aligned with reality, enhancement aims at extending the existing model, showing, for example, bottlenecks, throughput times productivity, and frequencies, by using timestamps stored in the event log.

Section 4.3 presents a method for identifying when a patient's treatment differs from that prescribed by their doctor using Process Mining techniques.

2.5 EXPLAINABLE SECURITY

Keeping information, data, processes, software, protocols, computers, networks, and systems secure is notoriously challenging (and often intractable). Security is a challenge. Achieving, reasoning about, applying, understanding, and teaching it is difficult. It is difficult to explain.

A new program called eXplainable Artificial Intelligence (XAI) was launched in 2017 by the Defense Advanced Research Projects Agency (DARPA). An emerging generation of AI systems can be understood, appropriately trusted, and effectively managed through AI techniques. Some research on explainable AI had already been published before DARPA's program but XAI encouraged a large number of researchers to take up this challenge. In the last couple of years, several publications have appeared that investigate how to explain the different areas of AI, such as machine learning [64], recommender systems [118], robotics and autonomous systems [152], constraint reasoning [67] and planning [66].

In [135], Pieters investigates the relation between explanation and trust, focusing in particular on expert systems and e-voting systems. Pieters observes that:

"In artificial intelligence, explanations are usually provided by the system itself. In information security, explanations are provided by the designers. Nonetheless, in both artificial intelligence and information security, the role of explanations consists for a major part of acquiring

and maintaining the trust of the user of the system."

He discusses how explanations are required for trust:

"Here, the question is how it is possible to communicate the analysis that experts have made of a security-sensitive system to the public. Why is it secure? Or, more appropriately: How is it secure?"

Explanations are thus:

"thought to bridge the gap between 'actual security' and 'perceived security'."

Pieters also discusses two main goals that an explanation may have: transparency (e.g., to allow users to understand what the designers have done to protect them) and justification (e.g., offering reasons for an action). He contrasts explanation-for-trust (i.e., explanation of how a system works, by revealing details of its internal operations) with explanation-for-confidence (i.e., explanation to make the user feel comfortable in using the system, by providing information on its external communications).

Developers, analysts, users, and attackers are among the stakeholders involved in XSec, which is multifaceted in nature (as reasoning is required about system models, threat models, properties of security, privacy, and trust, as well as concrete attacks, vulnerabilities, and countermeasures).

As a result, XSec is an exciting new paradigm that requires a full-scale and heterogeneous research program.

Vigano [167] defines a roadmap that identifies several possible research directions. To describe the challenges of XSec and how they could be tackled, we discussing the "*Six Ws*" of XSec summarized in Figure 2.15: Who? What? Where? When? Why? and How?

- *Who?*
  Consider a generic system, where we use here "system" to refer to a generic process, software, protocol, computer, network, cyber-physical system, critical infrastructure, etc. that processes information/data whose security must be protected, where "security" similarly generically refers to one or more of the security properties of interest, including confidentiality, integrity, availability,

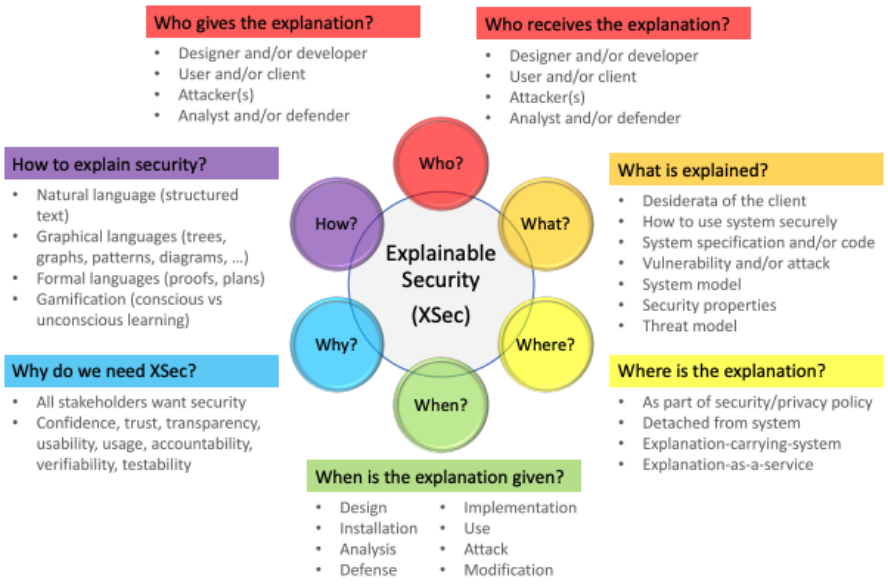

Figure 2.15: The Six Ws of Explainable Security.

authentication, authorization, nonrepudiation, accountability, unobservability, privacy, etc.

The dramatis personae of XSec are:

- the designer (and/or developer) of the system, who has designed (and/or developed) the system to guarantee a number of specified security properties;
- the user (and/or client) of the system, who can typically be assumed to be an honest non-expert who might commit mistakes that make the system vulnerable;
- the attacker (or more attackers) of the system, who searches for and exploits vulnerabilities of the system for reasons or profit, fame, reward, etc.;
- the analyst of the system, who carries out a semi-formal or formal analysis of the system at design time (or based on the specification of a deployed system) or tests the system at runtime (e.g., using penetration testing, vulnerability-based testing or model-based testing);
- the defender of the system, who attempts to protect the system, e.g., by monitoring the activities of the system and reacting to the attacker's actions.

For some situations, the recipient of the explanation will be an agent rather than a human, and we can then contrast internal explanations

(designed for software agents) and external explanations (designed for humans, which is what XAI research typically focuses on). Some of the above roles might actually be played by the same "principal" (agent or human), as the designer might for instance act also as analyst or defender, the analyst might also provide immediate defense and the attacker might be a user of the system. The literature is full of examples of vulnerabilities caused by mistakes by designers or users, along with details of the corresponding attacks. Some of these attacks could have been prevented by better explanations. In fact, all of these roles might require explanations or need to act as explainer, For instance:

- a designer and/or an analyst might need to explain to the user how to interact with the system, why the system is secure and why it carries out a particular action;
- a user or client might need to explain to the designer or the analyst how he expects the system to behave and how they typically interact with the system, to allow the designer to elicit the requirements for building the system in the first place and to allow the analyst to validate the security of user interactions;
- an attacker might need to explain the attack strategy to his accomplices so that they can attack in coordination, or he might have used a complex penetration testing tool to test the system for vulnerabilities and now needs the tool to explain to him the attack trace (or attack plan or strategy) that has been identified so that he can carry out the attack for real;
- an analyst might need to explain to the designer how to improve the system's security or to the defender how and what to defend;
- a defender might similarly need to understand possible attack traces in order to take action against them as well as explain to the users how they should behave to protect the system and themselves.

In addition to this, it is also necessary to tackle the research question of what actually constitutes a good (and secure) explanation, as we discuss in more detail in the following sections. Before we do so, let us consider a concrete example that arises from the observation that when dealing with sensitive data, classical authentication solutions based on username and password pairs are not enough. The General Data Protection Regulation (GDPR) [126] mandates that specific security

measures must be implemented, including multi-factor authentication (MFA), an authentication solution that aims to augment the security of the basic username and password authentication by exploiting two or more authentication factors.

In [European Central Bank, 2014], MFA is for instance defined as:

"a procedure based on the use of two or more of the following elements — categorised as knowledge, ownership and inherence: i) something only the user knows, e.g., static password, code, personal identification number; ii) something only the user possesses, e.g., token, smart card, mobile phone; iii) something the user is, e.g. a biometric characteristic, such as a fingerprint. In addition, the elements selected must be mutually independent [...] at least one of the elements should be non-reusable and non-replicable".

The underlying idea is that the more factors are used during the authentication process, the more confidence a service has that the user is correctly identified. This is the basic explanation provided by the designer to the user to justify a more involved authentication that the user might perceive as cumbersome. However, the user might also need to be told that choosing a weak password is a bad idea even in the case of MFA. Two attacker accomplices who carry out a coordinated attack against the two components of MFA might need to explain their sub-attack to each other to ensure their ultimate success. The analyst who has discovered an attack to the MFA system might need to explain to the designer why the attack succeeded and how to patch it. The analyst/defender might also need to explain to the users why they should, e.g., abandon the use of one of the elements they had been using so far and switch to using another pair of elements; for instance, because the new password that a user has chosen is too weak and thus easily guessable, or because the device on which the user is trying to authenticate does not include a biometric reader. There are thus many things to explain by/to many different stakeholders, which is one of the main reasons why XSec, even in the case of a relatively simple example such as MFA, is a challenging endeavor.

- *What?*
  It is not enough to explain the system in a generic way. First of all, the different stakeholders will need explanations at different levels of detail and with different aims:
  - designers/developers will need an explanation of the desiderata of the client that is detailed enough for them to be able

to realize the system in a satisfactory and secure way (e.g., if the client wants a system that replaces passwords with face recognition);

- non-expert users will need an explanation that increases their confidence and trust, and that also teaches them how to use the system correctly and securely (e.g., if passcodes are used as back-up in case face recognition fails as in the iPhone X, then the user should be made aware that the passcode ought to be strong enough and not guessable such as a date of birth or a phone number, since otherwise an attacker who steals the iPhone X will obviously fail face recognition but the iPhone X will allow him to get access by guessing the passcode);
- analysts will need access to the system's specification or to the system's code in order to be able to create a model to analyze or to be able to generate and execute test cases;
- designers/developers/defenders will need an explanation of a vulnerability and related attacks in order to implement patches or defenses;
- attackers will need an explanation of how to exploit possible vulnerabilities, of why their attacks failed and of the implications of new security techniques on their attack strategies.

Second, several different "things" will need to be explained, including:

- the system and the system model used for design, implementation and analysis, e.g., the model of how MFA actually works;
- the security properties that the system should guarantee. e.g., the authentication provided by MFA can be used as a basis to provide authorization, integrity, confidentiality, non-repudiation and so on;
- the threat model that has been considered by designers, developers and analysts, highlighting, in particular, the fact that a system might be secure against one threat model but insecure against another (e.g., a system might be secure against an outside attacker but insecure against insider attacks) or the fact that the successful MFA of a user won't prevent the system from being attacked when that user turns

out to be malignant and reveals, say, trade secrets of the company he works for;

- the actual vulnerability and related attack that has been discovered will need to be explained to the attacker (especially when the attacker used a tool to search for an attack and now needs to carry out it concretely), along with the costs and benefits of the attack;
- the possible countermeasures for the discovered vulnerability and related attack will need to be explained (along with the vulnerability and the attack) to the honest stakeholders who will need to understand the attack's impact, its risk and the mitigation strategies.

To that end, it will be helpful to answer a number of questions, including:

- What is actually secure? Which parts of the system? Which security properties are guaranteed and for how long? (For instance, authentication is typically granted for a session, which expires after some amount of time.) Which features are insecure and why and how can they be attacked? Are there different levels of security (e.g., for users with different rights)?
- What is the threat model considered? Does it include insiders and outsiders? Who are the potential attackers and what do they want? Why do they want it?
- How does the attack look like and how "difficult" is it? How expensive is it? (It does not make sense to use a one million dollar machine to mount an attack with a loot of a few thousand dollars.) How long will the attack take? (If students try to steal the questions of their next exam but their attack takes so long that they get hold of the questions only after the exam has been given by the professors, then there is actually no point.) This requires reasoning quantitatively about the economics of the attack (including costs, performance, time) but also about the trade-off between attacking and the risk of being discovered.
- Under which assumptions and conditions is the system assumed to be operating securely or has been proved to be secure? For instance, many security analyses (e.g., of protocols or web applications) typically assume a Dolev-Yao-

style attacker [58] who controls the network but cannot break cryptography, which is quite a strong assumption to make as cryptography might indeed be broken (by classical computers and even more so by quantum computers if and when they will be realized in their full capacity); on the other hand, relaxing this assumption and considering an attacker who might be able to break cryptography typically complicates the analysis (the problem is undecidable anyway) and the ability of an analyst to prove security guarantees.

  - What are the legal implications of the explanation? Is the explanation "binding"? This would require, for instance, explaining how the system works and what is expected of the user, possibly including a digital signature to acknowledge the receipt and understanding of the explanation. In case of an attack, this would also require explaining what happened and why, and what countermeasures can be taken (and by whom).

- *Where?*
  We have already addressed the question of which "parts" of the system need to be explained in the "What?" section. Now we focus briefly on the question of where the explanations should be made available. A number of different options are available here, including the following four main ones:

  - One could include the explanations to the users as part of the security/privacy policy, but it is well known that users typically ignore the policy and scroll down as quickly as possible so that they can get on with their interaction with the system.
  - One could completely detach the explanation from the system, e.g., by making it available on a different webpage, but it is unclear to us if and how the relevant stakeholders will be made aware of where to find the explanation and whether they will decide to trust it.
  - One could consider a sort of explainable security as a service, where stakeholders interact with an expert system to obtain and/or provide explanations.
  - One could proceed in the style of proof-carrying code [123], "appending" a possibly digitally signed explanation to the system to achieve a security explanation-carrying-system.

We believe that this is the most promising direction, but it will of course require considerable work to protect the explanation from attacks and actually explain to the stakeholder how they can access it and make use of it.

- *When?*
  We want Explainable Security and we want it now. Jokes aside, the many vulnerabilities that are reported daily, including some of our most widespread and supposedly secure systems (consider, e.g., recent attacks against: TLS; PGP; processors; dropbox, one drive, iCloud and other cloud systems; biometric authentication systems; e-commerce and e-banking systems; e-voting systems, etc.), are witness to the fact that security is indeed difficult to achieve (which is why security has been and still is one of the hottest research topics) but also that in many cases security systems are difficult to explain to the different stakeholders.

  We need to explain security when the system is:

  – designed;
  – implemented;
  – deployed and installed;
  – used;
  – analyzed;
  – attacked;
  – defended;
  – modified.

  and possibly even when the system is decommissioned and replaced, so that the different stakeholders understand why this decision was taken and how the new system will improve over the old one.

  In particular, explanations will need to be defined and provided at design time (when the system is developed) but also at runtime (when the system is running). For the runtime case, think, e.g., of a critical system, critical infrastructure or cyber physical system such as a nuclear power plant in which a supervisor is in charge of setting high/low security levels and of intervening in the case of an ongoing attack to estimate the success chances of the attack, understand its impact on the system and adopt possible countermeasures (see, e.g., [101]). The attack could have disastrous

consequences (e.g., manipulating the SCADA and PLC systems of a power plant as the Stuxnet Worm [63]) or (appear to) be non-threatening as it manipulates sensors and actuators of the system but without bringing them outside of their tolerance zone so that the supervisor actually decides not to intervene.

- *Why?*
  This is easiest question to answer: because all the different stakeholders of a system want it to be secure (well, with the exception of the attacker, of course). Explanations will help increase confidence, trust, transparency, usability and concrete usage (in the sense that users will be more keen to adopt the system), accountability, verifiability and testability.

- *How?*
  As we already remarked above, the different stakeholders will need explanations at different levels of detail and with different aims, and these explanations will need to be comprehensible, timely and accurate (among other properties). The explanations will need to be written in a language (and with a description strategy) suitable for the intended audience, including:

  – natural language (used to produce informal but possibly structured text written in English or any other language understandable by the audience);

  – graphical languages such as explanation trees, attack trees, attack-defense trees, attack graphs, attack patterns, message-sequence charts, use case diagrams ...;

  – formal languages including proofs and plans;

  – games that have been produced as the result of a gamification process to teach users how to interact with a system (although one could actually object that such games often provide for some "unconscious learning" in which the user learns how to interact but without really understanding why).

It should also be investigated whether one learns more by seeing a proof of the security of the system or by being shown an attack against an insecure system. Both are of course useful, but they explain different things in a different way, and they can both be traced back to the question asked by [135] of "how it is possible to communicate the

analysis that experts have made of a security-sensitive system to the public".

It will also be necessary to evaluate security explanations originating from applications of XAI and other areas of computer science, and possibly even social sciences, psychology and other disciplines. Careful subject studies will need to be designed to assess measurement categories such as: a priori measures of explanation quality, user satisfaction, mental model understanding, and user-machine task performance.

Moreover, the explanation processes will themselves have to be designed properly, tested thoroughly and deployed correctly, and it will be useful to investigate the trade-off between such explanation processes and the security threats. We expect that many of the How? questions posed in the specific case of security will actually be answerable by suitably adapting and extending the techniques and tools that have been and are being developed for XAI. Still, we conclude the discussion of the Ws by considering again the claim that we made above that XSec has unique and complex characteristics, and is more challenging than the pioneering research on explanations in security [24]. Let us illustrate this by an example that shows that XSec calls for proper extensions of the research on XAI and for novel investigations.

In Explainable Planning, one of the questions the planner should answer is why things cannot be done and why and how one needs to replan. Similarly, is the relational database system provides the principal with a concise explanation of why the query was rejected and what additional permissions the principal would need to be granted for a successful execution. If one considers a more general security system, however, such an explanation might make the system less secure. This is because the explanation itself might reveal security-sensitive information. For instance, the attacker might not know whether a certain person is indeed a user of the system: trying to login pretending to be that user and being told that the user does not exist, or that the password is wrong, or that the user needs more privileges to be able to carry out some specific action already constitutes a leak of information. 2 Hence, explanations need to be "relativized" and in some cases made less "powerful" by withholding certain details. But a less powerful explanation is essentially an incomplete explanation, which will be ignored or not fully achieve its purpose. The quest for a reasonable trade-off thus makes XSec particularly challenging.

3

SECURITY REQUIREMENTS AND PROTECTION PROFILE IN THE HEALTHCARE DOMAIN

OUTLINE

The EU Cybersecurity Act introduces cybersecurity certification framework for ICT products, services and processes.

Following ENISA's Common Criteria based European candidate cybersecurity certification scheme (EUCC), we provide the Security Problem and identify Security Requirementsof a healthcare specific product through a *Protection Profile*.

We consult ENISA's reports to identify the most impactful assets in healthcare that should be prioritized for certification. We select a sub-category system of Clinical Information Systems, such as PACS for Protection Profile.

Based on five use-cases of PACS, we define the Security Problem (assumptions, organizational security policies, threats) and we elaborate the Security Objectives.

In addition, we conduct a sector specific analysis of challenges and threats in healthcare sector to supplement the PACS specific threats. We detail Security Objectives from the Cybersecurity Act, and we offer a combination of these two elements, the broader scope of threats and objectives, as a baseline for future Protection Profiles of healthcare specific products.

We further provide PACS specific Security Functional Requirements, and we conclude with a guideline for selecting suitable Security Assurance Requirements.

Part of the content of this chapter has been presented in the papers [48, 80].

3.1 NEED OF PROTECTING THE HEALTHCARE SECTOR

The devastating effects of WannaCry [41, 94, 161] in 2017 crippled nearly two million unique devices worldwide [144]. The hardest hit was the *UK's national healthcare infrastructure*, completely locking out users and disrupting systems of nearly 80 healthcare facilities. Among the effected systems were acute medical units [116]. In 2020 one patient seeking emergency treatment for a life-threatening condition died

in Düsseldorf as the systems were paralyzed by a ransomware [45, 70]. Although, the German Federal Office for Information Security warned earlier in January about the critical vulnerability (CVE-2019-19781) for the US software manufacturer Citrix [1, 2], that very vulnerability was exploited in the Düsseldorf University Hospital.

Whilst the healthcare sector has been listed in the EU Commission's tentative list of Critical Infrastructure since 2005 [3], a strong motivation for establishing a level of security within the EU materializes with the Directive on security of network and information systems (NIS Directive)[4] and the EU Cybersecurity Act [5]. These two regulatory acts substantially contribute to molding of the cybersecurity resilience in the EU, particularly their relevance is timely for essential services provided in healthcare.

The EU Cybersecurity Act [137] with the help of European Union Agency for Cybersecurity (ENISA) [6] will operationalize sector specific certification scheme(s). This EU certification framework concentrates on ICT products, services, and processes.

Following the EUCC scheme by ENISA [7], our work concentrates on offering an approach for a sector specific candidate certification scheme for the healthcare sector products.

1 https://www.bsi.bund.de/DE/Presse/Pressemitteilungen/Presse2020/Citrix_Schwachstelle_160120.html
2 https://support.citrix.com/article/CTX267027
3 "Green Paper on a European programme for critical infrastructure protection | EUR-Lex - 52005DC0576," EU Commission, 17-Nov2005. [Online]. Available: https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=celex%3A52005DC0576.
4 "The Directive on security of network and information systems (NIS Directive)," the European Parliament and the Council, 06-Jul-2016. [Online]. Available: https://ec.europa.eu/digital-singlemarket/en/network-and-information-security-nis-directive.
5 "Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 52," the European Parliament and the Council, 17-Apr-2019. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2019/881/oj
6 https://www.enisa.europa.eu/
7 "Cybersecurity Certification: EUCC Candidate Scheme," European Union Agency for Cybersecurity (ENISA), 02-Jul-2020. [Online]. Available:https://www.enisa.europa.eu/publications/cybersecuritycertification-eucc-candidate-scheme

In this chapter, we combine a *top-down* and *bottom-up* approach to define a Security Problem and to unveil a predefined bucket of Threats and Security Objectives for a healthcare system that can serve as a basis for a healthcare product certification scheme. In addition, we explain how to select suitable Security Assurance Requirements.

3.2 CHALLENGES AND OPPORTUNITIES OF CERTIFICATION

Certification of a product, service and a process is a formal evaluation by an independent and accredited body against a defined set of evaluation criteria standards with a final output of issuing a certificate indicating conformance [8].

To build confidence and increase trust in security of product in the EU's internal market, cybersecurity certification is one step forward towards achieving that goal. A major contributing legal act to enforce the roll out of cybersecurity certification EU-wide is the Cybersecurity Act [9], which is not the silver bullet but a steppingstone that introduces a framework upon which certification scheme(s) for different sectors should be built.

The EU Cybersecurity Act establishes the framework for the cybersecurity certification of ICT products, services, and processes.

A *Cybersecurity Certification Scheme* is defined as:

Definition 4. "The comprehensive set of rules, technical requirements, standards and procedures defined at EU level applying to products or services falling under the scope of the specific scheme".

Afterwards, ENISA applied the Cybersecurity Act to release the first EU scheme specifically for ICT products [10]. Next expected step is to provide sector specific certification scheme for products based on the general EU certification scheme. On this path, however, one of the challenges is to elaborate the building blocks of a sector specific cybersecurity certification scheme that has the right level of abstraction and universal applicability so that it can be utilized for multiple products in a singular sector. In order to devise a flexible-enough methodology for a sector specific scheme for products, it would be helpful to look into the core challenges of both the certification ecosystem and the sector

8 https://eur-lex.europa.eu/legalcontent/EN/TXT/HTML/?uri=CELEX:52017PC0477

9 https://eur-lex.europa.eu/eli/reg/2019/881/oj

10 "Cybersecurity Certification: EUCC Candidate Scheme," European Union Agency for Cybersecurity (ENISA), 02-Jul-2020

specific challenges. To date, among the missing binding ingredients for a successful EU-wide certification ecosystem, the industry partners pointed out the following issues [59, 60] [11]:

- Harmonization issue of certificates;
- Costly, tedious and formal characterization of the certification process;
- Lack of security baseline definition;
- Issue of composite certification of several independent systems in interaction;
- Certificates are static, lack agility and don't address patching and software updates and changes in the initial system configurations;
- Lack of common language/vocabulary for certification and labelling;
- Varying ICT landscapes of systems and lifespan, vendors, protocols, and technologies.

3.3 SECTOR SPECIFIC SCHEMES

To be compliant with the official definition of a Certification Scheme, it is needed to identify rules, requirements and standards for sector specific schemes. Rules and procedures are the mandatory elements listed in Art. 54 "*Elements of European cybersecurity certification schemes*" of the Cybersecurity Act [137] that have been already defined by the EU Certification scheme but only for generic ICT products. Technical requirements are security requirements to be identified and we derived them complementing our sector specific analysis with a sectoral assessment carried on following ENISA guidelines [12].

Standards meaningful for specific subdomains are not only identified but we also suggest how to use them, whether to be used as an Organisational Security Policy or to refine Security Functional Requirements.

11 https://www.enisa.europa.eu/publications/challenges-of-securitycertification-in-emerging-ict-environments/

12 "Methodology for Sectoral Cybersecurity Assessments", European Union Agency for Cy- bersecurity (ENISA), 13-Sept-2021

We also define the pillars for the identification and definition of a Specific Scheme [13]: (i) *Rules and procedures*, (ii) *Technical Requirements* and, (iii) *Guidelines on CyberSecurity Onboard Ships Standards*.

3.3.1 *Rules and Procedures*

The following rules and procedures are established in the EU Certification Scheme, and they have been further reviewed by sector expert in Healthcare, to identify the need of sector specific customization and in such case a rationale for the customization is explained.

- Subject matter, scope, covered asset categories;
- Purpose;
- References to standards;
- Assurance levels;
- Conformity self-assessment;
- Evaluation standards, criteria ad methods;
- Information to be supplied by an applicant;
- Rules related to mark and labels condition of use;
- Rules for monitoring compliance;
- Rules for issuing, continuing, renewing certificates;
- Rules related to consequences of non-conformity;
- Rules related to handling vulnerabilities;
- Retention period;
- Correlation with another national scheme;
- Content and format of certificates.

It turned out that most of the elements of the list can be directly inherited as already defined by the EU Certification Scheme for ICT products - except for the rules related to the handling of vulnerabilities and to evaluation criteria and methods.

13 "D2.14 Update - ECHO Cybersecurity Certification Scheme," European Network of Cybersecurity Centres and Competence Hub for Innovation and Operations, 2022.

3.3.2 *Technical Requirements*

Among all the assets identified and classified in categories and subcategories, sector-experts have identified the most critical ones, to be certified in priority.

To build a complete overview of all potential assets, we used ENISA reports and official sector-specific reports to derive a sector-specific Asset Taxonomy. We mark with higher priority such Business Processes that are linked to Services that in turn are supported by the asset categories identified as critical. Security Requirements derivation is based on the sectoral risk assessment that involves a context assessment, *Attack Potential (AP)* assessment and risk-based *Common Security Levels (CSL)* for the asset to be certified. *Security Objectives (SO)* mitigate the risk and one SO is met by implementing a set of Security Controls, i.e., Security Requirements. A *Security Functional Requirement (SFR)* [14] is a requirement, stated in a standardized language, which is meant to contribute to achieving the Security Objectives for a *Target Of Evaluation (TOE)* of the certification.

Security Controls can be technical, operational or organizational. To establish a common set of controls it is suggested [15] to reuse Security Controls from ISO/IEC 27002 [159] or define new control controls which could be employed across sectors.

We provided an initial set of Security Functional Requirements integrating the following catalogues:

- Common Criteria Part 2 [16];
- GDPR [142];
- ISO27001:2022 [159];
- ISA/IEC 62443 [115];
- PIMS specific guidance for ISO27002 [62];
- The Guidelines on Cybersecurity onboard Ships [17].

14 "Common Criteria for Information Technology Security Evaluation: Part 2 - Security func- tional components"

15 "Methodology for Sectoral Cybersecurity Assessments", European Union Agency for Cybersecurity (ENISA), 13-Sept-2021

16 "Common Criteria for Information Technology Security Evaluation: Part 2 - Security functional components"

17 The Guidelines on Cybersecurity onboard Ships", IMO

This should be considered as a baseline to be extended according to other standards and regulation that the Protection Profile writer decide to use to refine Security Controls.

3.3.3 *Guidelines on Cybersecurity Onboard Standards*

According to Common Criteria, a standard can be used as an input to build an Organizational Security Policy (OSP) or a refinement of a security requirement. A standard used as an OSP aims to specify some aspects of the implementation of the ICT product or its operational environment, expressing requirements from national or sectoral regulations.

A standard used as a technical refinement of a security requirement aims to force conformance to the standard as part of the fulfilment of the security requirement, including fully or partially part of the standard's text.

3.4 APPROACH TO SECTOR-SPECIFIC SCHEME DEFINITION

It is worth establishing baseline Protection Profiles to provide a sector-specific scheme. The Common Criteria document allows users to express their security needs given that Protection Profiles are an implementation-independent set of security requirements for categories of ICT products that meet specific consumer needs. Consequently, we have provided a *baseline* for the pieces composing a *Protection Profiles*:

- Security Problem Definition;
- Security Objectives;
- Security Requirements.

In particular, the ENISA guidelines for sectoral risk assessment for certification [18] were integrated to derive the Security Problem Definition. Security Objectives baseline started with the mandatory objectives described in the Cybersecurity Act. They have declined in lower-level objectives according to sector needs and the ISO27001:2022 Control Objectives. Such Security Objectives are achieved by applying Security Controls, i.e., SFRs, which strongly depend on the assurance level established according to the risk level calculated following the ENISA

18 "Methodology for Sectoral Cybersecurity Assessments", European Union Agency for Cybersecurity (ENISA), 13-Sept-2021

guidelines. In the future, the Security Control catalog can be extended and deployed at suitable Security Levels according to their implementation strength. The implementation of the SFRs is assessed through the activities defined in the Security Assurance Requirements (SARs) of Common Criteria Part 3 [19] but also leveraging the Cyber Range technology for the Conformity Assessment dealing with the *AVA_VAN (Assurance Vulnerability Assessment – Vulnerability Analysis)* class of SARs. The table A.2 in the Appendix A, summarizes the main elements needed for a sector-specific scheme and our proposition on how to determine them for a specific sector. In the Table, we assume the following acronym: CCP - Common Criteria Part, SPD - Security Problem Definition, SFR - Security Functional Requirements, SSR - Specific Standards and Regulations, TR - Technical Requirements, CSL - Common Security Levels, OSC - Organisational, Security Controls, SSS - Sector Specific Standards and PR - Privacy Regulations.

3.4.1 *Security Problem Definition*

A Security problem definition needs to report what are the assumptions that are made on the operational environment (physical, personnel, and connectivity) to be able to provide security functionality. Secondly, it describes the set of security rules, procedures, or guidelines for an organisation to identify threats applicable to the Target Of Evaluation. It turned out [20] that assumptions identified as a baseline are mostly transversal and not sector specific. Concerning policies, they can be defined by leveraging a standard as described in Section 3.3.3. Additionally, policies are derived from ISO27001:2022 taking into consideration the Security Controls labeled as “organisational”. A baseline is proposed ready to use with the source of controls. Threats are identified after the Sectoral Risk Assessment (Section 3.4.1.1), which main steps are described in the following section.

3.4.1.1 *Sectoral Risk Assessment*

1. *Context Assessment:* identification of business-critical services, processes and supporting assets according to the sector-specific TOE taxonomy;

19 “Common Criteria for Information Technology Security Evaluation: Part 3 - Security assur- ance components”

20 “D2.14 Update - ECHO Cybersecurity Certification Scheme,” European Network of Cyber- security Centres and Competence Hub for Innovation and Operations, 2022.

2. *Risk Scenarios Assessment:* identification of threats applicable to asset categories through different attack vectors to be exploited by the typical attacker of the sector. For each tuple of asset category-threat it is calculated the Meta Risk Class (MRC) [21] giving a suitable probability and impact for such risk scenario;

3. *Attack Potential Assessment:* calculation according ENISA algorithm of the level of the AP of the most dangerous type of attacker in a specific sector;

4. *Assurance Level Assessment:* selection of the suitable Common Assurance Reference (CAR), defined in such a way that the integration of product certification schemes and Information Security Management System certification is possible. CAR concept based on ISO/IEC 15408's AVA_VAN approach to assurance levels and reuse the associated evaluation methodologies;

5. *Security Level assessment:* for each Control Objective needed to counteract the risk scenarios, more than one Security Control can be applied. If the MRC is at a lower level than the estimated AP, the AP level should determine the CSL that is used for selecting the strength of the controls employed for the treatment of risk.

Table A.1 in the Appendix A shows, relationships among the key elements of the sectoral risk assessment and the traditional certification assurance levels, while Table A.3 in the Appendix A shows security levels presented in relation of the risk level determined during the sectoral risk assessment.

3.4.2 *Security Objectives and risk-based Security Controls*

The Security Objectives intend to solve security problems. They can be traced to TOE and the Operational Environment (OE). The Security Objectives have a relationship with threats in terms of countering and/or mitigating them; they enforce the Organizational Security Policies and uphold the Assumptions.

The SOs detailed in Table A.5 in the Appendix A, derived from the Cybersecurity Act is high-level and were tailored in specific ones for the identified TOE categories: this can help in the subsequent selection of Security Controls/SFRs.

21 "Methodology for Sectoral Cybersecurity Assessments", European Union Agency for Cybersecurity (ENISA), 13-Sept-2021

Moreover, we provide for each subdomain a table highlighting the SFRs/SARs proposed by the EU certification scheme to fulfill Cybersecurity Act SOs, mapping them against the ones detailed to complete sector-specific cases.

We highlight the complete overview of how the threats can be counteracted by Security Objectives and guidance on which SFRs select to implement the SOs.

We provided an added value by detailing which sector-specific standard it can use to refine an SFR Class. Suppose a sector plans to deploy different variants of the asset category depending on its in- tended use and Operational Environment. In that case, the assignment of security and assurance levels should be carried out for each variant. Then, for each subdomain, we have the analysis flow:

Critical Business Process -> Critical Service -> Critical Asset Category -> applicable threats/attack vector -> MRC based on impact and probability of the identified threats -> applicable AP with respect the most dangerous typical attacker of the sector -> suitable CAR depending on the MRC plus the AP level -> suggested CSL for the iden- tified CAR.

The summary of the analysis has been reported for each sector [22] in the Table A.6 in the Appendix A, which shows all stages of the analysis from the beginning (leftmost column) to the end (rightmost column).

The SO are linked to the MRC: during risk assessment, any identified risk scenario is associated with an MRC. In the second step, the sector expert must define Control Objectives/Security Objectives to mitigate such risk. The SFRs are linked to the CSL: A SO is met by implementing a set of Security Controls (SFRs). For each SO, more than one control can be applied. The AP is linked to the suitable CSL: if the MRC is at a lower level than the estimated AP, the AP level should determine the CSL, which is used for selecting the strength of the controls employed for the treatment of risk. CSA requires that certificates reference technical controls (Art. 52.4 of Proposal for a Regulation of the European Parliament and of the Council on ENISA) and that these should be documented for each assurance level.

We have seen how a default relationship between the CAR level and CSL can be established using the MRC and AP as joint reference points.

22 "D2.14 Update - ECHO Cybersecurity Certification Scheme," European Network of Cyber- security Centres and Competence Hub for Innovation and Operations, 2022.

3.4.3 *Conformity Assessment with Cyber Ranges*

The CC has defined the assurance family Vulnerability Assessment, addressing the possibility of exploiting vulnerabilities introduced in the development or the operation of the TOE. This class has therefore been selected as the most representative class to fulfil the requirements of Article 52.1 of Cybersecurity Certification: EUCC Candidate Scheme:

Definition 5. "The assurance level shall be commensurate with the level of the risk associated with the intended use of the ICT product, ICT service or ICT process, in terms of the probability and impact of an incident."

Levelling (from 1 to 5) is based on an increasing rigor of vulnerability analysis by the evaluator and increased levels of attack potential required by an attacker to identify and exploit the potential vulnerabilities. According to EU Certification scheme:

Definition 6. "where no Technical Domain has been defined for a technology of ICT products, associated certificates shall not claim a vulnerability assessment level above the AVA_VAN.3 component. Certification AVA_VAN.3 for ICT products that are not covered by a Technical Domain shall only possible based on a specific Protection Profile defined and certified under EUCC scheme that includes mandatory guidance for the specific evaluation methodology and is annexed to the scheme for this purpose."

Consequently, asset categories for which the suitable MRC, CAR and CSL is above level 3 after the sectoral risk assessment they will need specific evaluation methodology to be accompanied to the PP.

The Conformity Assessment is carried on using a Cyber Range, implementing sector-specific scenarios where the product can be tested to define their conformance. The activities were organized as follows:

1. Drafting of the Security Target of a prototype belonging to MT and HC;
2. Develop/consolidate the prototype with respect Technological Readiness Level 6;
3. Description of Tests following the suitable AVA_VAN class;
4. Creation of a meta narrative for the Demonstration Case;
5. Cyber Range scenario development and testing tools deployment;

6. Testing of the prototypes using the scenario and the testing tools;
7. Creation of the Evaluation Technical Report.

Regarding the tests, benefits using Cyber Ranges are:

- Ease of design and deployment of realistic test environments;
- Ease of backup and restoration of the test environment in case that a test is irreversibly destructive (e.g., ransomware);
- Test environment isolation: with a cyber range no collateral effects on impact on production networks;
- Flexibility and adaptability: different test environment setups and configurations;
- Access control: with a cyber range it's easy to access to the activities of the range from everywhere, for authorised users;
- Orchestration and simulations capabilities: triggering things, generate specific network traffic;
- Content Management System: easy to produce related documentation and reports.

The Conformity Assessment is possible only if the Cyber Ranges are accredited environment labs. Not accredited Cyber Ranges can be used before requesting an official certification test from relevant authorities: the product owner may ask for simulations in an extended cyber range environment.

3.5 CHALLENGES OF CERTIFICATION FOR HEALTHCARE SECTOR

Besides the broad spectrum of sector agnostic challenges, every specific sector (e.g., healthcare, transportation, energy etc.) has its own functional and security challenges that make the elaboration of a sector specific certification scheme even more laborious.

Healthcare sector stands out for several reasons [23]:

23 https://www.enisa.europa.eu/publications/challenges-of-securitycertification-in-emerging-ict-environments/

- This sector falls under the category of essential services according to Article 4 of NIS Directive, therefore it merits special attention in terms of security [24];
- Healthcare nowadays greatly depends on ICT systems and interfaces;
- The ICT connected databases hold and/or exchange patients' sensitive health data for administrative purposes. Health data, under the General Data Protection Regulation (GDPR), is categorized as special category of data that solicits strict processing requirements and secure technical environment [25];
- The lifespan of medical devices affects their security (e.g., some of the Magnetic Resonance Imaging machines are nearly a decade old and new sophisticated vulnerability may arise).

3.5.1 *State-of-the-art Analysis of Healthcare Standards*

The healthcare service providers, as other service providers, should undergo information security evaluations due to the fact that these medical institutions process vast amount of personal and sensitive health data. Information security certifications are mainly based on ISO standards from series ISO 27000 [89] and ISO 20000 [88].

One of the most important standards covering several general aspects of information security is ISO 27001 [90]. It specifies the requirements for establishing, implementing, maintaining and continually improving an information security management system within the context organisation, along with the generic requirements for the assessment and treatment of information security risks tailored to the needs of any type of organisation. This standard can be complemented with the ISO/IEC 27002 [91] that guides organisations on selection, implementation and management of information security controls.

Within the ISO family, the ISO 27799 [86] is designed for information security management for health informatics, but it heavily relies on ISO/IEC 27002. The ISO 27799 determines guidelines to support the interpretation and implementation of the information security con-

24 https://ec.europa.eu/digital-singlemarket/en/network-and-information-security-nis-directive

25 https://eur-lex.europa.eu/eli/reg/2016/679/oj

trols under the 27002 in health informatics. The ISO 27779 standard can be applied by healthcare institutions and other possessors of health information to satisfy the requirements of confidentiality, integrity and availability of personal health information in their care. It applies to processing of any data format (e.g., words, sound and video recordings, digits, medical images), to any medium for storing (e.g., printing or writing on a paper or storing electronically) and transmitting (e.g., fax, computer networks).

Another generally applicable standard is the ISO 9001 [87]. This standard helps to govern the implementation of a Quality Management System (QMS) in companies, aiming at verifying customer satisfaction with the products and services provided, as well as the continuous improvement of company performance, enabling the certified company to assure its customers that the quality of its goods and services is maintained and improved over time. A similar Canadian regulation, the Canadian Medical Devices Conformity Assessment System (CMDCAS), requires the medical devices to be designed and manufactured according to a registered Quality Management System (QMS).

The ISO 62304 standard [82] provides a framework for safe design and maintenance of software for medical devices, and its requirements apply throughout the life cycle process, sub-activities and tasks.

The ISO 13485 [84] targets medical instrument(s) and machine(s) that are intended for use in the diagnosis, prevention and treatment of diseases or other medical conditions. This standard is designed to be used by organisations involved in the design, production, installation and servicing of such medical devices, as well as by the certification bodies, to help them with auditing processes in these organisations.

The ISO 14971 [85] standard helps medical device manufacturers identify the hazards associated with medical devices. It specifies terminology, processes for managing device risks, including the software itself and the medical diagnostic devices used.

The Medical Device Directive (MDD) 93/42/EEC [49] specifies the requirements for device manufacturers and importers for meeting the CE mark to legally market or sell their devices in the EU. There are specific requirements for devices depending on the classification and intended use of the device. In addition, to market access requirements, in

the healthcare domain, the In Vitro Diagnostic Medical Devices (IVDD) are subject to regulation. The IVDD 98/79/EC [56] regulates a subset of medical products, their market access, and their use.

In a different format an initiative is formed under the International Medical Device Regulators Forum (IMDRF) at an international level [83]. Representatives of medical devices regulatory authorities around the world come together to set standard requirements for auditing organisations that perform certification on the Quality Management Systems of medical device manufacturers. Similarly, the Medical Device Single Audit Program (MDSAP) represents requirements that apply to regulatory authorities as well as to third-party organisations performing this type of audit.

In addition, to all these standards, the IT Health Check (ITHC) [92] provides assurance that the organization's external and internal systems are protected from unauthorized access and/or change, and they do not provide an unauthorized entry point into systems that consume Public Services Network (PSN) services. A follow-up to unauthorized access to health data, the HIPAA [76] compliance standard can serve as an example. These regulations allow physicians or other health care professionals to share information directly with parties related to the patient (e.g., spouse and other family members, and/or friends).

3.5.2 *HIPAA: Health Insurance Portability and Accountability Act regulamentation*

This section describes HIPAA regulations [125, 173] and why it is essential to consider them, especially in the industrial setting. HIPAA is defined as follows [26]:

Definition 7. *"A U.S. federal law that defines requirements for the handling of individuals' protected health information. HIPAA compliance is regulated by the Department of Health and Human Services (HHS), and the Office for Civil Rights (OCR) is responsible for enforcing it. HIPAA compliance must be an integral part of the corporate culture of every entity operating in the health care industry to ensure the privacy, security, and integrity of protected health data".*

Compliance with the U.S. HIPAA regulations requires companies that process PHI data to adopt strict physical, network and procedural

26 https://www.cdc.gov/phlp/publications/topic/hipaa.html

security measures within company buildings. In order to comply with HIPAA regulations, anyone who provides treatment or provides financial or insurance services for medical care must comply at any level. Not only the so-called covered entities just mentioned. But all companies in business that have access to patients' health data, operating, providing support in treatments or health care payments, must also comply with HIPAA regulations, which also binds any subcontracting or otherwise professionally related partner companies to compliance.

THE HIPAA was passed by the U.S. Congress and became law with the signature of President Bill Clinton in 1996. It was enacted primarily to:

- Modernize the flow of health information;
- Regulate the processing of Personally Identifiable Information (PII) by health and health insurance companies to prevent theft and fraud;
- Overcome limitations in medical insurance coverage when it comes to portability and coverage of people with pre-existing medical conditions;
- HIPAA mandated the adoption of national standards to protect sensitive patient health information from disclosure without the patient's knowledge or consent [27].

To implement this legislation, HHS, the U.S. Department of Health and Human Services, published the list of rules to be complied with regarding confidentiality, which is called the HIPAA Privacy Rule [28].

The Privacy Rule contains 12 exceptions that provide for the sharing, among different entities, of health data without patient consent. It include:

- Victims of domestic violence or assault;
- Judicial and administrative proceedings;
- Organ explantation from cadavers, cornea or tissue donations;
- Workers' compensation.

27 https://www.hipaaguide.net
28 https://www.cdc.gov/nhsn/hipaa/index.html

Another key element of HIPAA is the so-called *Security Rule*, which is part of the Privacy Rule. This subgroup includes all those personal health data that any included entity generates, receives, retains or transmits in electronic format. Basic principles of the Security Rule are:

- Ensure the confidentiality, integrity, and availability of all protected health information in electronic format;
- Identify possible threats in order to safeguard confidential information;
- Protect data from possible disclosure or misuse;
- Certify the compliance of personnel who at any level come into contact with such data.

Protected Health Information (PHI) is any demographic information that can identify a patient or client of a HIPAA-covered entity. A few examples of PHI include names, addresses, phone numbers, social security numbers, medical records, financial information, and close-up facial photos.

3.5.3 *Opportunities of Certification for Healthcare Sector*

The state-of-the-art analysis showcases the vastness of the landscape of standards and regulations. It becomes evident that it is a tremendous effort to try certifying the healthcare sector as one due to the immense diversification of healthcare systems, components, infrastructures and medical devices. However, in order to build trust in these systems and components used in the healthcare sector, some acceptable level of security and privacy should be achieved. The goal of certification is to help reduce the potential societal risk that would have been otherwise overlooked without establishing baseline security for products, with the main approach of targeting first and foremost the critical services provided in the healthcare sector [29].

The healthcare systems' dependency on ICTs creates a vector for potential attacks or failures that can result in a much greater impact to the sectors' constituencies in contrast to other non-essential services. According to ENISA's study, picture archiving and control systems (e.g., Picture Archiving and Communication System), that fall under

29 "European Cybersecurity Certification: Challenges ahead for the roll-out of the Cybersecurity Act," ECSO, Dec. 2020.

the category of Clinical Information Systems, are listed among the most impactful equipment within the healthcare that should be prioritised for certification [30]. Safety and availability of services are important factors for hospitals, because of the importance of the integrity of health data and for the need of that data to remain private. Personal health data is valuable for various reasons: threat actors may be interested in having access to patient health data for a myriad of malicious purposes.

Fundamentally, the healthcare sector specific challenges can be credible for shaping the sector specific certification scheme bottom-up. In the meantime, to be able to apply a sector specific scheme to various products within the sector, the approach for the methodology of product evaluation should be based on a principle of re-usability and costeffectiveness. Any scheme developed should not jeopardize the safety of patients and their health data[14, 31].

By analysing this wide-ranging spectrum of standards and regulations, we arrive to an understanding that Security Problem definition is a feasible approach. This approach would allow us to describe the asset for evaluation and help scope the product boundaries, as well as help map the potential threats based on those boundaries defined. This approach will provide flexibility for security evaluation for any product in this specific sector. Additionally, Security Problem definition would allow elaboration of relevant Security Objectives which would lead to Security Requirements.

An opportunity that forms here, for this universally applicable scheme, is to identify a healthcare specific asset as a target of evaluation, detail the Security Problem and specific Security Requirements[14] to showcase this approach.

In this regard ENISA's guidance on the methodology for establishing the cybersecurity certification framework at EU level is pertinent. The ENISA's candidate cybersecurity certification schemes of ICT product(s), services and processes is based on Common Criteria (CC)[10] with a rationale that the CC have proven its efficiency previously with

30 "Challenges of security certification in emerging ICT environments," ENISA, 06-Feb-2017. [Online]. Available: https://www.enisa.europa.eu/publications/challenges-of-securitycertification-in-emerging-ict-environments/

31 "ICT security certification opportunities in the healthcare sector," ENISA, 31-Jan-2019. [Online]. Available: https://www.enisa.europa.eu/publications/healthcare-certification.

regards to certifying chips and smartcards.

Based on ENISA's EUCC scheme[10], we use the Protection Profiles to elaborate the Security Problem and derive Security Requirements for the PACS. We offer an initial Threat landscape and, concurrently, the Security Objectives specific for PACS. We then supplement the Threats, from the sector based analysis of threats and challenges, and Security Objectives, by detailing the Security Objectives from the Cybersecurity Act. The Threats and the Security Objectives with the broader scope can be potentially applied to other products in the healthcare sector. This system falls under the definition of the Article 2 of the Cybersecurity Act, that defines the product as *product* means an element or a group of elements of a network or information system[5].

3.6 USE CASE: PICTURE ARCHIVING AND CONTROL SYSTEMS

This section presents a healthcare use case defining the methodology for identifying security requirements through the protection profile.

Protection Profile (PP) is an "implementation independent" set of Security Requirements for a category of ICT product that meets specific consumer needs [32]. Identification of Security Requirements is reached through the steps described in this section (through A to E) and defined by Common Criteria methodology of building Protection Profile. In particular, the purpose of the PP is to state a Security Problem (SP) for a given system or a product category and specify Security Requirements to solve a problem. The SP is a formal statement defining the nature and the scope of the security that TOE is intended to address.

Although flexible structuring of PPs content is allowed, they however have an imperative content outline, that should record the description of the TOE; Conformance Claim; Security Problem Definition (Threats, Organisational Security Policies, Assumptions); Security Objectives; Definition of the Extended Components; Security Requirements (Security Functional and Security Assurance Requirements).

The PPs are not designed to have detailed security specifications but they describe the security needs at a high level of abstraction. Its purpose is to specify generic security evaluation criteria. The PPs should

32 "Common Criteria for Information Technology Security Evaluation Part 1: Introduction and general model," Common Criteria, Apr. 2017.

be used where it is necessary to define a common set of security requirements that will help the consumer, the IT developer and/or the regulatory entity to obtain, use and/or produce the evaluated information technology in accordance with the baseline Security Requirements. The identification of Security Requirements will contribute to achieving the Security Objectives for the TOE. All these characteristics of the Protection Profiles will help devise the Security Problem that, potentially, can meet both the generic and specific challenges related to certification and help develop security requirements for the healthcare sector.

3.6.1 *Target of Evaluation*

Under the Common Criteria, the scope of the target(s) for security evaluation is rather flexible: it may be an IT product or a part of an IT product, it may also be a set of IT products, or a combination of these [33]. For the purpose of our work, the TOE is Picture Archiving and Communication System with specific use-cases in scope. PACSs are nowadays a backbone in the effective management of imaging departments in the hospitals, where a large number of medical images and reports are being transmitted digitally on a daily basis.

The PACS is a complex and a hybrid system, comprised of both the software and the hardware. The system serves the purpose of transferring, storing and displaying medical images and reports. The PACS are integrated with the Radiology Information System (RIS) and Hospital Information System (HIS) [34], using the Digital Imaging and Communications in Medicine (DICOM) standard [35]. PACS are interoperable [108] systems capable of handling numerous medical imaging instruments: Computed Tomography (CT), Magnetic Resonance (MR), DR, Mammography (MG), Ultrasound (US), X-ray angiography, Endoscopy (ES), Computed Radiography (CR) and other types of imaging systems.

33 "Common Criteria for Information Technology Security Evaluation Part 1: Introduction and general model," Common Criteria, Apr. 2017.

34 H. Khaleel, R. Wirza, and D. Zamrin, "Components and implementation of a picture archiving and communication system in a prototype application," Reports Med. Imaging, vol. Volume 12, pp. 1–8, Dec. 2018

35 "Standard: PS3.21 DICOM PS3.21 2020 e-Transformations between DICOM and other Representations," National Electrical Manufacturers Association, 2020. [Online]. Available: http://dicom.nema.org/medical/dicom/current/output/pdf/part21.pdf.

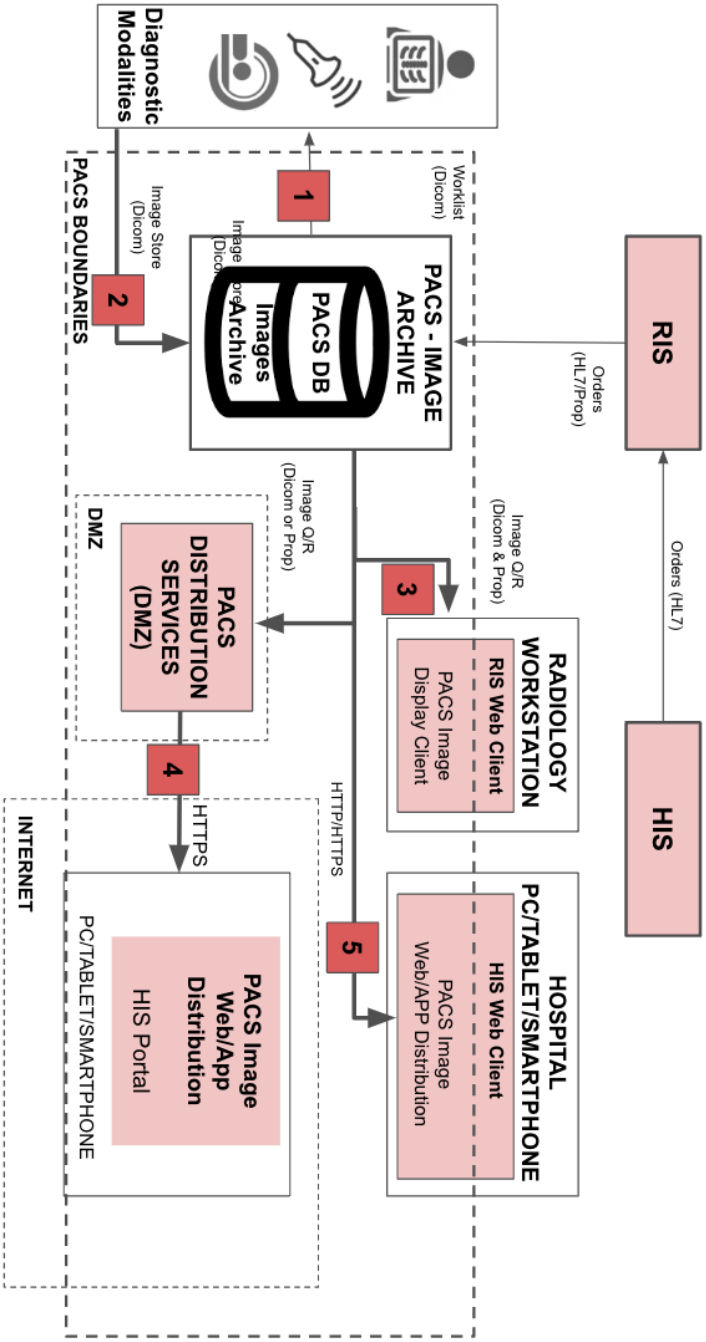

Figure 3.1: Scope of the Target of Evaluation (ToE) and Five Use-cases.

The five Use-cases are as follows:

1. *Worklist to Modalities:* After having received the information from the HIS, the RIS publishes a list of patients' demographics data and examination details to the Modalities (DICOM Worklist). All information passes through the hospital network in clear text.

2. *Image Store:* When the examination is executed, the Modalities acquire the images and send them (DICOM Study) with patient demographics to PACS (DICOM Store). The PACS stores the DICOM Study. The DICOM Study allows a patient to have *n* number (from 1 to *n*) of studies (examinations or other procedures). Each Study consists of N number of series. A series generally refers to a specific data type (modality), or the position of a patient on the acquisition device. Each series contains n number of DICOM object instances (mainly images, but also reports, signal objects, etc.). All of this information is contained in each DICOM object of a study. Therefore, if a study is performed on a patient, containing, for instance, of 2 series, all of the instances will contain both the patient and the study information. The instances will also point to the series they are located in, they will also provide information about itself.

3. *Image Display on Radiology Workstation:* The Radiologist, through a reporting worklist on the RIS, selects a patient examination that it wished to report. Then the RIS calls the PACS Workstation component with Patient/Study information. The PACS opens the related images and information (the DICOM Study). The Radiologist can then read the images on the PACS Workstation, work and writes, signs the Report on the RIS. After the radiologist has finalized the report, the RIS sends the signed Report to HIS, together with a reference, a unique global identifier to the DICOM Study (StudyUID).

4. *Image Display on Web Browser or App (Hospital):* The clinical reports are now available for the HIS. The Ward/Intensive Care Unit (ICU)/Emergency Room (ER) clinicians can view them; the StudyUID reference is the identifier through which the HIS can open the PACS Web Viewer.

5. *Image Display on Web Browser or App (Remote):* This use case deals with access through a Web Portal or a Mobile Service, of the

DICOM studies. The HIS can publish Reports and show DICOM Studies, by invoking PACS Web Viewer on a specific DICOM Study. It is then possible to access them from the outside, thus improving to Teleradiology/Second Opinion networks for teleconsultation.

3.6.2 *Conformance Claim*

The Conformance Claim indicates three main elements: (i) to which version of the Common Criteria the TOE or the PP claim conformance; (ii) to which Security Functional Requirements it conforms and, (iii) describes to which Security Assurance Requirements it conforms.

3.6.3 *Security Problem Definition*

The Security Problem is comprised of three elements. We commence defining the Security Problem by describing the Threats that the TOE is expected to address, assumptions about the operational environment, and any relevant OSP that the TOE is expected to enforce.

Then, a Security problem definition needs to report what are the assumptions that are made on the operational environment (physical, personnel and connectivity) in order to be able to provide security functionality. Secondly, it describes the set of security rules, procedures, or guidelines for an organisation to end up with the identification of threats applicable to the Target Of Evaluation. It turned out that assumptions identified as a baseline are mostly transversal and not sector specific. Additionally, policies are derived from ISO27001:2022 taking into consideration the Security Controls labeled as "*organisational*". A baseline is proposed ready to use with the source of controls. Threats are identified after the sectoral risk assessment, which main steps are described in the following paragraph.

3.6.3.1 *Security Problem in Healthcare Domain*

We formulate three OSPs, eight Assumptions and six Threats. We mark the Threats with $T.$, the Organisational Security Policies are marked with P, and the Assumptions are marked with A. The Objectives for the Operational Environment are marked with OE.

We use the five use-cases marked on Figure 3.1 to list potential Threats that the TOE might counter. These Threats are PACS specific.

We follow the methodology of the Protection Profile development and define the Security Objectives based on the identified Threats.

- For *Organisational Security Policies*, we list the following:
  P. USER; P. ROLES; P. ACCOUNTABILITY.
- For *Assumptions*, we list the following:
  A.RIS; A.HIS; A.DIAGNOSTIC_MODALITIES; A.UPS; A.PHYSICAL; A.EXTERNAL_COMMUNICATION; A.PROPER_USER; A.PROPER_ADMIN.
- For *PACS specific Threats*, we identified the following:
  T.DATA_MANIPULATION; T.DATA_LOSS; T.SERVICE_DISRUPTION; T.DATA_DISCLOSURE; T. ILLEGAL/UNAUTHORIZED_ACCESS; T.DATA_THEFT.
- For *PACS specific Security Objectives*, we define the following:
  O.SECURE_COMMUNICATIONS; O.SECURE_STORAGE_AND_BACKUP; O.AUTHORIZED_ACCESS_AND_PROCESSING; O. EVENT_MONITORING; O.SERVICE_RESILIENCE; O. HW_MAINTENANCE; O.SW_MAINTENANCE.

3.6.4 *Security Objectives*

The Security Objectives intend to solve security problem, they can be traced to TOE and to the operational environment. The Security Objectives have relationship with Threats in terms of countering and/or mitigating them; they enforce the OSP and uphold the Assumptions. There are two paths for tracing the security objectives:

1. Security Objectives of the TOE trace back to Threats and OSPs;
2. Security Objectives for the Operational Environment trace back to Threats, OSPs and Assumptions.

The Figure 3.2, showcases these relationships.

To enhance this mapping and to be able to provide a broader scope of threats for the healthcare sector, we use sector-based analysis of threats and challenges conducted within the ECHO Project [51–53] and the ENISA's report [36] to complement this initial bucket of threats, and offer a baseline of Threats for the healthcare specific scheme security problem definition. To ensure that these newly supplemental Threats gain Security Objectives, we detail these objectives from the Cybersecurity

36 "Procurement Guidelines for Cybersecurity in Hospitals," ENISA, 24-Feb-2020. [Online]. Available: https://www.enisa.europa.eu/publications/good-practices-for-thesecurity-of-healthcare-services.

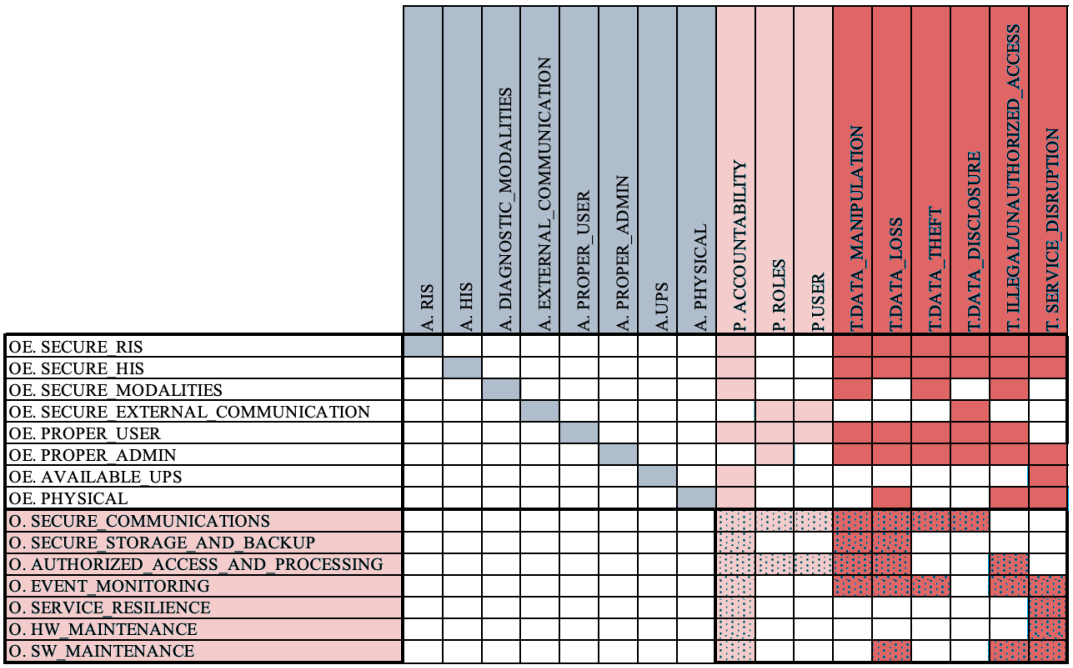

Figure 3.2: Tracing between Security Problem Definition and Security Objectives.

Act. Figure 3.3 summarizes both PACS specific Threats and Security Objectives, as well as the baseline Threats and Security Objectives that we offer for further consideration when building a healthcare specific certification scheme with the use of Protection Profile.

The supplemental Threats are:
T.DEVICE_THEFT; T.IDENTITY_THEFT; T.UNSECURE_COMMUNICATION; T.SOFTWARE_SYSTEM_FAILURE; T.SUPPLY_CHAIN_FAILURE; T.INSIDER_THREAT; T.PHYSICAL_THREAT; T.SECURITY_DEFAULT_AND _DESIGN_FAILURE; T.NO_REGULATORY_STANDARD_COMPLIANCE; T.HUMAN_ERROR; T.HUMAN_INJURY.

The Security Objectives detailed from the Cybersecurity Act are:
O.DATA_CONFIDENTIALITY; O.DATA_AVAILABILITY; O.DATA_INTEGRITY; O.ACCESS_CONTROL; O.VULNARABILITIES_ANALYSIS; O.EVENT_LOGGING; O.LOG_MANAGEMENT; O.VULNARABILITY_MANAGEMENT; O.BUSINESS_CONTINUITY; O.SECURITY_BY_DESIGN_AND_DEFAULT; O.SECURE_SOFTWARE_DEVELOPMENT_AND_MAINTENANCE.

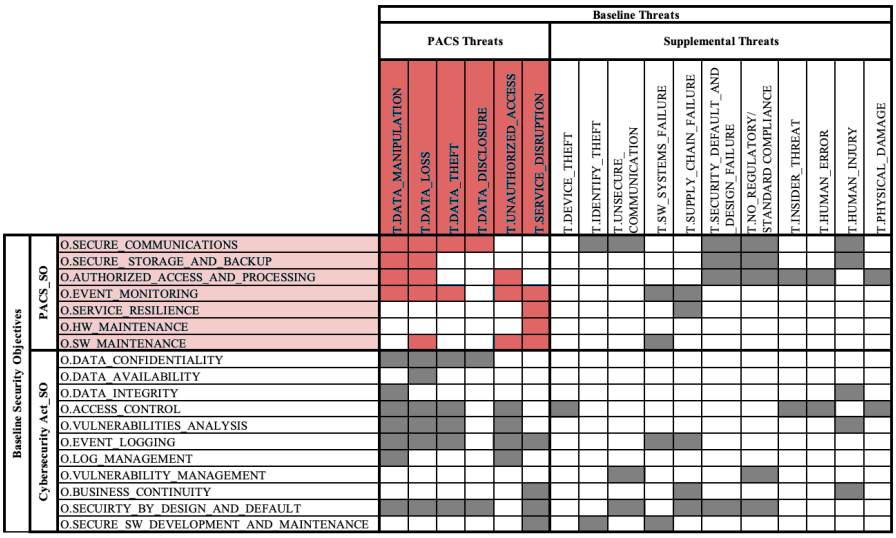

Figure 3.3: The Baseline Threats and the Security Objectives for a Healthcare Specific Certification Scheme.

3.6.5 *Security Requirements*

The goal for conducting security evaluation of the TOE is to ensure that the determined SFR are enforced on the TOE and its resources [46]. The SFRs may impose various security policies, each of them must specify scope of control (defining the subjects, objects, resources or information, and operations to which it applies). Each SFR is manifested via classes, families, and components.

For PACS and for the given scope of Threats and Security Objectives, we have applied the following SFRs as shown on Figure A.1 in the Appendix A: FDP: User Data Protection; FAU: Security Audit; FPT: Protection of the TOE; FIA: Identification and Authentication; FMT: Security Management; FTA: TOE Access; FTP: Trusted Path/Channel; FRU: Resources Utilisation.

As the healthcare sector processes primarily personal sensitive data, maintaining privacy is paramount. We use this opportunity to develop one of the elements of the PP, the Definition of the Extended Components. We develop a one distinct SFR Class named - FPP: Personal Data Protection, along with the families and customized components.

The Security Requirements for the Protection Profiles have two categories:

1. Security Functional Requirements (SFR);

2. Security Assurance Requirements (SAR).

The SARs are the descriptions of how TOE is being evaluated, and they are structured similarly in a hierarchical way by class, family and component. Before any of the SARs can be selected, the Evaluation Assurance Level (EAL) should be defined.

The Article 52 of the Cybersecurity Act mentions that each certification scheme should provide the assurance requirements matching to its respective assurance level.

The Cybersecurity Act defined three levels of assurance: *Basic*, *Substantial* and *High*. The following definition applies to these levels:

- If cybersecurity certificate or statement of conformity refers to assurance level *Basic*, the evaluation activities should include at least the review of technical documentation.
- If cybersecurity certificate refers to assurance level *Substantial*, the evaluation activities should include at least the review to showcase that the publicly known vulnerabilities are absent and conduct a testing to demonstrate that the SFRs are correctly implemented.
- If cybersecurity certificate refers to assurance level *High*, the evaluation activities should include at least a review to showcase that the publicly known vulnerabilities are absent, a test to demonstrate that the SFRs are implemented at the state of the art, and a penetration testing.

The selection of an assurance level may seem to be a straightforward process, however, in practice it is a challenging activity as several documents should be consulted simultaneously:

- ENISA: The EUCC scheme or products covers two assurance levels out of the three: the Substantial and the High. In the meantime, Common Criteria itself defines seven evaluation assurance levels and each of these levels have their minimum corresponding activities [37].

37 "Cybersecurity Certification: EUCC Candidate Scheme", European Union Agency for Cybersecurity (ENISA), 02-Jul-2020. [Online]. Available: https://www.enisa.europa.eu/publications/cybersecuritycertification-eucc-candidate-scheme.

- ENISA: Our next step is to follow the mapping of the assurance levels between the Cybersecurity Act and the Common Criteria (Part 3). ENISA notes that the selection of the assurance levels of the Cybersecurity Act should be based on the assurance components of a specific assurance class, the AVA: Vulnerability Assessment class (AVA_VAN), defined by Common Criteria Part 3. According to this formulation, the mapping of the assurance levels between these two documents are represented in Table A.4 in the Appendix A.4.

- ENISA: ENISA defines, that the products that do not fall under the "Technical Domains", such as the Smart Cards (and similar devices) and the Hardware Devices with Security Boxes, cannot apply SAR components AVA_VAN.4 and AVA_VAN.5.

- ENISA: Further, we exclude the entire assurance level High (including AVA_VAN.3), as ENISA suggest that assurance claim for level High originates from the authorization of a Governmental agency, leaving us with level Substantial with corresponding AVA_VAN components.

- Common Criteria Part 3: We are left with AVA_VAN.1 and AVA_VAN.2. We refer to Common Criteria Part 3, to determine which of these two components is more appropriate for our TOE. In the description of AVA_VAN.1 it is stated that the evaluator shall perform a search of public domain sources to identify potential vulnerabilities in the TOE. In case of AVA_VAN.2 the evaluator shall perform an independent vulnerability analysis of the TOE using guidance documentation, functional specification, TOE design and security architecture description to identify potential vulnerabilities in the TOE. Since PACS is rather more complex product, the use of its specific design and architecture information is definitely required. For that reason, we select AVA_VAN.2.

- ENISA: It is stated in the ENISA's guidance, that based on the selection of the AVA_VAN, the first EAL should apply along with its all dependencies of components that are associated with the selected AVA_VAN. In our case, the first EAL corresponding to AVA_VAN.2 would be the EAL 2 (refer to Table A.4 in the Appendix A.4. Evaluation assurance level summary of Common Criteria Part 3, p. 32) [47].

- Common Criteria Part 3: The dependencies marked by the Common Criteria with respect to AVA_VAN.2 are selected from the assurance classes of ADV (Development) and AGD (Guidance) and the components are the following: ADV_ARC.1 Security Architecture Description; ADV_FSP.2 Security-enforcing functional specification; ADV_TDS.1 Basic Design; AGD_PRE.1 Preparative procedures; AGD_OPE.1 Operational user guidance.

- Common Criteria Part 3: Based on the selected EAL, ENISA indicates what are the other further applicable SARs, apart from the direct dependencies indicated by Common Criteria Part 3. The additional classes are ALC (Life-cycle), ATE (Tests), ASE (ST Evaluation), APE (PP Evaluation) and the selected components are the following: ALC_CMC.2 Use of a CM system; ALC_CMS.2 Parts of the TOE CM coverage; ALC_DEL.1 Delivery procedures; ATE_COV.1 Evidence of Coverage; ATE_FUN.1 Functional Testing; ATE_IND.2 Independent Testing – Sample: ASE_TSS.1 TOE Summary Specification and these following SARs that refer to Security Target (ST), and not to Protection Profile: ASE_CCL.1 Conformance Claims; ASE_ECD.1 Extended Components Definition; ASE_INT.1 ST introduction; ASE_OBJ.2 Security Objectives; ASE_REQ.2 Derived Security Requirements; ASE_SPD.1 Security Problem Definition. Therefore, we find equivalent components that apply to PP, which are the following: APE_CCL.1; APE_ECD.1; APE_INT.1; APE_OBJ.2; APE_REQ.2; APE_SPD.1; APE_TSS.1.

Part II

THE SHOWCASE

Namely, my efforts, thus the successes and the failures. In a word, my results. The Showcase is here to respond to "*Where have you been to? Which have been your path, your climbs and descents?*".

4

A PROPOSAL FOR SECURE PATIENT MONITORING USING AI TECHNIQUES

OUTLINE

The reference scenario in which, using CP, we can monitor the compliance of a patient's clinical treatment provides ample room for value creation to support clinical trials and healthcare governance in response to the urgent need for:

1. Design interventions and policies to optimize treatment, improve prevention, epidemiological surveillance, and expenditure restructuring, especially concerning the growing incidence of high-risk and high-cost patients;
2. Accelerate the diffusion of CPs, so exploiting its benefits in terms of quality and continuity of care, prescriptive appropriateness, and containment of health spending;
3. Mitigate the cognitive and managerial overload of the healthcare staff in the treatment of highly complex patients.

We propose implementing a remote patient monitoring infrastructure using an edge computing approach. We design an edge architecture that we propose to use for patient monitoring.

We propose to equip the monitoring infrastructure with an anomaly detection module capable of identifying whether or not there is an intrusion during the transmission of parameters captured by the sensors. Using a running example, we illustrate the different types of attacks that can occur. We introduce an example of ECG monitoring and build on this data interpretation paradigm.

Furthermore, we propose an approach that, by exploiting process mining techniques, can measure the deviation of the patient's clinical path from that defined by the doctor.

Part of the content of this chapter has been presented in the papers [11, 13, 15, 16, 132]. The contribution [132] received the Best Paper Award at 2022 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA).

It is possible to identify four main problems related to patient care information systems:

- *P1*: Fragmentation of IT support in healthcare;
- *P2*: Expressive limitations of formalisms;
- *P3*: Information overload for physicians and care managers;
- *P4*: Low exploitation of clinical data.

In the following subsections are detail the problems mentioned above.

FRAGMENTATION OF IT SUPPORT IN HEALTHCARE. The adoption of CPs in health facilities and districts is experiencing rapid diffusion in Italy as a tool for rationalizing clinical guidelines and organizing the treatment of complex pathologies and chronic diseases such as diabetes, Chronic Obstructive Pulmonary Disease (COPD), rheumatoid arthritis, and heart failure. Some regions are implementing CPs as an extension of the Electronic Health Record (EHR), and there are many CP automation initiatives by individual local health agencies.

The prevailing direction is a *low automation* approach of the CP, e.g., support in the form of an *open* electronic document (Care Coordination) [93] shared among the clinicians involved in the patient's journey and coordinated/managed by the so-called case manager.

From this point of view, the electronic tracking document of a CP arises as an additional *informational debt* along with the traditional clinical documents such as the specialist report, the first aid report, and the hospital discharge letter.

However, the real potential of CPs, compared to clinical guidelines (which have a substantially descriptive/narrative nature), is in the broader possibilities offered by high automation.

The main technological barriers limiting the diffusion of IT tools for care continuity are the plurality of poorly interoperable software tools supporting clinical workflows, as well as the need to involve multi-disciplinary specialists and have many contacts with the patients when developing the systems assisting in the treatment of diseases [104].

EXPRESSIVE LIMITATIONS OF FORMALISMS. According to the definition of Edward Shortliffe [154]:

> "A learning health system is a system that is capable both of assuring that every decision is made with complete information and ensuring that every care instance can contribute a deeper understanding of care for individuals and populations".

Clinical/health workflows, in general, and CP, in particular, show a reduced level of causality. They do not guarantee to obtain the same outcome from the identical repetition of a very complex scheme of actions. Still, on the contrary, they are based on a principle of specificity of each individual "case," the uniqueness of the symptomatic, diagnostic, prognostic, etiological, and therapeutic response of each patient [71, 111].

A critical success factor for CP automation is the availability of a CP expression notation that overcomes the rigidity of process automation formalisms such as example, Business Process Model and Notation (BPMN) [69]. This approach incorporates design uncertainty, late choice, exception, and non-motivated adherence, as shown in Figure 4.1.

INFORMATION OVERLOAD FOR PHYSICIANS. The progressive spread of high-risk and high-cost populations is creating new challenges for the health government, both locally and regionally, in an already complex scenario of a constant reduction of public funding for health spending.

In this scenario, it would help the availability of data-driven clinical intelligence tools that allow statistical analyses based not only on the services provided and hospitalization but also on the details of the single treatment or clinical observation. Through this, evidence can be gathered, and detailed analyses can be conducted, such as latent patterns in a pathology's demographic and/or temporal distribution or hidden risk factors arising from syndromes or pathological conditions of high social importance [129].

Even though most of the data-generating streams have already been digitized, and despite the existence of technological and regulatory devices to standardize and centralise storage, such as the EHR and Pathology Network, clinical and health data are often

confined to silos and confined to forms that make them difficult to aggregate, access, and analyze.

The digital data containing patient-structured data is displayed in a context that enables the analysis of data collected in hospital discharge letters and reports. Text documents must be treated with semantic classifiers, i.e., with techniques that recognize and extract information, such as diagnoses, measurements, observations, and therapeutic indications.

On the other hand, the wide availability of clinical data relating to a sample of patients (considering all the clinical specialties and the temporal dimension) identifies new problems of:

- Conceptual modeling of the person (in the clinical sense);
- Population (in epidemiological and health governance);
- Technological management of masses of data that exceed the storage and processing capacity of traditional computer architectures.

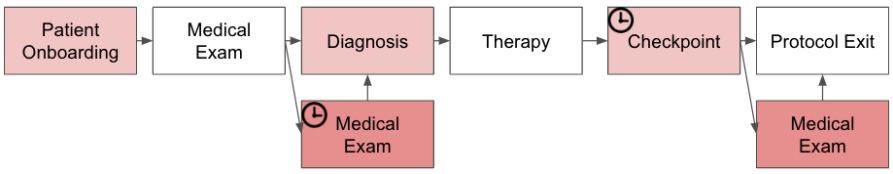

Figure 4.1: Example of healthcare process using BPMN approach.

Remote monitoring of a patients, uses health data transmitted electronically. Therefore, it is a healthcare delivery model that uses technology to connect patients and caregivers/professionals outside the clinic, doctor's office, or hospital. The use of modern equipment integrated with monitoring apps produces an overall positive impact on patient care and also reduces organisational costs. Remote Monitoring has become paramount to managing patients' clinical conditions and improving their course of treatment. Remote Monitoring facilitates the sharing of patient status information in real-time. However, it is of paramount importance to keep the sharing of this information secure. Tampering with it would be detrimental and even cause irreparable damage to the patient's health.

4.1 DESIGN OF AN ARCHITECTURE FOR REMOTE PATIENT MONITORING

Massive volumes of data are being generated at the edges of networks due to the proliferation of connected devices and IoT applications. Data is often processed locally due to intermittent connectivity and local quality of service expectations [153].

In edge computing, computation is done at the edge of the network, and the limitations associated with computation at the extreme periphery of the network are increasingly recognized [43].

According to NIST:

"Edge is the network layer encompassing the smart end devices and their users to provide, for example, local computing capability on a sensor, metering or some other network-accessible devices."

IoT applications require location awareness and low latency due to the wide geographic distribution of IoT smart end devices, which presents challenges for conventional cloud infrastructure. As a practical solution to these challenges, edge computing has been proposed to enable monitoring in patients' homes. Wearable devices are thus closer to edge devices and can provide limited computational power, allowing edge devices to process data more quickly. The continuous patient monitoring by an e-health monitoring system collects vast amounts of data that need to be analyzed in real-time without interruption. An e-health application that receives data late can compromise its efficacy and negatively affect the patient's health.

The proposed system below, thanks to the use of Bluetooth sensors, is able to monitor clinical parameters without the need of the physical presence of a healthcare professional. The system detects various clinical parameters (e.g., Electroencephalogram (EEG), Blood Oxygen Level (OXI), Electrocardiogram (ECG), Electromyography (EMG), ALT Blood Test (ALT), body temperature), processes captured data and generates the Clinical Pathway for the patient under treatment.

Since the smartwatch rose a few years ago, different smart wearable devices spread up in everyday human life. These intelligent objects are also employed in the medical scenario with notable results.

Imagining that we are in a state of remote home care, we describe two use cases (Section 4.2 and Section 4.3 respectively) to identify

possible tampering with the patient's CP and, thus, better manage the security of data transmission.

Figure 4.2 depicts a general overview of the system architecture for the continuous monitoring of a patient and safe management of his/her CP. Each smart device (e.g. headband, smartwatch) feeds the Infrastructure Edge Node with its specific data. As stated before, these signals are crucial to a convalescent patient or undergoing rehabilitation.

Thanks to machine learning solutions and related e-health techniques, these data are used to monitor a patient and produce an efficient clinical path, give continuous feedback about health conditions to the doctor's control unit and enable interaction between doctor, patient, and his/her relatives.

The data management is a challenging aspect because of the ingestion of heterogeneous data with low latency and zero downtime, alongside the generation of a proper clinical Pathway based on patient history. These crucial aspects are linked to some other issues: one of the most critical is the anomaly detection. The proposed architecture attempts to address this aspect, without forgetting all crucial aspects of the context where this one is employed.

The core of the architecture is a cluster of edge nodes that cooperate to perform a sort of Extract, transform, load (ETL) task. This cluster is intrinsically linked to the anomaly detection module, which monitors the general task step by step. Beyond this architectural schema, the idea is to equip a standard edge cluster of nodes with an intelligent anomaly detection module. The running example exposed in Section 4.2.0.2 and the architectural overview match the healthcare scenario, but all features are available in various domains. The main components of the architecture are described in the following.

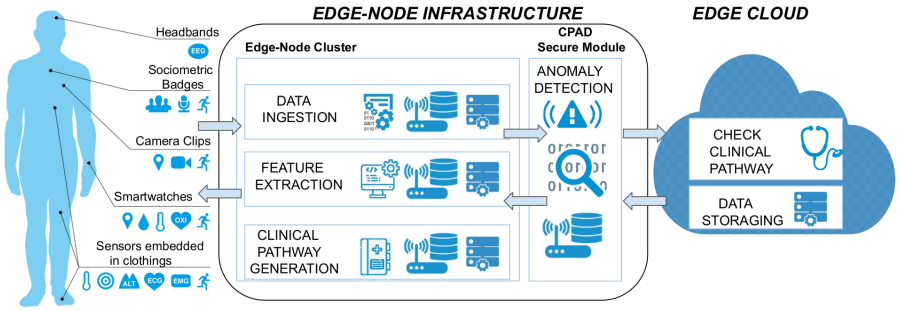

Figure 4.2: Edge architecture using AI techniques to monitor patients and manage the Clinical Path.

DATA INGESTION NODE. As stressed before, due to the massive employment of various smart devices, a data ingestion module or-

chestrates all data streams coming from these objects. This node is devoted to creating links between wearable medical devices and other modules belonging to the architecture. Data flows are injected into the Feature Extraction node for further specific elaborations and in the Clinical Path Anomaly Detection node to identify malformed data.

FEATURE EXTRACTION NODE. One of the most critical issues to address is to guarantee the data privacy of each patient. For this reason, data flow coming from smart medical devices are processed at the edge of the network by this node. All vital sign data are analyzed to extract notable characteristics from the stream. These features are then injected into the Clinical Path Anomaly Detection module to be confident about the goodness of the detected data or to identify some troubles inside them.

CLINICAL PATHWAY GENERATION NODE. Among the goals of the system, the generation of a personalized clinical pathway is a crucial task. Thanks to this node, the system learns from patient's history and combine this knowledge with that provided by doctors, thus producing a tailored therapy. Also this node interacts with the Clinical Path Anomaly Detection module to identify possible issues.

CLINICAL PATH ANOMALY DETECTION (CPAD) SECURE MODULE. This module makes the system less prone to anomalous situations, such as: (i) a specific malfunction related to vital sign and the therapy specified in the clinical pathway, (ii) hardware fail situations like battery degradation, (iii) system hacking by the patient or data tampering by someone not authorized to be involved in this process. The specific design of this module, is detailed in the next Section 4.2.

EDGE CLOUD. With this component, it is possible to manage two specific aspects of this scenario. First, it is possible to store all the data coming from the single edge- node cluster in a privacy-aware manner. Then, for every CP generated by the Clinical Pathway Generation module, a specific component performs a formal check for possible inconsistencies. In this way, a CP compliant to the anomaly detection module is double-checked with a specific component that validates his fairness.

4.2 USE CASE: AI TO IDENTIFYING ANOMALIES

Anomaly detection is of pivotal interest not only in network intrusion detection [28], fraud detection in financial domain [7], air pollution [151], but also in the healthcare domain concerning medical diagnosis [172].

As state in Section 4.1, sensors detect the vital parameters and send them at the Edge-Node cluster where the *Ingestion* node performs data orchestration. Then, the *Feature Extraction* node extracts key features. To check if data transmission is correct and that there have been no malfunctions (including system hacking), the proposed system is equipped with a module called *Clinical Path Anomaly Detection (CPAD)*.

The CPAD module analyzes all the data transmitted from the devices monitoring the patient to the Edge-Node cluster and eventually notifies detected anomalies. It using specifically implemented machine learning techniques, manages the security issues that could occur during the data transmission process. In this context, the anomaly could also consist of an attack to the monitoring of the patient's clinical parameters. The detected anomaly causes a dysfunction in the CP that in turn has a direct impact on the patient's health.

4.2.0.1 *Technological Approach*

The data collected in the *Ingestion* node can be seen as a queue and as organized into several sub-processes. Each sub-process represents the detection phase of a vital parameter from a single device worn by the patient. Thanks to the adoption of a recurrent sequential LSTM Autoencoder (described in Section 2.3.1.5), the CPAD analyzes the various sub-processes of the chain to perform the detection of anomalies on the steps of the chain [105, 113].

In particular, in our use case, the advantage of using sequential LSTM autoencoders is two-fold:

1. Taking advantage of the dimensionality reduction and extraction capabilities of the autoencoder (Section 2.3.1.4) to efficiently perform the data reconstruction process, and then detect the anomaly;

2. Using LSTM networks (Section 2.3.1.3) to manage the sequential nature of the data detected by the sensors.

The difference between a regular and recurrent autoencoders may be summarised as it follows: regular autoencoders work on sequential data by fixing the data size, usually by padding all sequences with

zero vectors to the length of the longest sequence [146]. In contrast, the recurrent autoencoders that were adopted in this proposal can compress variable-length sequences into fixed-length representations [61]. Therefore, they can generalize dependencies between nearby frames to other positions in the sequence.

In this way, the CPAD Module is able to define whether or not the patient's CP is correct. Otherwise, a specific machine learning algorithm adjusts the CPs according to the data currently detected.

The CPAD Module is able to detect three types of anomalies:

1. *Specific Malfunction*: it indicates a specific system malfunction. The module can detect whether the parameters that are transmitted from wearable devices to the edge node are reliable or not. It is also able to monitor whether the actions to be performed are those as per the CP.

2. *Hardware Malfunction*: it indicates a hardware malfunction. The module can detect the battery-charge status of the devices and the malfunctioning of the detection probes and patches. It also detects transmission errors at the Bluetooth protocol level.

3. *System Hacking*: it indicates system hacking. The module can detect if someone is trying to hack the system and if the user is trying to trick them.

4.2.0.2 *Running Example*

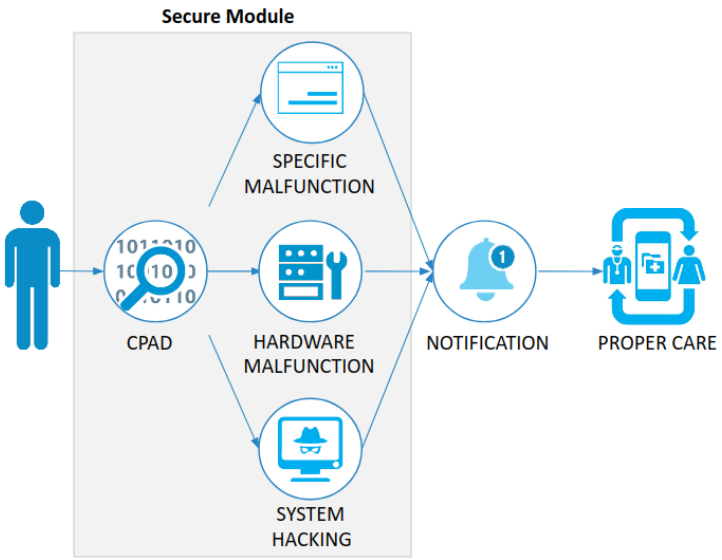

Figure 4.3: Running Example Use Case: CPAD Secure Module Functionality.

Figure 4.3 summarizes different kinds of possible anomalies that could be detected by the CPAD Module.

Suppose that the system is used to monitor a patient's in-home care. The patient suffers from a particular pathology that, among other problems, causes high blood pressure. To be able to lower the pressure, the doctor has prescribed two pressure pills a day, one at 07.00 *am* and the other at 09.00 *pm*. The doctor's diagnosis and the prescriptions for the medications to be taken are part of the CP. The pills are in a smart container (e.g. RxCap [1]) which indicates the time at which a pill is taken. If the doctor has prescribed that the patient should only take the pill twice a day, the CP knows that the sensor that controls the opening of the container should only be opened (or closed) twice a day and the pill can only be taken twice. The sphygmomanometer worn by the patient, according to the pressure monitoring instructions of the CP, performs pressure monitoring 5 times a day: 06.00 *am*, 09.00 *am*, 03.00 *pm*, 05.00 *pm*, and 08.00 *pm*. If the value of the pressure measurement is not in the range indicated in the CP, the CPAD detects a *Specific Malfunction*. This generates a notification that informs the actors involved (doctor, patient and relatives) of this event, and the correction flows are then appropriately generated.

It is possible to know the behavior of each sensor because the hardware specifications and operating details (and also malfunctioning) are available. For example, the sphygmomanometer measures blood pressure at predefined intervals, as specified in the CP. The pressure measurement process takes 30 seconds. If the measurement process lasted only 10 seconds, it detects a *Hardware Malfunction*, which is due to several factors, e.g. low battery.

In the same scenario, an example of *System Hacking* is the following one: the doctor has prescribed two blood pressure pills a day. This information is codified in the CP, thus it is displayed on the patient's tablet or programmed in the pill dispenser. The system could be hacked so that the number of pills is increased to 4.

Through the adoption of these AI-based Security techniques, which act as intrusion detection techniques, the objective is to prevent attacks before, after and during the CP management and to provide intelligent information to the physician who gives the treatment, and allow the reprogramming of the CP.

The previous Section 4.2 highlights the capability about the Edge-Node architecture of detecting different kinds of anomalies in the healthcare domain, which are useful in particular for patient assisted at home.

1 https://rxcap.com/

This approach exploits a novel ML technique that encapsulates two layers of LSTM into an Autoencoder structure. Indeed, the overall idea is to detect and keep track of anomaly situations with respect to the clinical history of a patient.

To best enable the interpretation and explanation of the outcome of the CPAD module we present in the Section 4.2.2 a Use Case in which, by monitoring the patient (specifically the heartbeat by ECG) we propose an Explainability system able to interpret the outcome. It is a module able to do explanation. As detailed in Section 2.5 the explanation of AI is a critical aspect in all the system that supports human decisions. This aspect represents a conjunction of different spheres: from the classical side of philosophical details to human-computer interaction.

4.2.0.3 *Monitoring Approach*

Security plays a major role in the healthcare domain. Preventing cyberattacks on healthcare infrastructures is no longer negligible. Compromising security in any e-Health system can lead to serious damage to patients' health. In particular, in a remote care context, the protection of telemonitoring systems of patients are essential to ensure that they follow their Clinical Pathway without any kind of external intrusion.

AI plays an important role in combating cyberattacks on the security of patient telemonitoring systems [27, 97, 127]. A system that monitors and prevents cyberattacks in healthcare not only must detect the attack, but should also be able to properly understand and report it to the user. In particular, Anomaly Detection systems are renowned approaches that are based on Machine Learning (ML) or Deep Learning (DL) methods to model normal activity in a such way as to easily detect abnormal deviations from the standards in a data-driven fashion. Therefore, in such a sensitive domain, where several healthcare professionals are involved, in addition to detecting threats, it is of paramount importance to represent and explain them through appropriate Explainability algorithms [79]. Moreover, current detection models and rules are not mature enough to recognise early breaches that have not yet caused any damage.

Intrusion analysts infer the context of the incident using prior knowledge to discover events relevant to the incident and understand why it happened [4]. Although security tools that provide visualization techniques and minimize human interaction have been developed to make the analysis process easier, too little attention has been given to making human-friendly the interpretation of security incidents. Simply report-

ing a cyberattack in written format is not enough to enable the healthcare professional to correct the patient's Clinical Pathway. This data must be represented in a graphical way, which can be grasped by the healthcare provider. The detection of the cyberattack must therefore be supported by systems that provide different forms of explanation, depending on the different end users, and that allow these users to have the possibility to interactively manipulate graphical representations based on Visual Data Mining techniques.

The Internet of Things (IoT) has transformed hospital settings and created a new moniker for the healthcare world, The Internet of Medical Things (IoMT). Ensuring a security mechanism for IoMT, which uses appropriate analytical tools in a distributed working architecture, also capable of analyzing huge data (i.e., big data) generated by IoMT devices in a distributed manner, is a challenging issue.

An approach using Process Mining techniques to identify variations in the clinical path from that defined by the physician is detailed in Section 4.3.

Section 4.2.1 detailed some e-health work related to the explanation task in the cyberattack detection domain. Subsequently, in Section 4.2.2 is present and discuss a cyberattack detection model that propones eXplainable Artificial Intelligence XAI approach to support caregivers in grasping that an attack has occurred to the telemonitoring system and its effect on the patient's Clinical Pathway.

4.2.1 *Detection Systems for e-health Domain*

Cyberattack detection can be defined as the problem of identifying individuals who are using a computer system without authorization, those who have legitimate access to the system but are abusing their privileges, and, in general, the identification of attempts to use a computer system without authorization or to abuse existing privileges.

In this landscape, modern cyberattack detection systems monitor either host computers or network links to capture cyberattack data. Host intrusion detection refers to the class of Intrusion Detection Systems (IDS) that reside on and monitor an individual host machine [157]. There are a number of system characteristics that a host intrusion detection system (HIDS) can make use of in collecting data [139]. A network intrusion detection system (NIDS), instead, monitors the packets that traverse a given network link. Such a system operates by placing the

network interface into promiscuous mode, affording it the advantage of being able to monitor an entire network while not divulging its existence to potential attackers [119].

Cyberattack Detection System (CADS) is software that automates the cyberattack detection process and detects possible cyberattacks. Cyberattack Detection Systems serve three essential security functions: they monitor, detect, and react to unauthorized activity by company insiders and outsider cyberattack. One of the major approaches to cyberattack detection is Anomaly Detection. It assumes that a cyberattack will always reflect some deviations from normal patterns. In this sense, Anomaly-based IDS compares a model of normal behavior against the incoming traffic in order to find anomalies [96, 170].

Once an intrusive incident has been reported by means of a correct detection, the reaction phase has to be fired, evaluating the impact of this event on the security level of the system [73]. It must provide the set of countermeasures to quickly eradicate the cyberattack and, at the same time, indicate the set of actions to heal the system and bring it back to its normal state. A possible way to react is via the use of Intrusion Response Systems (IRSs), as they are IDSs capable of counteracting suspicious activities [179]. Although intrusion response components are often integrated with the detection ones, they have received considerably less attention than IDS research.

Anomaly Detection typically operates on monitored networked traffic data. Actually, continuous monitoring is the main activity of modern e-Health technologies, ranging from devices that monitor health and deliver medication, to telemedicine delivering care remotely. Indeed, the integration of healthcare-based devices and sensors within IoT, led to the evolution of IoMT [175]. Therefore, IoMT-enabled devices have made remote monitoring possible in the healthcare sector, enabling the ability to keep patients safe, and inspiring doctor to provide superlative treatment [177]. As a result, the increasing demands and expansion of IoMT systems require advancements in data storage methods, data processing and cybersecurity related issues.

Healthcare providers can then provide efficient remote healthcare communication for monitoring and diagnosis services to the residents of these smart communities. Any security threat to these systems may cause a serious problem, such as imposing a false diagnosis or delaying the interaction. This leads to a violation of patients' privacy, health issues, and even death in extreme cases [65]. AI and Machine Learning (ML) have been largely employed for managing issues in healthcare systems as they are the most promising techniques to be used for previ-

ously unseen attacks [33]. It can identify attacks simply by monitoring data alteration or by detecting changes in the network's traffic characteristics. In particular, ML-based anomaly detection systems are crucial to ensure security and mitigate threats such as false data injection attacks [110]. IoMT systems are widely distributed and are collections of heterogeneous sensors. Attack detection in IoMT is entirely different from the present security mechanism, due to the special services required by IoMT such as: computing power, memory space, battery life, low latency, and network bandwidth, which cannot get fulfilled by the centralized conventional approach of standalone cloud computing [10]. In cloud computing architecture, data generated by IoMT devices is being transmitted to and from the cloud in order to provide services to the healthcare users. The limitations of traditional standalone cloud solutions is that the data recovery time is too high for a real-time emergency situation, such as fall detection or stroke prevention, which mostly needs rapid response time from medical professionals [156].

Therefore, designing a distributed security framework, for distributed IoMT applications is a challenging task due to the dynamic nature of IoMT network such as IoT devices, edge devices, and cloud. Moreover, the evolution of attacker behavior can intercept the transmission network in IoMT [138]. The line of research in this way is going towards constructing robust anomaly-based IDSs that efficiently distinguish attack and normal observations in IoMT environment, consisting of interconnected devices and sensors, with poor design and weak authentication measures. As stated in [2, 18], the collection of remote data from these sensors is a complex process due to the different types of devices that are involved to measure the parameters. For this reason, works in [130, 131] dealt with a clinical and operational context to develop integrated solutions for seamless care in which AI and IoMT are used at the Edge, with a people-centered approach that adapt to the needs of healthcare providers and that are embedded into their workflows. Recently, in [100] proposed an ensemble learning model that combines Decision Trees, Naive Bayes, and Random Forest to feed a final XGBoost classifier in order to identify normal and attack instances in an IoMT network. Also, authors in [74] have designed a a real-time Enhanced Healthcare Monitoring System (EHMS) test-bed that monitors the patients' biometrics and collects network flow metrics. Some recent works [54, 155, 166] have proposed to improve the performance of anomaly detection by incorporating a type of feedback from the user, called *User Feedback*. Nevertheless, all the studies lack of effective threat reaction phases that can be managed with appropriate

explainable modules of ML-based models and user feedback modules to ascertain that a detected anomaly is assumed to be malicious.

4.2.2 *CADS: Cyberattack Detection System*

Figure 4.4 depicts the architecture of the proposed *Cyberattack Detection System*. It is an improved version of the Figure 4.2 described in Section 4.1. This architecture focuses on the security of data transmitted from IoMT sensors to three different interconnected processing modules, namely the CPAD, the Explainer module, and the User Interaction Engine. The latter is made up of three sub-modules: (i) *Visualization Framework*, (ii) *User Interface* and, (iii) *User Feedback*.

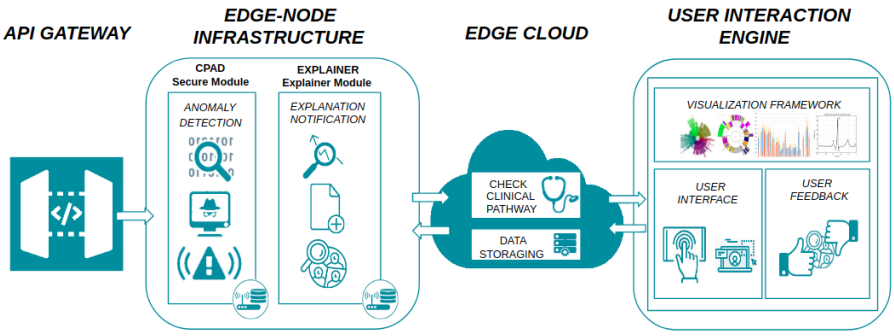

Figure 4.4: Architecture for the Cyberattack Detection System.

Figure 4.4 depicts the architecture of the proposed Cyberattack Detection System. An early version was already detailed in Section 4.2. This revised and refined architecture focuses on the security of data transmitted from IoMT sensors to three different interconnected processing modules, namely the CPAD, the Explainer module, and the User Interaction Engine. The latter is made up of three sub-modules, i.e.: Visualization Framework, User Interface, and User Feedback.

The system implements a methodological approach to the problem of anomaly detection by including, in addition to the identification of anomalous data, an explanation of the possible motivations for classifying such data as anomalous, and the possibility for a domain expert to validate data through their visual interactive representation. Anomalous data, which represent an intrusion in a hacked system, are explained according to the *Explainable Security (XSec)* paradigm [167]. As a result of a hacked home care telemonitoring system, the Cyberattack Detection System classifies some ECG instances as False Positives (FP) or False Negatives (FN). After the detection, the user will be able to analyze the characteristics of what caused the classification of some data

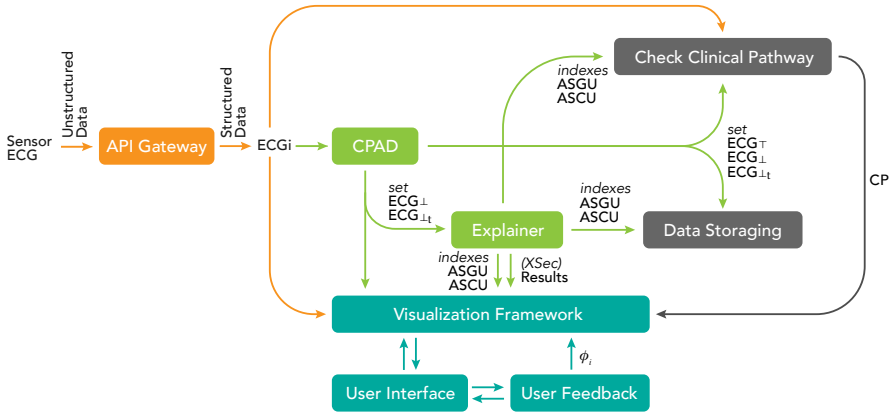

Figure 4.5: Data flow among system modules.

as FP or FN and interact with it. The interaction activity with the data is performed by means of the *User Interaction Engine*, which provides a dashboard through which to visually explore the data to get a clearer view of what happened in a time interval. Furthermore, thanks to the *User Feedback* sub-module, it is possible to implement a continuous improvement of the classification performances, and consequently of the anomaly detection, thus achieving a more robust identification of threats. Figure 4.5 shows the relationships between the various modules in terms of data flow.

Our system is beneficial in e-Health scenarios, supporting the patient who is in home-based healthcare. The technological infrastructure, based on the architecture presented in [12], promotes the care of a patient according to his or her CP, i.e. a set of diagnostic and therapeutic procedures related to the treatment of that specific patient. The CP represents a tool used to manage the quality in healthcare concerning the standardization of care processes. Its implementation reduces the variability in clinical practice and improves outcomes, aiming at promoting organized and efficient patient care based on evidence-based medicine, and to optimize outcomes in settings such as acute care and home care. A single CP may refer to multiple clinical guidelines on several topics in a well specified context. In this way, some activities can be managed by the health personnel of health structures; some others can be managed autonomously by the patient, in a sort of medical-unsupervised manner. The home care infrastructure, i.e. the *IoMT-Edge-Computing*, promotes a kind of distributed edge computing of the IoMT network, thus reducing latency and improving reliability. Patient monitoring devices are then connected in the IoMT network. In turn, edge devices communicate with a cloud infrastructure to store gathered clinical data

and keep in touch with the corresponding medical staff. Therefore, some vulnerabilities may arise regarding the security of patient and clinical data.

The proposed Cyberattack Detection System implements cybersecurity methods to identify which data were compromised after the system is hacked. Moreover, it provides an explanation of the cyberattack and enables interaction with the detected anomalies which may occur in the remote and continuous patient monitoring and care phases. In particular, the anomaly detection phase is carried out by the CPAD module, whose formalization, already tailored in the healthcare domain, was introduced in [14]. The CPAD detects deviations from the patient's CP and avoids the processing of inconsistent or false data, which could be life-threatening for a patient. After the detection phase, the Explainer module analyzes clinical data classified as anomalous and, through the User Interaction Engine, a validation request is sent remotely to the medical staff. A continuous telemonitoring of patient's clinical parameters is performed, without the need for a physical presence of the health operator. The system, by means of IoMT sensors connected to the Edge network, collects different clinical parameters, such as EEG, OXI, ECG, EMG, ALT, and body temperature. Afterwards, such clinical data are processed at the Edge and the patient Clinical Pathway is generated.

In the following, a running example in which an e-Health telemonitoring system has been hacked is reported. It describes how the proposed Cyberattack Detection System is able to highlight the cybersecurity threats and the related countermeasures to solve the hack and restore the system. A male patient who is following a certain CP is monitored. He suffers from Congestive Heart Failure and his CP requires that his heartbeats are monitored every 15 minutes. A smart end-device that measures the ECG is used. Hence, the proposed system, suitably connected to the end-device, receives data about the patient pulse. The gathered heartbeats are fed to the CPAD module that determines whether anomalous measurements occur, suggesting that the telemonitoring system has been hacked. Since heartbeats measures are a key factor in determining the clinical picture of a patient, a compromised flow of measurements would endanger the patient's CP and induce a wrong handling of the patient's health.

The Cyberattack Detection System is designed to be modular and can be integrated with various IoMT devices that use Bluetooth technology. An API gateway is responsible of the correct integration of the end-device with the Cyberattack Detection System. In particular, the API gateway checks the compliance of the gathered data to be fed into

Table 4.1: ECG device data table format.

ECG	HB_{ts_1}	HB_{ts_2}	$\cdots$	HB_{ts_u}
ECG_0	$x_{(0,1)}$	$x_{(0,2)}$	$\cdots$	$x_{(0,u)}$
ECG_1	$x_{(1,1)}$	$x_{(1,2)}$	$\cdots$	$x_{(1,u)}$
ECG_2	$x_{(2,1)}$	$x_{(2,2)}$	$\cdots$	$x_{(2,u)}$
$\vdots$	$\cdots$	$\cdots$	$\ddots$	$\vdots$
ECG_i	$x_{(i,1)}$	$x_{(i,2)}$	$\cdots$	$x_{(i,u)}$

the system. By means of specifically designed APIs, input data are normalized according to the system standards, allowing the correct exchange of information between the devices and the software modules, also converting unstructured data into structured ones.

Referring to the running example, in which a smart ECG monitoring end-device is connected via IoMT with the Cyberattack Detection System, the following definition to formally handle data inflow is proposed.

Definition 8. Let ECG be the data-flow of a smart ECG monitoring end-device, HB be the heartbeat information coming from ECG at a certain timestamp ts. The i-th heartbeat detection is defined as follows:

$$ECG_i = f(HB_{ts_u}), \text{with } HB_{ts_u} \in \mathbb{R} \text{ and } u \in [1, l], l \in \mathbb{R} \quad (4.1)$$

The variable HB_{ts_u} indicates the count-based feature representing the value of the u-th sampling step in a given timestamp, which can be assumed as a real value $x_{(i,u)} \in \mathbb{R}$, representing the amount in milliVolt (mV) of the count-based feature. Therefore, the representation of the ECG data can be formalized as in Table 4.1.

4.2.3 *ECG Anomaly Detection*

The proposed system, through the use of AI techniques, contributes to improve the security of the telemonitoring infrastructure. Once the data have been transformed into a structured form, they are given as input to the CPAD which analyzes the structured data according to Table 4.1 format, and checks in which point the cyberattack has been launched by retrieving the anomaly. The CPAD module is based on Robust Deep Autoencoders (RDA) [42]. In fact, some recent works demonstrated how these deep approaches perform quite well in detecting cyberattacks by carrying out anomaly detection by means of neural structures [19, 146,

176]. The main advantage of applying RDA for anomaly detection is the capability of discovering high-quality nonlinear features, while at the same time identifying and eliminating outliers and noise.

In the running example, in which it is assumed that a cyberattack on ECG measurements is going on, a delayed threat reaction would compromise the patient's care pathway. By exploiting RDA for anomaly detection, the CPAD modules ensure a rapid response in terms of inference time, quickly detecting anomalous ECG data. In particular, the CPAD module categorizes ECG data with a simple binary classification according to whether they have been identified as anomalous or not. Actually, in the medical literature [44], heartbeats can be classified into five different types:

1. (N) - Normal;
2. (RonT) - Premature Ventricular Contraction;
3. (PVC) - Premature Ventricular Contraction;
4. (SP) - Supra Ventricular;
5. (UB) - Unclassified Beat.

Therefore, the classification process can be further specified by dividing the normal heartbeats from the others, and then dedicate another RDA classifier to detect the remaining four different types of anomalies. In particular, the CPAD module defines three sets of classified ECG data:

- $ECG_{\top}$: instances classified by CPAD as normal, i.e. not anomalous;
- $ECG_{\perp}$: instances classified by CPAD as anomalous;
- $ECG_{\perp_t}$: instances of anomalous ECG, with $t \in \{$(RonT), (PVC), (SP), (UB)$\}$.

Once the structured data have been processed, the CPAD produces a CSV file containing all the instances classified by CPAD as anomalous $ECG_{\perp}$, i.e. all the heartbeats that are anomalous. A confusion matrix is also showed, in order to have a clearer view of performance classification, both in terms of predictability power and effectiveness of the learned model. To address the anomaly detection problem, the RDA method based on the autoencoder approach is exploited. It is based on a training pipeline, where examples of anomalies are provided during model training. The result is evaluated on sets consisting of anomalous

and normal (i.e., non-anomalous) data. In the end, the CPAD module separates the predicted instances into the three different sets of heartbeat anomalies, namely $ECG_{\top}$, $ECG_{\bot}$, and $ECG_{\bot_t}$.

4.2.4 *ECG Interpretation*

After the anomaly detection, the next step is to interpret and explain the sets that CPAD has created, to understand why certain instances have been classified as anomalous. In systems using Explainable AI (XAI) algorithms, additional data coming from the Machine Learning process may be useful explanations, produced by the system itself to enrich the predicted instances with a plausible rationale. In Information Security, instead, explanations are provided by the designers. However, the role of Explanations is crucial in AI field. Consequently, Information Security kept all advantages from that, for example, ensure the user trust concerning the system. Explanations are therefore designed to bridge the gap between "*actual safety*" and "*perceived safety*".

To this end, the Explainer module of our Cyberattack Detection System receives as input the i-th ECG instance (ECG_i) from the API Gateway, and also the output of the CPAD, appropriately separated in the three sets $ECG_{\top}$, $ECG_{\bot}$, and $ECG_{\bot_t}$. Exploiting the XSec paradigm (detailed in Section 2.5), the aim is to obtain its "*six W*" (Who? What? Where? When? Why? and How?) that give a complete view of the identified and perceived anomaly. XSec involves several actors (e.g., in this case we have security analysts, doctors, nurses, and the patient). It requires a dedicated reasoning tool to infer about the system model, the threat model, and properties of security, privacy, and trust, as well as concrete cyberattacks, vulnerabilities, and countermeasures.

In addition to the XSec paradigm, two specifically designed indexes, called respectively *Anomaly Score General Unsafe* (ASGU) and *Anomaly Score Class Unsafe* (ASCU), are introduced. First, we formally define the ASGU index as follows.

Definition 9. Let ECG be the set of heartbeat detections received from a smart ECG monitoring end-device, ECG_i be the i-th heartbeat detection, and $ECG_{\bot}$ the set of all anomalous instances. Then, the *Anomaly Score General Unsafe* index is a function $\text{ASGU}\colon ECG \mapsto [0, 2]$ which gives a general ranking of possibility of considering a heartbeat as anomalous, according to instances in $ECG_{\bot}$. Therefore, $\text{ASGU}(ECG_i) \in \mathbb{R}^{[0,2]}$ indicates how the instance ECG_i is currently considered an anomaly over the whole set $ECG_{\bot}$.

Hence, the higher the ASGU score value, and thus closer to 2, the more anomalous the heartbeats will be considered. The ASCU index, instead, is based on the comparison with one of the four classes of anomalies.

Definition 10. Let ECG be the set of heartbeat detections received from a smart ECG monitoring end-device, ECG_i be the i-th heartbeat detection, and $ECG_{\perp_t}$ a set of a class of anomalous instances, with $t \in \{(\text{RonT}), (\text{PVC}), (\text{SP}), (\text{UB})\}$. Then, the *Anomaly Score Class Unsafe* index is a function $\text{ASCU}: ECG \mapsto [0, 1]$ which gives a general ranking of probability of considering a heartbeat as anomalous, according to a class $ECG_{\perp_t}$. Therefore, $\text{ASCU}(ECG_i) \in \mathbb{R}^{[0,1]}$ indicates how the instance ECG_i is currently considered an anomaly over the whole set $ECG_{\perp_t}$.

In this case, the higher the value of the ASCU score, and therefore the closer to 1, the more abnormal the heartbeats will be in the set of heartbeats of the same class $ECG_{\perp_t}$. The goal of these two indexes is to quantify the strength of a feature in contributing to determine an anomaly in the ECG data. This would give a further useful information to characterizing the detected anomaly, and such information can be assumed as part of the explanation to be addressed together with the XSec approach.

In the running example, the Explaination module receives as input the CSV file containing the ECG_i instances and provides in output another CSV file composed of four columns:

- ECG_i: i-th instance classified by the CPAD as anomalous;
- CLASS: the corresponding class $t \in \{ECG_{\perp_t}\}$ of the i-th instance;
- ASGU: the Anomaly Score General Unsafe score;
- ASCU: Anomaly Score Class Unsafe score.

The Explanation module, therefore, acts in two conjunct phases: the first one takes place following the generation of the CSV file containing the information and indexes mentioned above (ECG_i, CLASS, ASGU and ASCU); the second one results from the information generated by the XSec approach, which will be displayed in the Visualization Framework. Thanks to the combination of the XSec paradigm and the two indexes ASGU and ASCU, the Cyberattack Detection System provides the doctor with a concise explanation of why the i-th detection has been classified as anomalous.

Following the Explanation, it is possible to check whether the Clinical Pathway that has been generated is correct or not. An incorrect CP has a double meaning: from a clinical point of view, it is a serious problem for the patient's health while, from the cybersecurity point of view, it means that the system has been hacked.

4.2.5 *ECG User Interaction*

In our case study, the contribution of the User Feedback to the system is the evaluation of a the detected anomaly and the embedding of a doctor's feedback. The User Feedback module will generate for each detection ECG_i a feedback coefficient ϕ_i that represents the doctor's feedback on a given instance.

Definition 11. Let ECG be the set of the heartbeat detections received from a smart ECG monitoring end-device, ECG_i be the i-th heartbeat detection. Then, the *feedback* coefficient is a function $\phi\colon ECG \mapsto \{-1, 1\}$ such that any i-th user feedback related to the heartbeat detection ECG_i, is defined as follows:

$$\phi_i = \begin{cases} +1 & \text{if } ECG_i \text{ is false positive or false negative} \\ -1 & \text{if } ECG_i \text{ is true positive or true negative} \end{cases} \tag{4.2}$$

Therefore, the *User Feedback* UF is a set of tuples such that, for any i-th pair of arguments (ECG_i, ϕ_i), a single element UF_i is defined as:

$$UF_i = (ECG_i, \phi_i) \tag{4.3}$$

In this way, the Cyberattack Detection System will become more robust to external cyberattacks, since the User Feedback would report the opinion of the caregiver which will confirm or not whether the i-the ECG detection is abnormal or not. In the User Interaction Engine, the *Visualisation Framework* represents the data orchestrator, handling and visualizing processed data coming from the various modules. It uses algorithms of Visual Data Mining (VDM) [77, 99] that allow, through different visualisation techniques, to interactively group data in a more efficient way, improving the data insight process.

Afterwards, the *User Interface (UI)* included in the User Interaction Engine allows the user to interact with the data. In the running example, the UI allows the caregiver to interact with the ECG instances. After the Explanation module has displayed the result of the *XSec* paradigm, it is possible to visually manage each ECG detection. For instance, one would be able to no longer consider an ECG instance as an anomaly, or,

more specifically, to improve the classifier performances by indicating the correct class of anomaly among the four types $ECG_{\perp_t}$ when a wrong one has been predicted. The interaction with the user, in this case a doctor, helps the system to be more and more reliable, as well as secure from cyberattacks.

Through the integration of CPAD, Explainer, and User Engine Interface modules, the Visualization Framework will be able to manage anomalies detected as *Threat Insight*. These will be appropriately displayed on the UI which, in addition to allowing interaction with the anomalous data (in this case the ECG detection), will be able to display the threat representation through a dashboard. Thanks to the threats graphical representation in the dashboard, the user's reaction to the threat is improved.

In conclusion, in the e-Health domain, new and continuous evolving threats emerge every day. The security of e-Health telemonitoring systems is no longer a negligible task. The proposed Cyberattack Detection System, thanks to the use of AI techniques, is able to protect the e-Health system from cyberattacks by automatically identifying the anomalies in the e-Health system, without the need of a dedicated security analyst.

The solution is focused on the task of cyberattack detection, in the particular case of exploiting a remote patient telemonitoring system that has been hacked. A specific running example, i.e. the heartbeat telemonitoring, has been considered.

The presented system is designed to automatically detect the anomaly by means of Deep Learning techniques. In particular, a Robust Deep Autoencoder detects anomalous heartbeats instances. The detected anomalous heartbeats are subsequently interpreted with a combination of state-of-the-art explainable security paradigms (XSec) (detailed in Section 2.5) and with two new explainable scores which have been introduced, showing to the user the reasons of a malicious activity interfering with the heartbeats telemonitoring.

4.2.6 *Visualization Framework*

In this way, the CADS will become more robust to external cyberattacks, since the User Feedback would report the opinion of the caregiver which will confirm or not whether the i-the ECG detection is abnormal or not. In the User Interaction Engine, the *Visualisation Framework* represents the data orchestrator, handling and visualizing processed data

coming from the various modules. It uses algorithms of VDM [99] that allow, through different visualization techniques, to interactively group data in a more efficient way, improving the data insight process. Afterwards, the *UI* included in the User Interaction Engine allows the user to interact with the data. In the case study, the UI allows the caregiver to interact with the ECG instances. In the CADS architecture, an Explanation module displays useful classification information, with which it is possible to visually manage each ECG detection. For instance, one would be able to no longer consider an ECG instance as an anomaly, or, more specifically, to improve the classifier performances by indicating the correct class of anomaly when a wrong one has been predicted. The interaction with the user, in this case a doctor, helps the system to be more and more reliable, as well as secure from cyberattacks.

Through the integration of CPAD, Explainer, and User Interaction Engine modules, the Visualization Framework will be able to manage anomalies detected as threat insights. These will be appropriately displayed on the UI which, in addition to allowing interaction with the anomalous data (in this case the ECG detection), will be able to display the threat representation through a dashboard. Thanks to the threats graphical representation in the dashboard, the user's reaction to the threat is improved. The visual process, whereby the healthcare professional is able to mark a detection as true or not, is a key scenario in which ML methods are combined with human feedback through interactive visualization. This process enables the fast prototyping of the ML model that can improve both the performance of the algorithm and human feedback. It will be also able to complete tasks where anomaly identification was not yet possible. The system then uses User Feedback to refine detection results and guide further analysis. Caregiver Feedback is therefore used as an essential source of ever-improving anomaly detection of ECG, which means that labeling the local environment will trigger global updates and thus guide further analysis.

4.3 USE CASE: PROCESS MINING TO DETECTION ANOMALIES

In this section we proposes an innovative method to measure adherence to therapy aimed at providing awareness of the patient's current situation in healthcare environments, specifically in-home care. Given the difficulty of quantitatively assessing a patient's behavioral rigor in following prescriptions, the idea is to exploit process mining techniques at the Edge in a smart home environment when a patient has to stick to therapy, in order to define the level to which the patient's actions (such

as drug intake, adherence to diets, physical activity) are in line with the physician's indications for the management of the metabolic syndrome.

Section 4.3.4 shows our proposal on a real-world dataset, taking into account process mining techniques to model the medical prescription as a process model and test the adherence to the therapy of patients with conformance checking. Section 4.3.4.1 outlines the experimental setting and assesses event log data with state of the art evaluation metrics.

4.3.1 *Metabolic Syndrome*

The metabolic syndrome is a markedly heterogeneous nosological and clinical entity, still in the process of being defined, represented by the simultaneous association of alterations such as obesity, impaired glucose tolerance, dyslipidemia, and arterial hypertension. In Italy, the prevalence of the metabolic syndrome is around 20% [68]. Progressive increases in the occurrence of metabolic syndrome are age-related, and the prevalence of insulin resistance and glucose intolerance usually increases with increasing age.

Generally, metabolic syndrome is diagnosed if at least three of the symptoms listed in Table 4.2 are present.

Table 4.2: Criteria used for the diagnosis of metabolic syndrome [9].

Criteria	**Value**
Waist circumference (cm)	$\geq$ 102 *men* $\geq$ 88 *women*
Fasting Blood Glucose (mg/dL)	$\geq$ 100
Blood Pressure (mmHg)	$\geq$ 130/85
Fasting Triglycerides (mg/dL)	$\geq$ 150
Cholesterol HDL (mg/dL)	$<$ 40 *men* $<$ 50 *women*

The primary goal of therapy for metabolic syndrome is to improve or normalize reduced insulin sensitivity. The initial therapeutic approach to metabolic syndrome may be an attempt to abolish its initial causes, e.g. atherogenic diet, sedentary lifestyle, overweight, and obesity. Therapeutic strategies include pharmacological interventions and supplemental home treatments. After diagnosing metabolic syndrome or any

of its components, making healthy lifestyle changes can help prevent or delay serious health problems, such as heart attack or stroke. The recommended treatment with home remedies for the management of the metabolic syndrome, validated by an Italian physician, is summarized in the Table 4.3.

Table 4.3: Recommended Treatment for Metabolic Syndrome Management Home Remedies.

Event	**Prescription**
Before Breakfast	- glucose measurement - weight measurement - diabetes pill
After Breakfast	- pressure pill
Mid-Morning	- pressure measurement
Before Lunch	- glucose measurement - diabetes pill
Afternoon	- exercise: walking, cyclette or tapis rulant
Before Dinner	- glucose measurement - pressure measurement - diabetes pill
After Dinner	- pressure pill

Therefore, the treatment described in the above table may be understood as a process model to be carefully followed. Deviations from such a model constitute a level of adherence to therapy that can be difficult to assess. Understanding the degree of adherence to a therapy and indicating an overall percentage of compliance can be a very effective tool for both the patient and the doctor, in order to be able to quickly intervene to help when this degree of adherence falls below a certain threshold.

4.3.2 *Data Collection*

Real-world data, containing the logs of the behavior of patients with metabolic syndrome, within the project of Italian Ministry of Education, University and Research MIUR, ‘Progetto Cluster Tecnologici

Nazionali - Tecnologie per gli Ambienti di Vita: Active Ageing At Home' (AAAH) [2].

The dataset recorded information of patients during telemonitoring at home, collected through IoT and medical devices, and describes various activities, such as meals, drug intake, physical exercises, measurement of vital signs, weight measurement. The data have been collected from a cohort of 19 continuously monitored patients with common characteristics located in the Italian Apulia region. The data collection process was carried out over 30 days, sampling all the actions performed by a patient during the daily routine from a series of medical devices which convey sensed data to an edge module in charge of gathering and supplying them to an event log in a standard format compatible with process mining techniques.

4.3.3 *Data Preprocessing*

Patients are equipped with an edge component capable of processing the behavior of the patient on site in a self-consistent and self-evaluative way, i.e. an autonomous module capable of making assessments without communicating with other systems. The proposed edge architecture downloads the medical prescription by connecting to the patient's medical record.

For this experimental analysis, the sensed data are stored in chronological order and in several different tables, i.e. a table containing the drugs taken by the patient, a dietary table, a table relating to blood pressure measurements, one relating to the measurements of blood glucose, one relating to body weight measurements, and finally a table with the different exercises performed. The presence of different tables is determined by the fact that the different equipped sensors store their data in the respective tables, given the consistency required for the data collection based on the patient taken into consideration.

An operation of data aggregation from the various tables in the database under consideration has been carried out, in order to create a single dataset in `.csv` format useful for subsequent processing. Hence, a standardization operation has been carried out. In particular, the timestamp has been standardized to the `dd/mm/yyyy hh:mm` format. Diet-related activities have been standardized in `breakfast`, `lunch` and `dinner`. Physical activity has been unified with a single

2 Avviso per lo sviluppo e il potenziamento di cluster tecnologici nazionali. Area TAV: Tecnologie per gli Ambienti di Vita. Active Aging At Home Project, PON Code: CTN01_00128_297061

`exercise` activity, and the type of performed activity has been indicated in one categorical value between "walking", "cyclette", and "tapis roulant". The blood glucose measurement activity has been standardized with `glucose_measurement` as well as the weight, standardized to `weight_measurement`. Also, the pressure measurement has been standardized with `pressure_measurement`, reporting the indication of the systolic and diastolic measurement in a "systolic/diastolic" way. While, the following notation has been adopted regarding the patient's drug intake during the day: `pressure_pill` and `diabetes_pill` for the intake of drugs related to blood pressure and diabetes, respectively.

Ultimately, the data were aggregated by patient using a unique id, and subsequently sorted according to the timestamp provided, in such a way as to obtain the sequence of activities performed by the same patient in chronological order during the period of monitoring. In this way the data stored in different tables have become rows of a single dataset with the respective label. Indeed, in Table 4.4 is described a fragment of the final event log including for each case: case ID, timestamp, activity, and value.

4.3.4 *Modelling Metabolic Syndrome*

The idea is to model the problem of metabolic syndrome in-home treatment as a process model and then feed the retrieved dataset to the process model in order to run the conformance checking technique, achieving, in this way, an awareness of some threat. The event log analysis is performed at the end of the day, in order to assess its compliance along the entire path.

We modeled the ideal process defined in Table 4.3 with the 3 Process Mining modeling algorithms, implemented in the Python library `PM4Py` [26]:

1. *Alpha Miner* [164]: With an event log as the input, the Alpha Miner algorithm derives various relations between the activities occurring in the event log. These relations are used to produce a Petri net that represents the log;

2. *Heuristic Miner* [169]: Heuristics Miner is an algorithm that provides a way to handle with noise and to find common constructs (dependency between two activities). The basic idea is that infrequent paths should not be incorporated into the model. The output of the Heuristics Miner is an object that contains the

Table 4.4: An event log fragment of the exploited dataset.

Case ID	Timestamp	Activity	Value
1	01/04/15 07:27	glucose_measurement	130
1	01/04/15 07:29	weight_measurement	85.1
1	01/04/15 07:33	diabetes_pill	diabetes
1	01/04/15 07:45	breakfast	breakfast
1	01/04/15 08:05	pressure_pill	pressure
1	01/04/15 11:15	pressure_measurement	135/90
1	01/04/15 12:27	glucose_measurement	203
1	01/04/15 12:33	diabetes_pill	diabetes
1	01/04/15 13:03	lunch	lunch
1	01/04/15 17:27	exercise	walking
1	01/04/15 19:45	glucose_measurement	300
1	01/04/15 19:49	pressure_measurement	140/91
1	01/04/15 19:57	diabetes_pill	diabetes
1	01/04/15 20:14	dinner	dinner
1	01/04/15 20:42	pressure_pill	pressure
2	02/04/15 07:20	glucose_measurement	133
2	02/04/15 07:29	weight_measurement	85
2	02/04/15 07:32	diabetes_pill	diabetes
2	02/04/15 07:40	breakfast	breakfast
2	02/04/15 08:09	pressure_pill	pressure
2	02/04/15 11:13	pressure_measurement	137/93
2	02/04/15 12:37	glucose_measurement	210
2	02/04/15 12:43	diabetes_pill	diabetes
2	02/04/15 13:13	lunch	lunch
2	02/04/15 18:27	exercise	cyclette
2	02/04/15 19:55	glucose_measurement	305
2	02/04/15 19:57	pressure_measurement	139/90

activities and the relationships between them, that can be then converted into a Petri net;

3. *Inductive Miner* [103]: The basic idea of Inductive Miner is to find a prominent split in the event log (there are different types of splits: sequential, parallel, concurrent, and loop). After finding the split, the algorithm recurs on the sub-logs (found by applying the split) until a base case is identified. Inductive miner can discover robust process models from noisy and incomplete data, and can produce a Petri net.

Process models modeled using Petri nets have a well-defined semantic: a process execution starts from the places included in the initial marking and finishes at the places included in the final marking. Directly-Follows graphs, instead, are graphs where the nodes represent the events/activities in the log and directed edges are present between nodes if there is at least a trace in the log where the source event/activity is followed by the target event/activity. On top of these directed edges, it is easy to represent metrics like frequency (counting the number of times the source event/activity is followed by the target event/activity) and performance (some aggregation, for example, the mean, of time inter-lapsed between the two events/activities).

In Figure 4.6 is depicted the Petri net of the process model for metabolic syndrome management. Specifically, the Figure describes the daily routine that a patient has to follow according to the prescription. The Petri net represented by Heuristic Miner algorithm contains hidden transitions and has the advantage of having different parameters that can be used for the elimination of unimportant clusters. The most important ones are the *dependency threshold*, with a default value of 0.5, and the *cleaning threshold* to remove weak edges, with a default value of default value 0.05. In our experimental setting we left these values by default.

4.3.4.1 *Conformance Checking*

Conformance checking is a technique to compare a process model with an event log of the same process. The goal is to check if the event log conforms to the model, and, vice versa. In `PM4Py`, two fundamental techniques are implemented: token-based replay and alignments.

Token-based replay is a heuristic technique, which uses four counters (produced tokens, consumed tokens, missing tokens, and remaining tokens) to compute the fitness of an observation trace based on a given process model [143]. A trace is fitting according to the model if, during

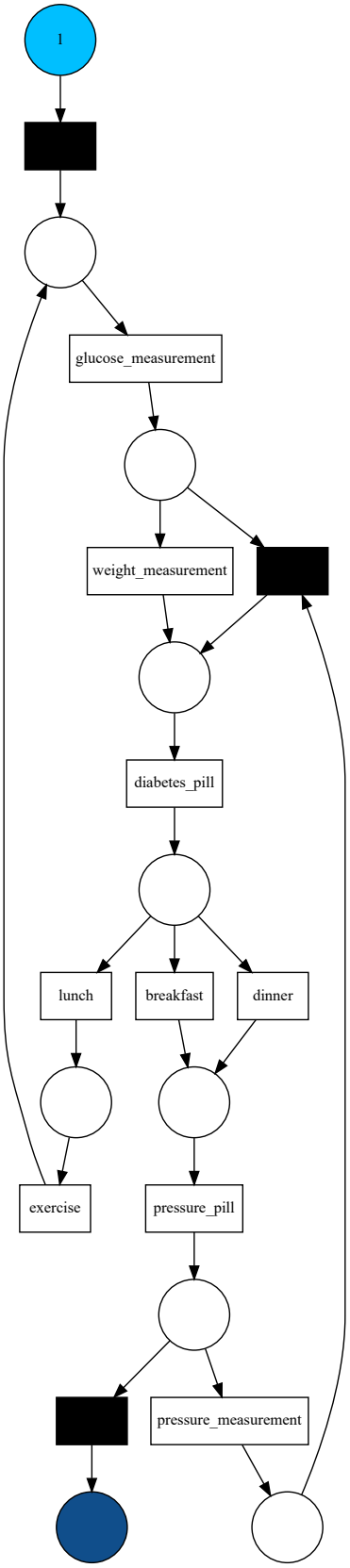

Figure 4.6: Modeling of Petri net related to the metabolic syndrome in-home prescription using Heuristic Miner algorithm.

its execution, the transitions can be fired without the need to insert any missing token. If the reaching of the final marking is imposed, then a trace is fitting if it reaches the final marking without any missing or remaining tokens.

Alignment-based replay is a technique, which performs an exhaustive search to find out the optimal alignment between the observed trace and the process model. Hence, it is guaranteed to return the closest model run in comparison to the trace [163].

It is possible to compare the behavior contained in the log and the behavior contained in the model, in order to see if and how they match. Four different dimensions exist in process mining, including replay fitness, precision, generalization, and simplicity.

- *Replay Fitness*: The calculation of the replay fitness aim to calculate how much of the behavior in the log is admitted by the process model. For token-based replay, the percentage of traces that are completely fit is returned, along with a fitness value that is calculated [25]. For alignments, the percentage of traces that are completely fit is returned, along with a fitness value that is calculated as the average of the fitness values of the single traces.

- *Precision*: The different prefixes of the log are replayed (whether possible) on the model. At the reached marking, the set of transitions that are enabled in the process model is compared with the set of activities that follow the prefix. The more the sets are different, the more the precision value is low. The more the sets are similar, the more the precision value is high. This works only if the replay of the prefix on the process model works: if the replay does not produce a result, the prefix is not considered for the computation of precision. Hence, the precision calculated on top of unfit processes is not really meaningful. There exist two approaches for the measurement of precision: *ETConformance* (using token-based replay) [121], and *Align-ETConformance* (using alignments) [3].

- *Generalization*: A model is general whether the elements of the model are visited enough often during a replay operation (of the log on the model). A model may be perfectly fitting the log and perfectly precise. Hence, to measure generalization a token-based replay operation is performed, and the generalization is calculated [34].

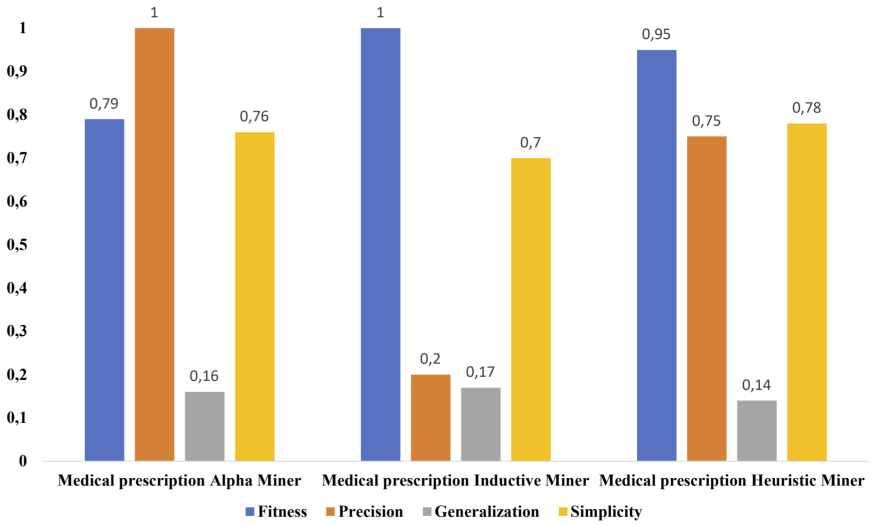

Figure 4.7: Evaluation metrics based on Medical Prescription log.

- *Simplicity*: A dimension that evaluates how simple the process model is to understand for a human. It is defined taking into account only the Petri net model; the criteria used for simplicity is the inverse arc degree [31].

4.3.4.2 *Model Evaluation*

For a better understanding of the evaluation phase, the results, for each of the three process mining algorithms, have been reported on the histogram in Figure 4.7.

In our investigation, it is crucial that a patient follows the doctor's prescription perfectly and should not reverse, introduce or forget activities. A small note can be made about certain activities, such as those relating to measuring blood pressure or blood sugar as if they were reversed they would not involve anything serious, compared to forgetting to take a drug, for example. This means that the evaluation metrics we want to be maximized are precision and replay fitness.

In particular, these two metrics behave differently depending on the inspected process mining algorithms. For a deeper understanding of how much patients have been adherent to the medical prescription for metabolic syndrome treatment, it is in general evident how the replay fitness addresses this compliance checking. Replay fitness is above 79% in all process mining modeling algorithms, suggesting that most of the patients observed the model behavior correctly. Precision, instead, is quite swinging, showing that, although the fitness reaches good results for all the algorithms, Alpha Miner and Inductive Miner algorithms

show opposite behaviors. This means that they do not rightly and fairly represent the actual behavior of patients. In fact, they are not able to capture some wrong activities, such as revers actions in the medical prescription, even though those actions are included in the medical prescription. A good trade-off between replay fitness and precision is given by the Heuristic Miner. It gives a more accurate representation of the patients' actual behavior. Selecting the Heuristic Miner as best model representation for conformance checking, it is possible to note that the 95% of patients followed the medical prescription entirely, while the 75% of them followed the medical prescription precisely.

The generalization metric aims to maximize model-supported behaviors that are not part of the system and are not present in the event log. This, in our case, is not good as the patient must follow the activities present in the medical prescription, without introducing new activities unless recommended by the doctor. As shown in Figure 4.7, the generalization values are typically low, below the 17% for all process mining algorithms, witnessing that model does not generalize. This is good news for our scenario, as it means that patients do not perform activities as different as those prescribed by the doctor

Simplicity assumes less importance as it is not linked to the behavior of the process model obtained but to its simplicity in terms of network and therefore of understanding. The value of simplicity for all algorithms used on the medical prescription event log is a high indication that the model processed by the three algorithms is quite simple.

Part III

CONCLUSION

In other words, an admission of where I have actually come to. Everybody can say you should go faster or go slower. But you have arrived, with your own legs. Conclusion is my "*Look back and ahead: what you have seen and what you see?*".

5

CLOSING REMARKS

This dissertation started with an analysis of the four most mentioned topics of our work: *monitoring system*, *security*, *explainability* and *security problem*. We ended up synthesizing all of them. The bunch of problems that motivated this three-year work have given us a chance not only to study and propose some potential solutions but also to unveil new opportunities.

In Chapter 3, we first introduced the reference domain of remote home care: Ambient Assisted Living ()AAL). It proved to be the choice in which to build an architecture for this purpose. It was essential to define the *Clinical Path (CP)* and understand how it can be technologically modelled. We also analyzed *Artificial Intelligence (AI)* and *Process Mining (PM)* algorithmic approaches that are most responsive to solving safety problems during remote patient monitoring. Furthermore, we described the *eXplainable Artificial Intelligence Security (XSec*) approach used to explain anomalies during monitoring.

Secondly, in Chapter 4, motivated by a general *Machine Learning (ML)* problem, we studied how to make remote patient monitoring more efficient. Sensor data can be manipulated for a variety of reasons. Moreover, if not correctly explained, the doctor can misinterpret them. We also presented two use cases concerning a patient's remote monitoring. In the first case, we collect ECG signals, and through the *Clinical Path Anomaly Detection (CPAD)*, we can tell when the data are accurate or compromised. We also use an Explainability paradigm to best interpret the collected data. In the second use case, by adopting *Conformance Checking* on a patient suffering from Metabolic Syndrome, we can model a system that can identify when the patient's behavior deviates from that of the doctor and, thus, from the actual prescription.

This is what we have tried to do, or instead, I have tried to do with the help, support, collaboration, and encouragement of my colleagues at the university and in the company. I tried to contribute to the research, aware that what I have done is not even a drop in the ocean. I have tried to consider research as an instrument of knowledge, not an object of competition and an agent of power, which is why my contributions are never without errors. Not for nothing, this three-year adventure of scientific research, although almost constantly guided by reasoning,

has always been an adventure. I conclude with a sentence from *Albert Einstein* that says: *"There is a passion for understanding just as there is a passion for music; it is a very common passion in children but one that most adults lose. Without it, there would be neither mathematics nor the other sciences."*

Part IV

APPENDIX

A

APPENDIX

A.1 UTILS TABLES

Table A.1: MRC, CAR, AVA VAN and assurance levels.

MRC	AP	CAR	AVA_VAN	EU Assurance Level
MRC1	AP1 Basic	CAR1	AVA_VAN.1	Substantial
MRC2	AP1 Basic	CAR2	AVA_VAN.2	Substantial
MRC3	AP2 Enhanced Basic	CAR3	AVA_VAN.3	High
MRC4	AP3 Moderate	CAR4	AVA_VAN.4	High
MRC5	AP4 High	CAR5	AVA_VAN.5	High
NA	AP5 Beyond High	NA	NA	NA

Table A.2: Elements of the Sector Specific Scheme.

ELEMENT	PROPOSITION FOR SECTOR DEFINITION
TOE categories	Asset Taxonomy of specific sector
Rules and Procedures	EU certification scheme, plus sector experts' customization
SFR/TR	Baseline according to CCP2, integration with ISO27001:2022 and SSR. Mapped in CSL
Standards	Sector specific, to be used in SFRs or as OSPs
Protection Profile (PP)	To be designed for a TOE category: it is comprehensive of SPD, SO, SFR, SAR
SPD	Core of PP, composed by OSP, assumptions and Threats
Organisational Security Policies (OSP)	Proposed baseline according to the OSC of the ISO27001:2022, SSS, PR and NIS directive
Assumptions	Proposed baseline according to sector experts
Threats	Proposed baseline according to sector experts, ENISA reports, sectoral risk assessment
Security Objectives (SO)	High level baseline from the CA
Security Assurance Requirements (SAR)	Proposed baseline according to CCP 3
Evaluation Methodology	Use of Cyber Range for Conformity Assessment

Table A.3: CSL and risk-based approach.

CSL	Description	Risk level	AP of attacker	Associated EAL	AVA VAN CLASS	CAR
CSL1	CSL 1 provides a basic level of security against unskilled adversaries.	MRC1	AP1 - Basic	EAL1	AVA_VAN.1 Vulnerability Survey	CAR 1
CSL2	CSL 2 adds requirements to CSL1, providing security against skilled adversaries with limited resources and opportunity to attack a system.	MRC2	AP1 - Basic	EAL2	AVA.VAN.2 Vulnerability Analysis	CAR 2
CSL3	CSL 3 extends the coverage of the security against skilled adversaries with significant resources and/or significant opportunity to attack a system.	MRC3	AP2 - Enhanced - Basic	EAL4	AVA.VAN.3 Focused Vulnerability Analysis	CAR 3
CSL4	CSL 4 provides security against a highly skilled adversary with significant resources and opportunity.	MRC4	AP3 - Moderate	EAL5	AVA_VAN.4 Methodical Vulnerability Analysis	CAR 4
CSL5	CSL 5 provides the highest level of security, capable of protecting against highly sophisticated adversaries with significant resources at their disposal and/or opportunity for an attack.	MRC5	AP4 - High	EAL6	AVA_VAN.5 Advanced Methodical Vulnerability Assessment	CAR 5
NA	NA	NA	AP5 - Beyond High	NA	NA	NA

Table A.4: Mapping of Evaluation Assurance Levels between Cybersecurity Act and Common Criteria.

Item Number	Mapping of Evaluation Assurance Levels		
	Cybersecurity Act	**Common Criteria (AVA_VAN)**	**Common Criteria Corresponding EAL**
1	Basic	-	-
2	Substantial	AVA_VAN.1	EAL 1
			EAL 2
		AVA_VAN.2	EAL 3
3	High	AVA_VAN.3	EAL 4
			EAL 5
		AVA_VAN.4	EAL 6
		AVA_VAN.5	EAL 7

Table A.5: Detailed SO.

CSA OBJECTIVES (Proposed naming)	**DETAILED POSSIBLE SO**
O.DATA_CONFIDENTIALITY	O.SECURE_ STORAGE O.SECURE_COMMUNICATIONS O.AUTHORIZED_ACCESS O.AUTHORIZED_PROCESSING
O.DATA_AVAILABILITY	O.SECURE_ BACKUP O.RELIABLE_TRANSMISSION O.ACCESSIBLE_DATA
O.DATA_INTEGRITY	O.INTEGRITY_CHECK O.SECURE_COMMUNICATIONS
O.ACCESS_CONTROL	O.STRONG_ROLES_MNG O.ACCOUNTABILITY O.TRUSTED_CONNECTION
O.VULNERABILITIES_ANALYSIS	O.UPDATED_COMPONENTS
O.EVENT_LOGGING	O.CONTINUOUS_MONITORING O.ACCOUNTABILITY
O.LOG_MANAGEMENT	O.SECURITY_AUDIT O.SECURE_LOGS
O.VULNERABILITY_MANAGEMENT	O.SW_TESTING
O.BUSINESS_CONTINUITY	O.SERVICE_RESILIENCE O.HW_MAINTENANCE O.SECURE_FAIL
O. SECURITY_BY_DESIGND&DEFAULT	O.SECURE_COMMUNICATION O.TRUSTED_COMPONENTS O.SECURE_ACCESS O.SECURE_PROCESSING O.ACCOUNTABILITY O.SECURE_ARCHITECTURE O.SECURE_INSTALLATION O.PRIVACY_SETTINGS O.DATA_MINIMISATION O.PSEUDONYMISATION O.PHYSICAL_SECURITY
O.SECURE_SW_ DEVELOPMENT_ AND_MAINTENANCE	O.SW_TESTING O.SW_MAINTENANCE

Figure A.1: PACS Specific Security Functional Requirements, Components are marked with numbers.

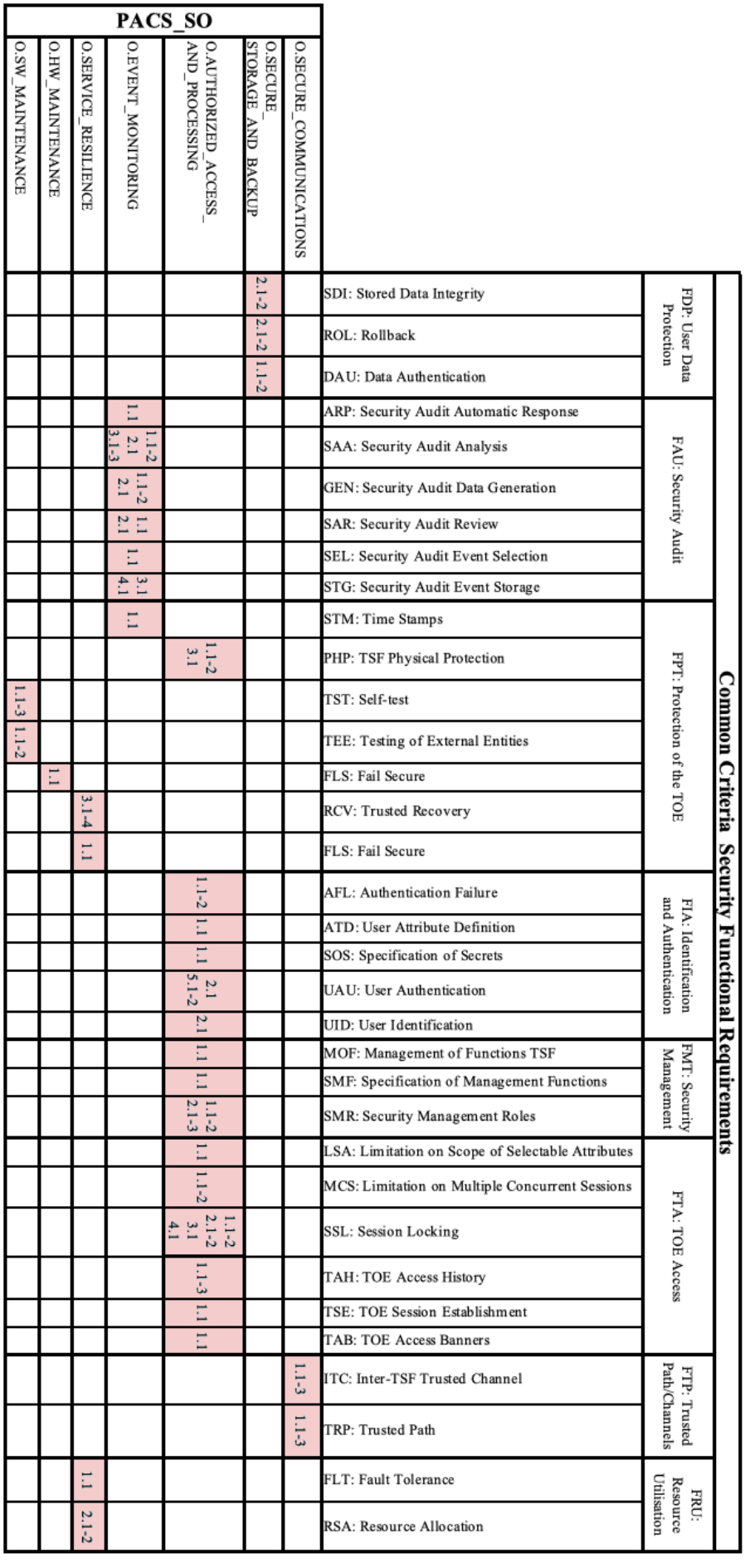

Table A.6: Summary table used for the outcome of the sectoral risk assessment.

Business Process	Services	Asset Category	Applicable Threat	MRC	AP	CAR	CSL
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...

BIBLIOGRAPHY

[1] Alaa Awad Abdellatif, Mohammad Galal Khafagy, Amr Mohamed, and Carla-Fabiana Chiasserini. "EEG-Based Transceiver Design With Data Decomposition for Healthcare IoT Applications." In: *IEEE Internet Things J.* 5.5 (2018), pp. 3569–3579.

[2] Alaa Awad Abdellatif, Amr Mohamed, Carla Fabiana Chiasserini, Mounira Tlili, and Aiman Erbad. "Edge computing for smart health: Context-aware approaches, opportunities, and challenges." In: *IEEE Network* 33.3 (2019), pp. 196–203.

[3] Arya Adriansyah, Jorge Munoz-Gama, Josep Carmona, Boudewijn F Van Dongen, and Wil MP Van Der Aalst. "Measuring precision of modeled behavior." In: *Information systems and e-Business Management* 13.1 (2015), pp. 37–67.

[4] Neda AfzaliSeresht, Qing Liu, and Yuan Miao. "An explainable intelligence model for security event analysis." In: *Australasian Joint Conference on Artificial Intelligence*. Springer. 2019, pp. 315–327.

[5] Charu C Aggarwal. "An introduction to outlier analysis." In: *Outlier analysis*. Springer, 2017, pp. 1–34.

[6] Mohiuddin Ahmed, Abdun Naser Mahmood, and Jiankun Hu. "A survey of network anomaly detection techniques." In: *Journal of Network and Computer Applications* 60 (2016), pp. 19–31.

[7] Mohiuddin Ahmed, Abdun Naser Mahmood, and Md Rafiqul Islam. "A survey of anomaly detection techniques in financial domain." In: *Future Generation Computer Systems* (2016).

[8] Leman Akoglu, Hanghang Tong, and Danai Koutra. "Graph based anomaly detection and description: a survey." In: *Data mining and knowledge discovery* 29.3 (2015), pp. 626–688.

[9] K George MM Alberti, Paul Zimmet, and Jonathan Shaw. "The metabolic syndrome—a new worldwide definition." In: *The Lancet* 366.9491 (2005), pp. 1059–1062.

[10] Arwa Alrawais, Abdulrahman Alhothaily, Chunqiang Hu, and Xiuzhen Cheng. "Fog computing for the internet of things: Security and privacy issues." In: *IEEE Internet Computing* 21.2 (2017), pp. 34–42.

[11] Carmelo Ardito, Francesco Bellifemine, Tommaso Di Noia, Domenico Lofu, and Giulio Mallardi. "A proposal of case-based approach to clinical pathway modeling support." In: *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*. IEEE. 2020, pp. 1–6.

[12] Carmelo Ardito, Tommaso Di Noia, Eugenio Di Sciascio, Domenico Lofú, Giulio Mallardi, Claudio Pomo, and Felice Vitulano. "Towards a Trustworthy Patient Home-Care Thanks to an Edge-Node Infrastructure." In: *International Conference on Human-Centred Software Engineering*. Springer. 2020, pp. 181–189.

[13] Carmelo Ardito, Tommaso Di Noia, Corrado Fasciano, Domenico Lofù, Nicola Macchiarulo, Giulio Mallardi, Andrea Pazienza, and Felice Vitulano. "An Edge Ambient Assisted Living Process for Clinical Pathway." In: *Italian Forum of Ambient Assisted Living*. Springer. 2022, pp. 363–374.

[14] Carmelo Ardito, Tommaso Di Noia, Corrado Fasciano, Domenico Lofú, Nicola Macchiarulo, Giulio Mallardi, Andrea Pazienza, and Felice Vitulano. "Towards a Situation Awareness for eHealth in Ageing Society." In: *Proceedings of the Italian Workshop on Artificial Intelligence for an Ageing Society 2020 (AIxAS), co-located with 19th International Conference of the Italian Association for Artificial Intelligence (AIxIA 2020)*. 2020, pp. 40–55.

[15] Carmelo Ardito, Tommaso Di Noia, Corrado Fasciano, Domenico Lofù, Nicola Macchiarulo, Giulio Mallardi, Andrea Pazienza, and Felice Vitulano. "Management at the edge of situation awareness during patient telemonitoring." In: *International conference of the italian association for artificial intelligence*. Springer. 2020, pp. 372–387.

[16] Carmelo Arditob, Tommaso Di Noiab, Corrado Fascianoa, Domenico Lofùa, Nicola Macchiaruloa, Giulio Mallardia, Andrea Pazienzaa, and Felice Vitulanoa. "Towards a Situation Awareness for eHealth in Ageing Society." In: (2020).

[17] Alaa Awad, Amr Mohamed, Carla-Fabiana Chiasserini, and Tarek M. El-Fouly. "Distributed in-network processing and resource optimization over mobile-health systems." In: *J. Netw. Comput. Appl.* 82 (2017), pp. 65–76.

[18] Alaa Awad, Amr Mohamed, Carla-Fabiana Chiasserini, and Tarek Elfouly. "Distributed in-network processing and resource optimization over mobile-health systems." In: *Journal of Network and Computer Applications* 82 (2017), pp. 65–76.

[19] R Can Aygun and A Gokhan Yavuz. "Network anomaly detection with stochastically improved autoencoder based models." In: *2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud)*. 2017, pp. 193–198.

[20] Fatemeh Azmandian, Ayse Yilmazer, Jennifer G Dy, Javed A Aslam, and David R Kaeli. "GPU-accelerated feature selection for outlier detection using the local kernel density ratio." In: *2012 IEEE 12th International Conference on Data Mining*. IEEE. 2012, pp. 51–60.

[21] Pierre Baldi. "Autoencoders, unsupervised learning, and deep architectures." In: *Proceedings of ICML workshop on unsupervised and transfer learning*. JMLR Workshop and Conference Proceedings. 2012, pp. 37–49.

[22] Dor Bank and Raja Giryes. "An ETF view of Dropout regularization." In: *arXiv preprint arXiv:1810.06049* (2018).

[23] Dor Bank, Noam Koenigstein, and Raja Giryes. "Autoencoders." In: *arXiv preprint arXiv:2003.05991* (2020).

[24] Gabriel Bender, Lucja Kot, and Johannes Gehrke. "Explainable security for relational databases." In: *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. 2014, pp. 1411–1422.

[25] Alessandro Berti and Wil MP van der Aalst. "Reviving Token-based Replay: Increasing Speed While Improving Diagnostics." In: *ATAED@ Petri Nets/ACSD*. 2019, pp. 87–103.

[26] Alessandro Berti, Sebastiaan J. van Zelst, and Wil M. P. van der Aalst. "PM4Py Web Services: Easy Development, Integration and Deployment of Process Mining Features in any Application Stack." In: *Proceedings of the Dissertation Award, Doctoral Consortium, and Demonstration Track at BPM 2019 co-located with 17th International Conference on Business Process Management, BPM 2019*. 2019, pp. 174–178.

[27] Sonu Bhaskar, Sian Bradley, Sateesh Sakhamuri, Sebastian Moguilner, Vijay Kumar Chattu, Shawna Pandya, Starr Schroeder, Daniel Ray, and Maciej Banach. "Designing futuristic telemedicine using artificial intelligence and robotics in the COVID-19 era." In: *Frontiers in public health* 8 (2020), p. 708.

[28] Monowar H Bhuyan, Dhruba Kumar Bhattacharyya, and Jugal K Kalita. "Network anomaly detection: methods, systems and tools." In: *IEEE Communications Surveys & Tutorials* (2013).

[29] Monowar H Bhuyan, Dhruba Kumar Bhattacharyya, and Jugal K Kalita. "Network anomaly detection: methods, systems and tools." In: *Ieee communications surveys & tutorials* 16.1 (2013), pp. 303–336.

[30] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*. Vol. 4. 4. Springer, 2006.

[31] Fabian Rojas Blum. "Metrics in process discovery." In: *Tech. Rep. Technical Report TR/DCC-2015-6, Computer Science Dept., University of Chile* (2015).

[32] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. "LOF: identifying density-based local outliers." In: *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. 2000, pp. 93–104.

[33] Anna L Buczak and Erhan Guven. "A survey of data mining and machine learning methods for cyber security intrusion detection." In: *IEEE Communications surveys & tutorials* 18.2 (2015), pp. 1153–1176.

[34] Joos CAM Buijs, Boudewijn F van Dongen, and Wil MP van der Aalst. "Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity." In: *International Journal of Cooperative Information Systems* 23.01 (2014), p. 1440001.

[35] Guilherme O Campos, Arthur Zimek, Jörg Sander, Ricardo JGB Campello, Barbora Micenková, Erich Schubert, Ira Assent, and Michael E Houle. "On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study." In: *Data mining and knowledge discovery* 30.4 (2016), pp. 891–927.

[36] Longbing Cao. "Coupling learning of complex interactions." In: *Information Processing & Management* 51.2 (2015), pp. 167–186.

[37] Piero Cappelletti. "Care pathways and Laboratory Medicine." In: *Italian Journal of Laboratory Medicine* 13.2 (2017), pp. 65–71. ISSN: 2039-6821.

[38] Piero Cappelletti. "PDTA e Medicina di Laboratorio." In: *La Rivista Italiana della Medicina di Laboratorio-Italian Journal of Laboratory Medicine* 13.2 (2017), pp. 65–71.

[39] Filip Caron, Jan Vanthienen, Kris Vanhaecht, Erik Limbergen, Jochen Weerdt, and Bart Baesens. "A Process Mining-Based Investigation of Adverse Events in Care Processes." In: *The HIM journal* 43 (Oct. 2013).

[40] Varun Chandola, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey." In: *ACM computing surveys (CSUR)* 41.3 (2009), pp. 1–58.

[41] Qian Chen and Robert A Bridges. "Automated behavioral analysis of malware: A case study of wannacry ransomware." In: *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE. 2017, pp. 454–460.

[42] Zhaomin Chen, Chai Kiat Yeo, Bu Sung Lee, and Chiew Tong Lau. "Autoencoder-based network anomaly detection." In: *2018 Wireless Telecommunications Symposium (WTS)*. 2018, pp. 1–5.

[43] Mung Chiang and Tao Zhang. "Fog and IoT: An overview of research opportunities." In: vol. 3. 6. IEEE, 2016, pp. 854–864.

[44] Ivalyo Christov and G Bortolan. "Ranking of pattern recognition parameters for premature ventricular contractions classification by neural networks." In: *Physiological Measurement* 25.5 (2004), p. 1281.

[45] Catalin Cimpanu. "First death reported following a ransomware attack on a German hospital." In: *ZDNet* (2020), pp. 16–17.

[46] . "Common Criteria for Information Technology Security Evaluation: Part 2. Security Functional Components, Common Criteria, Apr. 2017." In: ().

[47] . "Common Criteria for Information Technology Security Evaluation: Part 3 Security assurance components, Common Criteria, Apr. 2017." In: ().

[48] Colabuono Consuelo et al. "Approach to sector-specific Cybersecurity Schemes: key elements and Security Problem Definition." In: *2022 International Conference on Multimedia Communications, Services and Security (MCSS).*

[49] . *Council Directive 93/42/EEC of 14 June 1993 Concerning Medical Devices, EU Council, 14-Jun-1993*. URL: https://eurlex.europa.eu/legal-content/EN/TXT/?uri=CELEX$%5C%$3A31993L0042.

[50] Maria Manuela Cruz-Cunha. "Handbook of research on ICTs for human-centered healthcare and social care services." In: (2013).

[51] . "D2.4 Inter-sector Technology Challenges and Opportunities, European network of Cybersecurity centres and competence Hub for innovation and Operations, Jan. 2020." In: ().

[52] . "D2.5 Multi-sector Requirements Definition and Demonstration Cases, European Network of Cybersecurity centres and competence Hub for innovation and Operations, Mar. 2020." In: ().

[53] . "D4.2 Inter-sector Technical Cybersecurity Challenges Report, European Network of Cybersecurity centres and competence Hub for innovation and Operations, Jun. 2020." In: ().

[54] Shubhomoy Das, Weng-Keen Wong, Thomas Dietterich, Alan Fern, and Andrew Emmott. "Incorporating expert feedback into active anomaly discovery." In: *2016 IEEE 16th International Conference on Data Mining (ICDM)*. 2016, pp. 853–858.

[55] Emmelien De Roock and Niels Martin. "Process mining in healthcare–an updated perspective on the state of the art." In: *Journal of Biomedical Informatics* (2022), p. 103995.

[56] . *Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices, 1998*. URL: https://eur-lex.europa.eu/legalcontent/EN/TXT/HTML/?uri=CELEX:31998L0079.

[57] Onur Dogan. "Process mining for check-up process analysis." In: *IIOABJ* 9.6 (2018), pp. 56–61.

[58] Danny Dolev and Andrew Yao. "On the security of public key protocols." In: *IEEE Transactions on information theory* 29.2 (1983), pp. 198–208.

[59] . "European Cyber Security Certification: A Meta-Scheme Approach v1.0." In: *ECSO* (2017).

[60] . "European Cyber Security Certification: Challenges ahead for the roll-out of the Cybersecurity Act." In: *ECSO* (2020).

[61] Otto Fabius, Joost R. van Amersfoort, and Diederik P. Kingma. "Variational Recurrent Auto-Encoders." In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Workshop Track Proceedings*. Ed. by Yoshua Bengio and Yann LeCun. 2015.

[62] OM Fal. "Documentation in the ISO/IEC 27701 Standard." In: *Cybernetics and Systems Analysis* 57.5 (2021), pp. 796–802.

[63] Nicolas Falliere, Liam O Murchu, and Eric Chien. "W32. stuxnet dossier." In: *White paper, symantec corp., security response* 5.6 (2011), p. 29.

[64] Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss. *Computer Vision–ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part V*. Vol. 11209. Springer, 2018.

[65] Hossein Fotouhi, Aida Causevic, Kristina Lundqvist, and Mats Björkman. "Communication and Security in Health Monitoring Systems–A Review." In: *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*. Vol. 1. IEEE. 2016, pp. 545–554.

[66] Maria Fox, Derek Long, and Daniele Magazzeni. "Explainable planning." In: *arXiv preprint arXiv:1709.10256* (2017).

[67] Eugene Freuder. "Explaining ourselves: human-aware constraint reasoning." In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 31. 1. 2017.

[68] L Fusar-Poli, N Brondino, M Rocchetti, M Ballerio, M Vercesi, F Grasso, and P Politi. "Prevalence and predictors of metabolic syndrome in a sample of Italian psychiatric inpatients." In: *European Psychiatry* 41.S1 (2017), S472–S472.

[69] M. Geiger, S. Harrer, J. Lenhard, and G. Wirtz. "On the Evolution of BPMN 2.0 Support and Implementation." In: *2016 IEEE Symposium on Service-Oriented System Engineering (SOSE)*. 2016, pp. 101–110.

[70] D Goodin. "A Patient Dies After a Ransomware Attack Hits a Hospital." In: *Wired* (2020).

[71] Gianluigi Greco, Antonella Guzzo, Francesco Lupia, and Luigi Pontieri. "Process Discovery under Precedence Constraints." In: *ACM Trans. Knowl. Discov. Data* 9.4 (2015).

[72] Frank E Grubbs. "Procedures for detecting outlying observations in samples." In: *Technometrics* 11.1 (1969), pp. 1–21.

[73] Maheedhar Gunasekharan, Samik Basu, and Ganesh Ram Santhanam. "Selecting the minimal set of preferred responses to counter detected intrusions." In: *Proceedings of the 12th Annual Conference on Cyber and Information Security Research*. 2017, pp. 1–8.

[74] Anar A Hady, Ali Ghubaish, Tara Salman, Devrim Unal, and Raj Jain. "Intrusion detection system for healthcare systems using medical and network data: A comparison study." In: *IEEE Access* 8 (2020), pp. 106576–106584.

[75] Shah Ahsanul Haque, Mustafizur Rahman, and Syed Mahfuzul Aziz. "Sensor anomaly detection in wireless sensor networks for healthcare." In: *Sensors* (2015).

[76] *. Health Insurance Portability and Accountability Act of 1996 (HIPAA), The US Department of Health and Human Services (HHS), 1996.* URL: https://www.cdc.gov/phlp/publications/topic/hipaa.html.

[77] Alexander Hinneburg, Daniel A. Keim, and Markus Wawryniuk. "HD-Eye: visual mining of high-dimensional data." In: *IEEE Computer Graphics and Applications* 19.5 (1999), pp. 22–31.

[78] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory." In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[79] Andreas Holzinger, Georg Langs, Helmut Denk, Kurt Zatloukal, and Heimo Müller. "Causability and explainability of artificial intelligence in medicine." In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 9.4 (2019), e1312.

[80] Kristine Hovhannisyan, Piotr Bogacki, Consuelo Assunta Colabuono, Domenico Lofù, Maria Vittoria Marabello, and Brady Eugene Maxwell. "Towards a Healthcare Cybersecurity Certification Scheme." In: *2021 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*. IEEE. 2021, pp. 1–9.

[81] Zhengxing Huang, Xudong Lu, and Huilong Duan. "Anomaly detection in clinical processes." In: *AMIA Annual Symposium Proceedings*. 2012.

[82] . *IEC 62304 - Medical device software — Software life cycle processes*. URL: https://www.iso.org/obp/ui/#iso:std:iec:62304:ed-1:v1:en.

[83] . *International Medical Device Regulators Forum (IMDRF)*. URL: https://www.fda.gov/medical-devices/cdrhinternational-programs/international-med%20ical-device-regulatorsforum-imdrf.

[84] . *ISO 13485 - Medical devices — Quality management systems — Requirements for regulatory purposes*. URL: https://www.iso.org/standard/59752.html.

[85] . *ISO 14971 - Medical devices — Application of risk management to medical devices*. URL: https://www.iso.org/standard/72704.html.

[86] . *ISO 27799 - Health informatics — Information security management in health using ISO/IEC 27002*. URL: https://www.iso.org/standard/62777.html.

[87] . *ISO 9001 - Quality management systems — Requirements*. URL: https://www.iso.org/obp/ui/#iso:std:iso:9001:ed5:v1:en.

[88] . *ISO/IEC 20000 - Information technology — Service management*. URL: https://www.iso.org/standard/51986.html.

[89] . *ISO/IEC 27000 - Information technology — Security techniques — Information security management systems*. URL: https://www.iso.org/standard/73906.html.

[90] . *ISO/IEC 27001- Information technology — Security techniques — Information security management systems — Requirements*. URL: https://www.iso.org/standard/54534.html.

[91] . *ISO/IEC 27002 - Information technology — Security techniques — Code of practice for information security controls.* URL: https://www.iso.org/standard/54533.html.

[92] . *IT Health Check (ITHC): supporting guidance.* URL: https://www.gov.uk/government/publications/it-health-check-ithcsupporting-guidan%20ce.

[93] P. Jain, A. Agarwal, and R. Behara. "Care Coordination: A Systematic Review and a New Perspective." In: *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*. 2017, pp. 531–536.

[94] Da-Yu Kao and Shou-Ching Hsiao. "The dynamic analysis of WannaCry ransomware." In: *2018 20th International conference on advanced communication technology (ICACT)*. IEEE. 2018, pp. 159–166.

[95] Fabian Keller, Emmanuel Muller, and Klemens Bohm. "HiCS: High contrast subspaces for density-based outlier ranking." In: *2012 IEEE 28th international conference on data engineering*. IEEE. 2012, pp. 1037–1048.

[96] Ae Chan Kim, Won Hyung Park, and Dong Hoon Lee. "A study on the live forensic techniques for anomaly detection in user terminals." In: *International Journal of Security and Its Applications* 7.1 (2013), pp. 181–188.

[97] Ryan D Kindle, Omar Badawi, Leo Anthony Celi, and Shawn Sturland. "Intensive care unit telemedicine in the era of big data, artificial intelligence, and computer clinical decision support systems." In: *Critical care clinics* 35.3 (2019), pp. 483–495.

[98] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes." In: *arXiv preprint arXiv:1312.6114* (2013).

[99] M. Kreuseler, T. Nocke, and H. Schumann. "A History Mechanism for Visual Data Mining." In: *IEEE Symposium on Information Visualization*. 2004, pp. 49–56.

[100] Prabhat Kumar, Govind P Gupta, and Rakesh Tripathi. "An ensemble learning and fog-cloud architecture-driven cyber-attack detection framework for IoMT networks." In: *Computer Communications* 166 (2021), pp. 110–124.

[101] Ruggero Lanotte, Massimo Merro, Riccardo Muradore, and Luca Viganò. "A formal approach to cyber-physical attacks." In: *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*. IEEE. 2017, pp. 436–450.

[102] Aleksandar Lazarevic and Vipin Kumar. "Feature bagging for outlier detection." In: *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. 2005, pp. 157–166.

[103] Sander JJ Leemans, Dirk Fahland, and Wil MP Van Der Aalst. "Discovering block-structured process models from event logs-a constructive approach." In: *International conference on applications and theory of Petri nets and concurrency*. Springer. 2013, pp. 311–329.

[104] Richard Lenz and Manfred Reichert. "IT Support for Healthcare Processes – Premises, Challenges, Perspectives." In: *Data and Knowledge Engineering* 61 (May 2007), pp. 39–58.

[105] Kingsly Leung and Christopher Leckie. "Unsupervised anomaly detection in network intrusion detection using clusters." In: *Proceedings of the Twenty-eighth Australasian conference on Computer Science-Volume 38*. 2005, pp. 333–342.

[106] Weizi Li, Kecheng Liu, Hongqiao Yang, and Changrui Yu. "Integrated clinical pathway management for medical quality improvement - based on a semiotically inspired systems architecture." In: *European Journal of Information Systems* 23.4 (2014), pp. 400–417. ISSN: 1476-9344.

[107] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. "Isolation-based anomaly detection." In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 6.1 (2012), pp. 1–39.

[108] WP Lord, NJ Mankovich, et al. "PACS/information systems interoperability using Enterprise Communication Framework." In: *IEEE transactions on information technology in biomedicine* 2.2 (1998), pp. 42–47.

[109] Francisco J Lucas, Fernando Molina, and Ambrosio Toval. "A systematic review of UML model consistency management." In: *Information and Software technology* 51.12 (2009), pp. 1631–1645.

[110] Yuan Luo, Ya Xiao, Long Cheng, Guojun Peng, and Danfeng Daphne Yao. "Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities." In: *arXiv preprint arXiv:2003.13213* (2020).

[111] F. M. Maggi, A. J. Mooij, and W. M. P. van der Aalst. "User-guided discovery of declarative process models." In: *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. 2011, pp. 192–199.

[112] IPL McLaren and NJ Mackintosh. "An elemental model of associative learning: I. Latent inhibition and perceptual learning." In: *Animal Learning & Behavior* 28.3 (2000), pp. 211–246.

[113] Yair Meidan, Michael Bohadana, Yael Mathov, Yisroel Mirsky, Asaf Shabtai, Dominik Breitenbacher, and Yuval Elovici. "N-baiot—network-based detection of iot botnet attacks using deep autoencoders." In: *IEEE Pervasive Computing* (2018).

[114] Oleg Metsker, Sergey Kesarev, Ekaterina Bolgova, Kirill Golubev, Andrey Karsakov, Alexey Yakovlev, and Sergey Kovalchuk. "Modelling and analysis of complex patient-treatment process using graphminer toolbox." In: *International Conference on Computational Science*. Springer. 2019, pp. 674–680.

[115] Petr Mlynek, Radek Fujdiak, Pavel Mrnustik, Bohuslav Krena, and Ludovic Apvrille. "Co-Engineering Gap Analysis of ANSI/ISA-62443-3-3." In: *International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems* 9.1 (2020), pp. 1–9.

[116] Amyas Morse. "Investigation: WannaCry cyber attack and the NHS." In: *Report by the National Audit Office. Accessed* 1 (2018).

[117] Hassnaa Moustafa, Eve M Schooler, Gang Shen, and Sanjana Kamath. "Remote monitoring and medical devices control in eHealth." In: *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE. 2016, pp. 1–8.

[118] Khalil Muhammad, Aonghus Lawlor, and Barry Smyth. "Explanation-based ranking in opinionated recommender systems." In: *The 24th Irish Conference on Artificial Intelligence and Cognitive Science, University College Dublin, Ireland, 20-21 September 2016*. CEUR Workshop Proceedings. 2018.

[119] Biswanath Mukherjee, L Todd Heberlein, and Karl N Levitt. "Network intrusion detection." In: *IEEE network* 8.3 (1994), pp. 26–41.

[120] Jorge Munoz-Gama et al. *Conformance checking and diagnosis in process mining*. Springer, 2016.

[121] Jorge Munoz-Gama and Josep Carmona. "A fresh look at precision in process conformance." In: *International Conference on Business Process Management*. Springer. 2010, pp. 211–226.

[122] Vahideh Naderifar, Shahnorbanun Sahran, and Zarina Shukur. "A review on conformance checking technique for the evaluation of process mining algorithms." In: *TEM Journal* 8.4 (2019), p. 1232.

[123] George C Necula. "Proof-carrying code." In: *Proceedings of the 24th ACM SIGPLAN-SIGACT symposium on Principles of programming languages*. 1997, pp. 106–119.

[124] Jacquelyn K O'herrin, Norman Fost, and Kenneth A Kudsk. "Health Insurance Portability Accountability Act (HIPAA) regulations: effect on medical record research." In: *Annals of surgery* (2004).

[125] Jacquelyn K O'herrin, Norman Fost, and Kenneth A Kudsk. "Health Insurance Portability Accountability Act (HIPAA) regulations: effect on medical record research." In: *Annals of surgery* 239.6 (2004), p. 772.

[126] Marta Otto. "Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation–GDPR)." In: *International and European Labour Law*. Nomos Verlagsgesellschaft mbH & Co. KG. 2018, pp. 958–981.

[127] Danica Mitch M Pacis, Edwin DC Subido Jr, and Nilo T Bugtai. "Trends in telemedicine utilizing artificial intelligence." In: *AIP conference proceedings*. Vol. 1933. AIP Publishing LLC. 2018, p. 040009.

[128] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. "Deep learning for anomaly detection: A review." In: *ACM Computing Surveys (CSUR)* 54.2 (2021), pp. 1–38.

[129] Silvia Panzarasa, Silvana Quaglini, Giuseppe Micieli, Simona Marcheselli, Mauro Pessina, Corrado Pernice, Anna Cavallini, and Mario Stefanelli. "Improving compliance to guidelines through workflow technology: Implementation and results in a stroke unit." In: *Studies in health technology and informatics* 129 (Feb. 2007), pp. 834–9.

[130] Andrea Pazienza, Roberto Anglani, Giulio Mallardi, Corrado Fasciano, Pietro Noviello, Corrado Tatulli, and Felice Vitulano. "Adaptive Critical Care Intervention in the Internet of Medical Things." In: *2020 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS)*. IEEE. 2020, pp. 1–8.

[131] Andrea Pazienza, Giulio Mallardi, Corrado Fasciano, and Felice Vitulano. "Artificial Intelligence on Edge Computing: a Healthcare Scenario in Ambient Assisted Living." In: *Proceedings of the 5th Italian Workshop on Artificial Intelligence for Ambient Assisted Living 2019, co-located with 18th International Conference of the Italian Association for Artificial Intelligence, AI*AAL@AI*IA 2019*. 2019, pp. 22–37.

[132] Andrea Pazienza and Daniele Monte. "Introducing the Monitoring Equipment Mask Environment." In: *Sensors* 22.17 (2022), p. 6365.

[133] James L Peterson. "Petri nets." In: *ACM Computing Surveys (CSUR)* 9.3 (1977), pp. 223–252.

[134] Carl Adam Petri and Wolfgang Reisig. "Petri net." In: *Scholarpedia* 3.4 (2008), p. 6477.

[135] Wolter Pieters. "Explanation and trust: what to tell the user in security and AI?" In: *Ethics and information technology* 13.1 (2011), pp. 53–64.

[136] Elad Plaut. "From principal subspaces to principal components with linear autoencoders." In: *arXiv preprint arXiv:1804.10253* (2018).

[137] Ilaria Pretelli. "Proposal for a Regulation of the European Parliament and of the Council on Jurisdiction and the Recognition and Enforcement of Judgments in Civil and Commercial Matters (Recast)." In: *Available at SSRN 1963155* (2011).

[138] Rojalina Priyadarshini, Mohit Ranjan Panda, and Brojo Kishore Mishra. "Security in Healthcare Applications Based on Fog and Cloud Computing." In: *Cyber Security in Parallel and Distributed Computing: Concepts, Techniques, Applications and Case Studies* (2019), pp. 231–243.

[139] Jamal Raiyn et al. "A survey of cyber attack detection strategies." In: *International Journal of Security and Its Applications* 8.1 (2014), pp. 247–256.

[140] Sridhar Ramaswamy, Rajeev Rastogi, and Kyuseok Shim. "Efficient algorithms for mining outliers from large data sets." In: *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. 2000, pp. 427–438.

[141] Marc'Aurelio Ranzato, Fu Jie Huang, Y-Lan Boureau, and Yann LeCun. "Unsupervised learning of invariant feature hierarchies with applications to object recognition." In: *2007 IEEE conference on computer vision and pattern recognition*. IEEE. 2007, pp. 1–8.

[142] Protection Regulation. "Regulation (EU) 2016/679 of the European Parliament and of the Council." In: *Regulation (eu)* 679 (2016), p. 2016.

[143] Anne Rozinat and Wil MP Van der Aalst. "Conformance checking of processes based on monitoring real behavior." In: *Information Systems* 33.1 (2008), pp. 64–95.

[144] . "S. on Oversight, S. on R. and Technology, A. Committee on Science, Space, and Technology, and H. of Representatives. Serial No. 115-17: Bolstering the Government's Cybersecurity: Lessons Learned from Wannacry, Joint Hearing Before the Subcommittee on Oversight and Subcommittee on Research and Technology Committee on Science, Space, and Technology, House of Representatives, in Homeland." In: ().

[145] Abdel Mlak Said, Aymen Yahyaoui, and Takoua Abdellatif. "Efficient anomaly detection for smart hospital IoT systems." In: *Sensors* 21.4 (2021), p. 1026.

[146] Mayu Sakurada and Takehisa Yairi. "Anomaly detection using autoencoders with nonlinear dimensionality reduction." In: *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*. 2014.

[147] Paul M Salmon, Neville A Stanton, and Daniel P Jenkins. *Distributed situation awareness: Theory, measurement and application to teamwork*. CRC Press, 2017.

[148] H. Schlieter, M. Benedict, K. Gand, and M. Burwitz. "Towards Adaptive Pathways: Reference Architecture for Personalized Dynamic Pathways." In: *2017 IEEE 19th Conference on Business Informatics (CBI)*. Vol. 01. 2017, pp. 359–368.

[149] Guus Schrijvers, Arjan Hoorn, and Nicolette Huiskes. "The care pathway: Concepts and theories: An introduction." In: *International journal of integrated care* 12 (Sept. 2012), e192.

[150] Guus Schrijvers, Arjan van Hoorn, and Nicolette Huiskes. “The care pathway: concepts and theories: an introduction.” In: *International journal of integrated care* 12.Special Edition Integrated Care Pathways (2012).

[151] Norshahida Shaadan, Abdul Aziz Jemain, Mohd Talib Latif, and Sayang Mohd Deni. “Anomaly detection and assessment of PM10 functional data at several locations in the Klang Valley, Malaysia.” In: *Atmospheric Pollution Research* (2015).

[152] Raymond Ka-Man Sheh. “" Why Did You Do That?" Explainable Intelligent Robots.” In: *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*. 2017.

[153] Weisong Shi, Jie Cao, Quan Zhang, Youhuizi Li, and Lanyu Xu. “Edge computing: Vision and challenges.” In: vol. 3. 5. IEEE, 2016, pp. 637–646.

[154] E.H. Shortliffe and J.J. Cimino. *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*. Health Informatics. Springer New York, 2006. ISBN: 9780387362786.

[155] Md Amran Siddiqui, Alan Fern, Thomas G Dietterich, Ryan Wright, Alec Theriault, and David W Archer. “Feedback-guided anomaly discovery via online optimization.” In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2018, pp. 2200–2209.

[156] Saurabh Singh, Young-Sik Jeong, and Jong Hyuk Park. “A survey on cloud computing security: Issues, threats, and solutions.” In: *Journal of Network and Computer Applications* 75 (2016), pp. 200–222.

[157] Shailendra Singh and Sanjay Silakari. “A survey of cyber attack detection systems.” In: *International Journal of Computer Science and Network Security* 9.5 (2009), pp. 1–10.

[158] Xiuyao Song, Mingxi Wu, Christopher Jermaine, and Sanjay Ranka. “Conditional anomaly detection.” In: *IEEE Transactions on knowledge and Data Engineering* 19.5 (2007), pp. 631–645.

[159] International Organization for Standardization. *ISO/IEC 27002: information technology-security techniques-code of practice for information security management*. ISO, 2005.

[160] Tarik Taleb, Konstantinos Samdanis, Badr Mada, Hannu Flinck, Sunny Dutta, and Dario Sabella. "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration." In: *IEEE Communications Surveys & Tutorials* 19.3 (2017), pp. 1657–1681.

[161] Lawrence J Trautman and Peter C Ormerod. "Wannacry, ransomware, and the emerging threat to corporations." In: *Tenn. L. Rev.* 86 (2018), p. 503.

[162] Wil Van Der Aalst. "Process mining: Overview and opportunities." In: *ACM Transactions on Management Information Systems (TMIS)* 3.2 (2012), pp. 1–17.

[163] Wil Van der Aalst, Arya Adriansyah, and Boudewijn van Dongen. "Replaying history on process models for conformance checking and performance analysis." In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2.2 (2012), pp. 182–192.

[164] Wil Van der Aalst, Ton Weijters, and Laura Maruster. "Workflow mining: Discovering process models from event logs." In: *IEEE transactions on knowledge and data engineering* 16.9 (2004), pp. 1128–1142.

[165] P Patrick Van Der Smagt. "Minimisation methods for training feedforward neural networks." In: *Neural networks* 7.1 (1994), pp. 1–11.

[166] Kalyan Veeramachaneni, Ignacio Arnaldo, Vamsi Korrapati, Constantinos Bassias, and Ke Li. "AIˆ 2: training a big data machine to defend." In: *2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS)*. 2016, pp. 49–54.

[167] Luca Viganò and Daniele Magazzeni. "Explainable security." In: *2020 IEEE European Symposium on Security and Privacy Workshops(EuroS & PW)*. 2020, pp. 293–300.

[168] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. "Extracting and composing robust features with denoising autoencoders." In: *Proceedings of the 25th international conference on Machine learning*. 2008, pp. 1096–1103.

[169] AJMM Weijters, Wil MP van Der Aalst, and AK Alves De Medeiros. "Process mining with the heuristics miner-algorithm." In: *Technische Universiteit Eindhoven, Tech. Rep. WP* 166.July 2017 (2006), pp. 1–34.

[170] David J Weller-Fahy, Brett J Borghetti, and Angela A Sodemann. "A survey of distance and similarity measures used within network intrusion anomaly detection." In: *IEEE Communications Surveys & Tutorials* 17.1 (2014), pp. 70–91.

[171] Stephen A White and Derek Miers. *BPMN modeling and reference guide: understanding and using BPMN*. Future Strategies Inc., 2008.

[172] Weng-Keen Wong, Andrew W Moore, Gregory F Cooper, and Michael M Wagner. "Bayesian network anomaly pattern detection for disease outbreaks." In: *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*. 2003, pp. 808–815.

[173] Ruoyu Wu, Gail-Joon Ahn, and Hongxin Hu. "Towards HIPAA-compliant healthcare systems." In: *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*. 2012, pp. 593–602.

[174] Haifeng Xu, Jianfei Pang, Xi Yang, Liqun Ma, Huajian Mao, and Dongsheng Zhao. "Applying clinical guidelines to conformance checking for diagnosis and treatment: a case study of ischemic stroke." In: *2020 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE. 2020, pp. 2125–2130.

[175] Guojie Yang, Mian Ahmad Jan, Varun G Menon, PG Shynu, Mian Muhammad Aimal, and Mohammad Dahman Alshehri. "A centralized cluster-based hierarchical approach for green communication in a smart healthcare system." In: *IEEE Access* 8 (2020), pp. 101464–101475.

[176] Chong Zhou and Randy C Paffenroth. "Anomaly detection with robust deep autoencoders." In: *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 2017, pp. 665–674.

[177] Hongxu Zhu, Chung Kit Wu, Cheon Hoi Koo, Yee Ting Tsang, Yucheng Liu, Hao Ran Chi, and Kim-Fung Tsang. "Smart healthcare in the era of internet-of-things." In: *IEEE Consumer Electronics Magazine* 8.5 (2019), pp. 26–30.