Fundação
para a Ciência
e a Tecnologia

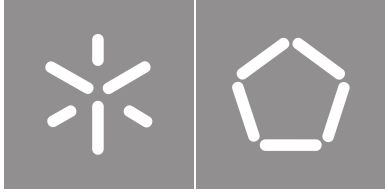
Ethics and New Trends in Artificial Intelligence in Intensive Care Units
Diana Ferreira
UMinho | 2024

Universidade do Minho
Escola de Engenharia

Diana Lisandra Azevedo Ferreira

Ethics and New Trends in Artificial Intelligence in Intensive Care Units

April, 2024



Universidade do Minho
Escola de Engenharia

Diana Lisandra Azevedo Ferreira

Ethics and New Trends in Artificial Intelligence in Intensive Care Units

Doctorate Thesis
Doctorate in Doctoral Program in Biomedical Engineering

Work developed under the supervision of:
José Manuel Ferreira Machado
António Carlos da Silva Abelha

April, 2024

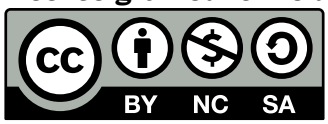
COPYRIGHT AND TERMS OF USE OF THIS WORK BY A THIRD PARTY

This is academic work that can be used by third parties as long as internationally accepted rules and good practices regarding copyright and related rights are respected.

Accordingly, this work may be used under the license provided below.

If the user needs permission to make use of the work under conditions not provided for in the indicated licensing, they should contact the author through the RepositoriUM of Universidade do Minho.

License granted to the users of this work



Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International
CC BY-NC-SA 4.0

https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en

Acknowledgements

Curiosity, the driving force behind intellectual growth and cognitive development, has always been a defining trait of human nature. Ever since we were children, we have been looking for answers to questions, regardless of scale or complexity, about ourselves and the world around us. As human beings, our minds are naturally drawn to the unknown and to the mysteries and enigmas that lie beyond our understanding. From the earliest days of our species, long before the advent of civilization, human curiosity has conferred our ancestors the ability to explore new territories, forage for resources, and decipher environmental cues, thereby facilitating survival and increasing reproductive success. Hence, stemming from innate inquisitiveness, curiosity has always been ingrained inside of my mind, propelling me to explore new paths, quest for knowledge, and engage in new intellectual pursuits. By embracing and cultivating curiosity, I have been able to embark on a transformative academic journey of exploration and discovery, which I hope will contribute to ever-expanding frontiers of knowledge and enlightenment in the topic of this thesis and promote scientific literacy in some way.

I want to express my deepest gratitude to all the human minds behind the curiosity ingrained in my cognition, who have nurtured my intellectual growth, and inspired me to embark on this academic journey of exploration and discovery.

First and foremost, I want to express a sincere appreciation to my mentor, Professor Doctor José Machado, whose guidance, expertise, and relentless support have been invaluable throughout this journey. Your enthusiasm for knowledge and your unwavering commitment to excellence have inspired me to push the boundaries of my intellectual capacities and step aside of my comfort zone. To Professor Doctor António Abelha, my second mentor, I am extremely grateful for the encouragement, constructive criticism, and insightful perspectives which have been fundamental to help me navigate the complexities and move forward during this challenging process. I will forever cherish the calm and composed demeanor that you have always shown, even in the most stressful situations. All the opportunities you have both provided me with, the challenges you have presented me, and the wisdom you have shared with me have been instrumental in shaping my academic and personal development. Thank you for believing in me and for guiding me through this transformative journey.

I extend my gratitude to the to the professionals of the Information Service of the ULSdSA institution

and to the members of the research project I have been involved with in the past years, particularly to nurse Nuno Abreu and doctor Pedro Leuschner, for their insightful opinions, clinical expertise, and invaluable suggestions, which have enriched the quality and rigor of this research.

I would also like to acknowledge the contributions of FCT, whose financial assistance has enabled the realization of this doctoral thesis. The investment in scientific research is crucial for the advancement of knowledge and the promotion of innovation, and I am grateful for the opportunity that has been given to me to contribute to this endeavor.

Furthermore, I want to express my appreciation to my laboratory colleagues, who have been a constant source of support, camaraderie, and inspiration. I am thankful to have had the opportunity to share the madness, the laughter, and the challenges of this academic adventure with you. The "hora da loira" has been the highlight of the last stressful weeks and I will forever remember the good times we have shared over a glass or two (or three) of sangria. To Cristiana Neto, my partner in crime, there are no words to describe how grateful I am to have you by my side, to share the ups and downs, the joys and sorrows, and the countless hours of hard work and dedication. I would not have made it this far without your sense of humor, never-ending support, and friendship.

A special word of gratitude to my friends, who have been a constant source of joy, laughter, complicity, and comfort throughout this journey. The inside jokes, the late-night walks, the spontaneous trips, the endless conversations, and the fun nights out we have experienced together will forever be cherished in my heart. Thank you for making this journey more bearable and for reminding me that life is not just about work and study.

To my soulmate, my best friend, and my biggest supporter, João, I want to thank you with all my heart for being my rock, the hug I run to when I need comfort, the shoulder I lean to cry, and the smile that lights up my day. There is no one in this world capable of understanding me as you do and making me forget all the worries of life, for that I am eternally grateful. A special appreciation to your family, quite literally my second family, for welcoming me with open arms and for all the love they have shown me. Soon, our families will become one, and I am looking forward to the new adventures that life has destined for us.

Finally, a heartfelt word to my family, who have always been the pillar of my strength, the nurturing voice of reason, and the everlasting source of support in my life. To my p arents, a million words would not suffice to describe the warmth of your love, the sacrifices you have made, and the unquestionable belief you have always had in me. I am forever glad for the values you have instilled in me, the lessons you have taught me, and the opportunities you have provided me with. To my brother, thank you for being the most annoying yet caring sibling, for always having my back and being there to protect me, and for being the best brother a sister could ask for. To my sister-in-law, thank you for your kindness and support, and for being part of our family. I am still waiting for the day you will surprise me with a positive pregnancy test, I cannot wait to be an aunt! To the rest of my family, I want to tell you how much I love you and how grateful I am for having you in my life. To my loved ones who watch over me from above, I hope I am making you proud.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the Universidade do Minho.

_____, _____

(Place) (Date)

_____

(Diana Lisandra Azevedo Ferreira)

”

“Tenho em mim todos
os Sonhos do mundo.”

— Fernando Pessoa
(Portuguese writer, literary critic, and philosopher,
considered one of the greatest poets.)

Resumo

Ética e Novas Tendências em Inteligência Artificial em Unidades de Cuidados Intensivos

Caracterizada pela rápida propagação do SARS-CoV-2, a pandemia catalisou uma mudança de paradigma nos sistemas de saúde em todo o mundo, revelando vulnerabilidades sistémicas há muito intrínsecas no que diz respeito à interoperabilidade dos dados, à sua estruturação e aos dilemas éticos. A necessidade de uma resposta rápida e eficiente por parte das autoridades de saúde pública era urgente mas não foi atendida devido à falta de Registos de Saúde Eletrónico (RSE) estruturados e normalizados, o que impossibilitou a interoperabilidade semântica e a tomada de decisões baseada em evidências. As Unidade de Cuidados Intensivos (UCI) foram particularmente afectadas por estas limitações, o que se refletiu num aumento das taxas de mortalidade. A presente investigação foca-se na transformação de sistemas de saúde, alavancando a norma openEHR para responder aos desafios enfrentados pela Unidade Local de Saúde de Santo António (ULSdSA). Fundamentada nos princípios da Design Science Research (DSR), a investigação culminou em três contribuições: o mapeamento dos princípios éticos com a norma openEHR, a conceção de uma solução baseada em openEHR que cria formulários de interface gráfica a partir de templates openEHR e o desenvolvimento de uma prova de conceito para a integração de algoritmos de Machine Learning (ML) na solução anterior. A COVID-19 actuou como um catalisador para as ideias de investigação, mas também como uma barreira à sua implementação. No alvoroço desta crise, surgiu uma confluência de desafios, incluindo a indisponibilidade dos profissionais de saúde da UCI para colaborar no projeto. No entanto, a natureza iterativa da metodologia DSR permitiu o desenvolvimento de uma solução de prova de conceito baseada em openEHR, que garante interoperabilidade semântica e conformidade ética. Através de uma abordagem de casos de estudo, a solução foi implementada de forma incremental em diferentes especialidades, estando atualmente integrada nas tarefas diárias de vários profissionais de saúde. Este processo de desenvolvimento permitiu obter uma solução robusta e flexível capaz de acomodar as necessidades específicas da UCI. A investigação da capacidade de integrar algoritmos de ML para suportar a tomada de decisão foi promissora, sugerindo que a solução poderá ser expandida para desbloquear novas possibilidades de melhoria da qualidade dos cuidados prestados.

Palavras-chave: Ética, Inteligência Artificial, openEHR, Registo Eletrónico de Saúde, Unidade de Cuidados Intensivos

Abstract

Ethics and New Trends in Artificial Intelligence in Intensive Care Units

Characterised by the rapid spread of SARS-CoV-2, the COVID-19 pandemic catalyzed a multifaceted paradigm shift in healthcare systems worldwide, shedding light on longstanding systemic vulnerabilities regarding data interoperability, structuring, and ethical dilemmas. The need for a fast and efficient response from public health authorities was urgent but unmet due to the lack of structured and standardised Eletronic Health Record (EHR)s, which hindered semantic interoperability and evidence-based decision-making. Intensive Care Unit (ICU)s were particularly affected by these shortcomings, which resulted in increased mortality rates. The present doctoral thesis focus on the transformation of healthcare systems, leveraging the openEHR standard to address the challenges faced by the ULSdSA, a Portuguese healthcare institution. Fundamented on the principles of DSR, the research culminated in three main contributions: the mapping of ethical principles with the openEHR standard, the design, development, and implementation of an openEHR-based solution for transforming openEHR templates into User Interface (UI) forms, and the development of a proof-of-concept for the integration of ML algorithms within the openEHR-based solution to provide decision support to clinicians. COVID-19 acted as a catalyst for the research ideas evolved within the thesis, but also as a barrier to their full implementation. Amidst the tumult of this crisis, a confluence of challenges emerged, including the unavailability of healthcare professionals from ICU to collaborate on the project, which constrained the scope of the research. The iterative nature of the DSR methodology, however, enabled the development of a proof-of-concept openEHR-based solution, that ensures semantic interoperability and ethical compliance. Through a case study approach, the solution was incrementally implemented in different specialties of the institution, being currently integrated into the daily tasks of several healthcare professionals. The iterative development process allowed the solution to become robust and flexible to later accommodate the specific needs of the ICU. The investigation of its ability to integrate ML algorithms to provide decision support to clinicians was promising, suggesting that the solution could be expanded to unlock new possibilities for improving the quality of healthcare services provided.

Keywords: Artificial Intelligence, Eletronic Health Records, Ethics, Intensive Care Units, openEHR

Contents

List of Figures **xiii**

List of Tables **xv**

Acronyms **xviii**

1 Introduction **1**

1.1 Context and Scope 2
1.2 Motivation 4
1.3 Objectives 6
1.4 Document Structure 9

2 State of the Art **11**

2.1 Information Technology in Healthcare 11
2.1.1 Health Information Systems and Eletronic Health Records 13
2.1.2 Clinical Decision Support 16
2.1.3 Artificial Intelligence and Machine Ethics 17
2.1.4 Interoperability 19
2.2 Health Data Standards 25
2.2.1 Overview of Existing Standards 25
2.2.2 Understanding openEHR Specifications 28
2.2.3 Summary Remarks 35

3 Research Methodologies **38**

3.1 Design Science Research Methodology 41
3.2 Case Study Approach 43
3.3 Analysis and Evaluation Segment 44
3.3.1 Proof of Concept 44

3.3.2 Strenghts, Weaknesses, Opportunities and Threats Analysis 45
3.3.3 Technology Acceptance Model 47
3.4 Research Design Overview 50

4 Research Contributions **52**
4.1 Spheres of Ethics, Health Information Systems, openEHR, and Artificial Intelligence: A Systematic Literature Review 53
4.1.1 Introduction and Problem Contextualization 53
4.1.2 Research Purposes 55
4.1.3 Design and Development 55
4.1.4 Results and Discussion 64
4.2 Transforming Eletronic Health Records: A Focus on openEHR Forms Development . 68
4.2.1 Introduction and Problem Contextualization 68
4.2.2 Research Purposes 69
4.2.3 Design and Development 74
4.2.4 Demonstration 102
4.2.5 Evaluation 118
4.2.6 Results and Discussion 124
4.3 From openEHR Data to Ethical Decision-Making in the ICU using Artificial Intelligence 126
4.3.1 Introduction and Problem Contextualization 126
4.3.2 Research Purposes 127
4.3.3 Design and Development 128
4.3.4 Results and Discussion 150
4.4 Connecting the Dots: A Consolidated Perspective 153

5 Conclusion **156**
5.1 Main Contributions 157
5.2 Future Work 160

Bibliography **162**

Appendices

A Questionnaire for Assessing the Acceptance of the FormBuilder Artefact **173**

B Scientific Contributions **176**
B.1 Journal Publications 176
B.1.1 Standardizing Corneal Transplantation Records with openEHR: Case Study . 176

B.1.2 Automation of Hospital Workflows Using International Standards: a Case Study With Imaging Exams . . . 177
B.1.3 An Interoperable Approach to Support Clinical Pathway Decisions for Nutritional Screening using openEHR . . . 178
B.1.4 Prediction of COVID-19 diagnosis based on openEHR artefacts . . . 179
B.1.5 Data Mining for Cardiovascular Disease Prediction . . . 180
B.1.6 Different Scenarios for the Prediction of Hospital Readmission of Diabetic Patients . . . 181
B.1.7 Predicting the Survival of Primary Biliary Cholangitis Patients . . . 182
B.1.8 Integrating a New Generation of Interoperability Agents into the AIDA Platform 183
B.1.9 Recommendation System Using Autoencoders . . . 184
B.1.10 Monitoring and Maintenance of Web Service Processes in Health Units . . 184
B.1.11 Improving the Decision-Making Process in a Hospital Environment With New Interactive Visualization Methods . . . 185
B.2 Conference Publications . . . 186
B.2.1 Recommendation of Medical Exams to Support Clinical Diagnosis Based on Patient's Symptoms . . . 186
B.2.2 OpenEHR based bariatric surgery follow-up . . . 187
B.2.3 Classification of Dementia in Adults . . . 188
B.2.4 Detecting Autism Spectrum Disorder Using Data Mining . . . 188
B.2.5 Prediction Models for Coronary Heart Disease . . . 189
B.2.6 A CRISP-DM Approach for Predicting Liver Failure Cases: An Indian Case Study 190
B.2.7 Data Mining Approach to Classify Cases of Lung Cancer . . . 191
B.2.8 Step Towards Predicting Patient Length of Stay in Intensive Care Units . . . 191
B.2.9 Diagnosis of Diabetic Retinopathy Using Data Mining Classification Techniques 192
B.2.10 Prediction Models for Polycystic Ovary Syndrome Using Data Mining . . . 193
B.2.11 Data Mining Approach to Understand the Association Between Mental Disorders and Unemployment . . . 194
B.2.12 Data Mining for the Prediction of Fetal Malformation Through Cardiotocography Data . . . 195
B.2.13 Review of Trends in Automatic Human Activity Recognition in Vehicle Based in Synthetic Data . . . 195
B.2.14 Review of Trends in Automatic Human Activity Recognition Using Synthetic Audio-Visual Data . . . 196
B.2.15 Bridging the Gap of Neuroscience, Philosophy, and Evolutionary Biology to Propose an Approach to Machine Learning of Human-Like Ethics . . . 197
B.2.16 Step Towards Monitoring Intelligent Agents in Healthcare Information Systems 198
B.2.17 Prediction of Mental Illness Associated with Unemployment Using Data Mining 198

B.2.18 Data Mining to Predict Early Stage Chronic Kidney Disease 199
B.2.19 Intelligent Nutrition in Healthcare and Continuous Care 200
B.2.20 Prediction of Neoadjuvant Chemotherapy Outcome in Breast Cancer Patients 200
B.2.21 Steps Towards Online Monitoring Systems and Interoperability 201
B.2.22 Improving Healthcare Delivery with New Interactive Visualization Methods . 202
B.2.23 Predictive Data Mining in Nutrition Therapy 203
B.3 Book Chapters 204
B.3.1 Classification of Polycystic Ovary Syndrome Based on Correlation Weight Using Machine Learning 204
B.3.2 Business Intelligence for Nutrition Therapy 204

List of Figures

1 Interoperability levels in Information Systems (IS). Adapted from (Muguira & Tolk, 2006). 22
2 Connecting the dots: clinical terminologies, data models, and interoperability standards. 27
3 The dual modelling approach of openEHR. Adapted from (openEHR Specifications, latest). 29
4 Archetype reuse across different templates. 35

5 The methodological layers: research onion representation. 41
6 Methodological steps intrisic to the DSR process. Adapted from (Peffers et al., 2007). . . 42
7 SWOT analysis matrix. Adapted from (Benzaghta et al., 2021). 46
8 Constructs of the Technology Acceptance Model (TAM)3. Adapted from (Holden & Karsh, 2010; Venkatesh & Bala, 2008). 51
9 The DSR process conducted for the case studies of the thesis. 51

10 Methodological steps carried out within the systematic literature review process. 56
11 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram illustrating the study selection process. 61
12 Proposed framework architecture for integrating openEHR standards within Health Information System (HIS) to promote ethical compliance and data quality in healthcare. 66
13 Scrum framework followed during the artefact development process. Adapted from (Schwaber & Sutherland, 2011). 76
14 The database schema that supports the artefact. 79
15 High-level architecture of the openEHR-based artefact. 86
16 Integration of FormBuilder within the openEHR modelling flow. 88
17 Logic flow that occurs in the FormBuilder process responsible for initiating rule requests. 96
18 Diagram of HTTP request and response interactions between the Form Builder, which acts as the front-end server, and the Decision Logic Module (DLM)Engine, which serves as the back-end server, in two distinct settings. 97
19 The design of the graphical interface in configuration/design mode for general customizations. 99

20 The design of the graphical interface in configuration/design mode for individual form element customization. 100
21 The design of the graphical interface in edit/filling mode. 102
22 The design of the graphical interface in view/consult mode. 103
23 The different stages involved in the development of the case studies. 104
24 Mind-map view of the "Service Request"archetype. 109
25 Mind-map view of the "Anatomical Location"archetype. 109
26 Mind-map view of the "Eye Surgery Details"archetype. 110
27 The "Corneal Transplantation Proposal"form generated with the FormBuilder tool. . . . 112
28 The top-level task plan view of the "Corneal Transplantation Workflow". 114
29 Single number indicators in the corneal transplantation process. 116
30 Proportional analysis of form conclusion and form availability within the corneal transplantation workflow. 117
31 Total number of corneal transplantation forms submitted over time. 117
32 Categorization analysis of submitted forms within the corneal transplantation workflow. . 118
33 Number of submitted tasks within the corneal transplantation workflow for each form over time. 119
34 CRISP-DM Methodology 128
35 Age distribution of patients in the COVID-19 dataset 142
36 Gender distribution of patients in the COVID-19 dataset 142
37 Temperature distribution of patients in the COVID-19 dataset 143
38 COVID-19 test result distribution in the COVID-19 dataset 143
39 Interconnected contributions of the research. 154

List of Tables

1 The seven interoperability levels of Levels of Conceptual Interoperability Model (LCIM) (Muguira & Tolk, 2006; Tolk et al., 2007) . . . 23
2 Summary characteristics of healthcare standards . . . 26
3 Description of TAM3 constructs . . . 48
4 Key concepts and synonyms used in the search string for the systematic literature review 57
5 Research Question (RQ)s that guided the systematic literature review . . . 57
6 Eligibility criteria of the literature research conducted . . . 59
7 Main ethical principles and values in EHR and HIS . . . 62
8 Mapping of openEHR Features to Ethical Principles . . . 63
9 Main objectives guiding the design and development of the artefact . . . 70
10 Scrum roles and respective descriptions. Adapted from (Garcia et al., 2022) . . . 75
11 Libraries used in the development of the artefact . . . 81
12 Description of some of the microservices that support the operational functioning of the artefact . . . 82
13 Main objectives guiding the design and development of the artefact . . . 88
14 Key events identified within the corneal transplantation process . . . 107
15 Description of the coded values assigned to each coded text item . . . 110
16 Default values assigned to the "Corneal Transplantation Proposal"template . . . 111
17 Association between forms, DLMs and rules in the corneal transplantation process . . . 115
18 Characterization of the dataset attributes for ICU length of stay prediction . . . 130
19 Algorithm's performance evaluation in the ICU Length of Stay (LOS) case study . . . 137
20 Description of the attributes of the dataset for the COVID-19 case study . . . 141
21 Data Mining Model (DMM)s with the highest accuracy for each Data Mining Technique (DMT) in the COVID-19 case study . . . 147
22 DMMs with the highest precision for each DMT in the COVID-19 case study . . . 148
23 DMMs with the highest sensitivity for each DMT in the COVID-19 case study . . . 148

24 DMMs with the highest specificity for each DMT in the COVID-19 case study 148
25 DMMs with the highest Area Under the Curve (AUC) for each DMT in the COVID-19 case study 149
26 Association between the contributions and the research questions 159
27 Structure of the survey questionnaire based on TAM3 constructors for assessing the FormBuilder 173

Listings

4.1 Search queries used in the systematic literature review 59
4.2 Structure schema of the new form request 83
4.3 Structure schema of an HTTP POST request to retrieve the refset values according to a specific field 94
4.4 Structure schema of the response for the request example used above 94

Acronyms

A Approach *(pp. 134, 135, 145)*
AD Archetype Designer *(pp. 87, 111)*
ADL Archetype Definition Language *(p. 34)*
AI Artificial Intelligence *(pp. 1, 4–11, 13, 17–19, 36, 40, 43, 50–54, 56–58, 60, 61, 64–66, 123, 124, 126–128, 151–153, 155, 157–161)*
AM Archetype Model *(pp. 33, 37)*
AMM Archetype Modeling Methodology *(p. 108)*
ANN Artificial Neural Network *(pp. 134, 135, 137–139, 145)*
API Application Programming Interface *(pp. 82, 86)*
AQL Archetype Query Language *(pp. 30, 35, 116)*
AUC Area Under the Curve *(pp. xvi, 146, 147, 149)*

CDA Clinical Document Architecture *(p. 24)*
CDSS Clinical Decision Support System *(pp. 9, 11, 13, 16)*
CKM Clinical Knowledge Manager *(pp. 34, 63, 87, 108, 161)*

DA Data Approach *(pp. 134, 135, 145, 147–149)*
DBMS Database Management System *(pp. 77, 78)*
DIM Demographic Information Model *(pp. 63, 64, 78)*
DL Deep Learning *(pp. 18, 134, 135, 137–139, 145, 147–149)*
DLM Decision Logic Module *(pp. xiii, 30, 93, 95, 97, 101, 103, 112, 113, 123, 124, 160)*
DMM Data Mining Model *(pp. xv, xvi, 130, 133–135, 137, 145, 147–149)*
DMT Data Mining Technique *(pp. xv, xvi, 134, 135, 145, 147–149)*
DOM Document Object Model *(p. 79)*
DSR Design Science Research *(pp. vii, viii, xiii, 9, 38, 40–43, 50, 51, 156)*
DSS Decision Support System *(p. 16)*

DT Decision Tree *(pp. 134, 135, 137–139, 145, 147–150)*

DTPB Decomposed Theory of Planned Behaviour *(p. 48)*

EHR Eletronic Health Record *(pp. viii, xv, 3, 7, 9–11, 13, 15, 18, 25, 28–37, 52–54, 56, 61–65, 68–70, 78, 93, 105, 106, 115, 122–124, 126, 153, 157, 160)*

EU European Union *(pp. 5, 25, 68, 124, 161)*

FAIR Findable, Accessible, Interoperable, Reusable *(pp. 62–64)*

FHIR Fast Healthcare Interoperability Resources *(p. 23)*

FN False Negative *(pp. 135–137, 139, 146, 150)*

FP False Positive *(pp. 135, 136, 139, 146, 147, 150)*

GBT Gradient Boosted Tree *(pp. 134, 135, 137–139, 150)*

GDPR General Data Protection Regulation *(pp. 25, 62–64)*

HIMSS Healthcare Information and Management Systems Society *(p. 20)*

HIPAA Health Insurance Portability and Accountability Act *(p. 24)*

HIS Health Information System *(pp. xiii, xv, 3, 9, 11, 13–16, 21, 22, 24, 25, 27, 28, 35, 40, 43, 50, 52–66, 153, 155)*

HL7 Health Level 7 *(pp. 23, 30)*

HTML Hypertext Markup Language *(p. 23)*

HTTP Hypertext Transfer Protocol *(pp. 82, 97)*

HTTPS Hypertext Transfer Protocol Secure *(p. 82)*

ICU Intensive Care Unit *(pp. viii, xv, 1, 3–6, 8, 10, 36, 39, 40, 43, 50, 53, 68, 126–131, 133, 137–139, 150, 151, 153, 156, 158–161)*

IEEE Institute of Electrical and Electronics Engineers *(p. 20)*

IQR Interquartile Range *(p. 132)*

IQRM Interquartile Range Multiplier *(p. 132)*

IS Information Systems *(pp. xiii, 12, 13, 16, 22, 23, 39)*

ISO International Organization for Standardization *(p. 20)*

IT Information Technology *(pp. 3, 11, 13, 17, 40, 43, 44, 47, 48, 50–53, 55, 75, 102, 106, 115, 118, 122)*

JDT JSON Data Template *(pp. 77, 87, 112)*

JSX JavaScript XML *(p. 81)*

kNN k Nearest Neighbours *(pp. 134, 135, 137–139)*

LCIM Levels of Conceptual Interoperability Model *(pp. xv, 22, 23)*
LOS Length of Stay *(pp. xv, 10, 127, 129, 130, 136–139, 150)*

ML Machine Learning *(pp. vii, viii, 18, 43, 44, 52, 123, 124, 127, 130, 132, 140, 150, 151, 154–156, 158–161)*
MVA Missing Value Approach *(pp. 134, 135, 145, 147–149)*

NB Naïve Bayes *(pp. 134, 135, 137–139, 145, 147–149)*
NB-K Naïve Bayes Kernel *(pp. 145, 147–149)*

OOP Object-Oriented Programming *(p. 80)*
OPT Operational Template *(pp. 77, 87, 111)*

PoC Proof of Concept *(pp. 40, 44, 45, 50)*
PoP Proof of Principals *(p. 44)*
PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses *(pp. xiii, 9, 54–56, 60, 61)*

RF Random Forest *(pp. 134, 135, 137–139, 145, 147–149)*
RM Reference Model *(pp. 30, 31, 33, 37, 108)*
RMSE Root Mean Squared Error *(p. 136)*
ROC Receiver Operating Characteristic *(pp. 146, 147)*
RQ Research Question *(pp. xv, 6, 7, 15, 55–57, 59, 154, 157, 159)*
RSE Registos de Saúde Eletrónico *(p. vii)*
RT Random Tree *(pp. 145, 147–149)*

S Scenario *(pp. 134, 135, 139, 145, 147–149)*
SM Sampling Method *(pp. 134, 135, 145, 147–149)*
SMOTE Synthetic Minority Over-sampling Technique *(pp. 145, 147–149)*
SNOMED-CT Systemised Nomenclature of Medicine – Clinical Terms *(p. 24)*
SOFT Satisfatory, Opportunity, Fault, and Threat *(p. 45)*
SQL Structured Query Language *(pp. 35, 77, 116)*
SWOT Strengths, Weaknesses, Opportunities, and Threats *(pp. 40, 44–47, 50, 119, 122, 125)*

T Target *(pp. 134, 135, 145)*

TAM Technology Acceptance Model *(pp. xiii, xv, xvi, 40, 44, 47, 48, 50, 51, 98, 121, 173)*
TN True Negative *(pp. 135, 136, 146, 147)*
TP True Positive *(pp. 135–137, 139, 146)*
TPlan Task Planning *(pp. 30, 113, 123, 124, 160)*
TRA Theory of Reasoned Action *(pp. 47, 48)*

UCI Unidade de Cuidados Intensivos *(p. vii)*
UI User Interface *(pp. viii, 66, 78–82, 85, 87, 88, 97, 98, 112, 122, 124)*
UK United Kingdom *(pp. 5, 124)*
ULSdSA Unidade Local de Saúde de Santo António *(pp. vii, viii, 39, 42, 52, 68, 74, 102, 103, 119, 121, 123–125, 141, 153, 154, 156, 158, 160)*
UTAUT Unified Theory of Acceptance and Use of Technology *(p. 48)*

WHO World Health Organization *(pp. 68, 104)*
WP Work Plan *(pp. 103, 112, 113, 115)*
WWW World Wide Web *(p. 2)*

XML eXtensible Markup Language *(p. 23)*

1

Introduction

"The real voyage of discovery consists not in seeking new landscapes, but in having new eyes."
— Marcel Proust

Significant attempts have been made with the aim of advancing ethical practices in the field of Artificial Intelligence (AI). However, in this research the idea is to address deficiencies at the foundational level: the data layer. Data serves as the raw material upon which AI algorithms operate, thus constituting the cornerstone of AI systems. It is therefore imperative to mitigate ethical concerns at the data level, as these concerns are propagated throughout the entire AI pipeline. Consequently, the mitigation of ethical concerns necessitates the rectification of underlying data issues, such as bias, inaccuracy, incompleteness, and inconsistency. Without such remediation, ethical dilemmas persist unabated regardless of the magnitude of the applied efforts.

The present doctoral thesis outlines a comprehensive investigation into the use of openEHR specifications as a means to standardise clinical records and support the integration of AI algorithms in healthcare, with a particular focus on the design of ethically compliant solutions for ICUs.

The research is framed within the scope of the doctoral thesis proposal entitled "Ethics and New Trends in Artificial Intelligence in Intensive Care Units" of the Doctoral Program in Biomedical Engineering of the University of Minho. This chapter presents a contextualization and a framing of the theme (Section 1.1), as well as the main motivations for its elaboration (Section 1.2). It also includes the research questions that serve as a motto for the development of the thesis, and the objectives outlined to answer these questions (Section 1.3). These questions will be answered during the course of the investigation and will be further debated in the conclusions chapter, in a clear and objective way. Finally, to properly end this chapter, the last section outlines the structure of the document (Section 1.4), providing a roadmap that readers can follow to navigate the thesis.

1.1 Context and Scope

Medicine is a primitive science that has been present in the daily life of human beings since the early days of mankind. The art of healing and discovering cures against disease is perhaps the oldest and greatest struggle that the human being has ever faced. Although its origin goes back to the beginning of the human race and despite the many scientific breakthroughs that have been accomplished over the years, this type of art is a universal practice of colossal complexity that is still far from being explored in all its fullness.

Hence, health organizations are among the most demanding as they are responsible for preserving and saving human lives. Life is an inherent gift of our nature, but it also poses a great challenge to the human being in the face of a multitude of variables that can jeopardise their well-being. The human body is a complex and intricate machine, composed of a multitude of systems and organs that interact with each other in a harmonious and synchronised way. The slightest imbalance in this system can lead to the onset of diseases and disorders that can have serious consequences for the individual's health and well-being. Throughout history, several agents have been known to adversely affect human life, like viruses and bacteria. In addition, different diseases are discovered and reported everyday, which makes it difficult to keep up with the pace of the evolution of the healthcare sector. The SARS-CoV-2 virus, responsible for the COVID-19 pandemic, is a clear example of how a new virus can disrupt the entire world, causing a global health crisis.

Historically, medicine has been made in a reactive way, and based largely on the interpretation of symptoms and clinical histories. The problem with this kind of approach is the fact that often human beings are unable to correctly express their symptoms or fail to communicate a detailed and complete description of their medical history. Such incompleteness of knowledge often brings undesirable outcomes in the provision of health care, such as misdiagnosis and inappropriate treatment, which can have serious consequences for the patient's health. In addition, the lack of a comprehensive and longitudinal patient record can hinder the continuity and quality of care, as well as the ability to make informed clinical decisions.

In this context, an increasing pressure triggered by a whole set of economic, social, technological, and political factors has been felt worldwide, leading to the need for a paradigm shift in the healthcare sector. This shift is characterised by a transition from a reactive to a proactive approach, where medicine is becoming more preventive and predictive, and clinical practice is a mixture of science and technology.

Computers play a fundamental role in almost all domains of life and their use has proved to be particularly striking in healthcare. Scientists produced the first digital computers in the 1940s. Within the following decade, physicians and other health professionals began to wonder about the positive effects that such technology would have if applied to medical practice. Those early predictions were followed by eight decades of remarkable progress in computing and most of the original prophecies have come to pass. Technology has been continually progressing and the extraordinary advances of the last two decades, like, for instance, the new methods for human-computer interactions, the innovations in mass storage of data, the Internet and the World Wide Web (WWW), and the wireless communications, have

all inevitably combined to make routine use of computers by healthcare professionals (Shortliffe et al., 2014). Always evolving, Information Technology (IT) has proved to be one of the main contributors to the advancement of medicine. Although only until recently computers and IT have been applied to healthcare, the influence of biomedical-computing systems is already remarkably broad.

The advent of the digital era has brought about a revolution in the healthcare sector, with the emergence of new technologies and innovative solutions that are transforming the way healthcare is delivered and managed. One of the most significant advancements in this regard is the widespread adoption of EHRs, which are digital versions of patients' paper registries that contain comprehensive and up-to-date information about their medical history, diagnoses, medications, treatment plans, immunization dates, allergies, radiology images, and laboratory test results.

Given the widespread adoption of EHRs in healthcare institutions, healthcare is one of the most data-intensive industries, generating vast amounts of data on a daily basis, particularly in the ICU. The ICU is a specialised unit within a hospital where patients are in critical conditions and whose clinical state requires extraordinary surveillance. The strict monitoring of the medical condition of these patients is ensured through electronic equipment to gather several measurements of physiological signals at an extremely high frequency. In addition to the significant number of variables that generally needs to be collected in ICU, each one being recorded at a high sampling rate, the vulnerable state of patients admitted means that most of them have to stay in this unit for several days, which leads to massive amounts of data per patient.

Although, at first glance, it may seem that this vast amount of data will be helpful in the decision-making required in ICU, in reality, the use of this data is not as straightforward because data are often unstructured, incomplete, and heterogeneous, making it difficult to extract meaningful insights and actionable information from it. Moreover, the continuous influx of data from multiple sources, such as EHRs, medical devices, and laboratory tests, further complicates the analysis process. This heterogeneity coupled with the lack of standardization and interoperability among these data sources poses significant challenges to the seamless exchange and integration of data across different healthcare settings, impeding the development of comprehensive and longitudinal patient records. These information silos not only hinder the continuity and quality of care but also limit the ability to make informed clinical decisions and to provide personalised and evidence-based treatments.

The systemic challenges of current healthcare systems were further exacerbated by the COVID-19 pandemic, which highlighted with lethal force the urgent need for reliable and efficient healthcare solutions. The integration of openEHR standards in HIS has emerged as a promising solution to address the lack of standardization and interoperability challenges associated with EHR data. openEHR provides a robust framework for structuring and representing clinical information in a standardised and computable format, facilitating the development of scalable and interoperable HIS solutions.

In addition, the rapid spread of the virus, coupled with limited resources and infrastructure, as well as limited knowledge about the virus and its effects, intensified the pressure on healthcare systems worldwide, particularly in critical care settings. With the intensified number of patients requiring intensive care and

the situation evolving rapidly, healthcare professionals were faced with the daunting task of making critical decisions under immense uncertainty and time constraints, often with limited access to comprehensive and up-to-date patient information. Under these circumstances, healthcare professionals are forced to allocate the medical resources and decide which patients should benefit from these resources. Decisions on therapeutics are also problematic as they affect the clinical condition of individuals, since the decisions of non-initiating or suspending certain treatments are often decisive in the death or survival of the patient. The ethical nature of these decisions is evident, as they involve the allocation of scarce resources and the prioritization of patients based on their prognosis and likelihood of survival, with limited information and time available to make these decisions. In this way, healthcare professionals might face conflicting feelings of incapacity or guilt as the code of medical ethics endorses that all patients should receive equal access to treatment, and, at the same time, that each patient should obtain the maximum care possible in order to save his/her life.

In light of these challenges, the consideration of ethical principles and values in this thesis is paramount to ensure the responsible and ethical use of healthcare technologies and to uphold the highest standards of patient care and well-being. The use of AI algorithms in the ICU holds great promise for enhancing clinical decision-making, optimizing resource allocation, and improving patient outcomes. However, the ethical implications of AI in the ICU are complex and multifaceted, requiring a nuanced understanding of the ethical principles and values that underpin healthcare practice.

Given the inability of critical care intensivists to dedicate time and effort to this investigation, the ethical exploration will be conducted through the lens of the data layer, focusing on the development ethically responsible openEHR-based solutions. Data, being the substratum of AI algorithms, constitutes the foundation upon which AI systems operate. Consequently, addressing deficiencies in data quality is essential for mitigating ethical dilemmas in AI. Bias, inaccuracy, incompleteness, and inconsistency within datasets can significantly compromise the fairness, transparency, and accountability of AI systems. These issues not only undermine the reliability of AI-driven decision-making but also perpetuate and exacerbate societal inequalities and injustices. Therefore, it is paramount to rectify these underlying data issues to effectively mitigate ethical concerns at the AI level.

Ethical considerations will therefore play a critical role in the design, development, and deployment of the openEHR-based solution developed within this thesis, committing to upholding the highest standards of ethical practice in healthcare. By exploring the ethical implications of integrating the openEHR standard and incorporating AI algorithms in this solution, this research aims to contribute to the development of ethically responsible healthcare solutions that prioritise patient safety, well-being, and autonomy.

1.2 Motivation

Healthcare will hardly be flawless considering it is a huge complex interacting system where there are too many human variables. However, as co-responsible for the practice of medicine, healthcare organizations

have the duty to foster the endless discovery of answers and solutions to enhance patient care. Furthermore, it is impossible to neglect the fast-growing amount of data being collected in ICU, given its wealth and the possibly lifesaving knowledge that can be extracted from it. Under these circumstances, it is necessary to look for ways to ensure that these data are used to its maximum potential in order to support the multidisciplinary teams working in ICU in the diagnosis of the severity of pathological situations as well as in the analysis of their prognosis and in the evaluation of the effectiveness of the treatment measures applied.

Although patient safety is one of the fundamental principles for the practice of medicine, research reveals that a considerable number of patients are harmed while receiving health care (Organization et al., 2017). Patient safety incidents can cause unnecessary suffering, prolonged hospitalization, permanent injury or even death. ICUs are susceptible to potential patient safety challenges as critically ill patients require more medical interventions, invasive procedures, and higher frequency of decision-making processes. In addition, these patients need to take several medications, enhancing the risk of missing medication intervals, providing incorrect drug dosages, or even switching medication with another patient (Orgeas et al., 2008). Moreover, critically ill patients often suffer concurrent problems, which requires knowledge from different system specialties. The heterogeneity of the medical condition of critically ill patients, the lack of specialists with interdisciplinary knowledge about different organ systems, the poor communication in critical care teams, and flawed processes of documentation and care further weaken the ability of ICUs to ensure safe health care delivery. Given their vulnerable condition, incidents involving critically ill patients have increased likelihood of being life threatening (Reader et al., 2009). In these conditions, the investigation of new ways to reduce the probability of occurrences that can compromise the well-being of these patients is paramount.

Traditional ways of conducting medicine within ICU is therefore limited by heterogeneity of illness, complexity of clinical data, vast amounts of deficient and heterogeneous data, and inability of healthcare providers to precisely analyse it given the constraints of the human mind. Furthermore, COVID-19 pandemic has underscored the interconnectedness of ethical concerns with data quality in critical settings, highlighting the imperative for ethical AI practices that prioritise social justice and human rights of all individuals, especially those most vulnerable to the pandemic's impacts. As a result, evidence base for critical care practice is underdeveloped. Thus, one of the leading motivations to bring this study forward is the absence of reliable and efficient decision-making tools regarding the care processes of critically ill patients.

In recent years, openEHR has gained increasing attention as a promising solution to address the challenges of data standardization and interoperability in healthcare. A growing number of healthcare organizations are adopting openEHR across the European Union (EU) and United Kingdom (UK) to improve the quality, safety, and efficiency of patient care. Hence, given the pressures of the COVID-19 pandemic and the increasing demand for digital health solutions, as well as the strain to keep up with the evolution pace of neighbour countries, the claimed beneficial impacts of openEHR standard serve as a strong motivation for this research.

Furthermore, the rapid influx of data generated during the pandemic, coupled with limited resources has heightened concerns about data quality, completeness, and consistency, leading to potential inaccuracies and biases in AI algorithms. In recent years, considerable efforts have been devoted to advancing ethical practices within the domain of AI. These endeavors have primarily focused on refining algorithms and implementing regulatory frameworks aimed at ensuring fairness, transparency, and accountability in AI systems. One of the key motivations driving this research is the acknowledgment that ethical dilemmas persist unabated despite the magnitude of efforts directed towards algorithmic refinement and regulatory compliance. This persistence reveals a often underestimated gap in addressing the underlying causes of ethical concerns in AI. This study seeks to disrupt the cycle of ethical issues that pervade the entire AI pipeline by focusing on correcting data deficiencies at the foundational level. Moreover, advancing ethical practices in AI through data rectification aligns with broader societal aspirations for technological progress that is both innovative and responsible. By addressing data-related ethical concerns, it is possible to cultivate a culture of responsible innovation that values the well-being and dignity of individuals and communities.

In conclusion, this research is motivated by the recognition of the critical role that data plays in shaping the ethical landscape of AI. By rectifying data deficiencies at the foundational level, the author believes it will be easier to mitigate ethical concerns that permeate throughout the AI pipeline and pave the way for the responsible development and deployment of technologies in the health domain.

All things considered, the leading motivation is to rethink the traditional ways in which critical care conducts medicine, focusing on what matters the most – patient's lives and their well being – to ultimately achieve meaningful research achievements.

1.3 Objectives

In light of a detailed exploration of the information mentioned in the previous sections of this chapter, which include the main particularities of the problem and the factors that motivated this investigation, it is possible to outline the objectives that will guide the development of the research process. Hence, the research objectives should be outlined in the beginning of the project as they will assume a preponderant role in the development of the whole research. The objectives are the main guidelines outlining a logical set of instructions that will direct the research process, ensuring that the RQ is answered in a clear and objective way.

Accordingly, the main RQ that supports this doctoral thesis proposal is the following:

RQ1 *- How can openEHR standards be leveraged to support the integration of AI algorithms in the ICU and promote ethical healthcare practices?*

Given the intricacy and multidimensionality of the main RQ, it is essential to break it down into more

specific RQs that will guide the development of the research process. In addition, each RQ will be associated with a set of objectives that will serve as a roadmap for the research process.

These RQ are as follows:

RQ1.1: *What is the level of agreement of openEHR standards with ethical principles and values in healthcare?*

Motivation: The integration of openEHR standards in healthcare systems is intended to improve interoperability, data standardization, and patient care. Understanding the alignment of openEHR standards with ethical principles and values is crucial to ensure that the adoption of these standards promotes patient safety, autonomy, privacy, and equitable access to healthcare services. By assessing the ethical implications of openEHR standards, healthcare stakeholders can make informed decisions about their implementation and usage, ultimately contributing to the development of ethically responsible healthcare solutions.

- *Objective 1.1.1: Identify the main data quality issues in traditional EHRs that impact the ethical development of AI algorithms.*
- *Objective 1.1.2: Indentify the main ethical principles and values in healthcare.*
- *Objective 1.1.3: Evaluate the level of agreement of openEHR standards with the principles and values in healthcare.*
- *Objective 1.1.4: Propose a conceptual framework for a comprehensive hospital ecosystem based on openEHR standards that addresses the shortcomings identified in the data layer and is compliant with ethical principles in healthcare.*

RQ1.2: *How can openEHR specifications be applied to improve and streamline clinical documentation processes while ensuring data quality and integrity?*

Motivation: Clinical documentation plays a critical role in healthcare delivery, facilitating communication, care coordination, and decision-making among healthcare professionals. However, traditional documentation systems are often inefficient, error-prone, and fragmented. The adoption of openEHR specifications offers an opportunity to modernise and standardise clinical documentation processes, thereby enhancing efficiency, accuracy, and comprehensiveness. Investigating the application of openEHR specifications in clinical documentation can provide insights into how healthcare organizations can optimise data quality and integrity while improving workflow efficiency and patient care outcomes.

- *Objective 1.2.1: Develop a solution for converting openEHR templates into user-friendly interface forms.*

- *Objective 1.2.2: Investigitate the potential of openEHR specifications to streamline clinical documentation processes through data validation and logical rule enforcement.*
- *Objective 1.2.3: Provide directions on how to implement openEHR in real-world healthcare settings.*
- *Objective 1.2.4: Analyze the transformative power of openEHR specifications in healthcare.*
- *Objective 1.2.5: Evaluate the usability and effectiveness of the solution developed in a real-world setting.*
- *Objective 1.2.6: Discuss the limitations of the solution developed and propose strategies to mitigate them.*

RQ1.3: *How can AI algorithms be used to predict patient outcomes and support clinical decision-making in the ICU?*

Motivation: ICUs are critical environments where timely and accurate clinical decision-making is essential for patient survival and recovery. AI algorithms have the potential to analyze vast amounts of patient data and identify patterns or trends that may not be apparent to human clinicians, thereby aiding in the prediction of patient outcomes and the selection of appropriate interventions. By exploring how AI algorithms can leverage the structured data provided by openEHR, it is possible to uncover opportunities to enhance the accuracy, reliability, and timeliness of patient outcome predictions in the ICU. In addition, understanding how AI algorithms can be effectively integrated into ICU tasks and workflows to support clinical decision-making is fundamental. Investigating the synergies between openEHR and AI not only highlights the transformative potential of these technologies but also underscores the importance of addressing data quality and interoperability challenges to realise their full benefits in improving patient care and clinical outcomes.

- *Objective 1.3.1: Develop and validate AI algorithms to predict patient outcomes in the ICU.*
- *Objective 1.3.2: Analyse the impact of problems at the data layer in the performance of AI algorithms.*
- *Objective 1.3.3: Develop and validate AI algorithms on data collected from the openEHR based solution.*
- *Objective 1.3.4: Discuss the limitations of the AI algorithms developed and propose strategies to mitigate them.*
- *Objective 1.3.5: Evaluate the potential of using AI algorithms to support clinical decision-making in the ICU by implementing decision logic modules within the openEHR-based solution.*

1.4 Document Structure

The present doctoral thesis is organised into five chapters, namely Introduction, State of the Art, Research Methodologies, Scientific Contributions, and lastly Conclusions. These chapters contain the following contents:

CHAPTER 1 – INTRODUCTION: the thesis begins with a chapter that provides a general contextualization and framing of the topics inherent to the research under investigation. In addition, it elucidates the motivations that led to its elaboration, as well as the objectives and the research questions that are intended to be answered throughout its execution. The chapter ends with the organizational structure of the document, providing readers a clear roadmap for navigating the document.

CHAPTER 2 – STATE OF THE ART: comprehensively describes, through bibliographic references, the current knowledge, advancements, and trends of all the theoretical and scientific concepts of interest for the concretization of the thesis, namely HIS and EHR, interoperability and its particularities in the health area, Clinical Decision Support System (CDSS) and AI, as well as health standards, emphasizing openEHR. Elucidating key findings, debates, and gaps in knowledge, the chapter offers not only an overview of the current body of knowledge but also provides a foundation for the research conducted in this thesis. Furthermore, the "State of the Art" chapter serves to contextualise the significance of the research findings within the broader landscape of ethical HIS, highlighting the relevance and potential implications of the thesis's contributions to advancing knowledge and informing practice within the field.

CHAPTER 3 – RESEARCH METHODOLOGIES: explains the methodological approaches used in the doctoral thesis to ensure transparency, clarity, and confidence in the research process carried out. The chapter contemplates the characterization and justification of the various methodological strategies used in the pursuit of the research objectives, including qualitative, quantitative, and mixed methods approaches. More specifically, the DSR methodology was employed in the development of the case studies, while the Proof of Concept methodology was used to assess and defend the viability and usefulness of the artefact. Each methodology is rigorously contextualised in regard to the research objectives, ensuring alignment with the epistemological and ontological underpinnings of the thesis.

CHAPTER 4 – SCIENTIFIC CONTRIBUTIONS: provides a comprehensive description of work developed in the context of the thesis topic, offering valuable insights into the application of theoretical concepts and methodologies explored in the preceding sections. The first contribution - Spheres of Ethics, HIS, openEHR, and AI: A Systematic Literature Review - Section 4.1 synthesises, trough PRISMA existing research on the ethical aspects surrounding the integration of openEHR standards within HIS. It concludes with a overall hospitalar ecossystem proposal sustained on openEHR to fix the shortcomings identified in the data layer and promote ethical healthcare practices. In turn, the

second contribution - Transforming EHRs: A Focus on openEHR Forms Development - Section 4.2 - makes a detailed examination of the design, structure, functionality, and implementation of a solution for converting openEHR templates into user-friendly interface forms. Furthermore, it uses one out of the five healthcare specialties in which the solution is integrated to illustrate and exemplify the practical application of the proposed solution in a real-world setting and to demonstrate its potential benefits and implications for healthcare practice. Finally, the last contribution - From openEHR Data to Ethical Decision-Making in the ICU using AI - Section 4.3 - delves into two case studies conducted to explore the potential of AI algorithms to predict patient LOS in the ICU, as well as to identify COVID-19 cases based on patient symptoms collected through the openEHR-based solution developed in the previous contribution. By offering concrete examples and practical illustrations, each case study serves as a rich source of insight, providing readers a deeper understanding of the intricate dynamics of the theoretical concepts intrisic to the research process while also validating and refining hypothetical theories formulated throughout the thesis.

CHAPTER 5 – CONCLUSION: the last chapter makes a comprehensive synthesis of the main findings, contributions, and limitations of the study, and answers all the research questions raised at the beginning of this dissertation. In addition, the chapter provides ideas and recommendations for future research directions. By consolidating the key insights gleaned from the extensive research process carried out, this chapter serves as a culmination of the work undertaken, paving the way for further advancements within the field.

2

State of the Art

"Somewhere, something incredible is waiting to be known."
— Carl Sagan

This is a deeply conceptual chapter that offers a comprehensive literature review and the contemporary state of the art pertaining to a consistent set of theoretical and scientific concepts that are deemed indispensable for the understanding of this doctoral thesis in all its scope. By critically analyzing and synthesizing existing literature, this profound conceptual exploration provides readers with a nuanced understanding of the forefront of knowledge and advancements relevant to the field.

Furthermore, the examination of the state of the art not only contextualises the current research but also highlights gaps, challenges, and opportunities for further exploration and contribution to the field. Through this exploration, the thesis aims to contribute to the advancement and enrichment of knowledge about ethical reasoning in healthcare systems.

First, an introduction to IT in healthcare is provided (Section 2.1), which entails a description of diverse dimensions, including EHR, HIS, CDSS, AI, and interoperability. This section A detailed exploration of health data standards is then made (Section 2.2), addressing the significance of standards in fostering interoperability across HIS, focusing more specifically on openEHR. The chapter concludes by synthesising the key insights and discussing the broader implications of leveraging this standard to advance healthcare.

2.1 Information Technology in Healthcare

In the last years, a profound paradigm shift has unfolded, announcing the foundation of a new world-wide era - the "Information Era" - being particularly remarkable the influence of manifold forces and trends. Notably, the acceleration of scientific and technological progress in the information domain has been instrumental in shaping this transformative era (Hawken et al., 2020). The emergence of the "Information Era", or alternately recognised as the "Digital Era" or "Technological Era", promotes a fundamental reconfiguration of societal norms, intellectual frameworks, and perceptual paradigms at an international scale (Prins et al., 2022).

In a global context of escalating dynamism, a steady revolution flourishes, catalyzing profound alterations in values, knowledge, and perceptions across all domains of human knowledge (Y. Wang et al., 2021). As the nomenclature suggests, the information is the structuring element of the evolving society.

When discussing information, it is imperative to discern and define a clear distinction between data, information, and knowledge.

In its native form, data are defined as raw, isolated, and unprocessed facts or observations that lack context or interpretation (Fisher & Thompson, 2022). In many ways, data can be viewed as representations or descriptions of the world. Fundamentally, these data points hold minimal intrinsic value on their own, merely representing discrete elements or values devoid of meaning or significance, which alone are of no use because they do not have meaning of themselves (Li et al., 2021; Van Meter, 2020).

Information and data are closely related, their bonding is analogous to that between raw materials and finished products. Similarly to how raw materials require processing to transform into a product, data must be processed and undergo interpretation to transition into information (Fisher & Thompson, 2022; Van Meter, 2020). Thus, information arises when data points are processed, organised, and contextualised to convey meaning and relevance through relational connections. This meaning can be useful but does not necessarily have to be (Li et al., 2021).

Finally, knowledge represents a deeper comprehension of patterns, relationships, and implications gleaned from information. In other words, knowledge is gained from the understanding of the information available (Fisher & Thompson, 2022; Van Meter, 2020).

Hence, data serves as the foundational building blocks, while information and knowledge represent successive layers of refinement and abstraction that enable human cognition and decision-making processes.

The escalating volumes of data, information, and knowledge being produced on a daily basis reflect the demand for systems capable of ensuring the complete lifecycle management of information assets, from collection, storage, and processing to retrieval and dissemination.

As a consequence, the terms "Information Technologies" and "Information Systems" have earned significant prominence, being its increasing utilization a direct manifestation of the significance that the objects of their meaning have acquired within the contemporary society.

An IS can be described as a socio-technical subsystem of an organization designed with the explicit purpose of acquiring, processing, storing, transmitting, and displaying data, information, and knowledge pertinent to organizational functions (Alter, 2008; Laudon & Laudon, 2004; Winter et al., 2023). This intricate concept embodies a synthesis of human and technological elements, where the "socio" facet corresponds to the human actors engaged in information processing, whereas the "technical" dimension refers to processing tools orchestrating information flows within the system (Alter, 2008; Laudon & Laudon, 2004). Hence, an IS operates as a dynamic entity that promotes data-driven decision-making, supports operational workflows, and enables strategic planning initiatives. This conceptualization aligns with Davis and Olson's (1985) vision of an IS, which advocates an integrated man/machine system that serves as a critical enabler to facilitate organizational operations and decision-making processes (G. B. Davis & Olson,

1984).

Technological advancements expand horizons, forge new pathways, and create unprecedented opportunities, transforming and reshaping traditional ways of conducting business. The continuous cycle of innovation drives technological progress, where emerging technologies build upon and surpass their predecessors, shaping contemporary industry practices and propelling organizations towards new heights of efficiency, competitiveness, and innovation.

Propelled by the cumulative advancements in both hardware and software as well as the increasing recognition of its potentialities, IT came out of a situation of gloom within organizational settings, where it practically only took place in task automation, to find itself as a key ingredient in shaping the competitive positioning of any organization (Berdik et al., 2021).

Through a strictly technological lens, IT encompasses all the technological means used to support the operation of an IS. These resources leverage computational capabilities that allow to perform the whole spectrum of tasks crucial to the work of an IS such as acquisition, storage, processing, transmission, access, retrieval, and display of available information (Alter, 2008; Vassilakopoulou & Hustad, 2023).

Nowadays, envisioning an organization that does not integrate IT seems almost impossible. As time progresses, the indispensability of IT becomes increasingly evident, as organizations recognise its profound effects are not merely advantageous, but imperative for maintaining competitiveness and achieving profitability. Moreover, relying on IT resources transcends a mere operational strategy and emerges as a critical determinant for modern enterprise survival.

This paradigmatic shift reflects the importance of IT within current organizational frameworks to address the evolving demands and complexities as well as to achieve efficiency, growth, and competitiveness across diverse spheres of human activity, including healthcare.

Informed by empirical research and theoretical frameworks, the following sections are dedicated to provide a comprehensive understanding of the different facets of IT in the healthcare domain, encompassing diverse dimensions such as EHR and HIS (Section 2.1.1), CDSS (Section 2.1.2), AI (Section 2.1.3), and interoperability (Section 2.1.4).

2.1.1 Health Information Systems and Eletronic Health Records

In the health sector, IT influences medical procedures insofar as they provide comprehensive and credible information to health professionals and support the decision-making process, both clinically and administratively, thereby reducing the medical errors associated with these decisions (Bonney, 2013).

The computer emergence and the associated innovative technology began by being used to record small notes. These first notes contained records as simple as the number of beds occupied, helping in the hospital planning process. These were the first indication of the occurrence of HIS. Since then, advances in IT have contributed to the development of better and innovative health information solutions. Existing applications are renewed to increase performance while other innovative applications are developed from

scratch with new capabilities and specificities, extending their presence to all medical areas and throughout the whole hospital process chain (Degoulet, 2014; Hasman et al., 1996; Haux, 2006a, 2010).

Any health care assumes the involvement and participation of multiple professionals: doctors, nurses, nutritionists, psychologists, physiotherapists, and others. The patient care activities are also performed in different places such as operating room, nursing homes, outpatient clinics, intensive care units, among others. To enable such health activities, multiple information about the patient's condition is required. On the other hand, the procedures performed by each health professional generate even more information. Together, all this information will ensure the continuity of the care process. To do so, these data need to be aggregated and organised in order to produce a context that will support decision-making, guiding the entire healthcare process. HIS were designed to meet this end.

The aim of HISs is as simple as relevant: to contribute to high-quality and efficient patient care. This aim is primarily centered towards the patient, so it is a patient-centered approach and towards medical and nursing care, and the administrative and management tasks needed to support such care. Accordingly, as main purposes, it can be said that a HIS should be used to manage the information that the health professionals need to perform their activities effectively and efficiently, to ease the communication, integrate the information and coordinate actions among the multiple members of the professional care team, providing resources for financial and administrative support (Haux, 2006a, 2006b; Haux et al., 2004).

HIS are information systems processing data, information, and knowledge in healthcare environments. In other words, they are support tools that ensure the efficient delivery of health care and improve the quality of services, by reducing the incidence of adverse events and clinical errors, supporting health professionals with better evidence and making it possible to conduct research. Thus, HIS congregate a set of data, information, and knowledge used in the health area to support the planning, refinement and decision-making process of multiple health professionals involved in the patient care and users of the entire health system. It is important to know that HIS should not automate the medical decision but reduce the cognitive overload of health professionals and improve the basis of their decisions (Ammenwerth et al., 2003; Lenz & Reichert, 2007).

The HIS act in the acquisition, processing, and representation of the information, whether it is relative to the users, doctors or nurses. They can be implemented in public or private hospitals, clinics, doctor's office, pharmacies, service providers related to diagnostic exams, nursing services and therapies to support treatments (Haux, 2010; Weber-Jahnke et al., 2012).

Progress in the field of HIS is high and will continue to be in constant development and improvement since it is directly related to higher quality and efficiency of care, which is a universal and omnipresent concern (Haux, 2006b; Haux et al., 2004). Obviously, without having appropriate access to relevant data, practically no decisions on procedures, diagnostic, therapy, and others can be made without occurring medical errors or other problems, which may result in fatal consequences for patients. These fatal consequences are not derived from the disease that led the patient to seek medical help, but rather derived from the medical conduct, being insufficient communication and lack of information some of the factors

that contribute the most in this sense (Haux, 2006a; Lenz & Reichert, 2007). Having this in mind, it is obvious to recognise the relevance of systematically processing data, information, and knowledge for the quality and efficiency of healthcare (Haux, 2006a).

The genesis of the EHR system marks a significant milestone in the convergence of medical and technological domains. Over the years, the adoption of EHR systems has gained momentum, becoming a ubiquitous presence in healthcare settings worldwide to modernise patient data management and streamline healthcare delivery processes. EHR systems serve as comprehensive repositories of patient health information, encompassing medical history, diagnoses, treatments, medications, and laboratory results, among other pertinent data. Hence, an EHR system represents digital versions of patients' medical data, enabling healthcare providers to access, update, and share essential health information more efficiently. These digital records offer numerous advantages over traditional paper-based systems, facilitating information exchange, enhancing clinical decision-making, and ultimately improving patient outcomes. However, the effective integration of EHR within broader HIS is paramount to maximise their utility and ensure interoperability across various healthcare settings.

While the integration of EHR with HIS offers significant benefits, it also presents challenges related to data standardization, privacy, and security concerns, interoperability issues, and financial investments required for system implementation and maintenance. Coming from a variety of sources and used with multiple purposes, clinical data is characterised by prevailing heterogeneity, often not suitable for the traditional HIS where integration is not always privileged. Moreover, when considering the characteristics of traditional HISs, it can be said that the predominance in most countries is of a fragmented vendor-specific system, that emphasis on super-specialties and in the treatment of diseases, with the hospital as the main center of care. In this scenario, the computerized solution that predominates is the departmental systems, which include particularities that assure answer to the individual needs of the medical services and often do not have integration and interoperability characteristics (Lippeveld et al., 2000). Thus, currently, a HIS is described as a complex environment, founded by heterogeneous, distributed and ubiquitous systems, which culminates in a wide range of tools available for manipulation of information in an individualized way (Duarte et al., 2010; Peixoto et al., 2012), leading to data silos and hindering data sharing across different systems. Moreover, the lack of standardized data formats and terminologies poses significant challenges to data exchange and interoperability.

However, these systems and their performance could be more comprehensive if there was an interaction between them. Thus, the need for interoperability, to establish communication between systems and to reinforce the available content, which assumes a process of integration and interaction that must consider the scalability, flexibility, portability, and security of systems (Miranda et al., 2012; Peixoto et al., 2012; Weber-Jahnke et al., 2012). This means that the focus changes from isolated systems in one healthcare institution to the patient-oriented care processes, encompassing diagnosis and therapy, spreading over institutional boundaries. These issues will be later discussed in this thesis with more detail as the RQ of this thesis is directly related to the integration of EHR and HIS systems, and the challenges and opportunities that arise from this integration. Hence, healthcare organizations must carefully navigate these

challenges and adopt strategies to ensure successful integration while safeguarding patient confidentiality and maintaining data integrity.

2.1.2 Clinical Decision Support

Decision Support System (DSS)s rely on decision-making models that examine a large set of variables to answer a specific question. In addition to being "a computational system that assists the decision-making process" (Finlay, 1989), it has also been described as "an interactive, flexible and adaptable information system specially developed to support the solution of an unstructured problem and to improve decision making" (Turban, 1993). As a result, a DSS can be defined as an IS that allows users to quickly access relevant information through a process of customisation and configuration according to the needs of each user and the problem at hand, as well as to perform analysis and establish comparisons (Turban et al., 2004).

In recent years, the recognition of the value and potential of CDSSs as practical tools has increased substantially (Castaneda et al., 2015; Musen et al., 2014). This trend is mainly due to the inexorable growth in complexity and inappropriate costs associated with healthcare, the challenge of offering personalized services for each patient, the increasing pressure to adopt and use HIS, as well as the increasingly frequent introduction of mandatory requirements for the implementation of this type of computing tools in the health sector (Butler et al., 2015; Mattila et al., 2012; Musen et al., 2014; Shirabad et al., 2012).

Conceptually, the main purpose is to create and study computerized ISs capable of supporting and improving the decision-making processes in a healthcare setting (Power, 2008). These systems provide specific information, recommendations, fundamental knowledge, and analysis required to assist decision makers (Butler et al., 2015; Musen et al., 2014). Thus, CDSSs help clinicians apply new information to patient care through the analysis of patient-specific clinical variables (Machado et al., 2016; Trowbridge & Weingarten, 2001).

In a more technical perspective, a CDSS can retrieve relevant documents, create and send alerts or recommendations to patients or healthcare professionals, and organise and present relevant information on dashboards, diagrams, documents, and reports to ease, speed up, and improve decision-making (Castaneda et al., 2015; Musen et al., 2014).

Hence, the aim of a CDSS is to assist, rather than replace, healthcare professionals in their daily tasks, reducing the incidence of medical errors, improving the effectiveness and efficiency of decision-making processes, enhancing the quality of services, as well as decreasing unnecessary costs and waste (Butler et al., 2015; Keen, 1978; Musen et al., 2014).

It is critical to make quick and quality decisions in the health domain because they are almost always, directly or indirectly, related to human life. Medical decisions are made based on the interpretation of patient-specific data taking into account the existing medical knowledge. Consequently, clinical decision-making must incorporate the best available evidence with the experience of clinical professionals and the specific values regarding patient health status (Foshay & Kuziemsky, 2014; Lenz & Reichert, 2007). In the

absence of timely access to high-quality information, medical decision-makers are forced to rely solely on their experience and intuition, disregarding important facts and information needed (Foshay & Kuziemsky, 2014). This problem is even more amplified in intensive care, where the fragility of the clinical situation of the patients makes them more vulnerable to decision-making, as small decisions can be decisive in patient's survival or death. One of the main purposes of this study is to contribute to conscious decision-making in order to improve the care provided to patients in critical condition.

2.1.3 Artificial Intelligence and Machine Ethics

The information revolution has been promoting, in a global context of increasing dynamism and changeability, a permanent transformation of values, knowledge, and perceptions in practically all areas of human knowledge, at a breathtaking pace (Floridi, 2008). Propelled, on one hand, by the successive technological advances in equipment and software and, on the other hand, by the growing recognition of its potentialities, IT is currently playing a prevalent and major role in almost every area of our daily lives, as pervasive computing begins to take place, moving from a futuristic idea to a reality. The gradual shift towards pervasive computing is enabling the creation of an omnipresent and intelligent atmosphere in which computers will be embedded in our natural movements and in our social and physical interactions. Such technological breakthroughs provide a multitude of extremely powerful tools and bring endless new opportunities hitherto unattainable, making possible unprecedent phenomena and experiences as well as offering potential to tackle some of the most pressing issues around the world. Technology is so embedded in our daily lives that it is now nearly impossible to conceptualize a world without it.

Indeed, the continuous technological advances have the power to broaden horizons, open new paths, create new opportunities and revolutionize traditional ways of conducting health. As healthcare organizations around the world are increasingly becoming digitized, there is daily huge amounts of data being collected on critical clinical variables, patient demographics, healthcare components and services. As the number grows, it is important to search and develop newer and increasingly efficient ways to mitigate Big Data problems and challenges. Most of these problems can be tamed using AI, which empowers the "brain" to analyze and make informed decisions as well as to carry out intelligent actions towards a deeper understanding of the healthcare environment.

The term AI was first mentioned in the Dartmouth Conference of 1956 by John McCarthy, consider the father of AI, together with Marvin Minsky, Claude Shannon, and Nathaniel Rochester (J. Moor, 2006). AI is a branch of computer science that studies intelligent behavior as well as the relationship between computation and cognition (Barr & Feigenbaum, 2014). The ultimate goal of AI is a theory of intelligence that guides the creation of programs that attempt to simulate artificial entities capable of some kind of intelligent behavior (Barr & Feigenbaum, 2014; Genesereth & Nilsson, 2012). Since the foundation of the AI term until now, several authors have dedicated to the field and, as a result, the scientific literature showcases an extensive number of studies and numerous applications of AI (Pannu, 2015). AI technologies have matured to the point of currently offering real practical benefits in almost every area of our daily

life, including healthcare.

In recent years, there has been an increase in the application of AI in healthcare, mainly because of the increasing availability of data from EHRs and the constantly expanding computational power. In particular, critical care is receiving an increasing interest, given the huge amount of clinical data that is collected everyday and the demanding complexity of patients' condition, which is often too overwhelming and difficult for health professionals to analyze. Although still evolving, AI already has a remarkable positive impact in health care, but it is expected that it will play a major and increasingly prevalent role in the future medicine.

This study aims to develop learning algorithms to offer an intelligent component to the solution proposed in this document. Such algorithms will rely in specialized AI techniques namely Machine Learning. In the future, it is intend to explore these concepts in depth and make a rigorous study and analysis of the state of the art of the different algorithms that exist in the scientific literature. These concepts have a natural connection with Big Data and Data Science, since the algorithms can solve the challenges of analysing large amounts of data, and then Data Science tools can be applied to further exploration of data.

Hence, the future of this investigation holds great promise for the design of a framework to assist healthcare professionals working in critical care units and beyond with smarter real-time decision-making.

Of all healthcare settings, critical care is one of the most time-sensitive, rich in terms of data, and with higher complexity. Such complexity extends for the most part from the immense clinical scenarios that arise every day in critical care that originate numerous ethical questions and concerns. Thus, it is of paramount importance to find ways to establish rigorous and appropriate ethic guidelines for these units.

In addition, from the moment of their invention, computers have raised a plethora of complex dilemmas and conceptual issues. As new and sophisticated forms of computer technology continue to evolve, certain concerns still persist, mainly privacy, accountability, social and ethical issues. At the same time, as major technological innovations are coming to the market, particularly in the field of AI and other related topics like ML and Deep Learning (DL), seemingly new and unique problems emerge. Hence the pressing need to regulate this new powerful form of smart agency and bring forth better ethical approaches (J. H. Moor, 2005). The process of implementing a moral attitude in computational entities, as with human beings, presupposes the existence of a mind.

For a long time, there have been deliberations about the nature of the mind within the field of philosophy. The main belief of the computational theory of mind is that a mind is a computer. Accordingly, the theory seeks to answer important questions such as the nature of mental phenomena and how mental phenomena fit into the causal structure of reality, to ultimately discover what kind of computational system a mind is. The philosophy of mind argues that the essential aspects of mentality are the structures and their functional organization instead of the material of which the elements of the structures are made. Although the answer is still unknown, if this theory is right, then it is logically conceivable to artificially create a mind. The field of AI presupposes the theory of mind to computationally model something with mental abilities out of something other than living tissue (Floridi, 2008; Kim, 2018).

Undeniably, AI is a powerful force, a new form of smart agency, that is already reshaping our lives,

interactions, and environments. Furthermore, AI promotes the development of machines increasingly more sophisticated and powerful. As modern machines make our daily lives more enjoyable, it is likely that future machines will have increased control and autonomy to do so. However, some citizens argue that increasing the reliance on autonomous agents may jeopardize humankind. In fact, there are many science fiction books and movies that portray human anxiety and concern towards the creation of autonomous forms of computational intelligence as, for example, a superior form of AI, like a sophisticated robot, enslaving or eradicating humanity. This kind of ideology is extreme and ignorant, causing unnecessary apprehension and impairing the recognition of the benefits offered by AI, which eventually can mean the use of AI techniques below their full potential (Anderson & Anderson, 2007; Floridi et al., 2018).

Perhaps human fear towards intelligent agents stems from the concern regarding whether the behavior of these agents will be ethical or not. Thus, the acknowledgement and acceptance of higher forms of AI by society directly depends on whether it will be possible to develop safeguards against unethical conduct. From an ignorant point of view, it is easy to assume that this accomplishment is nothing less than an unrealistic fantasy. Yet, when faced with the request for solid arguments and justifications to support this reasoning, there is no specific answer. Furthermore, while it is intuitively easy to realize that incorporating a moral dimension into autonomously functioning machines is a long and complicated process with many obstacles along the way, it is yet to be discovered a constraint that disables this feat. Unquestionably, as computers' decision-making roles grow and society increasingly relies on computer-based intelligence with reduced human supervision, ethical considerations are inevitable. Hence, the growing demand for computer scientists and engineers to research into ways by which intelligent systems can be guaranteed to follow a certain level of ethical behavior.

Dedicated to this subject is Machine Ethics, a field of expertise that seeks to implement a moral dimension in computational systems either by introducing moral principles in machines or discovering means to make machines function in an ethically responsible way on their own. Unlike computer ethics, which studies the ethical issues surrounding the use of machines by humans like fraud, identity theft and computer hacking, machine ethics is concerned with guaranteeing an ethically acceptable behavior of machines towards humans and other machines (Anderson & Anderson, 2007; Anderson et al., 2004).

Hence, successful implementation of machine ethics in intensive care would be a major technological step. This thesis is founded on the belief that the achievement of machine ethics is not possible until the problems that pervade the data layer are addressed. The data layer is the one that contains the data that the AI systems use to make decisions. As a consequence, problems at the root of the data layer will be reflected in the AI systems that use this data.

2.1.4 Interoperability

Society is witnessing a pronounced transition towards the era of ubiquitous computing, a world where computing resources are omnipresent, seamlessly integrated within our surroundings, and effortlessly accessible as part of our daily activities (Shakya & Nepal, 2020). As a consequence, it becomes imperative

to make the various computers and heterogeneous devices that populate these environments, exchange and interpret data among themselves, thereby arising the concept of interoperability.

Interoperability, in essence, is a multifaceted concept that finds application across a diverse spectrum of domains and contexts. As such, different definitions and approaches will be explored.

According to Nancy Ide and James Pustejovsky, “Broadly speaking, interoperability can be defined as a measure of the degree to which diverse systems, organizations, and/or individuals are able to work together to achieve a common goal” (Ide & Pustejovsky, 2010).

From an engineering viewpoint, the Institute of Electrical and Electronics Engineers (IEEE), determines interoperability as the ”ability of a system or product to work with other systems and products without any additional effort on the part of the user.”. The IEEE further asserts that interoperability is possible through the adoption and adherence of standards (Ahlfors & Jin, 2010; Mansoor & Majeed, 2010).

The definition provided by the International Organization for Standardization (ISO) adopts a similar perspective, expanding upon the one provided by IEEE, describing interoperability as the capacity of independent systems to not only exchange meaningful information but also to initiate actions in each other, thereby working together for mutual advantage (Cardoso et al., 2018). This definition emphasises the seamless interaction and cooperation between disparate systems, enabling them to function harmoniously and derive collective benefits from their connection.

A meticulous examination reveals a common thread among the definitions previously presented, particularly in the emphasis on the ability to work together, whether between humans or machines. Consequently, broadly speaking, interoperability can be described as the capacity of disparate and independent computer systems and/or software applications to accurately engage in communication and data sharing, using the exchanged information in a manner conducive to the execution of coordinated actions across multiple systems, all working towards a common goal aimed at overcoming heterogeneity (Jaleel et al., 2020; Kirch, 2008).

The need for interoperability is of utmost importance within health organizations given the intricate interplay of diverse stakeholders, systems, and data sources (Jaleel et al., 2020). IEEE defines interoperability in healthcare as the “ability to communicate and exchange data accurately, effectively, securely and consistently with different information technology systems, software applications, and networks in various settings and exchange data such that clinical or operational purpose and meaning of the data are preserved and unaltered” (Mansoor & Majeed, 2010).

On another perspective, Muhmmad Ehsan Mansoor argues that “Interoperability in healthcare information is the capability of health information systems to work together within and across organizational boundaries” (Mansoor & Majeed, 2010).

The concept of interoperability in the domain of healthcare has undergone continuous refinement by the Healthcare Information and Management Systems Society (HIMSS), which more recently proposes a refined definition, positing interoperability as ”the ability of different information technology systems and software applications to communicate, exchange data, and use the information that has been exchanged. Data exchange schema and standards should permit data to be shared across clinician, lab, hospital,

pharmacy, and patient regardless of the application or application vendor." (Angula & Dlodlo, 2018; Sensmeier, 2022).

These definitions converge in certain aspects, particularly in the claim that interoperability in health is the ability of heterogeneous HIS and computational applications to interact and exchange data. Nevertheless, Mansoor and HIMSS introduce a different dimension by considering organizational boundaries, a matter that remains subject to scrutiny and deliberation within the scientific community because it raises questions surrounding the safeguarding of patient data confidentiality and protection (Benson & Grieve, 2021).

The progressive evolution of HIS and its widespread adoption across health institutions have yielded higher quality services, but their development is made in an isolated way, often resulting in information silos. These silos can result in data loss and, consequently, restrict access to relevant information that may lead to misconceived decisions based on incomplete or mistaken information (Benson et al., 2021). The genesis of information silos in healthcare lies in the fragmentation of data management practices, stemming from disparate and independent HIS, diverse information sources, divergent data standards, and decentralised organizational structures within health institutions.

Moreover, HIS must contain accurate and up-to-date data presented in a format easily comprehensible to all healthcare practitioners and readily accessible regardless of geographical location. The availability of these data is conditioned by the degree of interoperability maintained in the HIS.

Furthermore, patients seek care in multiple providers and healthcare settings throughout their lifetimes, resulting in the registration of distinct episodes of care at each encounter, which means that patients accumulate multiple episodes within multiple healthcare units (Benson et al., 2021). The need to access historical clinical data about the patient is extremely important, as healthcare practice heavily relies on informed decision-making rooted on evidence. This need intensifies simultaneously with the need to consolidate patient data dispersed across the numerous HIS within healthcare organizations.

Interoperability has the potential to overcome these constraints, enabling seamless communication and exchange of vital clinical data among healthcare providers, institutions, and systems across local, regional, national, and international levels thereby promoting coordinated and comprehensive care delivery (Torab-Miandoab et al., 2023). This not only enhances patient safety but also improves clinical decision-making by providing healthcare professionals with a holistic view of the patient's health data, regardless of time or location (Canova-Barrios & Machuca-Contreras, 2022; de Mello et al., 2022).

As a result, interoperability has transcended the scope of optional functionality to become an indispensable requisite in healthcare. However, ensuring interoperability in a hospital setting is a challenging task (Torab-Miandoab et al., 2023). A plethora of significant obstacles arises in the integration of HIS, stemming from distinct factors, including but not limited to the intricacy and specificity of medical data, the heterogeneity caused by the variety of systems and application vendors, the proliferation of different solutions for sharing clinical and administrative information within a single organization, and the socio-political and ethical problems (Benson & Grieve, 2021; Benson et al., 2021; de Mello et al., 2022).

Indeed, interoperability in the healthcare domain constitutes a complex multidimensional problem

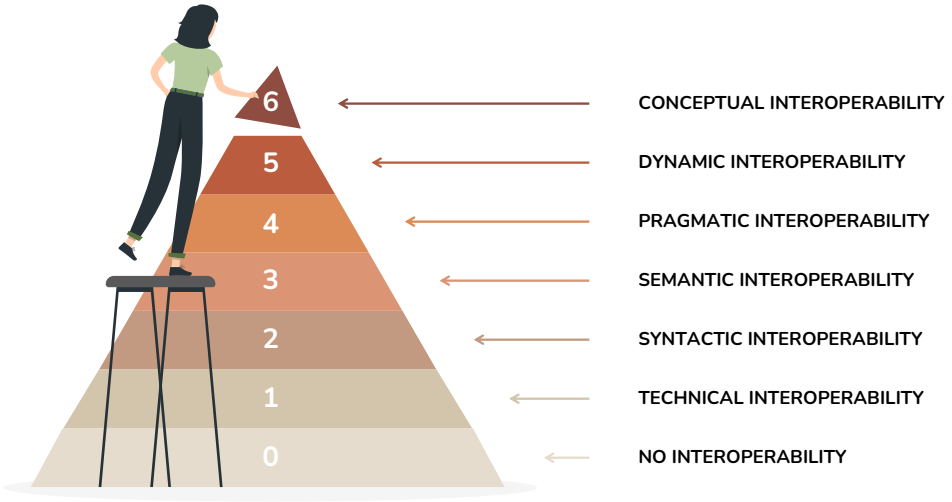


Figure 1: Interoperability levels in IS. Adapted from (Muguira & Tolk, 2006).

that concerns different layers of the healthcare infrastructure and encompasses a spectrum of dimensions that collectively shape the interoperability landscape (Benson et al., 2021; Torab-Miandoab et al., 2023). Hence, a paramount concern emerges with urgency surrounding the cultivation of a robust technological framework capable of enabling seamless interaction among all HIS (Cardoso et al., 2018).

The literature corpus contains several models that aim to categorise IS according to their level of interoperability. These models attempt to elucidate differing interoperability approaches through the definition of one or more sets of attributes, encompassing factors such as the extent of data exchange and abstraction, the technological implementation, the perspective from which interoperability is conceptualised and the underlying purposes (Cardoso et al., 2018).

Tolk and Muguira proposed in 2003 a hierarchical arrangement of seven tiers, the LCIM, that aims to distinguish the conceptual levels of interoperability according to the abstraction levels of the exchanged data, the technological implementations and the underlying objectives (Muguira & Tolk, 2006). As depicted in Figure 1, these levels delineate a progression from the absence of interoperability (Level 0) to the pinnacle of conceptual interoperability (Level 6).

In this model, abstract concepts and problem-solving methodologies are grouped based on their influence and capacity to engender the process of interoperability (Tolk et al., 2007; W. Wang et al., 2009). Each successive level within the interoperability hierarchy inherits the defining characteristics of its antecedent levels, thereby increasing the complexity of their implementation (Muguira & Tolk, 2006). Thus, upon achievement of conceptual interoperability at level 6, the systems are imbued with the all the characteristics delineated across the preceding tiers. It is noteworthy to mention that level 0 has no interoperability features. As one ascends along the categorization delineated by the LCIM, the degree of interoperability among systems also increases. Table 1 contains the designation and definition of each level (W. Wang et al., 2009).

Table 1: The seven interoperability levels of LCIM (Muguira & Tolk, 2006; Tolk et al., 2007)

Level	Designaton	Description
0	No Interoperability	Systems lack any ability to exchange data or communicate with each other, i.e., there is no point of contact between IS other than a human operator.
1	Technical Interoperability	Systems can exchange data over the network using well-defined communication protocols, such as HTML or XML.
2	Syntactic Interoperability	Systems can exchange data using standardised formats and syntax, but lack semantic clarity, i.e., the IS do not understand the meaning of the information shared.
3	Semantic Interoperability	Systems share data with the same meaning and interpretation.
4	Pragmatic Interoperability	Systems are aware of the methods and procedures that each system accomplishes and there is mutual knowledge of the context where the exchange is done.
5	Dynamic Interoperability	Systems can adapt to the change of meaning of information whenever it happens.
6	Conceptual Interoperability	Systems must be in accordance with the premises and constraints of the real environment. They can be implemented in different environments being aware of what happens to their peers at any time.

The categorization of interoperability levels provides a systematic approach for evaluating the extent to which different IS can seamlessly exchange and interpret data. In addition, the LCIM aids in identifying the specific attributes and functions required to achieve desired levels of interoperability, thereby informing decision-making processes related to system design, integration, and implementation. Moreover, interoperability levels are useful for benchmarking progress and setting targets for initiatives, advancing the interoperability agenda.

As healthcare continues to evolve in response to technological advancements and shifting patient needs, the need for interoperability remains paramount in accomplishing the vision of a connected and patient-centric healthcare system (de Mello et al., 2022). Fostering a unified ecosystem wherein information flows freely and transparently among all relevant stakeholders, not only promotes synergetic collaboration but also lays the groundwork for transformative advancements in healthcare delivery, clinical research, and population health management. Over the last decades, the development of a set of specifications for solving data exchange and consistency problems has become a major concern for governments throughout the world, specifically openEHR. The research available in scientific repositories in this field are mostly in the national scope and claims that Health Level 7 (HL7) Fast Healthcare Interoperability

Resources (FHIR), Clinical Document Architecture (CDA), Health Insurance Portability and Accountability Act (HIPAA) and Systemised Nomenclature of Medicine – Clinical Terms (SNOMED-CT) are among the most important requirements for implementing interoperability (Torab-Miandoab et al., 2023). In Section 2.2, a detailed exploration of health data standards will be made, elucidating their determinant role in fostering seamless interoperability across HIS.

2.2 Health Data Standards

In the earlier sections of the state-of-the-art, thorough examination has been undertaken regarding the intricacies inherent to HIS and the need for interoperability among them. Notwithstanding, achieving this target is a complex undertaking, constituting one of the main challenges encountered in the field of computerised HIS. Interoperability requires adherence to standards, which are norms and specifications that serve as foundational frameworks enabling disparate systems to exchange data effectively. These standards ensure integration among systems, fostering operational efficiency and information security within the technological health ecosystem (Canova-Barrios & Machuca-Contreras, 2022). Hence, the following sections will explore the domain of health standards and protocols, such as HL7, FHIR, and openEHR, comparing their scopes, architectures, and adoption (Section 2.2.1). A specific emphasis is made in the exploration of openEHR specifications (Section 2.2.2), which are gaining increasing prominence. Building upon this foundation, this section aims to synthesise the key insights and discuss the broader implications of leveraging these standards, particularly openEHR (Section 2.2.3), to advance healthcare.

2.2.1 Overview of Existing Standards

The proliferation of digital health technologies is transforming the healthcare industry towards a more interconnected and data-driven ecosystem. However, the operational integration of different healthcare systems and applications is often plagued by segregated data and legal regulations. As a consequence, healthcare personnel often find themselves compelled to undertake manual interventions to bridge the gaps between incompatible HIS (Torab-Miandoab et al., 2023). This manual labor not only demands significant time and effort, which comes at a high expense, but also introduces increased chances for errors and inefficiencies, further exacerbating the operational challenges faced by healthcare organizations. Furthermore, every year, over one million individuals are injured due to faulty health-care processes and system breakdowns De Vries et al., 2023.

Therefore, implementing health standards and protocols is instrumental in addressing the challenges associated with ensuring interoperability, safeguarding patient data privacy and security, mitigating risks and medical errors, promoting quality and consistency in healthcare delivery, upholding ethical principles, and fostering innovation and advancement within the healthcare ecosystem. Overall, interoperability stands as a foundational pillar for health institutions to build the technology infrastructure needed to maximise the value of EHR data.

Futhermore, the health standards serve as essential benchmarks against which healthcare organizations and providers are assessed for compliance with legal and regulatory requirements, thereby safeguarding patient safety, data security, and quality of care. Regulatory compliance in the health domain encompasses the protection of patient health information, and the General Data Protection Regulation (GDPR), which governs the handling of personal data in the EU.

As healthcare systems continue to evolve and diversify, the demand for seamless data exchange and

collaborative care delivery becomes increasingly evident, emphasizing the need to adopt standardised approaches to health information management.

In this way, healthcare relies on a plethora of health standards and protocols designed to standardise data exchange, promote interoperability, and ensure the security and privacy of patient information.

Health standards and protocols serve as the necessary frameworks and guidelines to ensure interoperability, semantic consistency, and data security across disparate healthcare systems and stakeholders.

The health industry operates within a complex ecosystem governed by a diverse array of standards encompassing various facets of healthcare delivery, information exchange, and regulatory compliance. These standards serve as indispensable guidelines and frameworks that dictate the protocols, procedures, and requirements necessary for ensuring quality, safety, and interoperability within the healthcare domain.

From clinical guidelines and best practices to technical specifications and data standards, the panoply of standards within the health industry encompasses a wide spectrum of disciplines and domains, including but not limited to clinical care, medical research, health information technology, and healthcare administration. These standards play a pivotal role in shaping the infrastructure and operations of healthcare organizations, influencing everything from patient care processes and electronic health record systems to medical device interoperability and healthcare policy formulation. Furthermore, adherence to these standards fosters consistency, transparency, and accountability within the health industry, facilitating effective communication, collaboration, and decision-making among stakeholders. Overall, the pervasive presence and adherence to standards underscore the commitment of the health industry to upholding quality, safety, and integrity in the delivery of healthcare services and the advancement of medical science.

HL7v2, FHIR, ISO 13606, CDA, and OpenEHR are prominent standards that play key roles in facilitating interoperability, data exchange, and standardization in healthcare. Hence, this section will explore and compare several health standards and protocols commonly used in healthcare, highlighting their significance, functionalities, and impact on healthcare delivery and patient outcomes.

Table 2: Summary characteristics of healthcare standards

Standard	Scope	Two-level	Flexibility	Granularity	Licensing	Adoption
HL7	messaging and data exchange	No	limited	limited	proprietary	widely adopted
FHIR	data exchange	No	high	moderate	open-source	increasing adoption
HL7 CDA	clinical document	No	limited	moderate	proprietary	widely adopted
openEHR	clinical data models	Yes	high	high	open-source	increasing adoption

(Table 2 continued on next page)

(Table 2 continued from previous page)

Standard	Scope	Two-level	Flexibility	Granularity	Licensing	Adoption
ISO 13606	full-meaning exchange of EHR extracts	Yes	moderate	moderate	proprietary	limited adoption
LOINC	laboratory and clinical observations	No	N/A	high	open-source	widely adopted
SNOMED	medical terminology	No	N/A	high	proprietary	widely adopted

According to the analysis of the literature surrounding health standards and informed by the comparative analysis presented in Table 2, Figure 2 categorises these standards into three distinct categories: data exchange standards, clinical terminologies, and data models and persistence. This categorization was conceived as a means to provide clarity to the complexity of healthcare standards, facilitating a nuanced understanding of their respective roles and functionalities within the realm of healthcare information exchange and management.

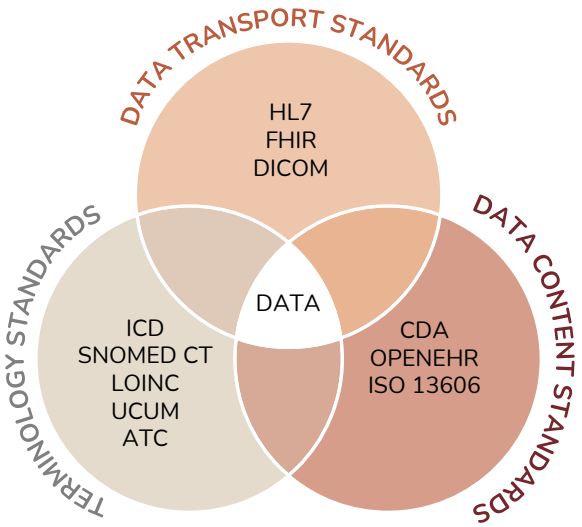


Figure 2: Connecting the dots: clinical terminologies, data models, and interoperability standards.

The analysis sheds light on different approaches and strategies for achieving semantically interoperable HIS , ranging from data exchange standards and models to clinical terminologies. In summary, there are several existing standards in healthcare informatics, each with its own strengths and weaknesses. Despite the absence of health standards consensus, there is a trend in recent studies demonstrating a predilection for the two-level standards (de Mello et al., 2022). ISO 13606 and openEHR both fall within

the scope of the two-level approach, but a comparative analysis between them reveals a fundamental distinction stemming from their underlying philosophies and governance structures. In terms of technical specifications, the adoption of ISO 13606 may be hampered by licensing constraints and limited availability of resources. Furthermore, the unavailability of ISO 13606 to the public poses a significant hurdle, exacerbated by the considerable time investment required for the development of novel archetypes tailored to specific system requirements. In contrast, openEHR distinguishes itself through its archetype-based approach, the backbone for its flexibility and adaptability. This archetype-based model not only enhances granularity and flexibility but also facilitates streamlined representation of complex clinical scenarios, thereby positioning openEHR as a more versatile and comprehensive solution. Following this line of reasoning, the next section makes a profound characterization of the specificities of the openEHR standard.

2.2.2 Understanding openEHR Specifications

There is a growing need for standardised EHR that can ensure the accuracy and consistency of patient data across healthcare organizations and locations to ultimately improve the management and documentation processes of health records (Yang et al., 2019a). One promising approach for standardizing EHR is the use of openEHR, a completely model-driven approach that separates data from applications (Min et al., 2021).

Established in 2000 by a virtual consortium of healthcare experts and information technology professionals from Europe, Brazil, and Australia (Ingram, 2016), openEHR is focused on enabling the development and deployment of open, interoperable, and computable patient-centric HIS. However, the more concentrated efforts towards comprehending the intricacies and functionalities of EHR systems trace its origins back to 1985, with a pronounced emphasis on the European landscape. In 2003, the openEHR foundation was established as a not-for-profit organization dedicated to combine international efforts in the design and implementation of comprehensive and interoperable EHR systems through the use of open-source components (Kalra et al., 2005).

openEHR is an open standard that provides a set of specifications and tools to support the creation of interoperable data structures and the long-term management of health data (openEHR Foundation, n.d.). One of the main challenges lies in the development of a scalable model capable of representing any conceivable health data, accommodating the diversity and intricacy of records arising from any profession, specialty, or service, which undergo frequent changes as medical practice evolves (Kalra et al., 2005).

In response to this, a foundational paradigm on which openEHR framework is based is the two-level modelling approach, separating domain semantics from technical aspects of software implementation. Under the model-driven approach, a stable reference information model constitutes the first level of modelling, while formal definitions of clinical content in the form of archetypes and templates constitute the second, as illustrated in Figure 3.

By separating the clinical content from the details of technical implementation, the dual modelling

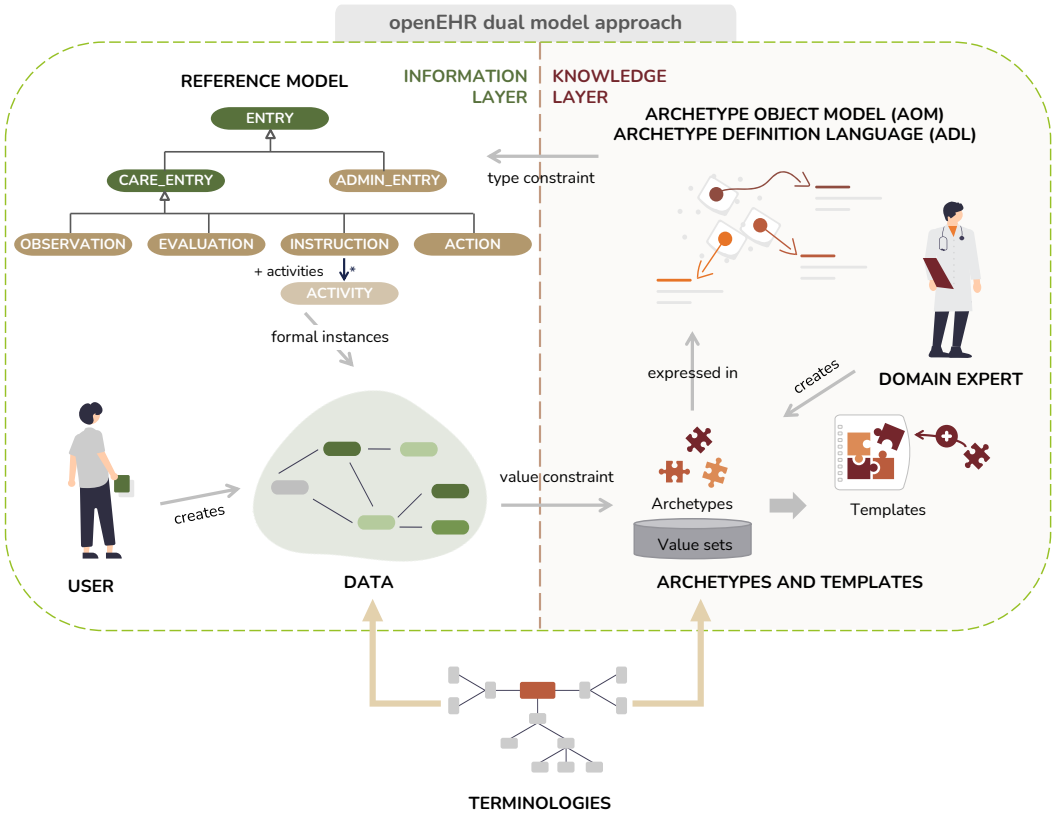


Figure 3: The dual modelling approach of openEHR. Adapted from (openEHR Specifications, latest).

approach of openEHR enables each stakeholder to contribute their expertise effectively. Clinicians and domain experts are empowered to focus their efforts on defining standardised clinical concepts without being constrained by technical considerations. In turn, software developers are able to develop EHR systems using standardised reference models without the need to understand the intricacies of clinical content. This separation of concerns fosters a collaborative environment between clinical and technical stakeholders and facilitates the development of interoperable, domain-specific EHR solutions.

In the pursuit of a life-long EHR, openEHR fosters a seamless exchange of health data among healthcare professionals and institutions by embracing a technology and data format independent approach, transcending barriers posed by different technologies and evolving medical guidelines. Moreover, the integration of medico-legal faithfulness, traceability, and audit-trailing mechanisms within the EHR framework ensures the accuracy and reliability of clinical data, meeting the stringent standards of medical practice and legal compliance. This endeavor not only improves patient care but also fosters collaborative care, empowering patients to actively engage in their healthcare journey (openEHR Specifications, latest).

In light of these features, openEHR currently has a huge international community and stands as the formalism with more robustness, capable of integrating the definition of clinical data, healthcare work processes, and a formal language for the specification of medical decision-making.

The specification program encompasses the definition of clinical data (archetypes), clinical records

(templates), clinical workflows (Task Planning (TPlan)), a formal language for defining clinical decision-making (DLM), combined with the openEHR Reference Model (RM) as well as openEHR own terminology, a specific language for querying data (Archetype Query Language (AQL)), among other components.

This program was thoroughly developed and continues to be an ongoing effort involving individuals all over the world to ensure the continuous enhancement, management, and maintenance of high-quality specifications that are both understandable and easy to compute and implement within a consistent change-management framework. All specifications adhere to the versioning rules at semver.org. Additionally, issue trackers are used to document problem reports and change requests. Major releases are reserved for breaking changes, particularly those that are disruptive or extensive in nature, and their potential impact is carefully evaluated before implementation proceeds to ensure the preservation of validity of clinical data created according to previous releases of the openEHR specifications (openEHR Foundation, n.d.).

Overall, the adaptability, flexibility, and scalability of openEHR's modular approach provides a powerful solution to the challenges facing the healthcare industry, and it is an ideal approach for healthcare systems of all sizes (Leslie, 2020; Min et al., 2021; Yang et al., 2019a).

2.2.2.1 The Reference Model

Recognizing the rapidly evolving context of healthcare technology, openEHR aimed to create a model that could transcend specific technologies and data formats, enabling interoperability and future-proofing EHR systems against technological obsolescence.

The openEHR information model is known as the RM and serves as a foundational framework for organizing and structuring EHR data in a standardised manner. This model plays a fundamental role in providing the basic information to the upper levels, defining various packages containing a repertoire of abstract data structures, data types, guiding principles, and global characteristics that inform the representation and storage of health record entries. In a nutshell, the RM represents the technical aspects of EHR systems, defining a hierarchical structure of classes that serve as the basic building blocks.

Additionally, the RM defines principles for versioning, querying, and accessing EHR data, enabling efficient retrieval and manipulation of health data.

By employing rigorous methodologies and semantic clarity, the openEHR RM sought to address inherent limitations in existing information models, particularly in minimizing errors and inconsistencies in data representation as well as reducing programmer errors and ambiguity.

Furthermore, it also facilitates integration with other international healthcare standards, such as ISO 13606, and existing messaging systems, including HL7v2, promoting interoperability and seamless exchange of EHR data across diverse healthcare environments on an international scale.

It is also important to note that a major focus was placed on enhancing the modelling of time and context within EHRs. Adopting a temporal/spatial approach, openEHR aimed to capture not only the

chronological sequence of events but also the contextual factors influencing healthcare decisions, thereby enriching the depth and accuracy of clinical data.

The EHR information model is the core part of openEHR and consists of the *EHR* and *COMPOSITION* packages, with the latter containing the *ENTRY* package. As mentioned previously, the classes of the RM are structured hierarchically, with higher-level constructs such as compositions, sections, and entries, providing context and organization for lower-level clinical data.

At the top level of the hierarchy is the Composition class, which represents a discrete clinical document or record within the EHR system. A Composition is based on the notion of a unit of information resulting from the interaction of a healthcare agent with the EHR. In other words, Compositions serve as containers for organizing and contextualizing clinical information, encapsulating data related to a specific event, encounter, or episode of care. There is a unique identifier assigned to each Composition, allowing for easy identification and retrieval of specific clinical documents.

The Composition class comprises several attributes, including:

- **Context:** captures metadata related to the context in which the Composition was created, including details such as the start and end time of creation, location, setting or healthcare facility where the event occurred, participating healthcare professionals, and any relevant clinical context or circumstances surrounding the event.

- **Content:** represents the substantive clinical content of the Composition, encompassing a collection of Entries that capture individual observations, assessments, interventions, or other relevant clinical data associated with the event or encounter.

- **Category:** classifies the Composition into a specific category based on its temporal context: event, persistent, or episodic. A persistent Composition is used to document continuant aspects of the patient's life such as family history and allergies. In contrast, event Compositions record occurrences, events, or states that occurred or were true at some moment in time. On the other hand, episodic Compositions capture relevant data for continuous care situations that extend over a significant period of time, such as pregnancy or inpatient care.

- **Language:** specifies the language in which the clinical content of the Composition is expressed, ensuring linguistic consistency and enabling multilingual support.

- **Territory:** denotes the territory in which the composition was written, coded from the openEHR countries code set, which is an expression of the ISO 3166 standard.

- **Composer:** identifies the individual or entity who is primarily responsible for creating or authoring the content of the Composition, providing attribution and accountability for the clinical content contained within the document.

Within each Composition, data is structured and organised using the Entry class. Entries contain the actual clinical data captured within compositions. Entries capture specific observations, assessments, or interventions related to patient care, enabling detailed and granular documentation of clinical events. Within the Entry class, a basic division occurs between clinical and non-clinical information. Accordingly, there are two top-level categories:

- **ADMIN_ENTRY:** pertains to data that lacks clinical relevance and is not generated during the care process, but rather concerns the logistical aspects of organizing it, such as appointments and admissions. This information is not about care and primarily focuses on the operational aspects of care delivery.
- **CARE_ENTRY:** this class serves as an abstract precursor of classes designed to express information pertaining to any clinical activity that may be recorded at any point in the care process around the patient.
  - **OBSERVATION:** encapsulates discrete clinical observations of any phenomenon or measurements, including pathology results, body temperature, blood pressure, patient responses to physician inquiries during a physical examination, and responses to psychological assessment questionnaires. It encompasses both quantitative and qualitative data pertaining to a patient's health status, symptoms, or vital signs.
  - **EVALUATION:** represents the assessment, opinion, or interpretation of clinical observations or data, allowing healthcare professionals to synthesise information and make judgments about a patient's condition or response to treatment. Evaluations include risk assessments, clinical goals/targets, diagnosing medical conditions, assessing the severity of symptoms, or determining the effectiveness of interventions.
  - **INSTRUCTION:** documents the provision of clinical directives, instructions, or orders aimed at guiding a clinical activity to be performed in the future. An Instruction, upon execution, elicits an Action in response. Instructions include medication prescriptions, service requests, laboratory tests, and procedural orders.
  - **ACTION:** captures the execution, undertaken by an individual or agent, of clinical activities based on instructions.

In addition, the Section class exists to represent a heading within a hierarchical structure or section tree. A given Composition may contain one or multiple Section(s), each contributing to the organization and structuring of the EHR's content alongside Entries.

As Compositions accumulate over time in the EHR, the Folder class serves as a useful tool to index compositions that share similar parameters or criteria. These parameters could include temporal aspects, such as compositions spanning specific dates or periods within certain clinical specialties. Additionally, folders can be organised based on individual physician folders containing patient information or clinical

team folders that aggregate data relevant to a particular healthcare team's activity. In the openEHR model, Folder structures do not contain Compositions, only references to them. Consequently, more than one Folder can therefore refer to the same Composition.

Overall, the RM represents the global characteristics of health records entries, how they are aggregated, and the context information required to meet the ethical, legal, and provenance requirements.

Due to its modular and customizable nature, the openEHR RM is often compared to the LEGO® approach. This analogy draws upon the inherent versatility of LEGO blocks, which can be assembled and combined in countless ways to create various structures, due to the adherence of LEGO bricks to standardised dimensions and connection points. Similarly, the RM provides a framework of modular components that can be configured and integrated to construct comprehensive clinical data models tailored to meet the specific needs of different healthcare domains or use cases. Mirroring the systematic approach employed in assembling LEGO pieces, the RM follows standardised data structures and semantics, ensuring interoperability and consistency across different implementations.

Furthermore, just as LEGO sets frequently include detailed instructions for building specific models, the openEHR RM offers comprehensive guidelines and specifications for creating clinical models using archetypes and templates. These directions help to ensure that clinical data is represented consistently and accurately across different implementations, emulating the LEGO approach to ensure uniformity in building LEGO sets.

The parallels between LEGO and openEHR RM emphasise the modular, customizable, and standardised nature of both systems, enabling the construction of coherent and consistent models, whether in the realm of physical construction or healthcare information management.

In general, the openEHR RM provides a standardised and interoperable framework for organizing, representing, and exchanging health information within EHR systems. By adhering to the principles and structures defined by the RM, healthcare organizations can achieve consistency, flexibility, and scalability in managing health records, ultimately supporting improved patient care and clinical decision-making.

2.2.2.2 The Archetype Model

While the RM lays the technical foundation for establishing health data, it is the implementation of archetypes and templates that enables the detailed specification and representation of specific care events and subjects of care within the EHR system. The Archetype Model (AM) contains various modules and packages that ensure the representation of the clinical content and the definition of standardised, domain-specific clinical concepts and relationships. Within the openEHR ecosystem, archetypes and templates serve as critical artefacts/resources in the standardised and structured representation of clinical concepts.

An archetype is a computable and reusable model that defines the structure and semantics of specific clinical concepts, such as blood pressure, medications, or procedures. These models encapsulate domain knowledge and specify how information about a particular clinical entity should be represented.

OpenEHR archetypes define the structure, constraints, and semantics of specific data elements of a clinical concept, ensuring consistency and interoperability across different healthcare systems. In addition, whenever necessary, these archetypes allow the incorporation of external terminologies, including but not limited to IDC-10, SNOMED, and LOINC, further augmenting their adaptability and semantic richness.

In the field of archetype modelling, nodes represent individual elements encapsulating distinct components, such as data fields, data structures, or constraints. Each node is associated with a specific data type, value range, and other attributes that govern its behaviour and usage. Furthermore, constraints define rules, conditions, and assertions that must be satisfied for the archetype to be valid and meaningful.

In traditional healthcare systems, the proliferation of multiple models aimed at representing a singular clinical concept poses a significant challenge to interoperability. Hence, within the openEHR framework, meticulous attention is paid to the construction of each archetype to encompass the maximum data set for representing a given clinical concept in its entirety and with utmost comprehensiveness. These models are crafted by domain experts and clinicians and are the result of collaborative endeavors to accurately capture the essential characteristics of a clinical concept.

An archetype is a computable model because it is constructed using Archetype Definition Language (ADL), a formal language that was designed to be both human-readable and machine-processable. ADL, which is the basis of ISO standard 13606-2, serves as a standardised syntax for domain experts to express archetypes, enabling their creation, validation, and utilization within EHR systems, promoting interoperability, reusability, and semantic consistency of clinical data.

Hence, archetypes are a powerful and intuitive means for domain experts to define computable models of clinical concepts and play a fundamental role in the development of interoperable EHR systems, as they enable standardised representation and meaningful exchange of clinical data across diverse healthcare systems in a consistent and semantically clear manner.

The Clinical Knowledge Manager (CKM) is an online structured environment where healthcare professionals, domain experts, and informaticians all over the world meet for the collaborative development, review, sharing, and governance of clinical content, including archetypes and templates. It is a central component of the openEHR ecosystem that fosters collaboration, standardization, and quality assurance in the development of clinical models.

Archetypes serve as building blocks and are the fundamental units for creating comprehensive and domain-specific clinical models, known as templates. In the LEGO metaphor, an archetype embodies the smallest building component, which can be combined with numerous other components to delineate a variety of clinical scenarios. Figure 4 illustrates this phenomenon through the reutilization of the same archetypes, for instance, Blood Pressure, across two different clinical situations: a cardiology encounter and an obstetric admission.

Templates are therefore higher-level constructs built upon the combination of one or more archetypes to create more comprehensive and specialised clinical models tailored for specific use cases or healthcare domains. A template is created using a Composition archetype that works as a parent entity that aggregates multiple child archetypes. In other words, templates define the composition of a specific type

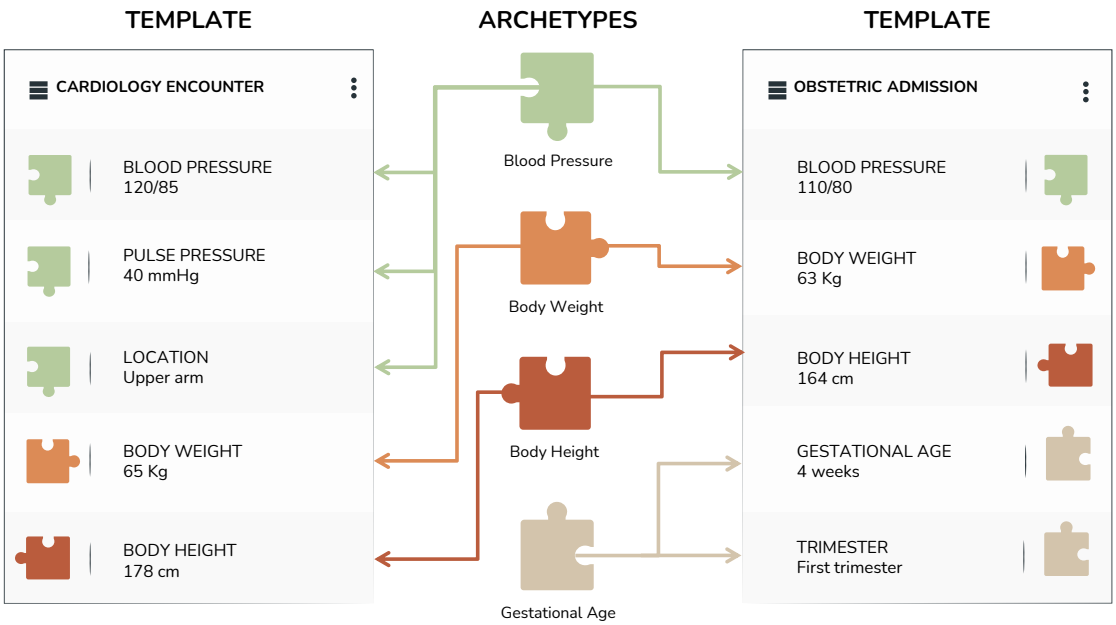


Figure 4: Archetype reuse across different templates.

of clinical document or record, specifying which archetypes should be included and how they should be organised.

In conclusion, templates encapsulate the clinical requirements of specific healthcare contexts while enforcing consistency in the generation of clinical documents, facilitating the creation of customizable and domain-specific EHR solutions.

Together, archetypes and templates form the backbone of the openEHR approach to clinical modelling, offering a robust mechanism to achieve standardised and interoperable representations of clinical concepts. The role of archetypes and templates in healthcare organizations is critical to achieve consistency, interoperability, and semantic clarity in the representation and exchange of clinical data.

Once the data has been entered and validated, it can be stored within compositions, which serve as structured containers, encapsulating validated clinical data in a coherent and standardised format. Moreover, compositions facilitate longitudinal tracking of patient health information, enabling healthcare providers to access and analyze historical data for informed decision-making and continuity of care. This access is ensured by AQL, a specialised query language very similar to Structured Query Language (SQL), which offers the possibility of querying and retrieving clinical data stored in EHR systems. It provides a standardised syntax that enables healthcare professionals, researchers, and developers to formulate complex against openEHR-compliant EHR data, facilitating efficient data retrieval and analysis.

2.2.3 Summary Remarks

In the previous sections, we have delved into the diverse landscape of information technology in healthcare, as well as the growing emphasis on the development and implementation of ethical HIS. With the

increasing integration of technology in healthcare and the astonishing evolution of AI in recent years, stakeholders across the entire healthcare spectrum are under even more pressure to race in the pursuit of imbuing algorithms and systems with ethical decision-making capabilities.

A key aspect of this process is the mitigation of bias and discrimination. In healthcare, biased algorithms can perpetuate disparities across the entire care pathway, disproportionately affecting marginalised communities. In recent years, several researchers have devoted the matter of their studies to shed light upon this topic and unravel the complexities of its nuances. The developers must implement measures to identify and mitigate bias at every stage of the AI lifecycle, from data collection and preprocessing to model training and validation. This includes diversifying training datasets, conducting bias audits, and incorporating fairness metrics to evaluate algorithmic performance across different demographic groups.

At the core of any algorithmic decision-making process lies the data it operates upon. This data serves as the raw material from which insights are drawn and decisions are made. Hence, if this foundational layer is compromised, no amount of improvement in the higher layers can repair the underlying issues.

Consider a scenario where an algorithm is tasked with determining patient risk and resource allocation within ICU based on historical data. If the dataset used to train the algorithm predominantly consists of patients from a certain demographic group or socioeconomic background, the algorithm may inadvertently learn to favor or prioritise patients with similar attributes, regardless of any ethical guidelines imposed at higher levels. Consequently, patients who do not fit the algorithm's learned characteristics may receive inadequate attention or delayed interventions. If left unaddressed, such bias can reinforce systemic inequities within the healthcare system, ultimately compromising patient welfare and exacerbating social injustices. Similarly, if the data is plagued with inaccuracies or lacks proper context and meaning, the resulting decisions will be flawed, potentially leading to erroneous assessments of patient conditions and unfair healthcare outcomes. Ultimately, these flaws in decision-making stemming from data deficiencies can perpetuate bias as well as undermine the quality of care provided to patients, jeopardizing their well-being and contributing to preventable medical errors.

Therefore, despite the numerous endeavors that have been undertaken to tackle ethical concerns, a hypothesis unfolds to address this issue: to truly ensure ethically compliant decision-making, focusing solely on refining the upper layers of these systems is insufficient; instead, it is necessary to delve deeper into the data layer, the very foundation upon which these systems are built.

The essence of the present thesis lies in the exploitation of this hypothesis through rigorous scrutiny and implementation of openEHR specifications to ensure that the data layer is structured, comprehensive, semantically clear, and devoid of biases.

Among the different health standards and protocols, openEHR stands out as a comprehensive and flexible solution for managing EHRs in a standardised and interoperable manner. Unlike traditional EHR systems, which often suffer from vendor lock-in, data silos, and interoperability challenges, openEHR offers a vendor-neutral, archetype-based approach that promotes semantic interoperability, information sharing, and data persistency.

The key strength of openEHR lies in its dual-modelling approach, which separates the clinical content - AM - from the technical implementation details - RM -, enabling healthcare organizations to create, customise, and exchange clinical models. The RM serves as the foundational framework for organizing and structuring EHR data, providing a common data structure and semantics that underpin interoperability and data exchange. Archetypes, on the other hand, define the content, structure, and semantics of specific clinical elements, allowing for domain-specific customization and extension. This division of concerns promotes a collaborative environment between clinicians and software developers, fostering the development of interoperable and domain-specific clinical models.

Furthermore, openEHR holds the ability to align with internationally recognised standards and principles, such as ISO 13606, allowing seamless integration with messaging standards and clinical terminologies. This interoperability extends across organizational boundaries, enabling healthcare organizations to exchange data seamlessly with external partners, specialists, and regulatory bodies.

Another noteworthy feature of openEHR is that, unlike traditional EHR systems that often fragment patient information across disparate platforms or institutions, it provides a unified framework to capture, store, organise, maintain, and access comprehensive health records throughout an individual's entire lifespan. The support for life-long EHRs enables a longitudinal perspective of health data that not only enhances continuity of care, allowing individuals to receive consistent and coordinated care across different healthcare settings, but also supports advanced analytics, population health management strategies, and research initiatives aimed at improving healthcare outcomes.

Theoretically, the adoption and integration of openEHR empowers healthcare organizations to derive actionable insights, overcome interoperability issues, improve care coordination, and unlock the full potential of digital health technologies to drive clinical decision-making. The current thesis will be dedicated to the transformative power of openEHR to enhance healthcare delivery, particularly focusing on its impacts in driving innovation in the ethical health landscape.

In conclusion, while efforts to imbue ethics into algorithms and systems are commendable, it is only by fortifying the foundational stratum that one can aspire to foster truly ethical decision-making within healthcare systems, thereby mitigating the risk of perpetuating harm or injustice. In essence, this thesis defends that the journey towards ethical HIS and AI begins not at the surface, but at the very core — the data itself.

3

Research Methodologies

"Do not give them a candle to light the way, teach them how to make fire instead. That is the meaning of enlightenment."
— Kamand Kojouri

Any research project requires the use of specialised methodologies to establish precise standards for a clear formulation of objectives and functionalities, as well as delineating the main steps and processes to be followed in an orderly manner throughout the thesis. In the pursuit of achieving a solid answer to the research question raised in Chapter 1, a new artefact is envisaged. artefact, from the Latin roots "arte" and "factus", means something crafted with skill and intentionality. It can be defined as an artifice designed for specific purposes, typically arising from deliberate planning or creative endeavor. artefacts are not restricted to physical objects, encompassing intangible intellectual artifices as well. Once the nature of the artefact is understood, the underpinnings of the investigation are clarified through a deep and exhaustive literature review. Then, research into paradigms, methodologies, philosophies, and approaches can take place.

Without assigning a paradigm to establish the purpose, motivation and prospects of the investigation, there is no ground for selecting methodologies, methods, literature, or research design (Mackenzie and Knipe, 2006). This research fits in the scope of engineering and will be conducted under the pragmatic paradigm. One of the foundational principles of this philosophy is the prioritization of the research question as the focal point of the investigation, guiding the selection of methodologies, methods, and data collection techniques (Mackenzie and Knipe, 2006) (Goldkuhl, 2004). Additionally, action is a fundamental backbone of the pragmatic paradigm and is guided by purpose and knowledge. From the pragmatic point of view, knowledge is shaped by action, i.e., there is an inherent connection between human knowledge and human action (Goldkuhl, 2004). Embracing the pragmatic view that knowledge is intricately linked to action, the methodology is designed to generate insights that inform practical outcomes and enhance real-world practices. As such, the research methodology emphasises a hands-on, action-oriented approach that is well-suited for research approaches intervening in the real world, such as the development of artefacts as in DSR. Furthermore, the philosophy of pragmatism was adopted since this research

entails the creation of an artefact – an IS – which is always part of a context of practice (action) that dictates its functionalities. Moreover, this paradigm is characterised by its flexibility and adaptability, allowing for iterative refinement and adjustment in response to emergent findings and changing circumstances. Accordingly, through the pragmatic approach, the thesis adopts a clear action focus, with data collection and analysis methods chosen as those most likely to provide tangible outcomes and actionable insights into the research question.

When conducting a research project, it is imperative to define the direction of theorizing. This investigation follows a deductive direction, in which, an investigation strategy is developed to move from an abstract level of ideas or theoretical relationships to a concrete level of observable empirical evidence (Neuman, 2007). Deductive reasoning begins with general premises, i.e., theoretical propositions among concepts, and transforms those propositions into testable empirical hypothesis that are yet to be tested or verified with empirical evidence or data, to later proceed to derive specific conclusions through logical inference. The certainty of a theoretical proposition intensifies upon repeated empirical testing of the corresponding hypothesis in different situations. In this way, the deductive approach is applied for the systematic literature review carried out in this thesis, as well as for the development of the artefact, which is grounded in established theories and concepts.

An inductive approach, on the other hand, moves in the opposite direction, starting from specific observations or empirical evidence and moving towards generalizations or theoretical propositions. This approach is particularly useful when exploring new phenomena or generating novel insights, as it allows for the discovery of patterns or relationships in the data that can inform the development of new theories or concepts. Hence, the inductive approach is well-suited for the final contribution of this doctoral thesis, in which algorithms and models are developed based on the empirical data collected from the ICU environment.

The investigation bounded to the doctoral thesis "Ethics and New Trends in Artificial Intelligence in Intensive Care Units" was conducted at the ALGORITMI Research Center, a research unit of the School of Engineering at the University of Minho, Braga, Portugal. The ALGORITMI Research Center has collaborations with several hospitals across the country. The artefact that guides this research was initially designed, implemented, and evaluated at the ULSdSA.

While the main focus of this doctoral thesis centers on the ICU due to its critical nature and the complexity of patient care within this environment, drawing parallels and extrapolating findings from other clinical contexts can enrich the depth and breadth of the investigation. As such, it is meant to expand the number of use cases with the intention to not only incorporate insights gleaned from other healthcare settings, but also to confirm and enhance the credibility of the resulting artefact, improving the generalizability and applicability of the findings. By incorporating use cases across different healthcare specialties and contexts, this thesis offers a comprehensive understanding of the pervasive principles and challenges inherent to healthcare decision-making and information management. Moreover, the greater the number of use cases, the greater the diversity and, consequently, greater are the chances of developing a robust and generalised artefact, fostering a holistic approach to addressing challenges and advancing knowledge.

Accordingly, this thesis fits in the spectrum of exploratory research as it seeks to find new perspectives and strategies to overcome inefficiencies and handicaps detected in the ICU environment.

The primary data was obtained through participant observation, survey questionnaires, semi-structured interviews, and focus groups. Ethical principles are a central aspect in this research and, logically, when collecting primary data ethical considerations were safeguarded through the assurance of privacy, confidentiality, and sensitivity concerns towards all the participants. In turn, the secondary data covers material from primary sources, such as theses and conference proceedings and secondary sources, such as journals and books. As mentioned, this investigation fits in the scope of engineering and, consequently, it is sustained by quantitative (numerical) data for the most part. In addition, a great portion of this research also resorts to qualitative (non-numerical) data. Hence, different types of quantitative and qualitative research methods were used both jointly and independently, revealing a mixed methods research.

In what concerns the time horizon, this research is both cross-sectional and longitudinal. It is classified as a longitudinal study, as it is conducted over an extended period of time, allowing to track the evolution of the artefact, assess its impact, and identify areas for improvement and refinement. Moreover, the longitudinal design is well-suited for evaluating the sustainability and scalability of the artefact, as it provides insights into its long-term implications. The research is also cross-sectional, as the systematic literature review examines a snapshot of the available research on ethics, HIS, openEHR, and AI within a defined period. In addition, the empirical data collection is conducted at a single point in time, providing a contemporary view of the ethical implications and emerging trends in the integration of AI in the ICU.

As mentioned, the development of this artefact was sustained by a set of methodologies and procedures that guided a logical direction for the transformation of ideas into something solid and concrete. Evidently, this involved a broad research and a careful analysis of the different methodologies available and feasible for the development of the artefact envisioned. After analyzing the existing possibilities, the DSR methodology was deemed the most appropriate to achieve the objectives initially outlined, since it is largely used in the development of IT solutions and artefacts. The DSR is therefore in perfect harmony with the aforementioned pragmatic perspective, which emphasises the importance of action and practical outcomes.

In addition, other evaluation measures were considered, which includes a Proof of Concept (PoC) as a critical precursor to the development of the artefact, system usability tests made by different types of users, a Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis to assess the competitive advantages and strategic implications of the artefact, and the TAM to gauge the acceptance and usability of the artefact.

Figure 5 contains a visual representation of the research onion framework established for the current thesis.



Figure 5: The methodological layers: research onion representation.

3.1 Design Science Research Methodology

DSR is a demanding and rigorous scientific research methodology that aims to develop and evaluate innovative artefacts to allow professionals to process the organizational information and the development of actions against real-world problems (Bilandzic & Venable, 2011; March & Storey, 2008). This is a problem-solving approach who has two main purposes: (1) to develop an artefact to solve a practical problem in a specific context and (2) to generate new technical and scientific knowledge.

Thus, the resulting artefact must be a viable and feasible technological solution able to respond to the initial goals and solve the problem at hand. Its usefulness, quality, and efficacy must be rigorously demonstrated by well-executed evaluation methods (A. Hevner & Chatterjee, 2010; A. R. Hevner et al., 2008). Furthermore, this epistemological approach must provide clear and verifiable contributions.

The model consists of six interconnected main stages: problem identification, definition of the objectives, design and development, demonstration, evaluation, and communication. The iterative nature of the DSR allows the refinement and enhancement of the artefact throughout the research process until the best possible solution is achieved. This iterative process is crucial to ensure that the artefact is continuously improved and refined to meet the requirements and objectives of the research project. Basically, the process associated with this methodology defines:

1. **Problem Identification and Motivation**: The first step in DSR involves identifying a problem or opportunity in a specific domain and understanding the motivation behind addressing it. This often involves reviewing existing literature, analyzing current practices, and identifying gaps or areas for improvement;

2. **Objectives Definition**: Once the problem is specified, clear objectives are defined to outline what the research aims to achieve, such as developing a new artefact, improving existing systems,

or advancing knowledge in the field. The precision and rigor with which objectives are delineated directly correlate with their achievement;

3. **Design and Development**: In this step, researchers design and develop a solution or artefact to solve the identified problem or achieve the objectives outlined. Its development must follow a rigorous scientific process based on the knowledge and the theory already explored. For this to happen, the initial research has to be as enlightening and efficient as possible so that the whole process flows as expected (A. Hevner & Chatterjee, 2010).

4. **Demonstration**: After the artefact is developed, its effectiveness and usability needs to be demonstrated. The demonstration may involve the utilization of the artefact, a simulation, a case study, or another pertinent activity;

5. **Evaluation**: This step involves testing the artefact in real-world or simulated environments, collecting data, and analyzing results to evaluate its performance and impact. In addition, the proposed objectives should be compared with the actual results observed when using the artefact;

6. **Communication and Propagation**: The findings from the evaluation step are communicated to the target audience through publications, presentations, or other means (March & Storey, 2008; Vaishnavi & Kuechler, 2015). Additionally, a reflection on the design process, lessons learned, limitations, and areas for future research is made.

Figure 6 illustrates the DSR model approach used throghout the current thesis, outlining the iterative process of artefact development and evaluation.

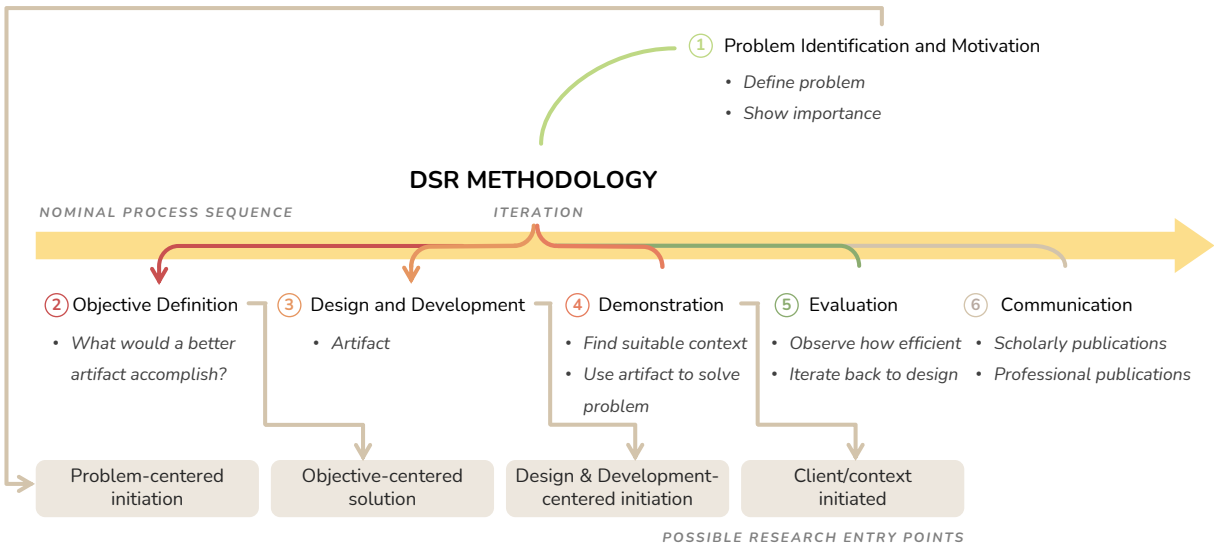


Figure 6: Methodological steps intrisic to the DSR process. Adapted from (Peffers et al., 2007).

By following the DSR model, this research aims to create a tangible and practical solution that addresses the challenges that currently exist in the ULSdSA hospital, particularly the ethical implications and

emerging trends surrounding the integration of AI and HIS in critical healthcare settings. To achieve this goal, the research will focus on the development of innovative intellectual and practical contributions that aim to enhance the quality and efficiency of ethical health decision-making and information management through standard-based IT solutions to ultimately provide actionable insights and ease clinical practice for health professionals.

3.2 Case Study Approach

The Case Study approach is a widely recognised and methodologically robust research method, particularly suited for an in-depth and multi-faceted investigation of complex phenomena within their real-life contexts (Flyvbjerg, 2011; Heale & Twycross, 2018). This approach is characterised by its ability to provide rich and detailed insights into the interactions between various factors and variables, offering a comprehensive understanding of the research topic (Takahashi & Araujo, 2020).

Typically, a case study has a well defined space and time frame. In terms of scope, case studies are an in-depth investigation focused on a specific phenomenon, event, situation, individual, group, or organization within its true context, i.e., they are used to explain, describe or explore events or phenomena in the everyday contexts in which they occur (Yin, 2009).

This methodological approach is particularly well-suited for research projects that aim to capture information at an explanatory level, providing insights into the questions 'how', 'what', and 'why' of a particular phenomenon or event (Takahashi & Araujo, 2020). By examining the interplay between multiple factors and variables, case studies can capture complex and nuanced insights into gaps, challenges, and opportunities that might not be readily apparent through other research methods (Wohlin, 2021).

In the context of the current thesis, the utilization of case studies provides a structured and systematic framework for the in-depth exploration of the artefacts developed under the DSR methodology in different clinical settings.

The initial case studies aim to investigate the practical implementation, usability, and efficiency of the openEHR-based artefact in real-world healthcare settings, focusing on its ability to improve the quality and efficiency of healthcare documentation, decision-making, and information management. In addition, the case studies aim to add value to the artefact and improve its robustness, reliability, and usability through iterative feedback and refinement.

In addition, the case study approach was used to investigate the performance of different ML algorithms across various clinical contexts, patient populations, and medical conditions before considering their application in high-stakes environments such as the ICU. The exploration of different datasets helps to identify nuances, challenges, and unique characteristics that affect algorithmic performance and adaptability. The comparison of the outcomes of several algorithms on diverse datasets allows to discern trends, patterns, and common challenges encountered in different medical scenarios. This comparative analysis also contributes to understanding the robustness and limitations of various ML approaches in real-world

applications. Additionally, studying algorithmic behaviour under different conditions, such as variations in patient demographics, disease prevalence, and healthcare practices, helps to distinguish algorithmic strengths and weaknesses, contributing to the development of more robust and adaptable models. Furthermore, by assessing the performance of multiple algorithms, it is possible to establish benchmarks and best practices for algorithm selection and deployment, ultimately improving the quality and reliability of ML applications in clinical settings.

3.3 Analysis and Evaluation Segment

The analysis and evaluation segment of this doctoral thesis is a critical component that underpins the development and validation of the artefact. This segment encompasses a range of methodologies and techniques that are employed to evaluate the proposed solution in a rigorous and comprehensive manner, laying its potential and shortcomings.

In terms of analysis and evaluation, the PoC is employed as a critical precursor to the development and implementation of the proposed solution. In addition, the SWOT analysis and the TAM emerge as indispensable tools to assess the viability, acceptance, and usability of the artefact developed in the context of contribution II - Transforming Eletronic Health Records: A Focus on openEHR Forms Development (Section 4.2). The SWOT analysis provides a structured framework for evaluating the strengths, weaknesses, opportunities, and threats associated with the openEHR-based solution, enabling a comprehensive assessment of its competitive positioning and strategic implications. Finally, the TAM is employed to gauge the acceptance and usability of the artefact.

A detailed description of the methodologies and techniques used in the analysis and evaluation segment is provided in the following sections.

3.3.1 Proof of Concept

In the context of technological innovation and scientific research, a PoC is a demonstration or prototype that demonstrates and ascertains the feasibility and viability of a proposed idea or technology before full-scale implementation. Therefore, the realization of a PoC is commonly reported as being one of the most important steps in the design, development, implementation, and proposal process of a prototype in the IT area (Schmidt, 2006).

The PoC or Proof of Principals (PoP) is a structured validation process that aims to verify if the theoretical concept can be proved through the practical model, i.e., it allows to demonstrate in practice the concepts, methodologies, and technologies involved in the development of a given artefact (Sergey et al., 2015). In this way, it bridges the gap between theory and practical application, paving the foundation for successful implementation and design of novel ideas.

Besides identifying potential failures or errors in the proposed solution, the utilization of a PoC enables the verification of whether the solution conforms to the requirements and objectives for which it was

originally conceived. This entails assessing if the purposes established prior to the beginning of the project's development process have been met or not, and in the latter case to understand what needs to be done towards this end (Schmidt, 2006).

Tipically, a PoC involves a limited number of users to determine within a controlled environment if the proposed concept or idea complies with predefined requirements and can be implemented to efficiently achieve the desired outcomes. This approach allows the early identification of benefits for target organizations, allowing for stakeholder buy-in before committing resources to a full-scale implementation. Accordingly, employing a PoC reduces costs and mitigates risks associated with non-compliant or defective solutions, as it requires decision-making in advance (Schmidt, 2006).

In light of the significance and intricate nature of the subject matter under examination, the PoC emerges in this thesis as an indispensable assessment tool to evaluate the efficacy and reliability of the proposed solution. Through this analytical framework, the PoC not only elucidates the viability of the concept but also furnishes invaluable insights to inform strategic decision-making processes.

3.3.2 Strenghts, Weaknesses, Opportunities and Threats Analysis

The SWOT analysis, as the designation suggests, is a strategic planning technique that allows the systematic evaluation of the main strengths and weaknesses intrinsic to a given product or organization (internal factors), as well as the opportunities and threats it faces (external factors) (Benzaghta et al., 2021).

Despite the fact that the SWOT analysis emerged through the collaborative efforts of multiple scholars, its roots can be traced back to the mid-20th century, with Albert Humphrey at the Stanford Research Institute in the 1960s (R. W. Puyt et al., 2023). Initially known as Satisfatory, Opportunity, Fault, and Threat (SOFT) analysis, it was later refined into SWOT analysis through the contributions of several domain experts (R. Puyt et al., 2020). Over time, the concept gained popularity and it has become a widely used tool in business strategy, academia, and other domains. Organizations can maximise their competitive advantage by identifying strengths and accelerating opportunities (R. W. Puyt et al., 2023). Moreover, recognizing weaknesses and addressing threats enables proactive risk mitigation and strategic adaptation. Hence, the SWOT analysis helps stakeholders to gain insights into the competitive positioning and potential challenges facing the organization, leading to informed strategic decisions.

This method evaluates both internal and external factors affecting a given product or organization, compiling these characteristics into a structured SWOT matrix, as illustrated in Figure 7. The matrix representation presents a concise snapshot of the strategic landscape of a given product or organization, easing the visualization and analysis of its attributes and substantiating decision-making processes and alternative courses of action. The characteristics represented in the matrix can be summarised as follows:

Strengths: refer to internal attributes and resources that confer a competitive advantage, i.e., the qualities and advantages that a given product or organization has or does not have in relation to its competitors;

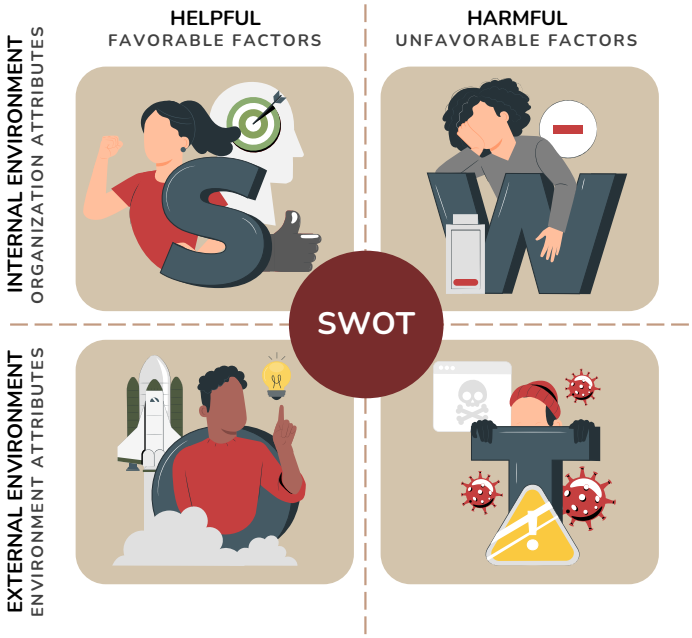


Figure 7: SWOT analysis matrix. Adapted from (Benzaghta et al., 2021).

Weaknesses: denote internal limitations or deficiencies that may interfere with or prevent the product or organization from achieving its objectives;

Opportunities: represent external factors that could be exploited, i.e., environmental conditions that could be beneficial to a particular product or organization;

Threats: encompass external factors that pose risks, i.e., environmental conditions that may jeopardise the development of the product or that may negatively influence the organization.

A sharp glance at Figure 7 reveals that helpful factors are those that contribute in a positive way to achieve the objectives initially outlined (left part of the matrix), while harmful factors have opposite repercussions, hindering the development and evolution of the product or organization and are therefore seen as adversities (right part of the matrix). These factors can be further divided into two distinct categories: the internal and external environment. The internal environment is the one whose organization or product can self-control (top part of the matrix) and relies on its respective strengths and weaknesses. On the other hand, the external environment refers to environmental factors beyond the control of the organization or product (bottom part of the matrix) and depends on opportunities and threats. Therefore, weaknesses and strengths are evaluated in terms of internal attributes, while opportunities and threats are assessed within the external factors.

In essence, the SWOT analysis supports strategic decision-making in several organizational contexts through a comprehensive understanding of the organization's current condition and its external environment. However, it is critical to recognise that it is susceptible to biases, such as overemphasis on internal perspectives or overlooking key external variables.

The decision to use the SWOT analysis for evaluating the artefacts developed within the current thesis holds different arguments. First, it allows to investigate the internal and external factors that influence the artefact's development, deployment, and potential impact. This includes assessing the strengths of the artefact, such as innovative features or technological advancements, which can highlight its competitive advantages and unique contributions. At the same time, weaknesses intrisic to the artefact, such as limitations in functionality or scalability, can be identified for strategic improvement. Moreover, opportunities in the external environment, such as emerging trends or market demands, can be explored to enhance the relevance and adoption of the artefact. Besides, threats such as technological obsolescence or competing solutions can be identified to mitigate risks and refine strategic positioning.

3.3.3 Technology Acceptance Model

The driving force behind the conception of the TAM was the growing concern regarding the acceptance and adoption of IT solutions made available to individuals (F. D. Davis, 1989; F. D. Davis et al., 1989). Davis (F. D. Davis, 1989) and Davis et al. (F. D. Davis et al., 1989) introduced the TAM in the 1980s to shed light on the reasons why users accept or reject IT. The inventors of the model reasoned that the key to increasing use was to first increase acceptance, which could be measured through the user's intent to use the system. The more accepting the users are of a new information system, the more likely they are to make changes in their behaviour and enforce their time and resources to learn and use the system (Pikkarainen et al., 2004). Consequently, organizations that can decipher the factors that shape user's intent possess the capacity to manipulate those factors, thereby promoting acceptance and subsequent use of the IT solution (Holden & Karsh, 2010). Hence, TAM seeks to elucidate the behavioural intention of potential users to adopt a new technology (King & He, 2006), which is a critical determinant of the actual use of the technology.

This theoretical model is based on the Theory of Reasoned Action (TRA) proposed by Fishbein and Ajzen (Fishbein & Ajzen, 1975), a well-established social-psychological theory that explains human behaviour through the attitudes and intentions of individuals. According to TRA, an individual's behaviour is determined by their intention to perform a given action, which is influenced by their attitude towards the behaviour and the subjective norms surrounding it (Gillenson, Sherrell, et al., 2002).

Originating in the late 1980s, the initial form of TAM relied on five major constructs: perceived usefulness, perceived ease of use, attitude toward using, behavioural intention to use, and actual system use. Despite being a groundbreaking model to explain information system acceptance and use at the time, the TAM was criticised for its simplicity, with many researchers suggesting it needed to be given additional variables to enhance its predictive power (Legris et al., 2003). In response to these critiques, Venkatesh and Davis (Venkatesh & Davis, 2000) proposed, in 2000, an extension of the original TAM, the TAM2, which introduced external variables such as social influence (subjective norm and image) and cognitive instrumental processes (job relevance, output quality, and result demonstrability). The attitude toward using was disregarded because it was a weak predictor of both behavioural intention to use and actual

system use. The TAM2 was later revised and refined by Venkatesh, Morris, and Davis (Venkatesh et al., 2003) into the Unified Theory of Acceptance and Use of Technology (UTAUT), in an attempt to unify the IT acceptance literature. The UTAUT is a comprehensive model that incorporates four key predictors of users' behavioural intention, namely performance expectancy, effort expectancy, social influence, and facilitating conditions (Lai, 2017; Sharma & Mishra, 2014). In addition, the model incorporates perceived usefulness into a performance expectancy construct and perceived ease of use into effort expectancy (Holden & Karsh, 2010; Lai, 2017).

Since its genesis, the TAM has been submittted to significant adjustments and refinement, propelled by advancements in technology, changes in user behaviour, and the emergence of new research paradigms. Furthermore, the proliferation of technology in organizational settings and the growing importance of social and cultural factors in determining technology acceptance have prompted researchers to expand the scope of the TAM beyond its original focus at individual-level technology adoption to encompass organizational and societal dimensions. In 2008, Venkatesh and Bala (Venkatesh & Bala, 2008) refined the TAM2, integrating other established theories, such as TRA and Decomposed Theory of Planned Behaviour (DTPB), to create the TAM3, an integrated model of technology acceptance that incorporates a complete nomological network of the determinants of individuals' IT adoption and use (Lai, 2017). The TAM3 expands the number of determinants that affect perceived usefulness and perceived ease of use (Venkatesh & Bala, 2008). The factors that determine perceived usefulness are subjective norm, image, job relevance, output quality, and result demonstrability. On the other hand, perceived ease of use is influenced by anchor variables (computer self-efficacy, perceptions of external control, computer anxiety, computer playfulness), and adjustment variables (perceived enjoyment and objective usability). The model also includes additional variables that moderate the relationship between perceived usefulness and perceived ease of use and behavioural intention to use, such as experience and voluntariness (Lai, 2017; Venkatesh & Bala, 2008). Figure 8 illustrates the TAM3 model, highlighting the key constructs and relationships that influence technology acceptance and use.

A complete list of the constructs included in the TAM3 is presented in Table 3, which includes a brief explanation for each construct (Venkatesh, 2000; Venkatesh & Bala, 2008).

Table 3: Description of TAM3 constructs

Construct	Name	Description
PEU	Perceived Ease of Use	The degree to which an individual considers that using the system is effortless.
PU	Perceived Usefulness	The degree to which an individual believes that using the system will improve their job performance.

(Table 3 continued on next page)

(Table 3 continued from previous page)

Construct	Name	Description
BI	Behavioural Intention	The degree to which an individual has the intention to perform a certain behaviour.
UB	Use Behaviour	The frequency, duration, and intensity to which an individual interacts with the system.
SN1	Subjective norm	The degree to which an individual feels social pressure to use or to not use the system.
SN2	Image	The degree to which an individual thinks that using the system improves their image or social status.
JR1	Job relevance	The degree to which an individual finds that using the system will be applicable and relevant to their job.
OQ1	Output quality	The degree to which an individual considers that the system will provide output that meets their job requirements.
RD1	Result demonstrability	The degree to which an individual believes that the results of using the system are observable and communicable to others.
	Anchor Variables	
AV1	Computer self-efficacy	The degree to which an individual believes in their ability to execute job-related tasks using the computer.
AV2	Perceptions of external control	The degree to which an individual considers that external factors such as organizational and technical resources, are available to support their use of the system.
AV3	Computer anxiety	The degree to which an individual feels anxiety and apprehension when faced with the possibility of using a computer.
AV4	Computer playfulness	The degree to which an individual has cognitive spontaneity to interact with a computer.
	Adjustment Variables	

(Table 3 continued on next page)

(Table 3 continued from previous page)

Construct	Name	Description
AJ1	Perceived enjoyment	The degree to which an individual finds the activity of using the system enjoyable.
AJ2	Objective usability	The degree to which an individual believes that the system is easy to use based on the actual level (rather than perceptions) of effort required to complete specific tasks.
	Moderating Variables	
MV1	Experience	The degree to which an individual has prior experience/contact with similar systems.
MV2	Voluntariness	The degree to which an individual believes that using the system is of free will.

The integration of all these constructs into a single survey allows for a comprehensive understanding of the aspects that influence healthcare professionals' acceptance of the technological artefact developed in this thesis, thereby informing the design, implementation, and management of a digital solution tailored to their specific needs. As human beings, healthcare professionals exhibit diverse preferences, attitudes, and skill levels, which directly impact their willingness to adopt new technologies. The TAM3 allows to enhance the validity and generalizability of the evaluation outcomes by considering variations in individual differences and complex contextual factors.

3.4 Research Design Overview

The previous sections of this chapter have outlined the main methodologies and approaches employed in this doctoral thesis, including the DSR methodology, the Case Study approach, the PoC, the SWOT analysis, and the TAM. All these methodologies and approaches were integrated into a comprehensive research design that guides the development and evaluation of the artefacts created in this thesis. The research design is structured around the DSR methodology, which is employed to develop and evaluate innovative artefacts to address or enlighten the ethical considerations of using HIS and AI in the ICU. The research process also incorporates the Case Study approach to provide in-depth insights into the ethical IT framework proposal and the challenges associated with the integration of ethically compliant AI in critical healthcare settings. Hence, the Case Study approach is employed to investigate the practical implementation of openEHR and AI technologies in ICUs, focusing on the development of innovative solutions that enhance ethical decision-making and information management. Furthermore, a PoC, a SWOT analysis and a questionnaire based on TAM3 to were used to assess the overall acceptance and

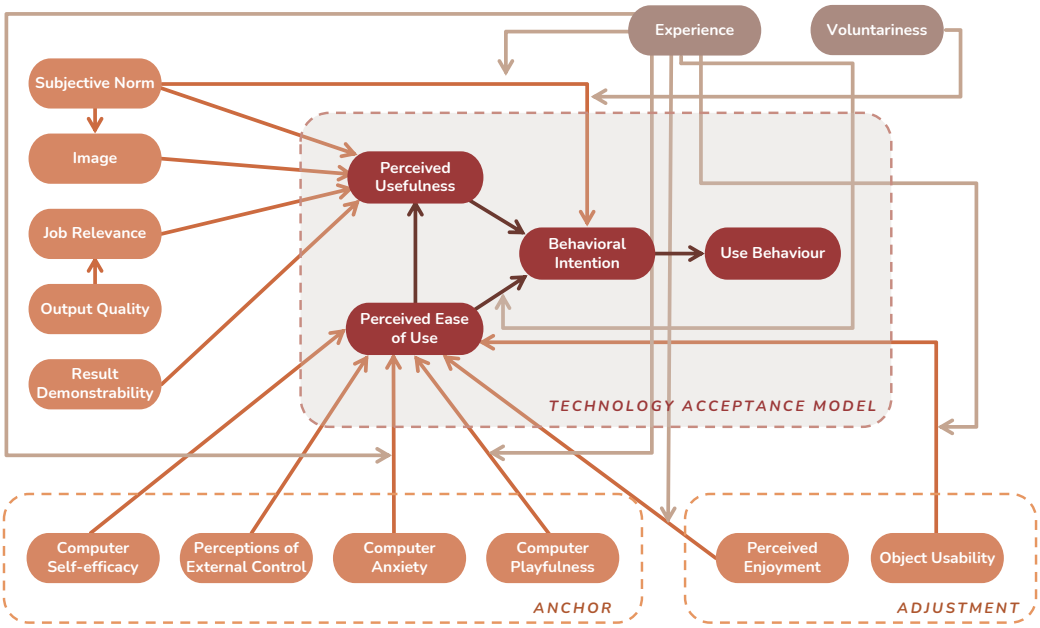


Figure 8: Constructs of the TAM3. Adapted from (Holden & Karsh, 2010; Venkatesh & Bala, 2008).

usability of the developed artefact. The integration of these methodologies and approaches in this thesis aims to generate actionable insights and recommendations for healthcare practitioners, policymakers, and stakeholders, thereby advancing the responsible and ethical use of IT and AI in critical healthcare contexts. Figure 9 provides an overview of the research design, illustrating the bigger picture of the methodologies and approaches integrated into the research process carried out.

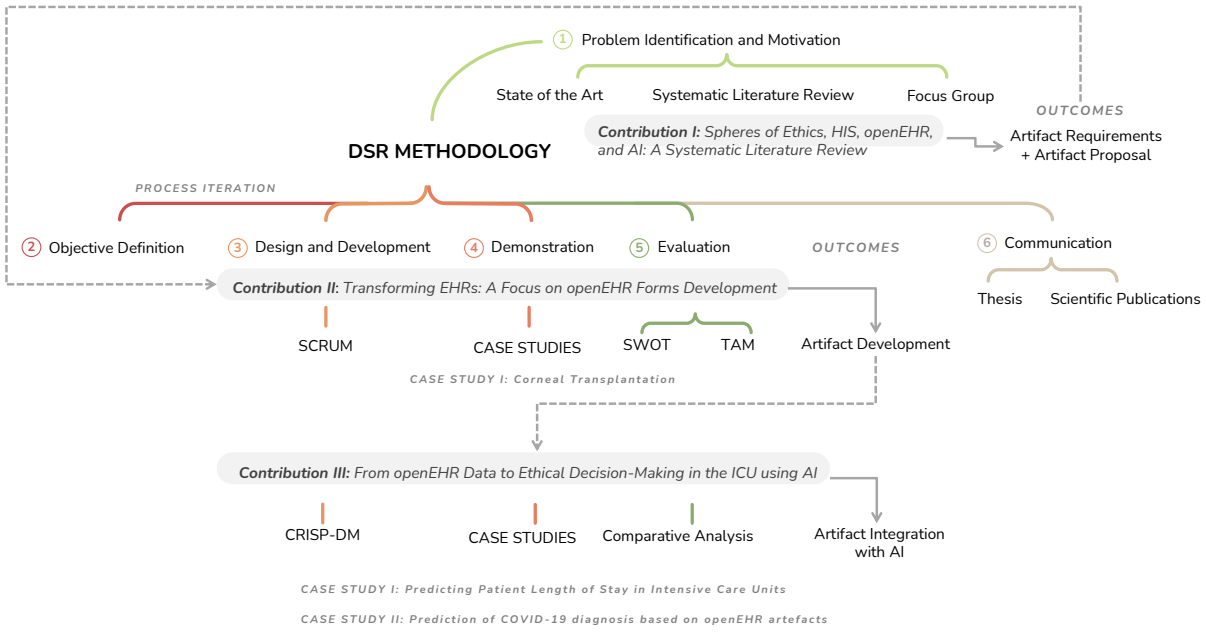


Figure 9: The DSR process conducted for the case studies of the thesis.

4

Research Contributions

"Curiosity has its own reason for existence. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality."

— Albert Einstein

In the pursuit of advancing knowledge and addressing critical challenges within the context of this PhD thesis, this chapter delves into the significant research contributions that have been made. The chapter is divided into three main sections, beggining with the key aspects and insights of a systematic literature review that explores the intersection of ethics, HIS, openEHR, and AI. From this review, the study identifies the requirements for an ethically compliant HIS and investigates the potential of openEHR specifications to achieve this goal. As a result, an artefact is proposed as a solution to the ethical concerns raised by HIS and AI in healthcare, promoting a synergetic environment between medical staff and IT professionals.

The second section focuses on the development aspects of a tool that automatically creates user-interface forms from openEHR templates, thereby facilitating the integration of clinical standards within EHR systems, fostering interoperability and data quality. This section also discusses the design, structure, and functionality of these forms, as well as their integration and application within simpler existing healthcare settings in the form of case study scenarios. These case studies were developed in collaboration with healthcare professionals at the ULSdSA with the aim of not only demonstrating the practical utility and benefits of the proposed artefact, but also to fortify its robustness and resilience, thereby enhancing its capacity to navigate the complexities inherent in intensive care contexts. For demonstration purposes, the implementation of the solution in the context of corneal transplant surgery will be presented, showcasing the potential of the tool to streamline data capture, improve documentation practices, and enhance the quality and safety of patient care.

As a consequence, these preliminary investigations lay the groundwork for a seamless transition towards the application of the tool in critical care scenarios, which are investigated in the third section. Furthermore, since the artefact is designed to be able to be integrated with AI algorithms, a initial exploration of the potential of using the data extracted from the openEHR forms as data source for the development of ML models is also presented, thereby enabling the automatic generation of decision support that can

assist healthcare professionals in making informed decisions when registering data regarding the care of critically ill patients in real-time.

Collectively, these sections contribute to a comprehensive and nuanced understanding of how ethical considerations and emerging trends in AI can be combined within ICUs, shedding light on pertinent issues, advancing theoretical frameworks, and offering actionable insights and recommendations for healthcare practitioners, policymakers, and interested parties. This chapter elucidates the multifaceted dimensions of the research process carried out, paving the way for deeper understanding and meaningful impact in the responsible and ethical use of AI in critical healthcare contexts, both within the scholarly community and beyond.

4.1 Spheres of Ethics, Health Information Systems, openEHR, and Artificial Intelligence: A Systematic Literature Review

4.1.1 Introduction and Problem Contextualization

The reactive nature of medicine has been a long-standing issue in healthcare, with diagnoses and treatments often based on the interpretation of miscommunicated symptoms and incomplete clinical histories. However, health institutions are living the momentum of the information era, having witnessed a significant growth of digitization of patient records and the widespread implementation of computerised systems worldwide over the last years, revolutionising the way medical data is managed and stored (Atasoy et al., 2019). HIS are computerised systems and processes that are routinely used by medical personnel to manage and store healthcare data, including EHR, which contain patient medical history and treatment information, as well as systems for managing appointments, billing, and other administrative tasks (Atasoy et al., 2019; Wager et al., 2021).

Always evolving, IT has proved to be one of the main contributors to the advancement of medicine and its influence is already remarkably broad with benefits ranging from the improvement of quality, safety, and efficiency of healthcare to the recommendation of diagnosis and treatment options using AI techniques (Razzak et al., 2020).

AI is a computer science field that refers to the ability of computers and machines to perform tasks that mimic human intelligence, such as learning, problem-solving, and decision-making, usually at a speed and scale that transcend human capability (Maddox et al., 2019; P. Wang, 2019). With the ability to process and analyze large amounts of data, AI has the potential to revolutionise the healthcare industry, automating routine tasks and assisting healthcare professionals identifying patterns and insights that would otherwise be difficult for humans to discover (Davenport & Kalakota, 2019). Despite the impressive returns of AI in medicine, the opacity of algorithms and the deficient explicability of the predictions yield uncertainty and skepticism, particularly when these black box reasonings neglect ethical considerations.

As machines are given increased power and autonomy in the medical landscape, ethical considerations must be included in their development process. Although hospitals have ethics committees governed by the laws and decrees of each country to ensure that ethical standards are followed, these committees only deal with medical ethics, failing to include issues related to machine ethics. This is a frightening reality that can compromise healthcare, since HIS play a critical role in patient delivery of care, and AI algorithms are increasingly being investigated to assist in decision-making processes. However, the lack of ethical guidelines in the development of these systems can hinder their adoption in clinical practice, as well as raise concerns about the potential for bias, discrimination, and unintended consequences in patient care, putting human life in jeopardy.

Ethical issues surrounding HIS and AI in healthcare are vast and have not been given full attention. One of the main ethical concerns related to HIS is the protection of patient privacy and the security of the personal health information stored in EHR. Because these systems are used to manage and store sensitive personal health data, it is critical to ensure that these data are handled in a way that respects patient privacy and confidentiality by confirming that only authorised personnel have access to it and that it is only used for legitimate healthcare purposes. Moreover, it is essential to guarantee that the data stored in HIS is complete, accurate and up to date, as data accuracy and reliability have significant impacts on the quality and safety of healthcare. This includes ensuring that data is entered correctly and that errors are promptly corrected, which the current thesis advocates that it can be achieved using clinical standards. In addition, there are also concerns about the potential for AI algorithms and decision-making processes to perpetuate biases and discriminatory treatment against certain factions of patients. Furthermore, questions concerning healthcare access and informed consent have earned considerable attention as well. These problems are just a part of a myriad of ethical issues that have been perpetuated over time within health institutions.

Although the scientific community has raised awareness and drawn attention to this matter, it fails to cover the characteristics that make a HIS ethically compliant, as well as to clarify in which ways it can be achieved. Hence, this contribution uses the PRISMA method to conduct a systematic literature review in order to find the requirements for an HIS to be ethically compliant and discover if openEHR specifications can be used to achieve such goal.

Hence, the literature review also focuses on openEHR and the level of agreement between openEHR principles and the requirements of an ethically compliant HIS. OpenEHR is an open-source standard for the representation, management, storage, retrieval, and exchange of health data in EHR that helps to ensure that EHR systems are able to interoperate in a consistent and standardised way (Kalra et al., 2005).

Overall, the ethical concerns raised by HIS and AI in healthcare highlight the importance of carefully considering the design and use of these systems to ensure that they are used in a morally acceptable manner. It is critical to raise awareness about machine ethics as well as to investigate how machine ethics can be implemented in clinical settings in order to demystify the dangers of AI and, consequently, unleash its power for the revolution of medical practice, while ensuring the delivery of safe and quality healthcare. The findings of the systematic literature review can encourage and assist future development

and implementation.

After collecting the results of the literature review, the hypothesis of using clinical standards, specifically openEHR guidelines, as the nucleus for the creation of an ethically compliant IT hospital ecosystem is investigated. This ecosystem promotes a synergetic environment between medical staff and IT professionals in order to address some of the ethical concerns highlighted with the study.

4.1.2 Research Purposes

The main purpose of this investigation is to answer one of the primary RQs of the study, which is:

RQ1.1 - What is the level of agreement of openEHR standards with ethical principles and values in healthcare?

Hence, a set of objectives were elaborated to guide the research process.

- **Identify Data Quality Issues:** Investigate the data quality issues prevalent in traditional EHR systems that impact the ethical development of AI algorithms, including inaccuracies, inconsistencies, incompleteness, and bias.
- **Analyze Ethical Principles in Healthcare:** Identify and analyze the main ethical principles and values in healthcare, considering their implications for data management, patient care, and clinical decision-making.
- **Evaluate Ethical Alignment:** Assess the alignment of openEHR standards with ethical principles and values in healthcare, focusing on data integrity, patient privacy, transparency, and accountability.
- **Develop Conceptual Framework:** Propose a conceptual framework for a comprehensive hospital information technology ecosystem based on openEHR standards, addressing data layer shortcomings and promoting ethical decision-making.
- **Provide Implementation Directions:** Offer guidance on the implementation of openEHR standards in real-world healthcare settings, considering technical, organizational, and cultural factors to promote successful adoption and integration.

4.1.3 Design and Development

A systematic literature review was conducted in order to understand current gaps and opportunities in the development of ethically compliant HIS. At the same time, the present literature review also situates the readers within the context of the study. The 2020 PRISMA guidelines were used to ensure that the systematic review is comprehensive, transparent, and reproducible (Page, McKenzie, et al., 2021). The

PRISMA approach specifies the inclusion of a clear definition of the RQ, a description of the inclusion and exclusion criteria for studies, a search strategy for identifying relevant studies, and a detailed description of the characteristics of the included studies (Page, Moher, et al., 2021). PRISMA also includes guidelines for assessing the risk of bias in the included studies and for synthesizing the results of the review (Page, Moher, et al., 2021). Figure 10 depicts the different stages involved in the literature review process.

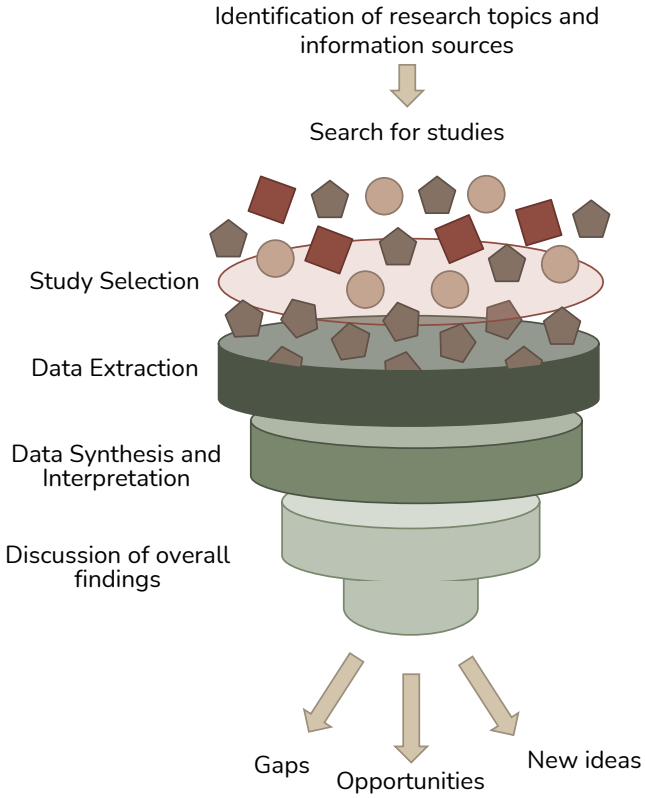


Figure 10: Methodological steps carried out within the systematic literature review process.

4.1.3.1 Topic Selection and Keywords

The first step was to select the research topics and the keywords that will be instrumental in limiting the scope of the literature review in order to narrow the problematic being addressed. As a result, HIS, ethical principles, AI, and openEHR were identified as primary research topics to learn the information available on those subjects and to answer some RQs (which will be described next), as well as to generate new ideas to solve specific problems encountered during the literature review.

The topics of this research affect health organizations and institutions all over the world and are a matter of increasing concern over the years, given the rapid digitization of medical records and the widespread adoption of EHRs. Ethics in HIS is not only a matter of safeguarding patient rights and well-being but also essential for upholding the integrity and trustworthiness of the healthcare system as a whole.

After selecting the topics, the authors then carefully explored and organised a set of synonyms that capture the essence of the research. The inclusion of synonyms in the search string broadens the search to capture variations in terminology, ensuring that it covers the various facets of the study. The complete list of key concepts related to the study is presented in Table 4.

Table 4: Key concepts and synonyms used in the search string for the systematic literature review

Research Topic	Keywords/Synonyms
Health Information Systems	Electronic Health Records, Hospital Information Systems, Health Data Management Systems, Clinical Information Systems, Health Informatics, Digital Health
Ethics	Ethical Principles, Ethical Guidelines, Medical Ethics, Bioethics, Ethical Considerations, Ethical Issues, Ethical Challenges
Artificial Intelligence	Machine Learning, Machine Intelligence, Computational Intelligence, Knowledge Engineering, Deep Learning
openEHR	openEHR Standards, openEHR Specifications, openEHR Guidelines, openEHR Models, openEHR Templates

The main objective of the proposed investigation is therefore to understand which characteristics must be fulfilled by HIS and AI systems to be ethically compliant in the healthcare industry and whether clinical standards, specifically the openEHR standard, support their construction. The main objective of this investigation is to answer one of the primary RQs of the study, which is:

RQ1.1 - What is the level of agreement of openEHR standards with ethical principles and values in healthcare?

Hence, a set of secondary RQs were elaborated to guide the research process. Table 5 describes the RQs that were elaborated, which compel the research purpose.

Table 5: RQs that guided the systematic literature review

#RQ	Description	Objectives
RQ1.1.1	What are the ethical principles and values that should be considered in the development of an ethically compliant HIS?	• Identify the ethical principles and values that should be considered in the development of an ethically compliant HIS; • Identify gaps in the development of ethically-compliant HIS.

(Table 5 continued on next page)

(Table 5 continued from previous page)

#RQ	Description	Objectives
RQ1.1.2	What is the level of agreement between openEHR principles and the requirements of an ethically compliant HIS?	• Prove openEHR's compliance with the principles of an ethical HIS; • Determine if there are any ethical requirements that cannot be achieved through openEHR and if so, propose ideas to accomplish those points.
RQ1.1.3	To what extent do health data quality issues, such as inconsistent data formats and unstandardised terminology, contribute to unethical behavior of AI-based healthcare solutions?	• Identify which clinical data issues underlie or exacerbate AI-based healthcare application problems; • Investigate the impact of using openEHR specifications on overcoming these problems.

4.1.3.2 Information Sources, Eligibility Criteria, and Search Strategy

In order to select the most appropriate data sources, six electronic databases were used, including PubMed, Science Direct, IEEE explore, Google Scholar, and ACM Digital Library.

Besides selecting the data sources, a search strategy was defined to ensure that the search was comprehensive and that all relevant studies were included. Hence, a set of eligibility criteria was defined for the literature review. The definition of eligibility criteria is an important step because it must strike a balance between being sufficiently broad to account for the diversity of studies and sufficiently narrow to ensure a meaningful result (Barrett et al., 2021). As usual, the eligibility criteria were divided into two categories:

- **Inclusion Criteria:** Criteria that must be met for a study to be included in the review.
- **Exclusion Criteria:** Criteria that, if met, will result in the exclusion of a study from the review.

In this work, it was considered that even if a study meets all of the inclusion criteria, it only needs to meet one of the exclusion criteria to be disregarded from the research. Factors such as geographic location and study setting were not considered as eligibility criteria. Table 6 lists the eligibility criteria outlined for this study.

Table 6: Eligibility criteria of the literature research conducted

Identifier	Category	Criteria Description
IC1	Inclusion Criteria	Studies published in article format.
IC2	Inclusion Criteria	Papers written in English language.
IC3	Inclusion Criteria	Open access or full-text published articles.
IC4	Inclusion Criteria	The years of research publication must be from 2018 to 2024.
IC5	Inclusion Criteria	Studies related to ethical-compliant HIS and/or openEHR.
IC6	Inclusion Criteria	Studies concerning all healthcare settings.
EC1	Exclusion Criteria	Webpages, presentations, posters, dissertations, essays, opinion papers, and news articles.
EC2	Exclusion Criteria	Unpublished studies, such as conference abstracts, working papers, and preprints.
EC3	Exclusion Criteria	Duplicated studies to avoid redundancy.

The concepts and keywords identified in the previous step were used to formulate search queries for each RQ, which in turn were used to find relevant articles for each segment of the scope of this study. The Listing 4.1 shows the search queries used in the literature review.

Listing 4.1: Search queries used in the systematic literature review

```
1  RQ1.1.1: ("Health Information Systems" OR "Electronic Health Records" OR "Hospital
       ↪ Information Systems" OR "Health Data Management Systems" OR "Clinical
       ↪ Information Systems" OR "Health Informatics" OR "Digital Health") AND (Ethic
       ↪ * OR Moral* OR "Bioethics")
2  RQ1.1.2: "openEHR" AND (Ethic* OR Moral* OR "Bioethics")
3  RQ1.1.3: (artificial* OR auto* OR intelligen* OR machine* OR comput* OR "Knowledge
       ↪ Engineering" OR "Deep Learning") AND (("health data quality" OR "data
       ↪ quality issues" OR "data inconsistency" OR "inconsistent data formats" OR "
       ↪ data standardization" OR "unstandardised terminology" OR "bias") OR (Ethic*
       ↪ OR Moral* OR "Bioethics"))
```

The corpus of papers was collected between September 1st, 2022 and February 29th, 2024, using the aforementioned electronic databases and search queries. The search strategy was carried out considering a descending order of publication date, i.e, from the most recent publications to the oldest ones, and with a publication date limit ranging from 2018/01/01 to 2024/02/29.

Once collected, the papers were filtered according to the initial eligibility criteria to eliminate those that were not directly related to the scope of the study.

The title, abstract, and keywords reading was the primary strategy for analyzing each scientific publication. When these parameters were insufficient, a complete reading of each article was made for further examination of potentially eligible studies.

4.1.3.3 Data Extraction and Synthesis

The screening process was carried out by two authors independently, by reading the titles, abstracts, and keywords of the publications of the search results. When the criteria were met, the full text was analysed. In cases of disagreement, a third reviewer was consulted to reach a consensus.

Data extraction was performed through the textual analysis of the eligible articles and focused on the following aspects:

- Ethical principles and values in healthcare;
- Data quality issues that hinder AI-based healthcare solutions;
- Ethical dimensions of openEHR;

These aspects were considered as the most relevant for the purpose of this study. In addition, the following information was extracted: clinical setting, study design, recipient of intervention, and purpose of care.

4.1.3.4 Quality Assessment

All the reviewed studies portray at least two of the topics of interest of this research, which are HIS, ethics, openEHR, and AI. Papers focusing isolatedly on one of these topics were not considered for the review. The key points for conducting this systematic review were to identify the gaps and opportunities in the development of ethically compliant HIS, to understand the requirements for an HIS to be ethically compliant, and to investigate the potential of openEHR specifications to achieve this goal. Furthermore, it was also intended to identify the level of agreement between openEHR principles and the requirements of an ethically compliant HIS. The investigation of how openEHR specifications can set the base for mitigating some of the ethical problems founded at the AI level in the literature was also a key point.

The reporting quality was made following the PRISMA checklist. Publications considered eligible for the analysis were assessed for the methodological quality, meeting the inclusion criteria, risk of bias, assumptions, and simplifications, and clarified evidence-based results. The reviewers did not apply any methodological assessment tool.

After defining the six data sources, the reviewers selected a set of publications within the scope of the study, through the search strategy and the inclusion criteria previously defined. The outcomes presented through tables and narrative summary characterise the different aspects of the research presented in the selected studies, in order to trace global patterns and perspectives.

4.1.3.5 Study selection

The search strategy resulted in a total of 105 articles, which were then screened for eligibility based on the inclusion and exclusion criteria. After the screening process, 18 articles were included in the final review.

The low number of articles included in the review reveals the specificity of the research topics and the limited availability of studies that address the intersection of ethics, HIS, openEHR, and AI in healthcare.

The selection process is illustrated in Figure 11.

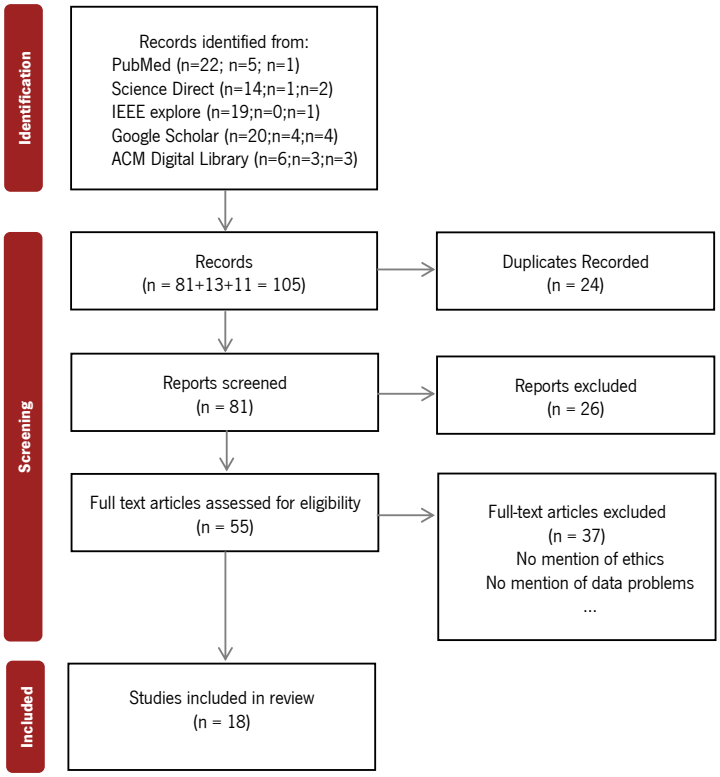


Figure 11: PRISMA flow diagram illustrating the study selection process.

4.1.3.6 Results Synthesis

Requirements of an Ethically-Compliant HIS

Among the pivotal components of Health 4.0 are EHR and HIS, which serve as the backbone for digitized healthcare delivery. However, alongside the transformative potential of these systems, ethical considerations loom large, demanding meticulous attention. The research conducted reveals that relevant codes of ethics are satisfactory on traditional clinical environments, but do not address the ethical challenges of current and potential digital environments. Most studies around ethics pertain to AI systems, leaving less attention to the ethical implications of HIS in healthcare. Despite the scarce body of knowledge, there are some studies dedicated to the ethical dilemmas in HIS and EHR. These studies highlight the importance of patient privacy, data security, informed consent, data ownership, and the responsible use of health data for research and decision-making. The findings suggest that ethical considerations are paramount

in the development and deployment of HIS to ensure patient safety, data integrity, and accountability in healthcare delivery. The lack of ethical guidelines in the development of HIS can lead to unintended consequences, such as data breaches, privacy violations, and discriminatory practices, which can undermine the trust and credibility of the healthcare system. Therefore, it is essential to establish ethical standards and guidelines for the design, implementation, and use of HIS to ensure that they are used in a responsible and ethical manner.

Table 7 summarizes the main ethical principles and values that should be considered in the development of an ethically compliant HIS and the studies that have addressed these issues.

Table 7: Main ethical principles and values in EHR and HIS

Ethical Principles	Studies
Privacy and Confidentiality	(Bani Issa et al., 2020; Brall et al., 2019; Feeney et al., 2020; Milosevic, 2019)
Integrity and Reliability	(Feeney et al., 2020; Milosevic, 2019; Piasecki et al., 2021)
Security	(Bani Issa et al., 2020; Brall et al., 2019; Vayena et al., 2018)
Transparency and Accountability	(Brall et al., 2019; Milosevic, 2019; Vayena et al., 2018)
Trust	(Brall et al., 2019; Feeney et al., 2020; Piasecki et al., 2021; Vayena et al., 2018)
Equity and Accessibility	(Brall et al., 2019; Piasecki et al., 2021)
Autonomy and Informed Consent	(Brall et al., 2019; Piasecki et al., 2021)
FAIR principles	(Meredith et al., 2023; Parciak et al., 2023; Plug et al., 2022)
GDPR	(Agyei & Oinas-Kukkonen, 2020; Carvalho et al., 2021; Feeney et al., 2020; Ruotsalainen & Blobel, 2019)

Ethical Dimensions of openEHR

There are very few studies in the literature on the use of clinical standards as a foundation for the development of ethical HIS. Focusing specifically on the openEHR specifications, only four studies were found

on the evaluation of this standard in relation to certain ethical principles.

In 2019, Bönisch et al. examined openEHR archetypes and templates using the FAIR Principles (Bönisch et al., 2019). The study concluded that all fifteen FAIR principles were met, with only two of them being partially fulfilled, both of which were related to data accessibility. The authors mention that by using a clear ID management and meaningful keywords, the required archetypes and templates can easily be found. Hence, the openEHR approach and the CKM as a collaborative library are compatible with the FAIR principles and are well suited for the exchange of data.

Frexia et al. conducted a similar study a few years later, in 2021, by theoretically assessing the compliance of openEHR specifications with each of the fifteen FAIR principles (Frexia et al., 2021). The research was carried out by establishing a direct mapping between the openEHR foundational principles and the FAIR dimensions. According to the results of this study, openEHR is inherently compatible with all FAIR principles and is a promising implementation strategy for developing FAIR-compliant clinical data repository.

A study carried out in 2018, focused on the identification of requirements for an HIS to be GDPR compliant and investigated the level of agreement between the principles of openEHR and the requirements of an HIS compliant with the GDPR (Sousa et al., 2018). The authors identified a total of 16 requirements from the regulation and 8 openEHR design principles and defined that one specification could match more than one requirement and that a requirement could be matched by more than one specification. In the end, all of the GDPR requirements identified for HIS were matched with the openEHR specifications. The study concludes by proposing the implementation of openEHR based systems as a solution for the development of HIS compliant with the GDPR requirements. A year later, the authors have published an updated research on this topic (Gonçalves-Ferreira et al., 2019). In their recent study, the authors found a total of 50 GDPR requirements and 8 openEHR design principles. The openEHR principles conformed to 30% (15/50) of GDPR requirements. All the openEHR principles were aligned with GDPR requirements. The study concludes by mentioning that by using an openEHR-based EHR, the institutions are closer to becoming compliant with GDPR while safeguarding the medical data.

Accordingly, the findings of the literature review suggest that openEHR specifications are compliant with FAIR principles and GDPR, which are essential for ensuring the ethical use of health data in healthcare. Table 8 summarizes the mapping of openEHR features to ethical principles.

Table 8: Mapping of openEHR Features to Ethical Principles

Ethical Principles	openEHR Features
Privacy and Confidentiality	Access control, Audit trailing, Separation of EHR and DIM
Integrity and Reliability	Version control, Digital Signature, UID, Two-Level Modelling, Clinical and technical review in CKM, Compatibility with clinical terminologies

(Table 8 continued on next page)

(Table 8 continued from previous page)

Ethical Principles	openEHR Features
Security	Access control, Audit trailing, UID, Separation of EHR and DIM
Transparency and Accountability	Audit trailing, Digital Signature
Trust	Version control, Digital Signature, Two-Level Modelling, Separation of responsibilities
Equity and Accessibility	Separation of EHR and DIM, notion of shared EHR and life-long EHR
Autonomy and Informed Consent	Self report data, Consent by data subject, patient-centred care

Data Quality Issues that Hinder AI-Based Healthcare Solutions

The literature review reveals that data quality issues within healthcare data significantly hinder the efficacy of AI-based healthcare solutions. The primary challenges identified include inconsistent data formats, unstandardized terminology, and data silos. These issues manifest in various forms, including inaccuracies, inconsistencies, incompleteness, and bias within health datasets. Such deficiencies pose significant obstacles to the performance and reliability of AI algorithms (Goodman, 2020; Müller, 2022).

While existing literature highlights these data quality issues as critical impediments to the effectiveness of AI in healthcare, notably, there is a gap in concrete strategies or solutions to address these challenges directly at the data layer. Hence, the findings of the literature review underscore the importance of addressing data quality issues at the source to enhance the ethical compliance, accuracy, and reliability of AI-based healthcare solutions.

4.1.4 Results and Discussion

The results of the systematic literature review are presented in this section, providing an overview of the current state of research on the intersection of ethics, HIS, openEHR, and AI in healthcare.

In what regards ethics and HIS, researchers have extensively explored ethical issues surrounding the use of HIS in healthcare settings. This includes considerations such as patient privacy and confidentiality, data security, informed consent, data ownership, and the responsible use of health data for research and decision-making. Studies have also focused on ethical dilemmas arising from the adoption of emerging technologies within HIS, such as cloud computing, big data analytics, and telemedicine.

openEHR has gained traction as a standardised approach to structuring and managing EHRs in healthcare systems. However, research examining the ethical implications of openEHR implementation, including its impact on data integrity, interoperability, patient-centered care, and clinician-patient relationships is underrepresented in the literature. Although there are three studies that prove that openEHR specifications are compliant with FAIR principles and GDPR, there is still no work that places openEHR at the base

of HIS as an incremental step to achieve ethical decision-making. Nonethless, the findings of the systematic review suggest that openEHR specifications align with ethical principles and values in healthcare by promoting data standardization, semantic interoperability, and patient empowerment. By adhering to openEHR guidelines, HIS can enhance transparency, accountability, and data quality while addressing concerns related to data silos, vendor lock-in, and proprietary formats.

Literature shows that the ethical use of AI in healthcare has emerged as a pressing concern. Research in this area has focused on ethical principles such as fairness, transparency, accountability, privacy, and bias mitigation in AI applications. Studies have explored ethical frameworks and guidelines for the development, deployment, and regulation of AI systems in healthcare, aiming to balance innovation with ethical considerations and societal values.

Overall, while research at the intersection of ethics, HIS, openEHR, and AI in healthcare is still evolving, there is growing recognition of the need for interdisciplinary collaboration, stakeholder engagement, and ethical oversight to ensure the responsible development and deployment of technology-driven solutions that prioritise patient welfare and ethical integrity. Hence, this study contributes to the ongoing dialogue on the ethical implications of HIS and AI in healthcare, highlighting that the shortcomings at the data quality level can hinder the ethical compliance of AI-based healthcare solutions and leveraging openEHR standards to promote ethical compliance, data quality, and patient-centered care in clinical settings.

As a result, the gaps highlighted set the breach for further research on approaches able to intertwine clinical information systems, openEHR and ethical decision-making.

Against this backdrop, an artefact is proposed with the aim of bridging the gap between ethical considerations and technological innovation in healthcare, offering a practical solution for integrating openEHR standards within HIS to promote ethical compliance, data quality, and patient-centered care. By leveraging the power of openEHR specifications, the artefact seeks to empower healthcare professionals to capture, store, and exchange structured health data in a standardised format that is interoperable, transparent, and ethically sound. Furthermore, the ability of this solution to integrate with AI algorithms holds promise for enhancing clinical decision support, improving patient outcomes, and advancing the quality and safety of healthcare delivery in the future.

The architecture proposed aims to create a unified and efficient hospital information technology ecosystem that can improve patient care, streamline clinical and administrative processes, and enhance communication between healthcare providers. This ecossystem will consist of various software applications, such as EHR, appointment scheduling, laboratory information systems, and telemedicine platforms. These applications will be integrated with a centralised database that for security reasons will be divided into two parts - one for storing demographic and sensitive patient data and another for storing non-sensitive data. The proposed architecture will also include clinical decision support that can provide real-time guidance and recommendations to healthcare professionals based on structured health data extracted from openEHR forms. This system will be integrated with the EHR system to ensure that the recommendations are aligned with the patient's medical history, preferences, and treatment plan. The EHR system will be the backbone of the system, providing a comprehensive view of patient information, including medical

history, test results, and treatment plans.

Figure 12 provides a high-level representation of the proposed framework architecture, including its major components and inherent interactions. The architecture is designed to be modular, scalable, and interoperable, allowing for the seamless integration of new applications and technologies as needed. The proposed framework aims to address the ethical concerns raised by HIS and AI in healthcare by promoting transparency, accountability, and patient-centered care. By leveraging openEHR standards, the framework seeks to ensure data quality, interoperability, and semantic consistency across healthcare settings, fostering a culture of ethical awareness and adherence to professional codes of conduct.

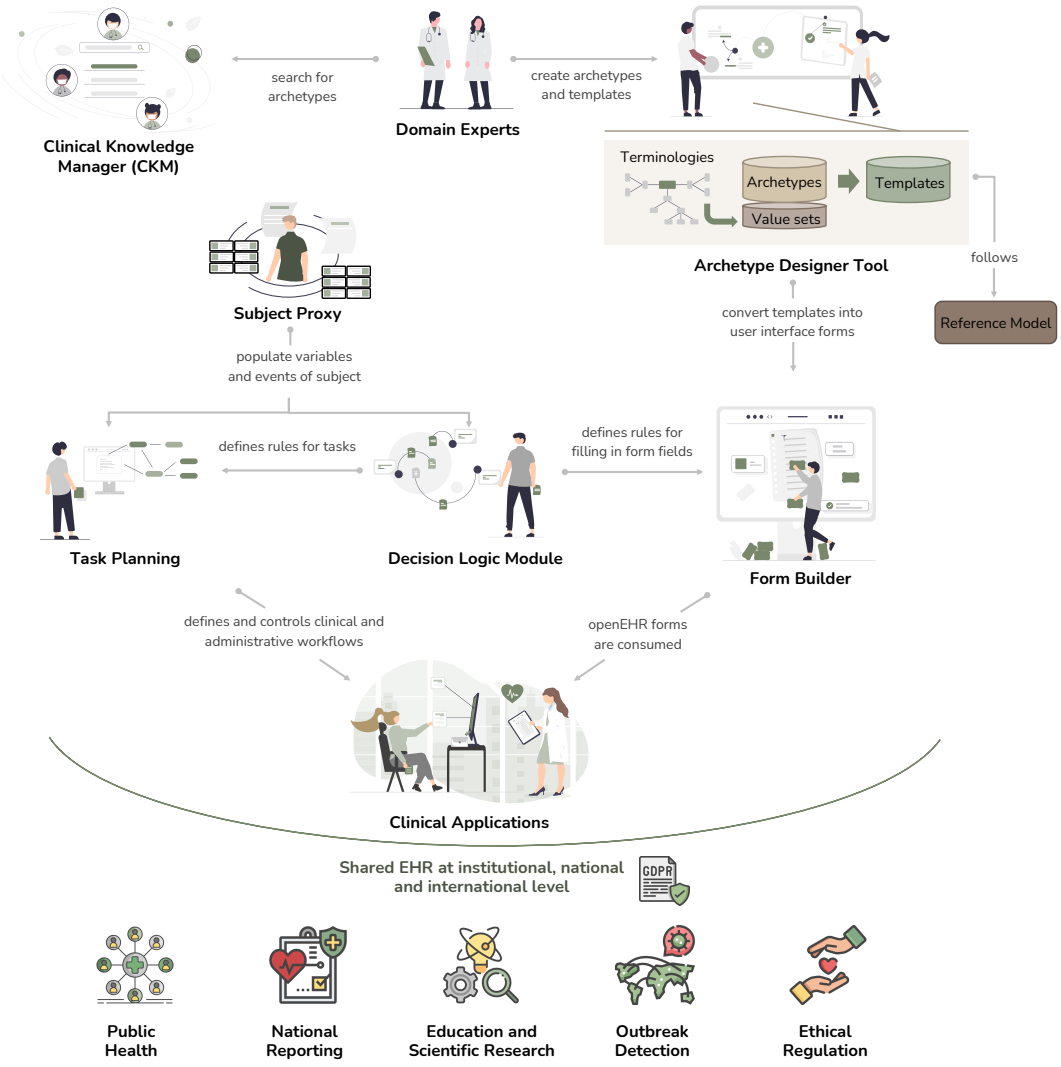


Figure 12: Proposed framework architecture for integrating openEHR standards within HIS to promote ethical compliance and data quality in healthcare.

This architecture proposal sets the stage for the development of the artefact, which focuses on a fundamental element of the framework, namely the FormBuilder tool. The FormBuilder tool is designed to automatically generate UI forms from openEHR templates, enabling healthcare professionals to capture

structured health data in a standardised format that is interoperable, transparent, and ethically sound. The next section will delve into the particularities of the design, development, and implementation of the FormBuilder tool, highlighting its key features, functionalities, and potential applications in healthcare settings.

4.2 Transforming Eletronic Health Records: A Focus on openEHR Forms Development

This section provides a detailed description of all the steps carried out in the development, implementation, and evaluation process of the proposed artefact with the aim of facilitating the integration of clinical standards within EHR systems, fostering interoperability and data quality. The section is divided into several subsections, each of which addresses a specific aspect of the design, structure, and functionality of the forms, as well as their integration into existing healthcare systems at the ULSdSA.

4.2.1 Introduction and Problem Contextualization

The lack of interoperability and standardization in EHR systems across the entire EU poses significant challenges to healthcare delivery, data sharing, and patient care coordination.

The results of the "MonitorEHR" study, which aimed to evaluate the state of interoperability capabilities of EHR systems across the EU in 2020, revealed a heterogeneous picture of EHR development among member states (Thiel et al., 2021). Although the majority of countries have implemented EHR systems, cross-border interoperability remains an idea. Out of the 29 countries surveyed, only 2 reported having nearly all electronic data in a structured format. According to the study, four out of five countries have a level of structured data below 75% and two out of five countries falling below 25% (Thiel et al., 2021).

The lack of standardization and the absence of a common structured format for EHR systems are the main obstacles to achieve interoperability. This fragmentation hampers the seamless exchange of patient data between healthcare providers, resulting in inefficiencies, medical errors, and compromised patient safety. Furthermore, the absence of standardised data models and terminologies complicates data aggregation and analysis for research and public health purposes.

The infectious disease caused by SARS-CoV2, designated as COVID-19 by the World Health Organization (WHO) in February 2020, has highlighted with lethal force the shortcomings of existing EHR systems. The intricate demands of the pandemic have led to an unprecedented strain on healthcare systems around the world, which struggled to rapidly adapt to the evolving needs of the crisis as well as to cope with the escalating number of patients arriving at hospitals on a daily basis. The availability of high-quality and structured health data was recognised as a decisive factor in the fight against the pandemic, enabling collaboration and seamless data exchange between healthcare providers, researchers, and public health authorities to ultimately track the spread of the virus, identify high-risk populations, and evaluate the effectiveness of interventions.

These challenges were exacerbated in the intensive care setting, where critically ill patients required specific resources and specialised care. The amount of resources available and the capacity of ICUs to accommodate the influx of patients affected by COVID-19 were limited in comparison to the number of patients in need, leading to a situation where healthcare professionals were forced to make difficult decisions regarding the allocation of resources and the prioritization of care. The lack of standardised

data models and the absence of interoperability between EHR systems further complicated the situation, making it difficult for clinicians to access comprehensive patient records and discuss critical information with other healthcare providers, which could have a significant impact on the management and outcome of patients in need for critical care services. Under these extraordinary circumstances, a myriad of ethical dilemmas surfaced, which were often not given the attention they deserved considering the burden of the crisis, as the primary focus was on saving lives and containing the spread of the virus.

The dynamic nature and devastating effects of the new coronavirus pandemic have brought to light the urgency of a robust and interoperable healthcare ecosystem capable of rapidly adapting to the demands and intricacies of a public health crisis, while ensuring ethical considerations are safeguarded. However, the lack of interoperability and the absence of standardised data models in EHR systems across the globe have hindered the timely sharing of critical information between hospitals, laboratories, and public health authorities, impeding surveillance efforts and response coordination.

By ensuring that critical patient information is accurately documented, accessible, and shared across healthcare settings, standardised EHRs empower clinicians to make informed, ethically sound decisions that prioritise patient well-being and respect individual preferences. Moreover, standardised EHRs contribute to the cultivation of a culture of ethical awareness and adherence to professional codes of conduct, thereby safeguarding the rights and dignity of patients while enhancing the overall quality and integrity of healthcare services.

In light of the critical issues discussed above, the present study aims to address the lack of interoperability and standardization in EHR systems by proposing a novel artefact that automatically generates user-interface forms from openEHR templates. By leveraging the power of openEHR specifications, the proposed artefact aims to facilitate the integration of clinical standards within EHR systems, fostering structured and semantically interoperable high-quality data, which can be readily understood by technically interoperable EHR systems.

4.2.2 Research Purposes

In light of the challenges highlighted in the previous section (Section 4.2.1), the primary goal of this study is to build a tool that automatically generates user-interface forms from openEHR templates, allowing healthcare professionals to efficiently capture, store, and exchange structured health data in a standardised format. The research purposes that guided the development of the artefact are as follows:

- **Data Standardization and Consistency:** Research the extent to which the tool facilitates data standardization and consistency across different healthcare settings and specialties by enforcing adherence to openEHR templates and terminology standards.
- **Impact on Documentation Practices:** Study the impact of using the tool on healthcare documentation practices, including documentation completeness, accuracy, and timeliness, as well as the perceived burden on healthcare professionals.

- **Persistent and Lifelong EHRs:** Examine the potential of the artefact to support the creation of persistent and lifelong electronic records that can be easily updated, shared, and accessed by patients and healthcare providers throughout the patient's entire care journey.
- **Interoperability and Data Exchange:** Investigate the role of the artefact in interoperability and data exchange between healthcare providers, systems, and institutions by ensuring that structured health data can be seamlessly shared and understood across different platforms.
- **Clinical Decision Support Integration:** Explore the potential of integrating the tool with clinical decision support systems to enhance the quality and safety of patient care by providing real-time guidance and recommendations based on structured health data.
- **Clinical Workflow Integration:** Assess the integration of the solution into clinical workflows and processes to streamline data capture, reduce duplication of efforts, and improve the efficiency and effectiveness of healthcare delivery.
- **Patient Engagement and Empowerment:** Evaluate the impact of the artefact on patient engagement and empowerment by enabling patients to access and contribute to their health records, fostering shared decision-making and collaborative care.

Accordingly, a set of objectives was defined to guide the design and development of the artefact, ensuring that it meets the needs and expectations expressed by the healthcare professionals that participated in the definition of the user stories created for the project. These objectives are detailed in Table 9.

Table 9: Main objectives guiding the design and development of the artefact

ID	Objective	Description
	User Experience	
UE1	Intuitive Design	Provide a user-friendly interface with an intuitive layout that allows users to easily navigate, input data, and access relevant information.
UE2	Customization	Allow the interface to adapt to specific user requirements such as colour, font size, field order, visibility, editability, customization of field size (vertically and horizontally), definition of input types to better suit their preferences, etc.
UE3	Automation	Minimise user input and streamline processes to reduce cognitive load and improve efficiency.

(Table 9 continued on next page)

(Table 9 continued from previous page)

ID	Objective	Description
UE4	Performance	Optimise form rendering and submission processes to minimise loading times and improve page responsiveness, ensuring a smooth and efficient user experience.
UE5	Responsive Design	Ensure that the interface adapts smoothly to different screen sizes and devices, providing a consistent user experience across desktops, tablets, and smartphones.
UE6	Validation Feedback	Provide real-time informative validation error messages to guide users in identifying and resolving validation errors, reducing frustration and improving form completion rates.
UE7	Guidance and Assistance	Provide informative tooltips and help features to guide users through the form completion process.
	Performance	
P1	Load Time	Optimise form loading times to ensure a seamless user experience and minimise waiting times.
P2	Lazy Loading	Implement lazy loading to defer the loading of non-essential components or resources until they are needed.
P3	Bundle Size Reduction	Minimise the size of bundled JavaScript and CSS files by removing unused code, optimizing dependencies, and employing compression techniques to reduce load times and improve performance.
	Modularity	
M1	Component Reusability	Design components with a high degree of reusability across different parts of the application, reducing redundancy and promoting code efficiency.
M2	Dependency Injection	Implement dependency injection patterns to allow components to receive dependencies or configuration options from their parent components.
M3	Version Control	Implement a version control system such as Git to track changes and manage code revisions.

(Table 9 continued on next page)

(Table 9 continued from previous page)

ID	Objective	Description
M4	Organization and Separation of Concerns	Organise component files and directories into separate modules based on their distinct nature in a logical and consistent manner, following established conventions and best practices to enhance code readability and maintainability.
	Scalability	
S1	Load Balancing	Implement load balancing strategies to distribute incoming traffic evenly across multiple servers or instances, ensuring optimal performance and availability.
S2	Microservices Architecture	Adopt a microservices architecture to break down complex functionalities into smaller, loosely coupled services that can be developed, deployed, and scaled independently.
S3	State Management	Implement efficient state management solutions such as Redux to handle complex data flows and application states, ensuring consistency and reliability across components.
S4	Distribution	Design the application as an npm package to facilitate easy distribution and consumption by other applications.
	Flexibility	
F1	Extension Capabilities	Design a flexible architecture that allows for easy customization and extension of features to meet evolving user needs.
F2	Configuration Options	Provide users with configurable options to adjust settings, preferences, and functionalities according to their requirements.
	Compatibility	

(Table 9 continued on next page)

(Table 9 continued from previous page)

ID	Objective	Description
C1	Cross-Browser Support	Ensure compatibility with major web browsers such as Chrome, Firefox, Safari, and Edge, providing consistent appearence and behaviour across different platforms.
C2	Operating System Compatibility	Ensure compatibility with different operating systems such as Windows, macOS, and Linux, allowing users to access and use the app regardless of their preferred platform.
C3	Internationalization and Localization	Support internationalization and localization features to make the app compatible with different languages, cultures, and regional preferences, enhancing its usability for a diverse global audience.
C4	Third-Party Integration Compatibility	Ensure compatibility with third-party APIs, libraries, and services commonly used in web development, enabling seamless integration with external tools and platforms.
	Security	
Se1	Authentication	Implement robust authentication mechanisms, such as OAuth, JWT (JSON Web Tokens), or multi-factor authentication, to verify user identities and prevent unauthorised access to restricted areas or functionalities.
Se2	Authorization	Enforce strict authorization policies to control user access rights and permissions based on roles, groups, or specific criteria, reducing the risk of unauthorised actions or data breaches.
Se3	Secure Communication	Use secure communication protocols such as HTTPS (Hypertext Transfer Protocol Secure) and SSL/TLS certificates to encrypt data transmitted between the client and server, protecting sensitive information in transit from eavesdropping or tampering.

(Table 9 continued on next page)

(Table 9 continued from previous page)

ID	Objective	Description
Se4	Data Privacy Compliance	Ensure compliance with data privacy regulations such as GDPR (General Data Protection Regulation) by implementing data protection measures, obtaining user consent for data processing, and providing transparency on data collection, storage, and usage practices.

4.2.3 Design and Development

The present section describes the design and development process of the artefact, focusing on the key aspects of the tool's architecture, functionality, and user interface. The development of the artefact was guided by the principles of user-centered design, which emphasises the importance of understanding user needs, preferences, and workflows in order to create effective and usable solutions. The artefact was developed in close collaboration with healthcare professionals at the ULSdSA, who provided valuable insights and feedback throughout the design and development process.

4.2.3.1 Research Methods

The development of the artefact was guided by the principles of user-centered design, which emphasises the importance of understanding user needs, preferences, and workflows in order to create effective and usable solutions. The artefact was developed in close collaboration with healthcare professionals at the ULSdSA, who provided valuable insights and feedback throughout the design and development process. Due to the ever-changing nature of the healthcare domain and the complexity of the problems faced by healthcare professionals coupled with the community's high demand for rapid and flexible solutions, an agile approach was adopted to guide the entire development process of the artefact. The Scrum framework was used to enable iterative and incremental development, allowing continuous refinement and improvement based on clinicians' feedback to ensure that the artefact met the needs and expectations of end-users (Garcia et al., 2022).

The Scrum framework is an agile project management methodology that has been widely adopted in the software development given its iterative and incremental nature, which allows for the rapid ability to respond to changing requirements (Bolloju et al., 2017).

In terms of team functions, there are three main roles in the Scrum framework, which are the Product Owner, the Scrum Master, and the Development Team. The Product Owner role was assumed by two healthcare professionals, appointed by the hospital administration board, who possessed a deep understanding of the hospital's needs and priorities. The Scrum Master role was assumed by the responsible of

the research laboratory in which the Development Team was inserted. The Development Team consisted of a multidisciplinary group of five researchers from the ALGORITMI Centre, which I am part of, who were responsible for the design, development, and testing of each artefact in the openEHR hospitalar ecossystem. The development team was self-organised and cross-functional, with each member dedicated to a specific area of openEHR. Each individual artefact was developed in parallel, with regular meetings held to discuss progress, share insights, and address any challenges. The artefact concerning the automatic generation of user-interface forms from openEHR templates was my responsibility, and it will be the object of study in this section. In addition, the IT department of the hospital was also involved in the development process, providing technical support and guidance on the integration of the artefact into the existing hospitalar technological environment. Table 10 briefly describes each role and respective responsibilities.

Table 10: Scrum roles and respective descriptions. Adapted from (Garcia et al., 2022)

Role	Description
Product Owner	Responsible for defining the features and user requirements as well as their prioritization. The Product Owner is the bridge between the stakeholders and the development team, deciding the release date and content.
Scrum Master	Responsible for ensuring that the development team is functional and productive as well as for solving eventual problems that may disrupt their work, enabling the team to self-organise and make changes quickly.
Development Team	Responsible for delivering potentially functional product increments at the end of each sprint, transforming the Product Owner's requirements into a potentially releasable part of the product at the end of each iteration.

The Scrum framework has a routine characterised by the division of the project into small iterations known as sprints, in which an incremental part of the product is delivered at the end of each sprint. Each sprint is tipically described as a time-boxed horizon of one to four weeks, during which a *Daily Scrum* of 15 minutes is held to discuss the progress, potential obstacles, and the plan for the day (Garcia et al., 2022). In addition, before the sprint can start, it must be carefully planned with the collaboration of the entire team in a sprint planning meeting, where the team decides and understands exactly what tasks will be completed during the sprint. The general direction of a sprint is defined by the *Product Backlog*, which is an ordered list managed by the *Product Owner* of all the *User Stories* to be addressed in the project from the most prioritised to the least prioritised. The *Product Backlog* is a living document that is constantly being updated and refined as the project progresses and new requirements emerge. A *User Story* is a simple and concise high-level definition of a requirement containing appropriate information for the *Development Team* to provide an estimate of the effort required to implement it (Bolloju et al., 2017). Therefore, the *Product Backlog* lists all the features, functions, enhancements, and modifications that can be implemented in the product in future releases, from the perspective of the individual who desires the

new capability (Bolloju et al., 2017). The *Sprint Backlog* is a subset of the *Product Backlog* that contains the tasks that the *Development Team* has committed to complete during the sprint (Garcia et al., 2022). At the end of the sprint, a *Sprint Review* meeting is held to demonstrate the work done. Afterwards, a *Sprint Retrospective* meeting is done to discuss success, areas for improvement, and strategies to make the next sprint more effective (Garcia et al., 2022). This is a cyclical process, in which after the sprint retrospective, a new sprint planning meeting is held, a new sprint begins, and the process is repeated until the project is completed. Figure 13 illustrates the overall functioning of the Scrum framework carried out during this investigation.



Figure 13: Scrum framework followed during the artefact development process. Adapted from (Schwaber & Sutherland, 2011).

By involving end-users in the design and evaluation of the artefact, this collaborative approach aimed to ensure that it meets the needs and expectations of healthcare professionals, thereby enhancing its usability, acceptance, and impact in real-world clinical settings.

4.2.3.2 Technologies and Tools

This section describes the technologies and tools used in the development of the proposed artefact, including the programming languages, frameworks, libraries, and software used to build the user-interface forms from openEHR templates. The choice of technologies was based on several criteria, including their compatibility with the openEHR specifications, their ease of use, their flexibility, and their ability to meet the requirements defined in the early stages of the project.

Database

A relational database is a structured set of data that is stored and organised in a way that allows for easy access, retrieval, and management. In 1970, Dr. Edgar F. Codd introduced the concept of the relational database model, which is composed of bi-dimensional tables (consisting of rows and columns) that are linked through relationships that define how the data is related and connected (Beaulieu, 2009; Greenwald et al., 2013). The data of the tables are related through primary and foreign keys, which establish the relationships between the data in different tables. A primary key is a unique identifier that unequivocally identifies each record in a table maintaining data integrity and consistency. In turn, a foreign key is a field in a table that links to the primary key of another table, establishing a relationship between the two tables and maintaining referential integrity (Beaulieu, 2009; Greenwald et al., 2013). A database is projected using a process known as modelling, which involves the creation of a conceptual representation of the data structure through the identification of entities, the definition of the relationships between those entities, and their characterisation through attributes. The database model also includes the logical design, which specifies the structure of the tables, assigning primary and foreign keys, domains, and constraints that govern the data integrity and consistency (Ramakrishnan & Gehrke, 2000). Finally, the physical design involves the implementation of the logical model in a specific Database Management System (DBMS) using SQL, defining the storage structures, indexing, and optimization strategies to ensure efficient data retrieval and manipulation (Ramakrishnan & Gehrke, 2000). SQL is a standardised programming language originally based on relational algebra and relational calculus, that is used globally to interact with relational databases, enabling users to insert, retrieve, update, and delete data from tables, as well as create and manage database schemas, constraints, and security settings (Coronel & Morris, 2016). All the processing logic is secured by the DBMS, which is responsible for managing the database, ensuring data integrity, enforcing security policies, and optimizing query performance (Connolly & Begg, 2005).

Databases are essential components of modern software applications, enabling the storage and manipulation of data in a secure and efficient manner. Hence, it was necessary to define the database model and schema for the artefact, ensuring that the data is structured, normalised, and optimised for efficient data retrieval and storage. The database model was designed based on the requirements identified, taking into account the relationships between the different , their relationships, and the data attributes that needed to be stored.

The artefact lays on an Oracle Database, which is a relational DBMS introduced by Oracle Corporation. Since its creation in 1977, Oracle Database has become one of the most widely used and trusted database management systems in the world. Oracle Database is known for its scalability, reliability, and performance, making it the preferred choice for many organizations, from small businesses to large enterprises (Greenwald et al., 2013).

Essentially, the database schema was designed to support the artefact, consisting of several tables that allow to store templates in Operational Template (OPT) format, templates in JSON Data Template (JDT) format, user-interface forms, and compositions. Figure 14 illustrates the database model that supports

the artefact, showing the tables, their relationships, and the attributes that define the data structure.

It is important to highlight the creation of check constraints in the database schema to limit the values that can be entered in a column, enforcing domain integrity and ensuring that the data stored in the database is valid and consistent. Another aspect to consider is the creation of indexes to improve query performance by speeding up data retrieval operations. Indexes are data structures that store a sorted list of values from one or more columns in a table, allowing the DBMS to quickly locate and retrieve the data based on the indexed columns, improving the overall performance of the system.

One of the foundational principles of openEHR is the complete separation of clinical data and demographic information in order to ensure the privacy and confidentiality of patient identity within EHR systems. In adherence to the philosophy of openEHR, there is a clear physical separation between demographic information and clinical data. The database schema was designed to reflect this separation, where there is no information regarding the personal details of patients or healthcare professionals stored in the database. Instead, the tables contain only references to entities in whichever DIM repository they belong to. A special kind of object known as PARTY_SELF in openEHR is used to refer to the subject in the EHR, using an optional external reference. This solution is viable under the circumstances of the research as the hospital has a secure environment with well established security policies and procedures. This approach facilitates the exchange and integration of health data across different systems and domains while safeguarding individuals' privacy and confidentiality. Security is also a critical aspect, as to steal an identified EHR, an attacker would need to compromise two systems: the DIM repository and the clinical data repository, which are physically separated.

Front-End Development

As the technological landscape evolves and the versatility of the online space grows at an accelerated pace, the development of feature-rich and user-friendly web applications has become considerably more complex. At the moment, it is not easy to satisfy user's expectations with basic applications that only provide static information, given the widespread desire for products that align with the latest technological trends. With a plethora of web development frameworks available today, discerning the most suitable framework for a given project poses a notable challenge. The choice of a framework is a critical decision that can significantly impact the development process, the performance of the application, and the overall user experience.

In this context, the front-end development of the artefact was carried out using ReactJS, a popular JavaScript library for building modular UIs. Launched in 2013 by Jordan Walke, a software engineer at Facebook, ReactJS was born due to the need to create scalable and faster UIs for the social media platform (Minnick, 2022). A UI is a critical component of any web application, as it is the space through which users interact with the system. UIs are everywhere, from the simple buttons on a microwave to the complex dashboards of a space shuttle (Buna, 2017). The quality of the UI directly impacts the user experience, which in turn influences the success of the application.

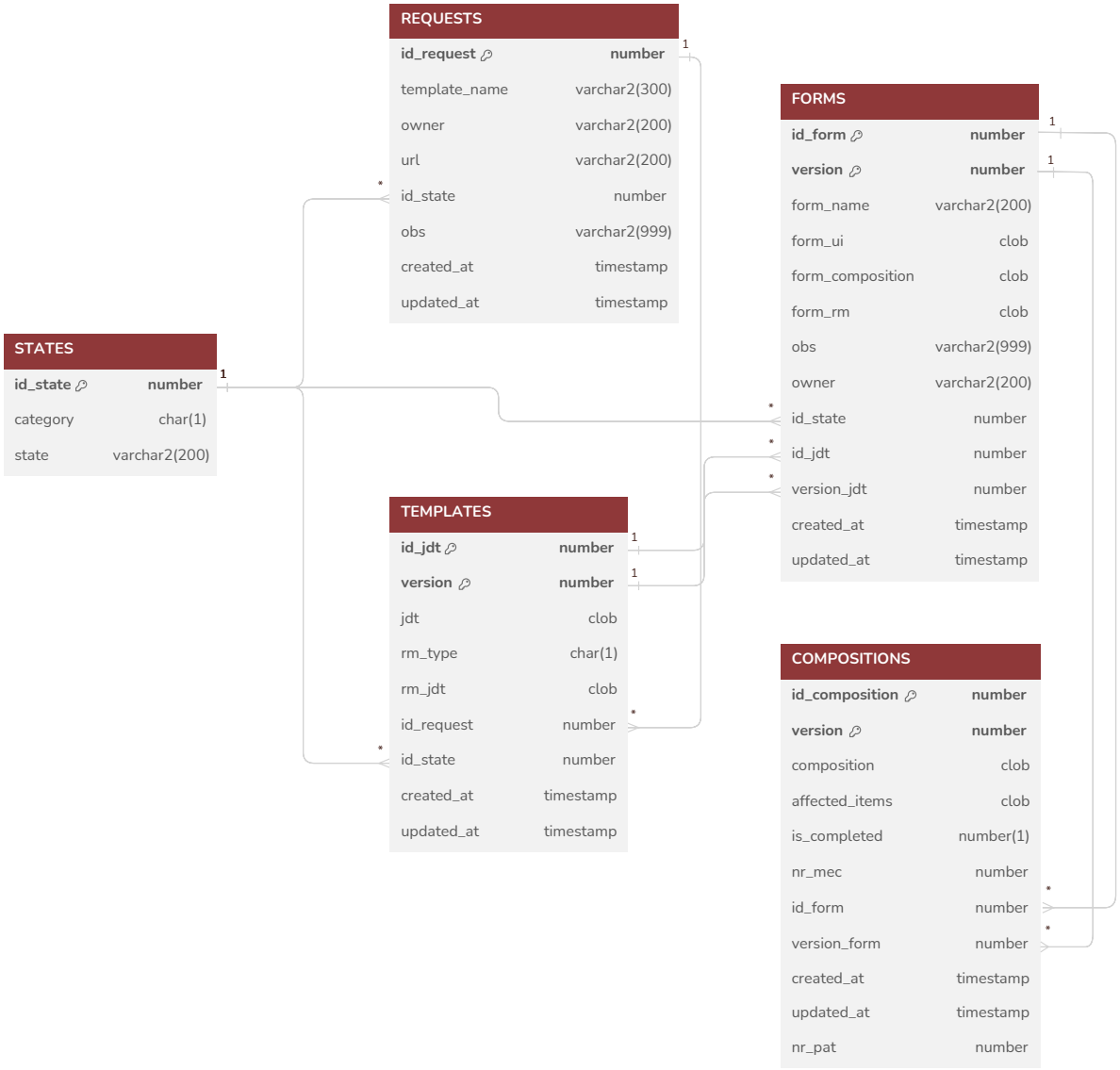


Figure 14: The database schema that supports the artefact.

Since its inception, ReactJS has gained immense popularity and widespread adoption in the web development community to build interactive and dynamic UIs (Minnick, 2022). ReactJS is taking the web development world by storm mainly because of its fast rendering capabilities with the virtual Document Object Model (DOM), which enables behind-the-scenes updates to the UI without requiring a full page reload (Buna, 2017). The virtual DOM is a lightweight copy of the actual DOM that ReactJS uses to make virtual representation of views in memory and sync it with the real DOM in an efficient way. Thus, any changes made are first reflected in the virtual DOM and then compared with the real DOM to determine the best way (minimum number of updates required) to apply the changes to the DOM. This feature is particularly useful when dealing with complex and data-intensive applications, as it improves performance and responsiveness, making the loading process of the page less time consuming, since it causes the browser to only re-render a small part of the page, instead of the entire page (Buna, 2017).

ReactJS has a component-based architecture, where the UI is divided into small blocks of reusable code that can be composed together to create complex interfaces. Each component encapsulates a specific piece of functionality independently and can be reused across multiple parts of the application. Thus, these components can be called at any time, combined, and nested within each other to create the desired UI layout. This modular approach simplifies the development process, enhances code reusability, and improves maintainability, making it easier to manage and scale large codebases (Banks & Porcello, 2020).

In Object-Oriented Programming (OOP), the concept of inheritance is used to create variations of classes based on existing ones, allowing the new classes to inherit the properties and methods of the parent class and extend or override them as needed. In ReactJS, the logic is different, as components are created using a composition model rather than inheritance (Minnick, 2022). This means that components are created more broadly to be reusable and composable, configuring them though props and state to achieve the desired behaviour or appearance in specific contexts. This approach is more flexible and scalable than inheritance, as it allows for greater customization and adaptability of components without the need to modify the original source code. Interestingly, ReactJS components use a philosophy similar to the concept of an archetype in openEHR, where a base archetype is built according to the maximal data set that can be represented, and then it can be constrained to meet the specific requirements of a particular use case. This analogy highlights the synergy between ReactJS and openEHR in terms of their component-based and archetype-based approaches, respectively.

Hooks are a powerful feature introduced in ReactJS 16.8 that allows developers to use state and other React features in functional components, which were previously only available in class components. Hooks provide a more concise and readable way to manage state, side effects, and other features in functional components. The most commonly used hooks are useState, useEffect, useContext, and useReducer, which enable developers to manage component state, perform side effects, share data between components, and manage complex state logic in a more efficient and declarative way (Rawat & Mahajan, 2020).

There are several features of ReactJS that make it an ideal choice for building user-interface forms from openEHR templates, among which are:

- **Component-Based Architecture:** ReactJS is based on a component-based architecture, where the UI elements are built upon reusable and modular components. This approach aligns well with the structure of openEHR templates, which are composed of reusable archetypes that follow a specific structure and a finite set of data elements.

- **Virtual DOM:** ReactJS uses a virtual DOM to optimise the rendering process of updating the UI based on data changes, improving performance and reducing the time required to synchronise the UI with the underlying data model. This is important when handling dynamic data from openEHR templates, as it minimises unnecessary re-renders and improves overall responsiveness.

- **Declarative Syntax:** ReactJS uses a declarative syntax that makes it easier to understand and write code. Using an analogy, it is similar to ordering food at a restaurant by simply stating your preferences (declarative), rather than instructing the chef on how to cook each ingredient (imperative). A declarative syntax simplifies the process of defining form elements and their behaviour.
- **State Management:** ReactJS provides a simple and efficient way to manage the state of components, allowing developers to create interactive and dynamic user interfaces that respond to user input and data changes. When building UI forms from openEHR templates, state management capabilities is crucial for keeping track of form data and handling user interactions effectively, validating user inputs and triggering updates to the UI based on changes in the data model.
- **JavaScript XML (JSX):** ReactJS uses JavaScript XML (JSX), a syntax extension of JavaScript that allows developers to write HTML-like code directly within JavaScript files. This declarative approach simplifies the creation of UI forms, making it easier to represent and maintain complex structures and layouts defined in openEHR templates.
- **One-Way Data Binding:** ReactJS uses one-way data binding, which means that data flows unidirectionally from parent to child components using props. It is like a waterfall, where the water (data) flows from the top (parent component) to the bottom (child components). In the context of openEHR forms, unidirectional data flow ensures that changes to the underlying data model are reflected in the UI seamlessly.
- **Rich Ecosystem:** ReactJS has a thriving ecosystem of libraries, tools, and extensions that enhance its functionality and streamline the development process, enabling developers to build sophisticated and feature-rich applications. This includes libraries for form validation, state management, and UI design, which can enhance the functionality and aesthetics of UI forms built from openEHR templates.

The richness of the ReactJS ecosystem is a determinant aspect for the development of the artefact, as it provides a diverse toolbox of libraries, tools, and extensions that can be leveraged to enhance the functionality, performance, and user experience of the application. Table 11 lists the most notable libraries used in the development of the artefact.

Table 11: Libraries used in the development of the artefact

Library	Description
Tailwind	Utility-first CSS framework for rapid UI development
Axios	Promise-based HTTP client for making API requests
Formik	Form library for React applications with easy form handling
Yup	Schema validation library for form validation in JavaScript

Font Awesome	Icon library providing scalable vector icons
react-datepicker	Datepicker component for React applications
react-dnd	Drag and drop library for React applications
pdfmake	Client-side PDF generation library for web applications
object-path	Utility library for accessing nested object properties

Back-End Development and API Integration

In terms of back-end, the artefact is sustained by a distributed microservices architecture composed of several small and independent services. Each microservice is designed to perform a single task or specific function, such as retrieving a particular form.

The ability to scale horizontally is one of the key advantages of a microservices architecture, as it allows individual services to be replicated and deployed across multiple servers or containers to handle increased load and improve performance. In addition, the independent nature of microservices offers resilience and fault tolerance, preventing failures from cascading throughout the entire application.

The communication between the client and the microservices is facilitated through well-defined Application Programming Interface (API)s using RESTful principles, specifying how data and requests should be structured and exchanged. The front-end client interacts with the microservices using Hypertext Transfer Protocol (HTTP) or Hypertext Transfer Protocol Secure (HTTPS), which are standard communication protocols that enables the transfer of data between the client and server over the web. Whenever the client application needs to perform an operation or retrieve data from the database server, it sends an HTTP request to the appropriate microservice API, specifying the desired action and any eventual parameters required. The microservice processes the request, retrieving the necessary data from the database or executing a certain operation, and returns a response to the client, which is then displayed on the UI. This client-server communication flow is essential for the artefact to function properly, enabling the client to interact with the back-end services and retrieve the necessary data to render the user-interface forms in a decoupled and modular manner.

Table 12 lists the main microservices that support the artefact, along with their respective functionalities and endpoints.

Table 12: Description of some of the microservices that support the operational functioning of the artefact

HTTP Method	Endpoint	Description
GET	/forms	Service that returns all existing forms.
POST	/forms	Service that allows to create a new form.
PATCH	/forms	Service that allows to update a specific existing form.
GET	/forms/idForm	Service that returns a specific existing form.
DELETE	/forms/idForm	Service that allows to ”delete” a form.

POST	/forms/analysis	Service that updates the state of a specific form and adds notes to it.
POST	/forms/idForm/newVersion	Service that creates a new version of a form.
POST	/updateInitialComposition	Service that updates the initial composition values for a given composition ID.
POST	/getRM	Service that returns the reference model attributes according to the reference model class.
GET	/getProductionTemplates	Service that returns all templates with production state.
POST	/getItemTemplate	Service that returns the items of a given template.
POST	/addFormConfig	If it receives a form ID, updates the FORM_SCHEDULING_CONFIG table; if not, inserts a new record.
POST	/addFormScheduling	Inserts into the FORM_SCHEDULING table according to the schedule in the FORM_SCHEDULING_CONFIG table.
POST	/getAgendamento	Returns the scheduling forms for a patient ID.
POST	/getFormsByTemplate	Returns forms originated from a particular template.
POST	/updateFormScheduling	Updates the SCHEDULE_FORM table for adding the correct url to open the form.
POST	/getPatientFormHistory	Returns all forms available for a given patient.
POST	/getInfoTemplate	Returns the information of a given template, including the composition and the reference model.

For demonstration purposes, the microservice responsible for creating a new form in the database will be used as an example. This microservice is accessed through a POST request to the /forms endpoint, which expects a JSON payload containing the form data to be created. The listing 4.4 contains the structure shema of the POST request. The microservice processes the request, validates the data, and inserts the form into the database, returning a response that confirms the successful creation of the form or provides an error message if the operation fails.

Listing 4.2: Structure schema of the new form request

```
{
  "idJDT": "example_template.v1::c7ec861c-c413-39ff-9965-a198ebf44747",
  "idState": 1,
  "idInitialComposition": "example_composition::1288",
  "composition": [],
  "formUI": {
    "formOrder": {
      "[openEHR-EHR-COMPOSITION.encounter.v1]/[openEHR-EHR-OBSERVATION.respiration.v1]"
          ↪ : 1,
```

```
    "[openEHR-EHR-COMPOSITION.encounter.v1]/[openEHR-EHR-OBSERVATION.respiration.v1]/
        ↪ [at0004]": 2,
    "[openEHR-EHR-COMPOSITION.encounter.v1]/[openEHR-EHR-OBSERVATION.respiration.v1]/
        ↪ [at0024]": 3
  },
  "datatypes": {},
  "hsizes": {
    "[openEHR-EHR-COMPOSITION.encounter.v1]/[openEHR-EHR-OBSERVATION.respiration.v1]/
        ↪ [at0004]": "50%",
    "[openEHR-EHR-COMPOSITION.encounter.v1]/[openEHR-EHR-OBSERVATION.respiration.v1]/
        ↪ [at0024]": "50%"
  },
  "fileExtensions": {},
  "fileMaxSize": {},
  "multimediaOrientation": {},
  "fontSize": {
    "button": "12px",
    "field": "11px",
    "fieldTitle": "11px",
    "formTitle": "16px",
    "header": "10px",
    "sectionTitle": "12px"
  },
  "fontStyle": {
    "button": [],
    "fieldTitle": [
      "font-bold"
    ],
    "formTitle": [],
    "header": [],
    "sectionTitle": []
  },
  "fontAlignment": {
    "fieldTitle": "justify-left",
    "formTitle": "justify-center",
    "sectionTitle": "justify-left"
  },
  "fontLetters": {
    "button": "normal-case",
    "fieldTitle": "normal-case",
    "formTitle": "normal-case",
    "sectionTitle": "normal-case"
  },
  "themePalette": {
    "$100": "#e4f1fc",
```

```
      "$200": "#aad7f8",
      "$300": "#8dcaf6",
      "$400": "#63b3ed",
      "$500": "#4299e1",
      "$600": "#3573a5"
    },
    "limitCharacters": {},
    "sectionArchetype": {},
    "textHeight": {},
    "showSectionTitle": {
      "[openEHR-EHR-COMPOSITION.encounter.v1]/[openEHR-EHR-OBSERVATION.respiration.v1]"
          ↪ : false
    }
  },
  "formName": "New Form Example",
  "owner": "dianaferreira",
  "formRM": []
}
```

All the microservices are deployed in a containerised environment using Docker and developed with Go programming language. It is important to emphasise that although the operational logic behind the microservices has been defined within the scope of this research, the actual implementation and coding of these services fall outside the research's extent.

Overall, this architecture is designed to be modular, scalable, resilient, and flexible, allowing the system to adapt to changing requirements and handle a large number of concurrent users.

4.2.3.3 Architecture and Functionalities

As mentioned in the above sections, the openEHR based-artefact developed in this research is a web application that is designed to generate UI forms from openEHR templates designated as FormBuilder. The FormBuilder is a web application that allows healthcare professionals who possess modelling knowledge to create, customise, and manage forms based on openEHR templates. The database is a relational database that stores the forms, templates, and compositions created by within FormBuilder, enabling data persistence and retrieval. The interaction between these components is facilitated through a set of microservices that handle the communication between the client and the server, ensuring the seamless operation of the artefact.

The architecture of the artefact is based on a distributed microservices model, where each microservice is responsible for a specific functionality or task, such as creating a new form, updating a form, retrieving form data, or managing form configurations. This modular approach allows for greater flexibility, scalability, and maintainability, as each microservice can be developed, deployed, and managed independently, without affecting the overall system. The microservices communicate with each other

through well-defined APIs using RESTful principles, enabling the front-end client to interact with the back-end services and retrieve the necessary data to render the user-interface forms. Figure 15 illustrates the overall architecture of the artefact, showing the components, their interactions, and the flow of data between the client and the server.

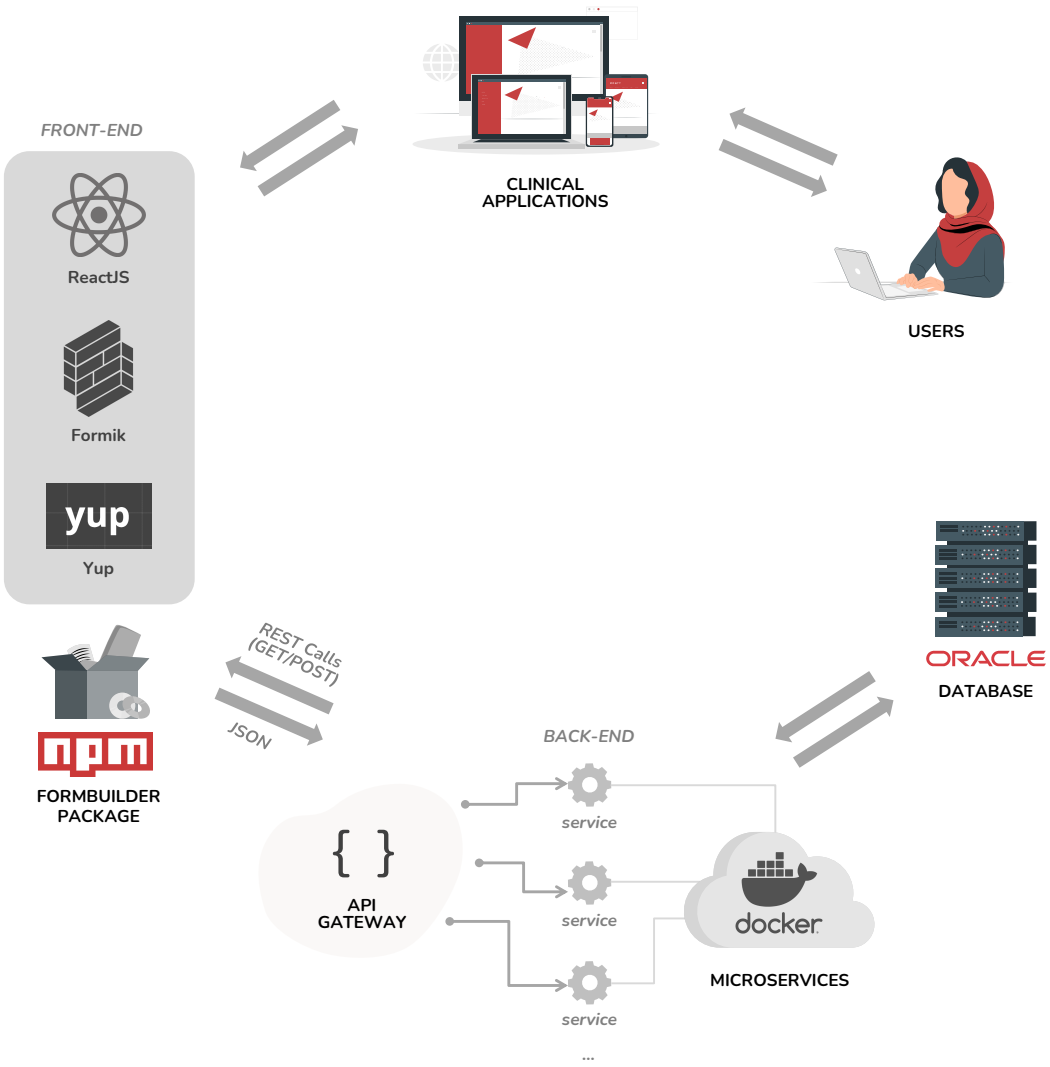


Figure 15: High-level architecture of the openEHR-based artefact.

It is important to note that the artefact was bundled into a npm package for distribution purposes. Using npm as a distribution platform offers unparalleled convenience and accessibility to developers within the JavaScript ecosystem. The npm package contains all the necessary files, libraries, and dependencies required to run the FormBuilder, making it easy to install and integrate into existing projects. Additionally, distributing the solution via npm facilitates version management and dependency resolution, ensuring that users can easily update to the latest releases and resolve any dependencies required by the package. In fact, the solution is already deployed within four other applications in the hospital, specifically:

- **EHRportal**: An openEHR structures management application, which allows the creation, edition,

and deletion of openEHR templates, forms, rules, and refsets;

- **adminEHR**: A healthcare staff professional application, which allows the creation, edition, and deletion of persons, teams, roles and capabilities within the openEHR environment;
- **patientEHR**: An application for the deployment of self-registered forms for patients, which allows the completion and edition of clinical data forms that can be filled by patients;
- **EHRcare**: An application for the management of healthcare professional's tasks, which allows the consultation, execution, and edition, of form-related tasks according to the hospital's clinical workflows. This application has a modular interface that allows to consult the tasks by patient, by professional, by team, by role, by form, by status, by date, and by priority.

The integration of FormBuilder with openEHR modelling flow represents a significant advancement in the realm of healthcare informatics. The integration process is a collaborative effort between healthcare professionals and software developers, with the goal of creating a user-friendly and efficient solution that aligns with openEHR standards and best practices. Figure 16 illustrates the different stages involved in the integration process, from the selection of relevant archetypes to the creation of UI forms. This synergetic approach promotes the exchange of knowledge and expertise between domain-specific professionals through the extensive knowledge base housed within openEHR's CKM. The integration process begins with the identification and selection of relevant archetypes from the CKM repository. These archetypes serve as foundational building blocks, defining the structured data elements and clinical concepts essential for capturing patient information. The archetypes are then imported into the Archetype Designer (AD) editor to create a template that encapsulates the desired arrangements of archetypes tailored to specific clinical scenarios. The template is then exported in OPT format and imported into the FormBuilder as JDT format. This conversion process is crucial for transforming the template into a format that is easier to work with and manipulate within the FormBuilder. The FormBuilder then processes the JDT file and generates a UI form based on the template's structure and content. By leveraging the modelling capabilities of openEHR formBuilder empowers users to construct intuitive forms aligned with openEHR standards. This integration not only ensures adherence to best practices in clinical modelling but also streamlines the process of form creation, enabling healthcare professionals to efficiently capture and manage patient data in a standardised and interoperable manner.

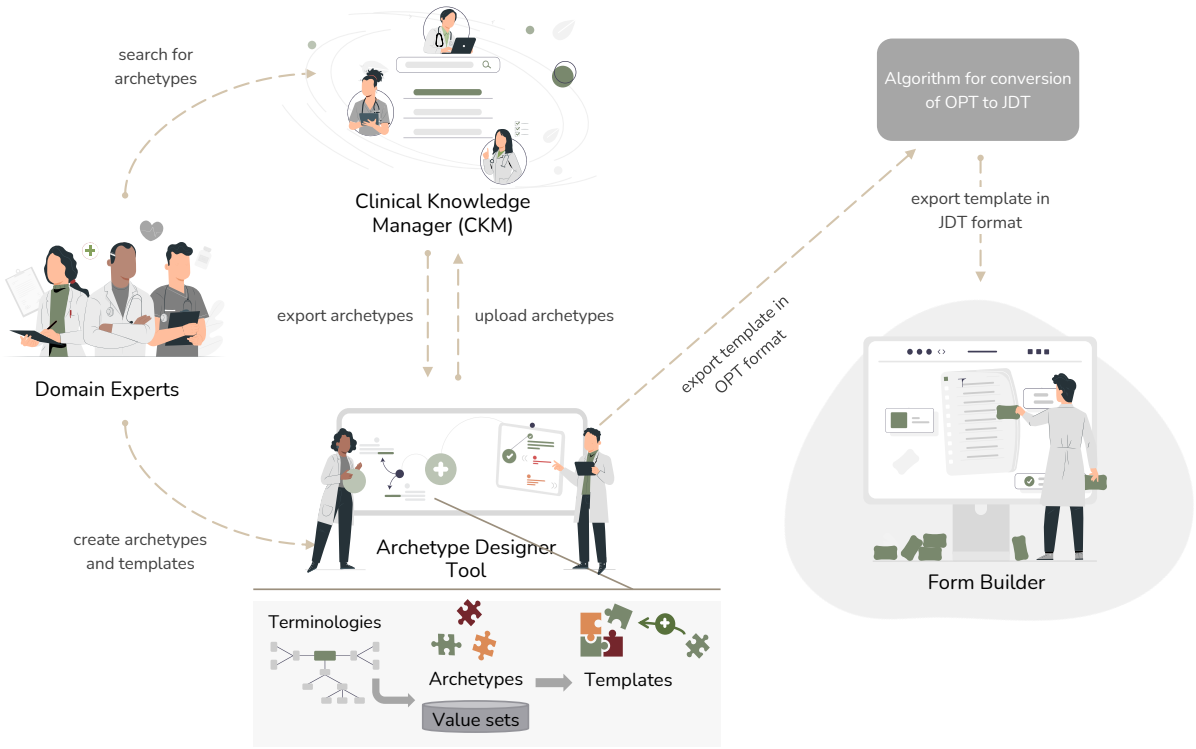


Figure 16: Integration of FormBuilder within the openEHR modelling flow.

By utilizing FormBuilder, healthcare professionals can streamline the process of data entry by creating forms that are intuitive and optimised for their specific clinical workflows. It provides a platform for clinicians with modelling knowledge to customise the UI forms by adjusting the formatting, such as choosing colours, fonts, and determining if fields should be shown or hidden. Additionally, it allows for the association of functions and refsets to form fields, which can add further functionality to the form.

In terms of functionalities, the FormBuilder provides a wide range of features, which are described below:

Table 13: Main objectives guiding the design and development of the artefact

Feature/Functionality	Objective ID	Description
Intuitive Design	UE1	The interface of formBuilder is designed to be user-friendly and easy to navigate, ensuring a seamless user experience.

(Table 13 continued on next page)

(Table 13 continued from previous page)

Feature/Functionality	Objective ID	Description
Customization	UE2	Users can customise various aspects of the form interface, including colours, fonts, field order, and input types, to better suit their preferences and requirements.
Automation	UE3	Automation features in formBuilder minimise user input and streamline processes, reducing cognitive load and improving efficiency.
Performance Optimization	UE4, P1, P2, P3	formBuilder optimises form rendering, submission processes, and loading times to ensure a smooth and efficient user experience, with lazy loading and bundle size reduction techniques implemented for enhanced performance.
Responsive Design	UE5	The interface of formBuilder adapts seamlessly to different screen sizes and devices, providing a consistent user experience across desktops, tablets, and smartphones.
Validation Feedback	UE6	Real-time validation error messages in formBuilder help users identify and resolve errors during form completion, improving completion rates and user satisfaction.
Guidance and Assistance	UE7	formBuilder provides informative tooltips and help features to guide users through the form completion process, enhancing usability and reducing errors.

(Table 13 continued on next page)

(Table 13 continued from previous page)

Feature/Functionality	Objective ID	Description
Component Reusability	M1	Components in formBuilder are designed for high reusability across different parts of the application, reducing redundancy and promoting code efficiency.
Dependency Injection	M2	formBuilder implements dependency injection patterns to allow components to receive dependencies or configuration options from parent components, enhancing flexibility and modularity.
Version Control	M3	A version control system such as Git is implemented in formBuilder to track changes and manage code revisions, ensuring efficient development and maintenance.
Organization and Separation of Concerns	M4	Components and files in formBuilder are organised into separate modules based on their distinct nature, following best practices to enhance code readability and maintainability.
Load Balancing	S1	formBuilder implements load balancing strategies to distribute incoming traffic evenly across multiple servers or instances, ensuring optimal performance and availability.
Microservices Architecture	S2	formBuilder adopts a microservices architecture to break down complex functionalities into smaller, loosely coupled services that can be developed, deployed, and scaled independently.

(Table 13 continued on next page)

(Table 13 continued from previous page)

Feature/Functionality	Objective ID	Description
State Management	S3	Efficient state management solutions such as Redux are implemented in formBuilder to handle complex data flows and application states, ensuring consistency and reliability across components.
Distribution as npm Package	S4	The application is designed and distributed as an npm package, facilitating easy distribution and consumption by other applications, promoting scalability and interoperability.
Extension Capabilities	F1	formBuilder is designed with a flexible architecture that allows for easy customization and extension of features to meet evolving user needs.
Configuration Options	F2	Users are provided with configurable options in formBuilder to adjust settings, preferences, and functionalities according to their requirements, enhancing flexibility and usability.
Cross-Browser Support	C1	formBuilder ensures compatibility with major web browsers such as Chrome, Firefox, Safari, and Edge, providing a consistent appearance and behaviour across different platforms.
Operating System Compatibility	C2	formBuilder is compatible with different operating systems such as Windows, macOS, and Linux, ensuring accessibility and usability across various platforms.

(Table 13 continued on next page)

(Table 13 continued from previous page)

Feature/Functionality	Objective ID	Description
Internationalization	C3	formBuilder partially supports localization features to make the app compatible with different languages, cultures, and regional preferences, by enabling the languages available in the template.
Third-Party Integration Compatibility	C4	formBuilder ensures compatibility with third-party APIs, libraries, and services commonly used in web development, enabling seamless integration with external tools and platforms.
Authentication	Se1	Robust authentication mechanisms such as OAuth and JWT are implemented in formBuilder to verify user identities and prevent unauthorised access, ensuring data security.
Authorization	Se2	Strict authorization policies are enforced in formBuilder to control user access rights and permissions based on roles or specific criteria, reducing the risk of unauthorised actions or data breaches.
Secure Communication	Se3	formBuilder uses secure communication protocols such as HTTPS and SSL/TLS certificates to encrypt data transmitted between the client and server, protecting sensitive information from eavesdropping or tampering.
Data Privacy Compliance	Se4	formBuilder ensures compliance with data privacy regulations such as GDPR by implementing data protection measures and providing transparency on data collection, storage, and usage practices.

Given their importance and particularities, some of the functionalities are described in more detail in the following sections.

Field Validation

Field validation is a critical aspect of FormBuilder, as it ensures that the data entered by users is accurate, complete, and consistent with the expected format. The validation rules are defined for each field based on the data type, constraints, and requirements specified in the openEHR template. These rules are enforced in real-time as users interact with the form, providing immediate feedback on the validity of their inputs. For example, if a user enters an value out of a predefined range or leaves a mandatory field blank, an error message is displayed to alert the issue and guide its correction. This approach is essential to help prevent errors, reduce data entry time, and improve the overall quality of the data captured in the form. Furthermore, it ensures that the data conforms to the standards and guidelines established by openEHR, allowing a seamless integration within a EHR server.

Technically, the validation process is implemented using the Yup library, a schema validation library for JavaScript that provides a simple and declarative way to define validation rules for form fields. Yup allows developers to create validation schemas that specify the data type, constraints, and conditions that must be met for a field to be considered valid. These schemas are then applied to the form fields, triggering validation checks whenever the field value changes. If the field value fails to meet the validation criteria, an error message is displayed to the user, indicating the issue and guiding them on how to correct it. Yup supports a wide range of validation rules, including required fields, data types, string lengths, numerical ranges, regular expressions, and custom validation functions, providing a comprehensive set of tools to ensure data integrity and accuracy in the form.

Moreover, validation rules can also be defined at the DLM level, allowing for more complex and context-specific validation logic to be enforced based on the data entered in multiple fields and the relationships between them. This validation mechanism enables to make requests to external services, databases, or APIs to perform additional checks and verifications, ensuring that the data captured in the form is compliant with the clinical requirements.

Refset Association

The association of refsets with form fields is a powerful feature of FormBuilder that enhances the functionality and usability of the forms. Refsets are predefined sets of codes or terms that represent specific clinical concepts, such as diagnoses, procedures, medications, or observations. Usually, these refsets are defined based on standard terminologies, such as SNOMED CT, LOINC, or ICD, and are used to standardise the representation of clinical data and ensure interoperability between different systems. However, local refsets are also commonly used to represent institution-specific codes or terms that are not part of standard terminologies.

Furthermore, refsets can be dynamic, meaning that they can be updated or extended according to other fields in the form. Hence, a certain value within the form can change the value sets available in another field. Listing 4.3 shows an example of an HTTP POST request to retrieve the refset values according to a specific field. The response structure is shown in Listing 4.4.

Listing 4.3: Structure schema of an HTTP POST request to retrieve the refset values according to a specific field

```
POST /refsetConcepts HTTP/1.1
Host: http://XXX.XX.XXX.XX:4010/refsetConcepts
Content-Type: application/json
Content-Length:
    {
    "id": "123456::6432",
    "code": "number",
    "text": "description",
    "var":
    {
           "code": "local_terms::2",
           "text": "Right"
       }
    }
Cache-Control: no-cache
Authorization: Bearer XXX
```

Listing 4.4: Structure schema of the response for the request example used above

```
  {
"success": true,
"concepts": [
  {
    "code": "local_terms::08R83KZ::total::right",
    "text": "08R83KZ - Total Transplant (right eye)"
  },
  {
    "code": "local_terms::08R8XKZ::DALK::right",
    "text": "08R8XKZ - Anterior Transplant -DALK (right eye)"
  },
  {
    "code": "local_terms::08R83KZ::DSAEK::right",
    "text": "08R83KZ - Posterior Transplant -DSAEK (right eye)"
  },
  {
    "code": "local_terms::08R83KZ::DMEK::right",
    "text": "08R83KZ - Posterior Transplant -DMEK (right eye)"
  }
```

```
20   ]
21 }
```

When a user interacts with a form field that is associated with a refset, FormBuilder triggers a request to a microservice to fetch the corresponding refset values. These values are then displayed to the user in a dropdown list, allowing them to select the desired term and insert it into the field. The refset values are retrieved in real-time, ensuring that the list is up-to-date and reflects the latest clinical terminology available. This dynamic lookup mechanism enhances the usability and efficiency of the form, enabling users to quickly find and select the appropriate clinical terms.

Hence, by associating refsets with form fields, users can easily select and insert relevant clinical terms into the form. This feature simplifies the data entry process, reduces errors, and ensures that the data captured in the form is standardised and consistent with the clinical terminology.

Integration with Decision Logic

In openEHR, forms and DLMs are closely related, as the data captured in forms can trigger rules in the DLM that provide decision support to healthcare professionals when filling out the form. The DLM defines the logic that governs the behaviour of the form, such as displaying or hiding certain fields, alter the editability of fields, make fields mandatory, perform calculations, display validation messages or even automatically populate fields based on the data entered by the user.

The DLMengine is responsible for executing clinical or operational rules defined in the DLM and triggering the corresponding actions/events when the predefined conditions are met. Hence, a seamless integration between the artefact developed in this thesis - FormBuilder - and the DLMengine is essential to ensure that the rules defined in the DLM are triggered and executed correctly and consistently when the form data meets certain conditions.

The integration between FormBuilder and the DLMengine is achieved through a series of computational tasks that are executed sequentially to verify the need to trigger any rule requests. Basically, the process starts when a healthcare professional updates a form field. Whenever the value of a field changes, FormBuilder performs an internal processing to check whether that field is a condition of any rule within the DLM associated with the current form. If so, FormBuilder identifies each associated rule and for each of them verifies if there are more associated conditions and whether they are conjunctions or disjunctions. If the conditions are conjunctions, FormBuilder checks if all the conditions have a value assigned to them. When all the conditions of a rule have an assigned value, FormBuilder triggers a POST request to the DLMEngine with the input variables. On the other hand, if the conditions are disjunctions, FormBuilder triggers a POST request to the DLMEngine with the input variables for each condition value available.

As mentioned, this is a well-defined sequence of tasks that need to be followed to properly implement the specific set of rules that have been modeled for each form. For better understanding, a graphical representation of the processing that occurs in the FormBuilder to verify the need to trigger any rule requests is presented in Figure 17.

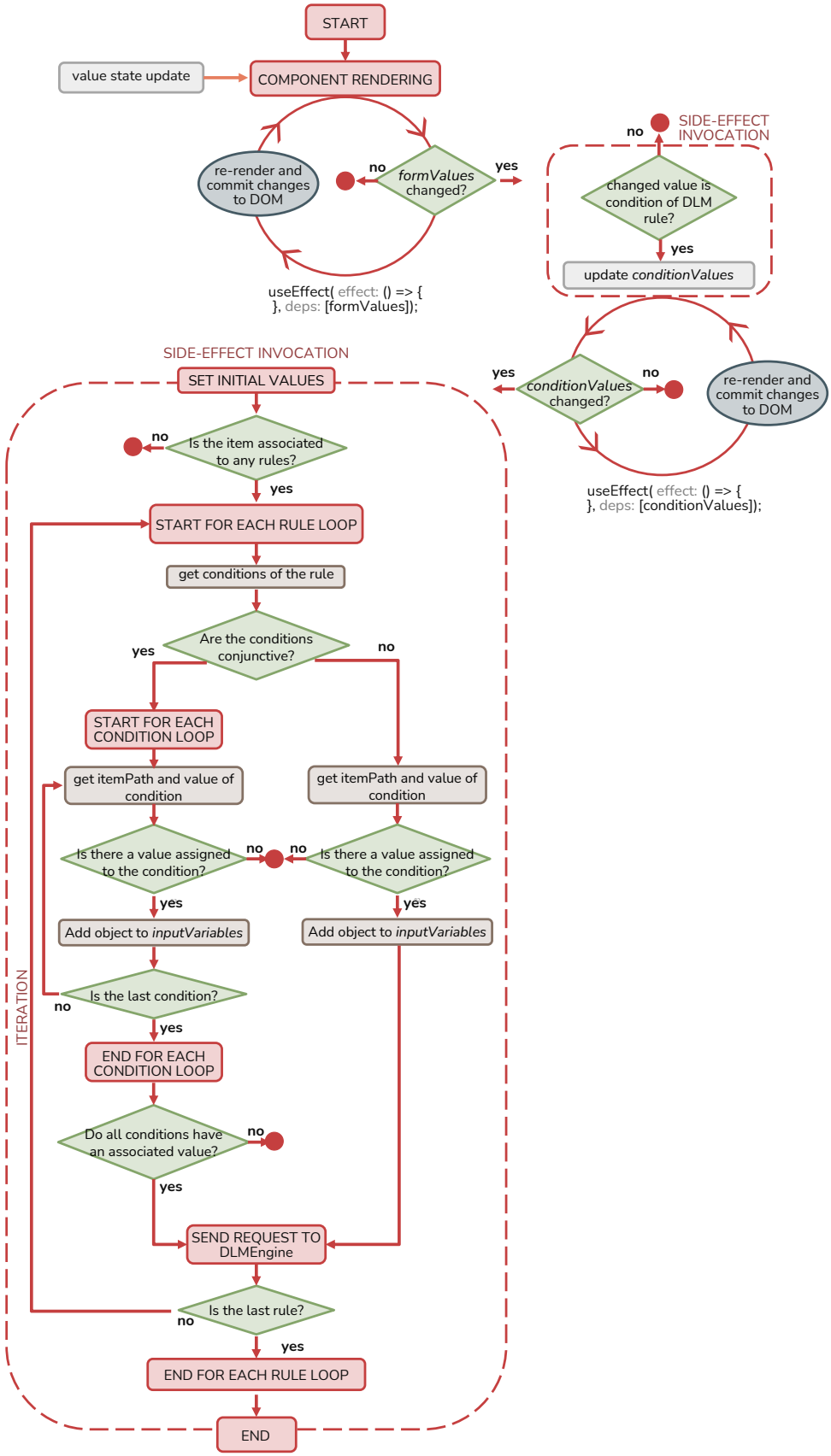


Figure 17: Logic flow that occurs in the FormBuilder process responsible for initiating rule requests.

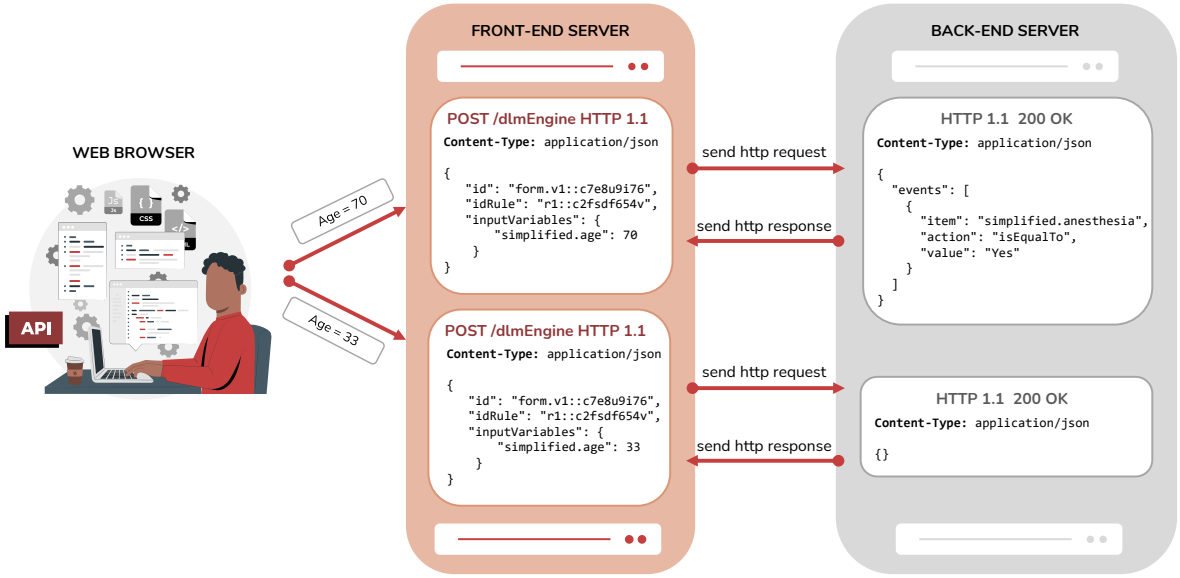


Figure 18: Diagram of HTTP request and response interactions between the Form Builder, which acts as the front-end server, and the DLMEngine, which serves as the back-end server, in two distinct settings.

Finally, FormBuilder executes the actions/events returned by the DLMEngine, updating the form UI accordingly.

Upon receiving the input variables, the DLMEngine processes the rule requests and checks whether the conditions are met for executing a given rule. If the values assigned to the form fields do not fulfil a rule, an empty object is returned. On the other hand, if the values meet the conditions for executing a rule, the engine returns the path of the affected field and the type of action/event to be executed by FormBuilder. Finally, FormBuilder executes the necessary actions/events for each field, updating the form UI according to the response returned by the engine. Figure 18 serves as an exemplification of the HTTP requests that are exchanged between the FormBuilder, front-end server, and the DLMEngine, back-end server, in two distinct scenarios. In the first scenario, the health professional enters a value that triggers the execution of a rule, while in the second scenario the value entered does not involve the activation of any rule. This diagram provides a clear representation of the data flow and communication between the two servers during the execution of the rule-based system. It is worth mentioning that the identifiers used in the diagram are fictitious and simplified for illustrative purposes.

4.2.3.4 Graphical Interface Design

The graphical interface of the artefact was designed to be user-friendly, intuitive, and visually appealing, providing healthcare professionals with a seamless and efficient experience when creating UI forms from openEHR templates. The design aims to ensure that healthcare professionals can easily navigate the application, interact with the form elements, and customise the form layout according to their preferences. The interface is consistent with the principles of user-centered design, focusing on the needs and expectations of healthcare professionals to ensure that the end application meets their requirements. Thus, all

the objectives outlined in section 4.2.2 were taken into account during the design process, ensuring that the interface is aligned with the needs and expectations of healthcare professionals. Later in this chapter, the artefact will be evaluated by healthcare professionals against these objectives in agreement with the TAM model to assess its usability, functionality, and impact on healthcare documentation practices.

As mentioned earlier, the artefact can be consumed in three different ways: in configuration/design mode, in edit/filling mode, and in view/consult mode. Hence, each mode has a specific set of functionalities and features that cater to the needs of healthcare professionals in different contexts. The design of the graphical interface was tailored to accommodate these modes, providing a consistent and intuitive user experience across all modes. The following sections explain the design of the graphical interface for each mode.

Configuration/Design Mode

In configuration/design mode, healthcare professionals can create user-interface forms from openEHR templates by customizing the form layout, formatting, and behaviour. The interface is designed to be visually appealing and highly interactive, allowing users to drag and drop form elements, adjust the styling, and associate functions and refsets with form fields. With a clean and modern aesthetic interface that aligns with the expectations of the new generation of healthcare professionals, the artefact provides a platform to design and customise forms that are tailored to their specific clinical workflows and documentation needs.

The design of the interface was designed to be as clean and minimalistic as possible, with a focus on usability and functionality. In this way, the configurations are accessible on the left side of the screen, inside a collapsible sidebar, which can be expanded or collapsed as needed. This design choice was made to maximise the available screen space for the form editor, providing users with a larger workspace to design and customise the form. The form elements are displayed in an invisible grid layout, allowing users to easily position and align the elements according to their preferences. The form elements can be resised, moved, , making it easy to create complex and dynamic form layouts. The sidebar is divided into several sections, each corresponding to a specific aspect of the form configuration.

These sections can be broadly categorised into general settings, which apply to the entire form, and specific settings that are specific to individual form elements. The general settings include options to change the form title, theme colour, font style, and other global settings that affect the overall appearance of the form. The specific settings allow users to customise the properties of individual form elements, such as the field label, placeholder, component type, visibility, editability, among others.

The application provides real-time actualizations of the UI to users as they make changes to the form. The real-time display of the form preview, enables users to visualise the impact of their changes immediately, facilitating the design process and reducing the need for frequent changes, reducing the need for versioning.

Overall, the design of the interface prioritises simplicity, clarity, and consistency, ensuring that users

can focus on the task of designing the form without being distracted by unnecessary elements or complex interactions.

Figure 19 illustrates the design of the graphical interface in configuration/design mode, showing the panel for overall form configurations, in this case, the theme colour, which is available to healthcare professionals when creating user-interface forms from openEHR templates. On the other hand, Figure 20 illustrates the design of the graphical interface in configuration/design mode for individual form element customization, in this case, showing the panel for specific datatype configurations.

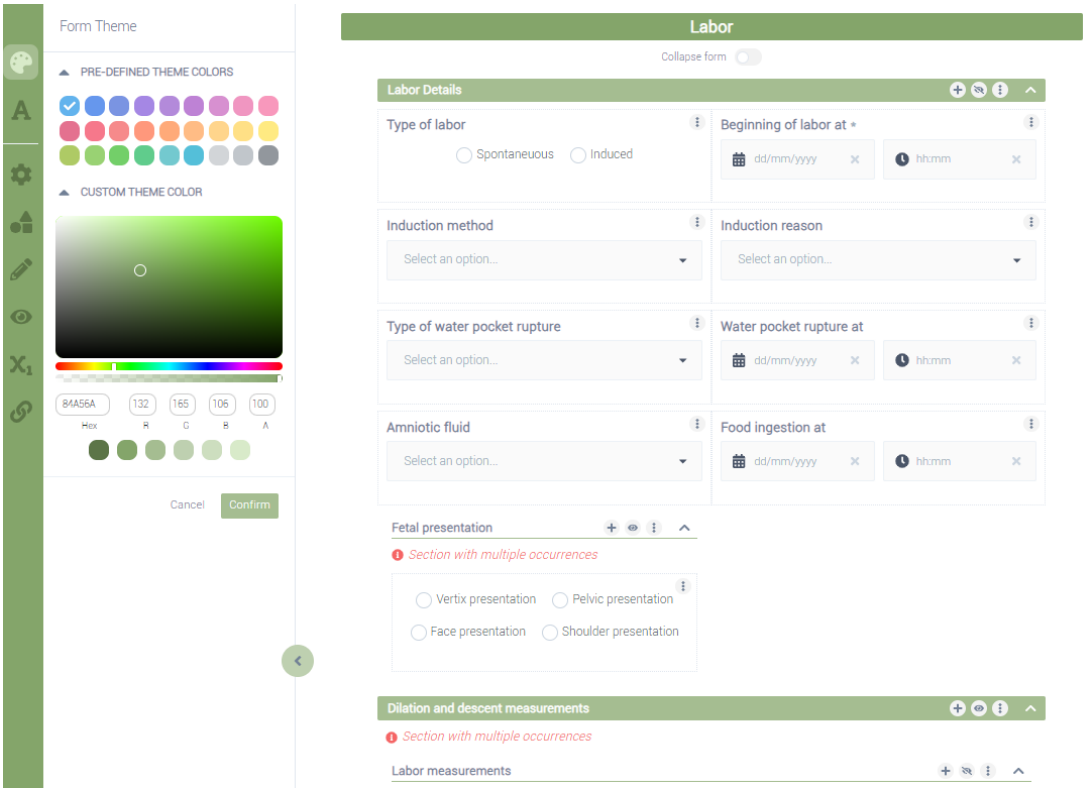


Figure 19: The design of the graphical interface in configuration/design mode for general customizations.

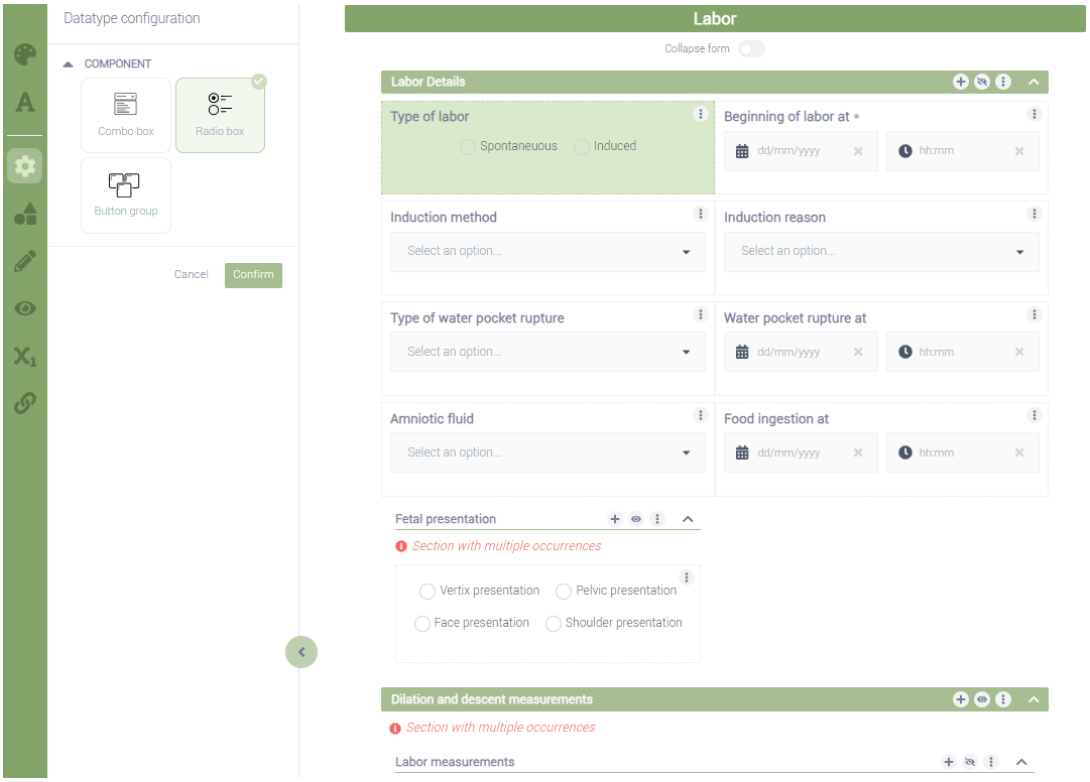


Figure 20: The design of the graphical interface in configuration/design mode for individual form element customization.

Edit/Filling Mode

In edit/filling mode, healthcare professionals or patients, in the case of self-reporting forms, can interact with the user-interface forms created in configuration/design mode, by entering or updating information, and saving or submitting the form for storage in the database. In the majority of cases, this mode is used during the clinical encounter to capture patient data, record observations, and document clinical findings. The interface in edit/filling mode is designed to be clean, uncluttered, and focused on the task of data entry, providing users with a distraction-free environment to input information efficiently.

As mentioned in the previous section, the form elements are arranged in a logical and intuitive manner by healthcare professionals with expertise in openEHR clinical modelling in configuration/design mode. This hands-on approach ensures that the forms are implemented in a manner that precisely aligns with clinicians' requirements and their cognitive processes. By directly involving healthcare professionals in the design process, the forms are tailored to mirror the specific workflows, terminology, and data collection needs inherent to a particular clinical practice. This alignment not only enhances the usability and efficiency of the forms but also fosters greater acceptance and adoption among clinicians. Additionally, healthcare professionals can iteratively refine and optimise the forms based on real-world feedback and evolving clinical needs, ensuring continuous improvement and relevance over time.

The interface in edit/filling mode is optimised for data entry and form completion, providing clear labels, input fields, and buttons that guide users through the process of completing the form, ensuring

that they can enter information accurately and efficiently. The interface also includes features such as auto-population of fields, dynamic field visibility, and conditional logic, which enhance the user experience and streamline the data entry process by reducing manual input and minimizing errors. Furthermore, the integration with the DLM ensures that the forms are dynamically updated based on predefined rules and conditions, providing real-time guidance to users as they fill out the form and ensuring that clinical guidelines and best practices are followed.

In addition, the fact that the forms are constrained and validated according to the openEHR templates ensures that the data entered is accurate, complete, and compliant with the clinical standards and terminologies defined in the templates. This standardization of data entry not only improves the quality and consistency of health data but also facilitates data sharing, interoperability, and reusability across different healthcare settings and systems.

By providing a user-friendly and intuitive environment for data entry, this interface aims to reduce the burden on healthcare professionals when documenting clinical information, empowering them to effectively capture, manage, and utilise accurate and structured health data, thereby enhancing the overall quality of healthcare documentation practices and, as a consequence, the delivery of patient care.

Figure 21 illustrates the design of the graphical interface in edit/filling mode.

View/Consult Mode

In view/consult mode, healthcare professionals can access and review the forms completed in edit/filling mode by themselves or other users, according to their access rights and permissions. This mode is typically used for clinical decision-making, patient consultation, and data analysis, allowing healthcare professionals to retrieve, consult, and analyze the information captured in the forms.

The interface in view/consult mode is optimised for data visualization, providing users with a clear and structured overview of the form data, including labels, values, and metadata, enabling them to review the clinical information and make informed decisions based on the data collected.

The form data is displayed in a readable and organised format, allowing users to quickly scan and interpret the information without being overwhelmed by unnecessary details or complex layouts. The interface includes features such as collapsible sections and expandable fields that enable users to navigate and explore the form data efficiently, ensuring that they can access the information they need quickly and easily.

It is intended in the future to integrate the interface with tools for searching, filtering, and sorting the form data, enabling users to locate specific information. Moreover, it is planned to incorporate data visualization tools, such as charts and graphs, to enable users to analyze and interpret the form data more effectively. The visual representation of the data allows users to identify patterns, trends, and outliers, facilitating complex data analysis tasks, data-driven decision-making and derivation of meaningful conclusions from the form data.

Figure 22 illustrates the design of the graphical interface in view/consult mode.

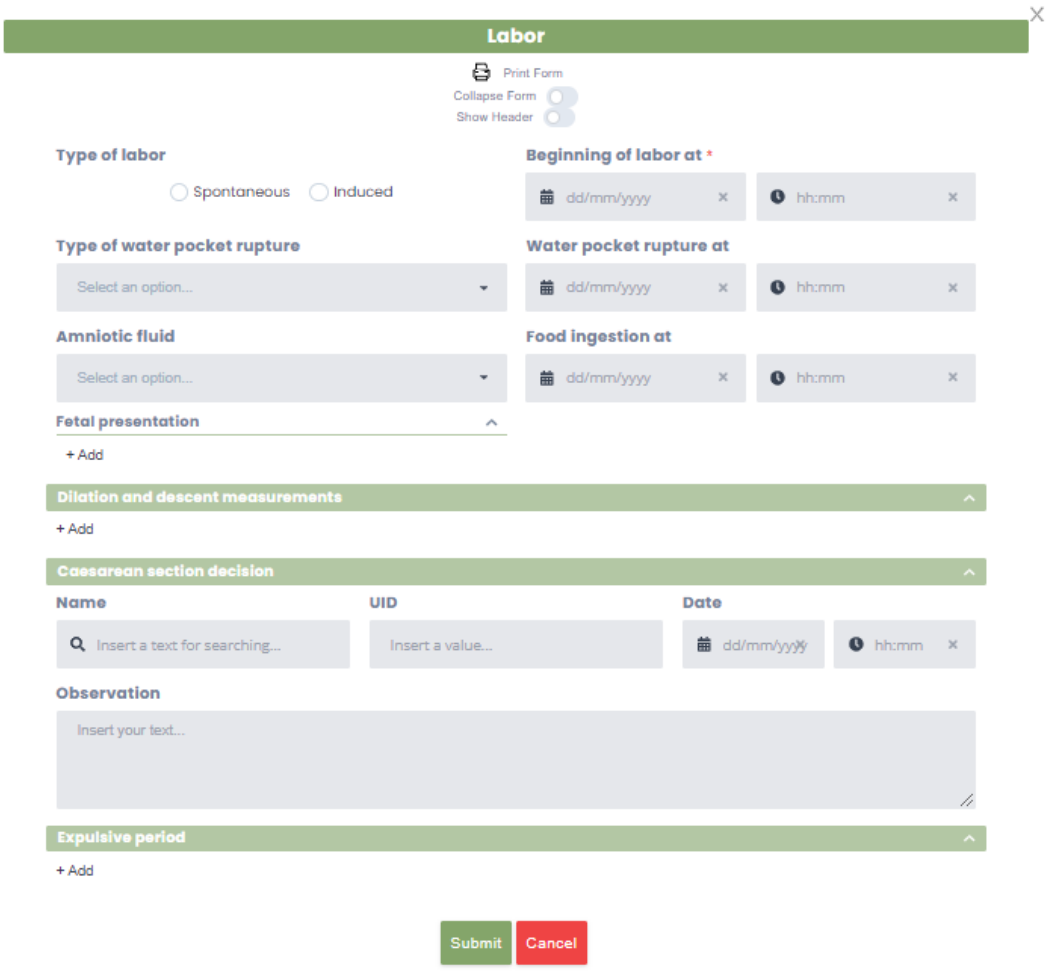


Figure 21: The design of the graphical interface in edit/filling mode.

4.2.4 Demonstration

To demonstrate the functionality and usability of the proposed artefact, a series of case study scenarios were developed in collaboration with healthcare professionals at the ULSdSA as well as members of the IT department. These case studies refer to real-world clinical scenarios and are designed to showcase the application of the artefact in different healthcare settings as well as to prove its relevance and impact on healthcare documentation practices. The case studies were developed based on the user stories and requirements identified during the Scrum meetings and were refined through iterative feedback and testing with end-users. The goal of the case studies is to illustrate how the artefact can be used to create user-interface forms from openEHR templates in a variety of clinical contexts, including corneal transplantation, imaging exams, and nutritional risk assessment before considering the artefact's integration within the complexity of the critical care environment. Through the implementation of the solution in simpler clinical scenarios, it is possible to evaluate and validate its functionality, usability, and impact by healthcare professionals from wide-ranging specialties and backgrounds, paving the way for its integration into more complex and demanding healthcare settings.

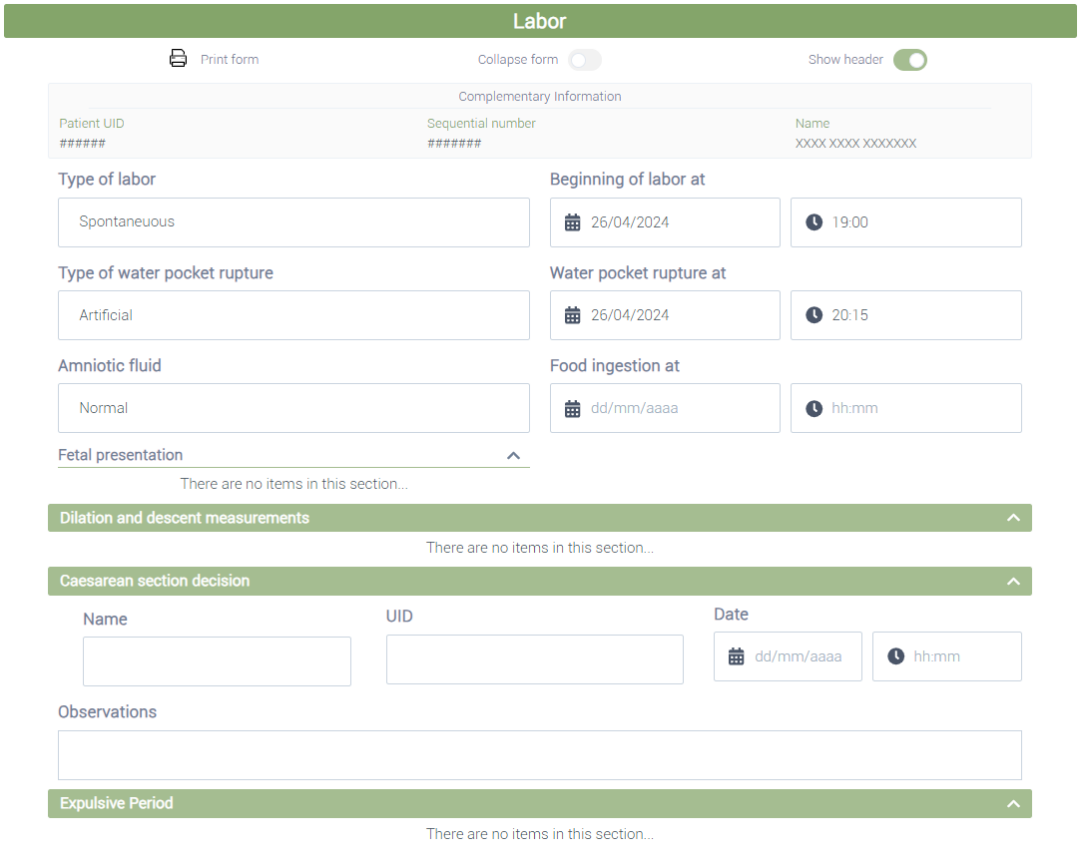


Figure 22: The design of the graphical interface in view/consult mode.

The artefact is integrated in several healthcare contexts across different specialties at the ULSdSA, demonstrating its versatility and adaptability to various clinical scenarios, which include neurophysiology, obstetrics, imaging exams, nutrition, and corneal transplantation. Given its extensive applicability, the artefact has the potential to be used by healthcare professionals in a wide range of clinical scenarios and healthcare settings. At the moment, the solution is in production at the ULSdSA and available for daily use by healthcare professionals in the context of neurophysiology, imaging exams, and corneal transplantation.

As demonstrability is a key aspect in assessing the artefact's functionality and usability, a case study focusing on the corneal transplantation process was developed to showcase the application of the artefact in a real-world clinical scenario. The case study was designed in collaboration with healthcare professionals from the Ophthalmology Department at the ULSdSA, who provided valuable insights and feedback on the requirements, archetypes, and templates used in the case study.

The implementation process of the artefact in the case studies followed a structured methodological approach, which included the collection of requirements, the search and construction of archetypes, the template development, the creation of the user-interface forms, the integration with other openEHR artefacts, specifically DLM and Work Plan (WP), and finally the evaluation of the implementation outcomes. Figure 23 illustrates the different stages involved in the case study process, which will be further detailed in the following sections.

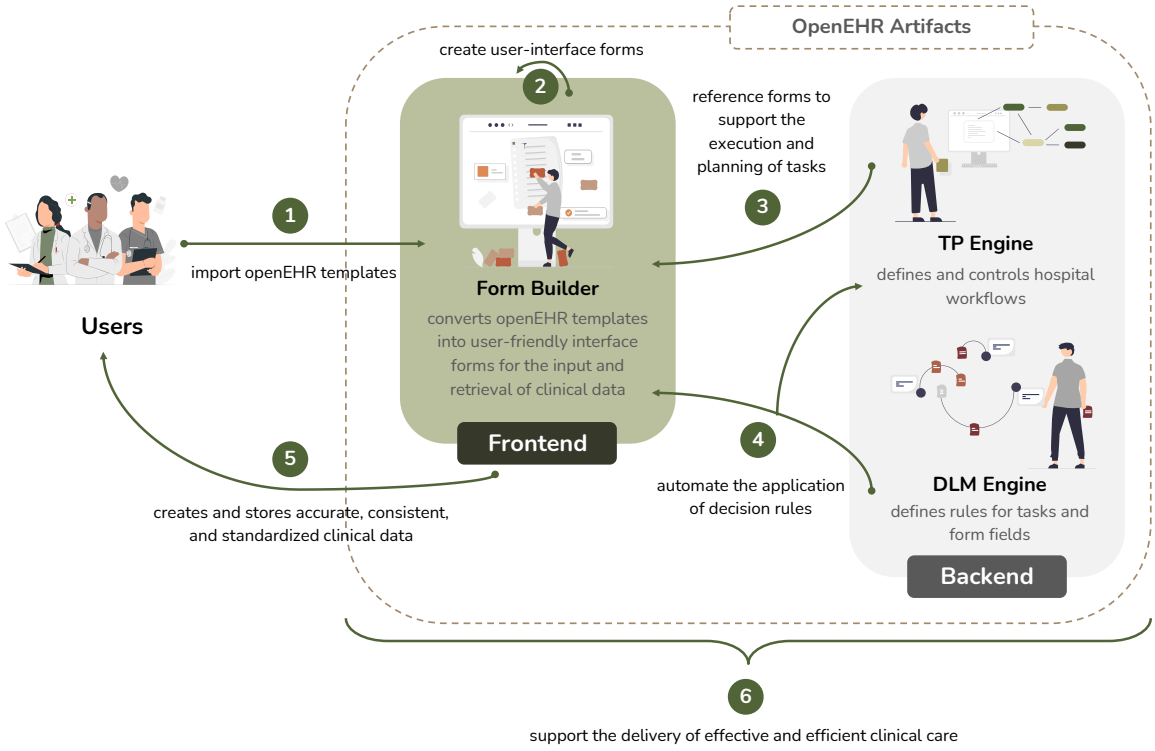


Figure 23: The different stages involved in the development of the case studies.

The following sections delve deeper into the details of the corneal transplantation case study.

4.2.4.1 Case Study I: Corneal Transplantation

Background

The eye is a highly evolved and complex sensory organ possessed by a wide range of species, enabling organisms to perceive and interpret visual information from their surroundings. Vision is one of the most valuable senses for humans and plays a critical role in every facet of an individual's life (World Health Organization, 2019). The sense of vision is the result of an intricate interaction between the eyes, the brain, and the nervous system (Hollins, 2022).

Visual impairment occurs when a pathological condition disrupts the visual system and one or more of its associated functions (World Health Organization, 2019). Blindness is a major public health issue, particularly in developing and underdeveloped countries where access to healthcare and resources is scarce (R. Bourne et al., 2021; Organization et al., 2022). The loss of sight can severely impact an individual's daily life, hindering their ability to perform routine tasks, interact with others, and preserve their independence (Brown & Barrett, 2011). On many occasions, blindness can also lead to social isolation, depression, and decreased quality of life (Brown & Barrett, 2011; World Health Organization, 2019).

In 2019, the WHO estimated, in the world report on vision, that there were approximately 2.2 billion

people worldwide with vision impairment or blindness (World Health Organization, 2019). The prevalence of visual disability is alarming and a source of growing global concern.

The apprehension surrounding blindness is not only rooted in the physical limitations it imposes, but also in its social and economic consequences (R. Bourne et al., 2021). From an economic point of view, the loss of sight can cause reduced workforce participation, decreased productivity, and increased health expenses (Brown & Barrett, 2011). Consequently, the loss of income and the higher healthcare costs can drain governments with additional financial pressures, exacerbating poverty and slowing economic growth (World Health Organization, 2019). Hence, the impacts of blindness are far-reaching, affecting not only the individual, but also their families, communities, and society as a whole.

Without more assertive measures, the escalating demand for eye care services worldwide is projected to persist and intensify in the next few decades, posing a meaningful challenge to the healthcare industry and requiring innovative solutions to meet the increasing pressure for quality eye care services (R. R. Bourne et al., 2017; Steinmetz et al., 2021).

The eye is a complex organ composed of several structures that work together in the perception of the world in all its lights, colours, shapes, and movements (Hollins, 2022). One of the most vital structures of the visual system is the cornea, which is the clear outermost layer located at the front of the eye (Singh et al., 2019). A transparent cornea acts as a clear window to allow light to enter the eye and reach the retina, a layer of neural tissue at the back of the eye where light is converted into electrical signals and transmitted to the brain for interpretation as visual information (Hollins, 2022). For an individual to have clear vision, the cornea must be transparent and free of any obstructions, such as scars or opacities, in order to allow light to cross the eye and access the retina (Pineda, 2020; Singh et al., 2019).

The main causes of visual impairment are cataract, glaucoma, macular degeneration, detached retina, diabetic retinopathy, and retrolental fibroplasia (R. Bourne et al., 2021; Flaxman et al., 2017; Steinmetz et al., 2021). Some of these eye conditions, such as cataract, diabetic retinopathy, and retrolental fibroplasia, can negatively impact the clarity of the cornea and lead to vision impairment (Hollins, 2022; Pineda, 2020). According to the WHO, corneal opacities are the fourth leading cause of blindness on a global scale (Pineda, 2020).

In many cases, visual rehabilitation is possible with corneal transplantation (Organization et al., 2022; Singh et al., 2019). Corneal blindness can be effectively reversed with a healthy donor cornea transplant. Because the cornea lacks blood vessels, the risk of graft rejection is significantly reduced, making corneal transplantation one of the most successful forms of organ transplantation in the human body (Singh et al., 2019). Corneal transplantation, also known as Keratoplasty, is a surgical procedure that replaces a damaged or diseased cornea with a healthy one to restore vision and improve the quality of life of patients (Organization et al., 2022).

The success of corneal transplantation heavily relies on the accurate and timely management of patient data, including preoperative evaluation, surgical planning, and postoperative care. EHRs have evolved as indispensable means for handling clinical information, providing a centralised repository for patient records, and facilitating communication between healthcare providers (Dobrow et al., 2019; Uslu,

Stausberg, et al., 2021). The appropriate use of EHRs can substantially bolster the positive outcomes of corneal transplant surgeries, culminating in improved patient results and a more seamless and efficient care journey. By minimizing errors and inconsistencies in patient data management, EHRs can further diminish the likelihood of unfavorable events, ultimately improving the success of corneal transplants.

Despite its widespread use, traditional EHRs are prone to lack of standardization and consistency (Dobrow et al., 2019). As a consequence, inconsistency and fragmentation plague the management and documentation of corneal transplantation records, causing difficulties in accessing and sharing information, increased medical errors, and decreased patient safety. Furthermore, the lack of standardization and organization creates an obstacle for healthcare professionals to access complete and accurate information about patients who are due to undergo or have undergone corneal transplantation, resulting in inefficiencies and suboptimal patient care.

To address these issues, there is a growing need for standardised EHRs that can ensure the accuracy and consistency of patient data across healthcare organizations and locations to ultimately improve the management and documentation processes for corneal transplantation records. One promising approach for standardizing EHRs is the use of openEHR structures (Min et al., 2021; Yang et al., 2019b).

Hence, the present case study focuses on the implementation of an openEHR-based solution as a means of standardizing records in the field of corneal transplantation.

Requirements Collection

Initially, as the objective was to develop a case study on the transformation of corneal transplantation records of a particular hospital institution, a direct line of communication with the medical team was deemed indispensable in order to conduct an in-depth inquiry on the necessary requirements.

The requirements collection is a crucial step in the development of healthcare systems and applications, as it lays the foundation to ensure that the needs and expectations of all parties involved are understood and incorporated into the final solution.

Accordingly, the first step of the methodology started with stakeholder identification, including healthcare professionals and IT personnel. Overall, the process of collecting requirements for enhancing the corneal transplantation records with openEHR was a collaborative effort between medical staff, the IT personnel of the hospital's information systems department, and the developers.

After identifying a work group, a series of meetings were organised to initiate the requirements-gathering phase. This stage entailed the active participation of stakeholders, who served as the primary source of knowledge for modelling, guiding the design and development of the openEHR structures. During these collaborative meetings, a thorough analysis of the current documentation processes and data management practices was conducted to identify areas for improvement. The meetings were structured to stimulate an active feedback process from the stakeholders regarding existing workflows, specific needs and preferences, and any identified pain points and suggestions for improvement.

Over the course of these conversations, it became apparent that the information system that the

hospital used for the management of corneal transplantation records was plagued by significant shortcomings, including the possibility of errors in data entry, loss of information, difficulty in sharing data between healthcare providers, and lack of standardization.

Therefore, the main goal was to address these issues and implement a more reliable and comprehensive information management system that utilises openEHR structures to mitigate the problems acknowledged. Through a collaborative effort, the stakeholders were able to delimit the scope and define the requirements and use cases of the clinical domain to be modeled.

The work group identified five key events in the corneal transplantation process where data recording actions could potentially take place. These moments were carefully analyzed to ensure that all necessary data are captured with the highest level of accuracy and consistency. Table 14 provides a description of each key event.

Table 14: Key events identified within the corneal transplantation process

Event	Description
Corneal Transplantation Proposal	To insert the data concerning the corneal transplantation proposal such as type of transplant, laterality, diagnosis, priority, motive, among others.
Contact for Corneal Transplantation	To record the three possible contact attempts for corneal transplantation, including information such as phone number, date of contact, result, reason, and date of next contact.
Schedule anaesthesia Consultation	To enter data regarding the scheduling of the anaesthesia consultation. It contains the requesting service, the date of the appointment, the motive, and additional information.
Perform anaesthesia Consultation	To record the result of the anaesthesia consultation, the executing service, the execution date of the appointment, and observations.
Manage Suspended Proposal	To register the result of the decision regarding suspended corneal transplantation proposals, either to reinstate to the list or to abandon, and the respective reason.

Modelling and Development of openEHR Archetypes, Templates, and Forms

After collecting the requirements, the modelling of the openEHR structures that will support the register and management of corneal transplantation records began, which included archetypes, templates, and forms.

The development process of the openEHR forms in which data can be introduced in some tasks of the corneal transplantation workflow involves transforming archetypes into templates and later into forms. For

representation purposes and to simplify the demonstration of the development process, from now on the description will focus on illustrating the development of the openEHR structures, archetypes, template, and form, related to the corneal transplantation proposal.

As previously stated, the first step involved meetings with domain experts to define the scope of the clinical domain to be modeled and collect data requirements, as specified by the Archetype Modeling Methodology (AMM) (Moner et al., 2018). The main stakeholders involved in this process were healthcare professionals who were familiar with openEHR, the archetype development process, and clinical terminologies.

After determining the scope of the modelling process and identifying the clinical concepts and information elements involved, the second step entailed searching the CKM for existing archetypes that fit the scope of the modelling scenario under consideration. As mentioned in the state of the art, CKM is an openEHR community pillar that enables worldwide governance of domain knowledge artefacts as well as collaborative development, management, and publishing (S. Rubi & L. Gondim, 2019). It is an open-source library of openEHR archetypes and templates that lays the foundation for both semantic and syntactic interoperability (Frade et al., 2022).

Some archetypes were used directly, while others did not fully represent the data elements and had to be adapted through specialization. When no corresponding archetypes existed, new ones were created. The openEHR RM defines four major categories of archetypes: COMPOSITION, SECTION, ENTRY, and CLUSTER. A COMPOSITION is a container class, whereas a SECTION is an organizing class, both of which contain ENTRY objects (de Moraes et al., 2016). The ENTRY class is further specialised into ADMIN_ENTRY, OBSERVATION, EVALUATION, INSTRUCTION, and ACTION subclasses, of which the latter four are kinds of CARE_ENTRY. CLUSTERS are reusable archetypes that can be used within any ENTRY or CLUSTER (Yang et al., 2019b).

For the corneal transplantation proposal, two archetypes were found suitable for use, namely, the service request (Figure 24), which is part of openEHR's INSTRUCTION subclass of the ENTRY class, and the anatomical location (Figure 25), which is part of the openEHR's CLUSTER class.

In addition, a third archetype was created to contemplate some procedure aspects, namely the risk, the complexity, the number of previous surgeries, and the need for anaesthesia consultation. This archetype was named "Eye Surgery Details" and belongs to the CLUSTER class. Figure 26 depicts the archetype's mind map.

At the end of this step, information concerning applicable data constraints, such as datatypes, cardinality, occurrences, and specific data values (e.g., terminologies for coded values, ranges for numerical values, etc.) are stipulated for each archetype. It is worth mentioning the use of SNOMED-CT terminologies for mapping the coded values of the "Laterality" item belonging to the "Anatomical Location" archetype. Local terms were used for the remaining coded text items. Table 15 provides a description of the coded values used for each coded text item.

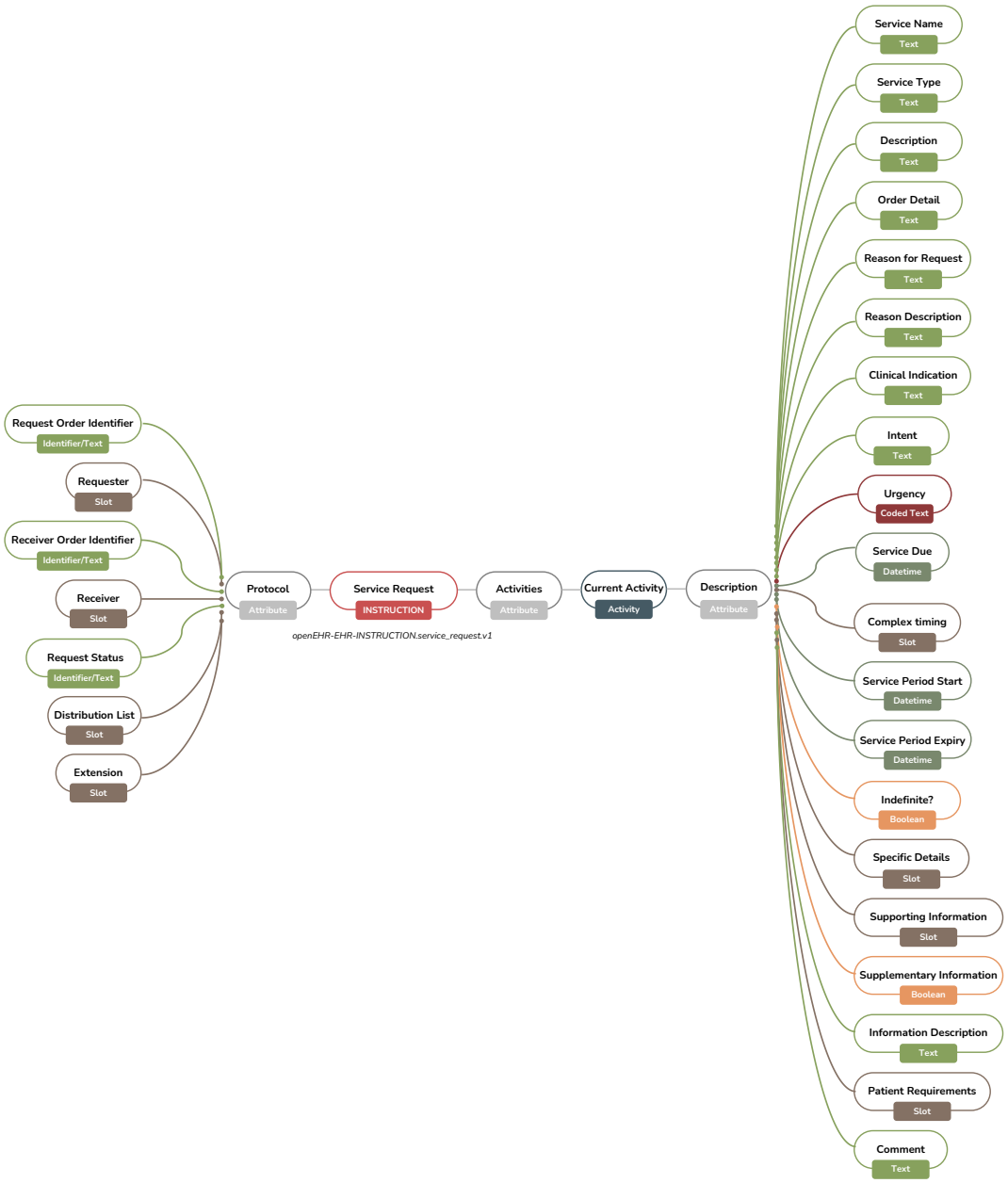


Figure 24: Mind-map view of the "Service Request" archetype.

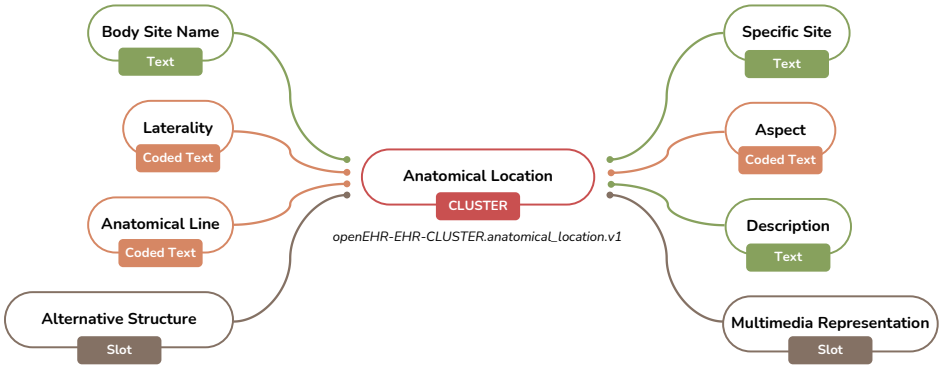


Figure 25: Mind-map view of the "Anatomical Location" archetype.

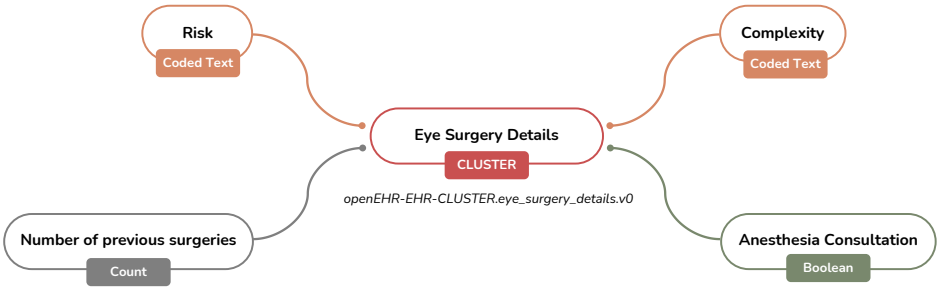


Figure 26: Mind-map view of the "Eye Surgery Details" archetype.

Table 15: Description of the coded values assigned to each coded text item

Archetype	Item	Code	Value
Anatomical Location	Laterality	SNOMED-CT::362503005	Left eye
Anatomical Location	Laterality	SNOMED-CT::362502000	Right eye
Anatomical Location	Laterality	SNOMED-CT::362508001	Both eyes
Eye Surgery Details	Risk	at0004	High
Eye Surgery Details	Risk	at0005	Moderate
Eye Surgery Details	Risk	at0006	Low
Eye Surgery Details	Complexity	at0007	High
Eye Surgery Details	Complexity	at0008	Moderate
Eye Surgery Details	Complexity	at0009	Low
Service Request	Diagnosis	H16319	Corneal abscess
Service Request	Diagnosis	H1830	Corneal membrane alterations
Service Request	Diagnosis	H52219	Irregular astigmatism
Service Request	Diagnosis	H18739	Descemetocele
Service Request	Diagnosis	H18519	Corneal endothelial dystrophy
Service Request	Diagnosis	H18719	Corneal ectasia
Service Request	Diagnosis	H1820	Corneal edema
Service Request	Diagnosis	T868499	Corneal graft (complication)
Service Request	Type of Transplant	08R83KZ_tt_d	Total transplantation (right eye)
Service Request	Type of Transplant	08R83KZ_tt_e	Total transplantation (left eye)

(Table 15 continued on next page)

(Table 15 continued from previous page)

Archetype	Item	Code	Value
Service Request	Type of Transplant	08R83KZ_dalk_d	Anterior Transplantation - DALK (right eye)
Service Request	Type of Transplant	08R83KZ_dalk_e	Anterior Transplantation - DALK (left eye)
Service Request	Type of Transplant	08R83KZ_dsaek_d	Anterior Transplantation – DSAEK (right eye)
Service Request	Type of Transplant	08R83KZ_dsaek_e	Anterior Transplantation – DSAEK (left eye)
Service Request	Type of Transplant	08R83KZ_dmek_d	Anterior Transplantation – DMEK (right eye)
Service Request	Type of Transplant	08R83KZ_dmek_e	Anterior Transplantation – DMEK (left eye)
Service Request	Urgency	at0136	Emergency
Service Request	Urgency	at0137	Urgent
Service Request	Urgency	at0138	Routine

After the discovery and development of the required archetypes, the AD tool was used to assemble and constrain the archetypes into a template that represents the requirements of the corneal transplantation proposal. A template of the COMPOSITION type was created using the "Request for Service" archetype and named "Corneal Transplantation Proposal". To begin, the "Service Request" archetype was incorporated into the content attribute. The template was then modified to remove items that were irrelevant to the clinical context being modeled, such as "Service Type", "Order Detail", and "Intent". Both the "Anatomical Location" and the "Eye Surgery Details" archetypes were imported into the "Specific Details" cluster. The items "Aspect", "Anatomical Line", and "Description" of the "Anatomical Location" archetype were also excluded from the template. Subsequently, some items were assigned specific default values, the content of which can be found in Table 16.

Table 16: Default values assigned to the "Corneal Transplantation Proposal" template

Archetype	Item	Default Value
Service Request	Service Name	Corneal Transplantation
Anatomical Location	Body Site Name	Eye
Anatomical Location	Specific Site	Cornea

To facilitate interpretation and manipulation, the template was exported in the OPT structure and later

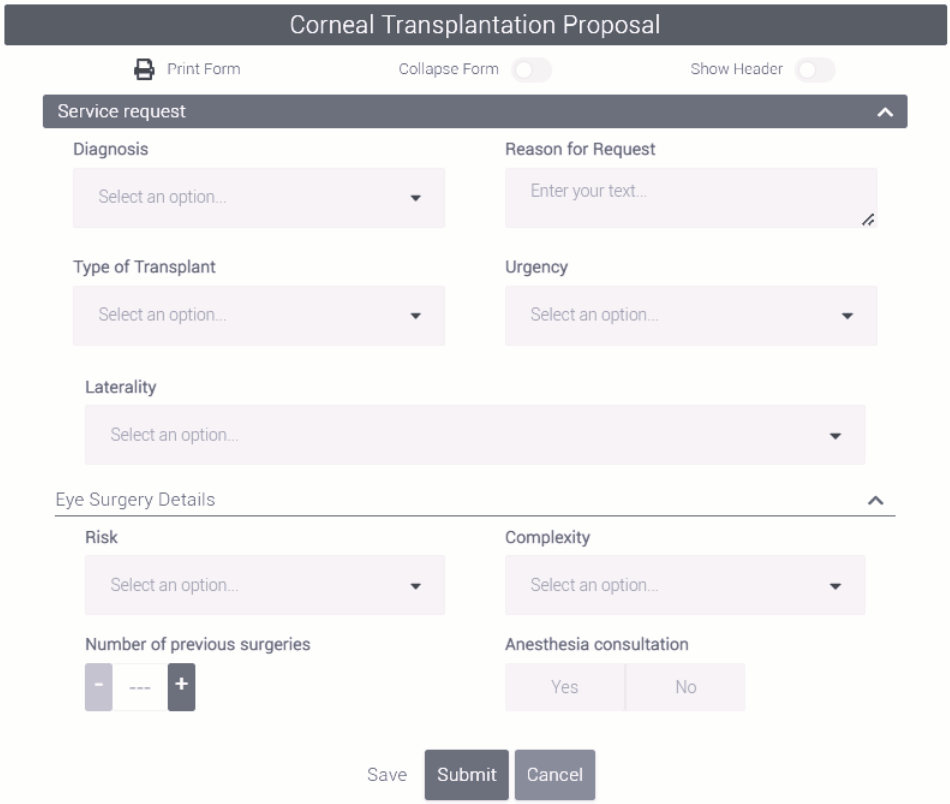


Figure 27: The "Corneal Transplantation Proposal" form generated with the FormBuilder tool.

converted into the JDT structure. Finally, the JDT template was injected into the FormBuilder tool to format the UI form, which can be consulted in Figure 27.

At the end of this stage, five forms were created. Following that, each of these forms was associated to specific tasks in the corneal transplantation workflow, acting as storage schemes for different patient tasks, which will be described with more detail in the next section.

Integration with Decision Rules and Clinical Workflows

The openEHR ecosystem provides a comprehensive solution for managing healthcare data by combining different tools and technologies. In openEHR, forms, DLMs, and WPs interact in a complementary manner to support the provision of effective and efficient clinical care.

The integration of forms with these modules provides data validation, automated actions, and real-time decision support to ensure that the best course of action is taken for each individual patient. Furthermore, it allows the organization and tracking of the progress of clinical form submission tasks to ensure that all necessary clinical tasks are performed in a timely and efficient manner.

In this study, a WP regarding the corneal transplantation was modeled to ensure that the implementation of the solution proceeds smoothly and is completed within the specified timeline. This WP contains five TPlans, specifically:

- Corneal Transplantation Proposal: This task is available to certain ophthalmologists in specific contexts.
- Schedule anaesthesia Consultation: This task is available to administrative staff.
- Perform anaesthesia Consultation: This task is available to anesthesiologists.
- Contact for Transplantation: This task is available to administrative staff.
- Transplant Execution: This task is available to the doctor who submitted the proposal in the first instance.

Figure 28 shows the top-level TPlan for the corneal transplantation WP in a simplified way.

After submitting the form "Corneal Transplantation Proposal", the execution of the anaesthesia consultation relies upon the value assigned to the DV_BOOLEAN field labeled "anaesthesia Consultation" within this form. If this field indicates a true value, the subsequent steps for that patient involve scheduling the anaesthesia consultation, conducting the consultation, and eventually establishing contact for the transplantation procedure.

On the contrary, if the anaesthesia consultation is not necessary, the patient goes to the active waiting list represented by the corneal transplant contact tasks. If the contact is successful, the patient proceeds to perform the transplant, if not, the proposal is suspended and the suspension management task becomes available.

In order to ensure the efficiency of the solution, decision rules were established for forms and TPlans to define the logic and actions that should be taken based on specific data inputs. The rules are stored in a standardised structure that can be applied to patient data at runtime.

According to the requirements gathered in the first stage, two DLMs were created for guiding healthcare professionals in the process of filling in the forms F1 – "Corneal Transplantation Proposal" and F2 – "Contact for Corneal Transplantation". Table 17 describes the association between the forms, DLMs, and rules, including the condition(s) necessary to trigger a certain action.

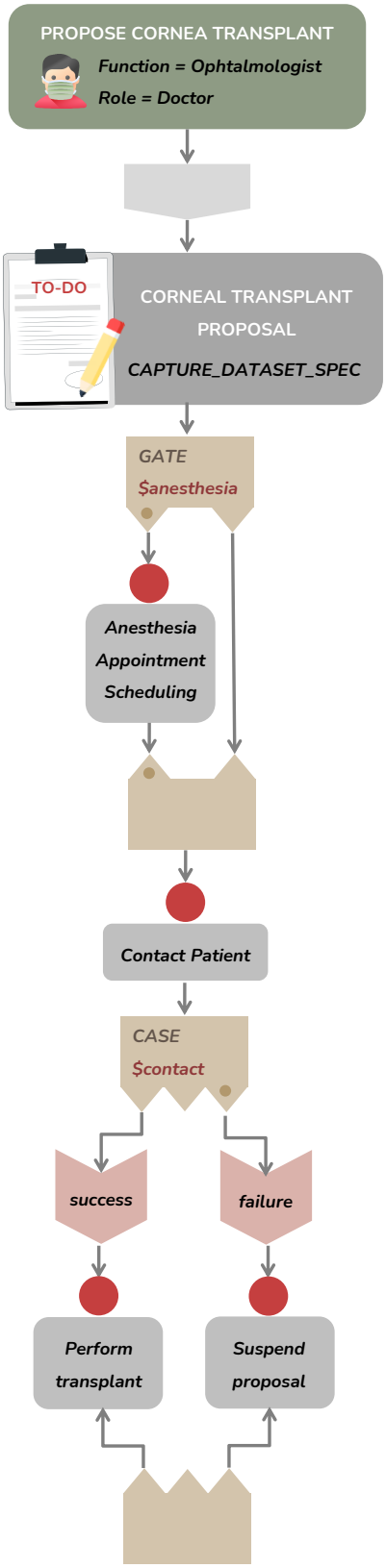


Figure 28: The top-level task plan view of the "Corneal Transplantation Workflow".

Table 17: Association between forms, DLMs and rules in the corneal transplantation process

Form	DLM	Rule Description
F1	DLM1	R1: If the patient's age is equal or superior to 75 years, anaesthesia is mandatory.
Condition: Age greaterThanInclusive 75		Rule: anaesthesia isEqualTo Yes
F2	DLM2	R1: If the patient intends to leave the transplant list, the motive is mandatory.
Condition: Result equal Cancellation		Rule: Motive isMandatory
F2	DLM2	R2: If the patient needs to postpone the contact, the next contact is mandatory.
Condition: Result equal Postpone contact		Rule: Next contact isMandatory

In terms of rules for the WP, a rule was created to decide whether or not the patient should proceed to an anaesthesia appointment. If the priority of the proposal is urgent, the patient does not need an anaesthesia appointment. If it is not urgent, the need for an anaesthesia consultation is decided by the "anaesthesia" field filled in by the physician in the "Corneal Transplantation Proposal" form.

Results and Discussion

In a data-driven world, statistical analysis has become critical to gain insights and draw conclusions that may not be immediately apparent through simple visual inspections. It holds particular significance in the realm of health research, where it can be used to evaluate the effectiveness and efficiency of IT solutions, such as EHRs. Data analytics in the healthcare has the potential to significantly impact the lives and well-being of countless individuals. In this way, it is possible to identify areas for improvement, including the streamlining of workflows, the enhancement of usability, and the resolution of technical glitches, ultimately leading to improved patient outcomes and enhanced quality of care.

Hence, to verify the efficiency and performance of the solutions offered to manage the corneal transplantation process, this section presents an analysis of the data gathered over the period of study from May 1st, 2022 to March 31st, 2023. Prior to presenting the data analysis, a brief overview of the data collection process will be provided.

The data collection process for this case study required careful planning, attention to detail, and adherence to ethical guidelines to ensure the accuracy and validity of the data. First, the databases and tables of interest were selected. Then, it was necessary to verify the anonymization of the data to preserve the privacy and confidentiality of the patients. Given openEHR separation of demographic and clinical data, there was no personal identifying information in the data. After ensuring patient's sensitive data is

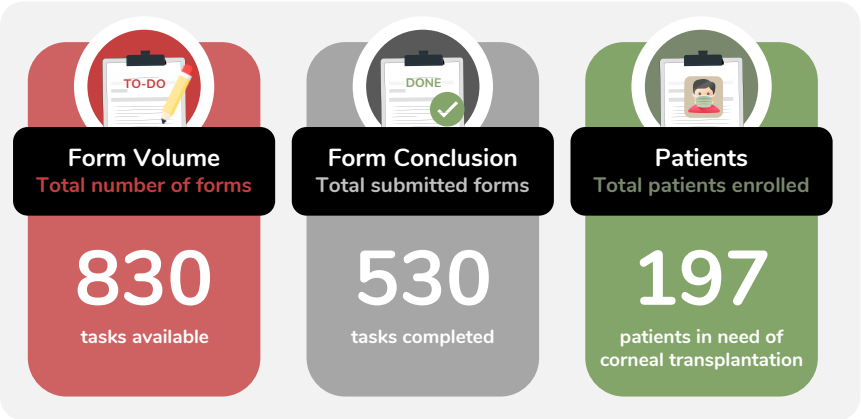


Figure 29: Single number indicators in the corneal transplantation process.

protected, the relevant AQL and SQL queries were developed to extract the data required. Finally, after the data had been extracted, it was meticulously organised into descriptive statistics, in the form of indicators, charts, and graphs, to ease interpretation and help convey the findings.

Figure 29 includes the graphical representation of key indicators in the corneal transplantation process, including form volume, form conclusion and patients enroled. The form volume represents the number of tasks available for healthcare professionals to fill out the corresponding forms, while form conclusion corresponds to the number of tasks successfully submitted by healthcare professionals. On the other hand, patient enrolment indicates the number of patients registered in the corneal transplantation list.

A donut chart was used in Figure 30 to visually depict the proportion of concluded forms in comparison to the number of available forms. This graphical representation provides a clear and concise overview of the overall form completion rate and offers a general picture of how well forms are being managed and completed. The number of available and completed forms can help to gauge the workload of healthcare professionals and assess the capacity of the system to handle the demands of the workflow. A higher number of completed forms relative to the number of available forms indicates that the system is functioning efficiently and effectively. Meanwhile, a lower completion rate could suggest potential bottlenecks or areas for improvement in the solution's design or implementation.

As a result, a 64% completion rate falls short of expectations. This lower rate could be explained by a number of factors. First, the hospital may have limited resources to complete the tasks within the workflow, leading to delays in completing the tasks. Furthermore, the healthcare professionals responsible for completing the tasks within the workflow may not have received enough training or may be inexperienced, which may imply some reluctance in adopting the new technologies.

In addition, since the majority of the available tasks within the studied workflow regard the "Contact for Corneal Transplantation", which represents the active waiting list for corneal transplantation, the workflow completion depends on external factors such as patient health and availability of donor organs. These dependencies directly affect the time required to complete the tasks and, consequently, the total workflow.

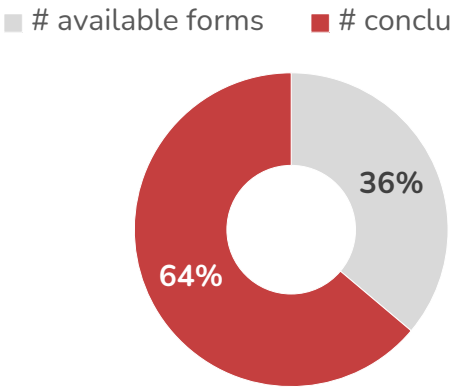


Figure 30: Proportional analysis of form conclusion and form availability within the corneal transplantation workflow.

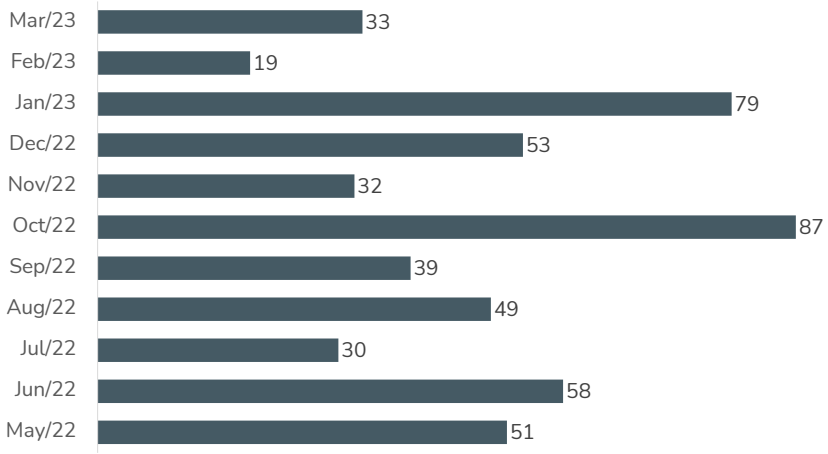


Figure 31: Total number of corneal transplantation forms submitted over time.

Furthermore, a multidisciplinary service team is required, namely the anaesthesia team and the ophthalmology team, which can cause communication issues among the teams and, as a result, delays in task completion.

Moreover, the enrolment of 197 patients is a significant aspect of the data as it serves as a critical contextual element in assessing the workload.

These factors may also explain the fluctuations in the total number of corneal transplantation form submissions over time depicted in Figure 31. This graphical representation allows the visualization of trends and patterns in form submission and provides insights into the volume and frequency of tasks completed.

As a result of the influence of these factors, it was not possible to establish patterns in the data over time, as expected given the non-seasonal nature of the data.

In this sense, to gain a more comprehensive understanding of workload management within a hospital setting, a more specific study on the different types of forms available was required. This analysis enables more detailed scrutiny of the specific challenges and constraints associated with each task type and allows

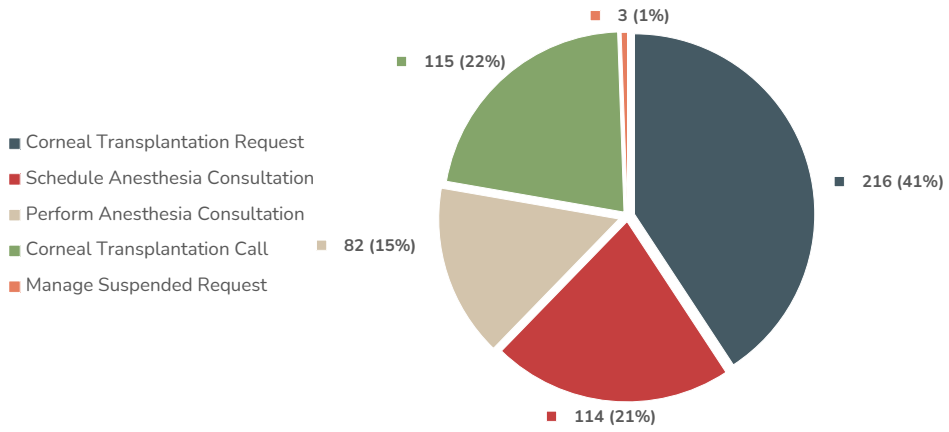


Figure 32: Categorization analysis of submitted forms within the corneal transplantation workflow.

for the identification of more targeted solutions to enhance efficiency and productivity.

Since the corneal transplantation workflow is a complex process involving the completion of a variety of different forms, a pie chart was used in Figure 32 to help visualise the distribution of these tasks across the different form categories. This chart displays both the numerical and percentage distribution of submitted tasks according to each respective form category.

A quick look into Figure 32 reveals a higher rate, 41%, of completed forms associated with the "Corneal Transplantation Proposal" in comparison to the remaining tasks. This rate can be explained by the lack of resources to perform the transplants at a quicker pace. It is also worth noting that the form with the lowest submission rate is associated with the "Management of Suspended Proposal", indicating the fewer cases in which the corneal transplantation proposals were suspended.

Last but not least, to provide a comprehensive visualization of the number of forms submitted over time for each form category, a stacked bar chart was employed in Figure 33. This chart displays the total number of tasks completed and submitted for each form category over the period of time under consideration, helping to assess the relative contributions of each form category to the overall workflow.

In the first months, it is possible to observe that the tasks associated with the "Contact for Corneal Transplantation" increased, whereas the number of "Corneal Transplantation Proposal" decreased. From August of 2022, it can be seen a higher number of submitted tasks pertaining to the "Schedule anaesthesia Consultation". As a result, a month later, it is possible to observe an increase in the number of tasks pertaining to the form "Perform anaesthesia Consultation".

4.2.5 Evaluation

Evaluation is a critical stage in the development process of any IT artefact. The solution must be submitted to rigorous assessment before being made available to its users, being first installed, tested and evaluated in a non-production environment. Therefore, it is crucial to follow a series of tests in order to evaluate the tool in question, namely with respect to the fulfilment of the pre-defined objectives.

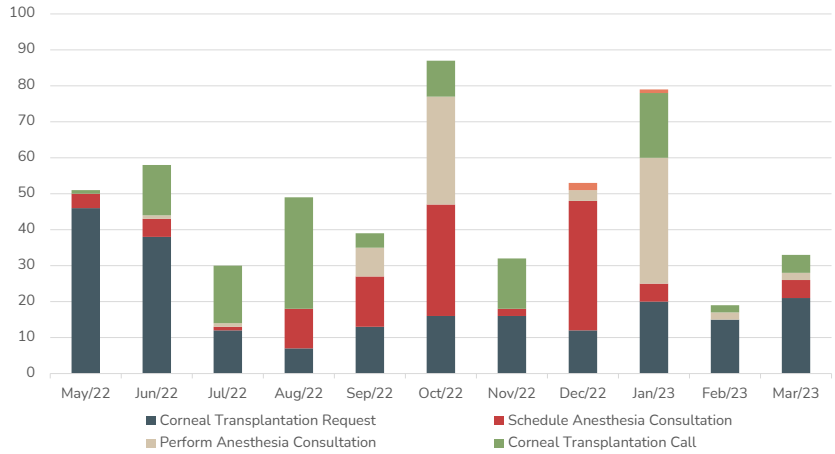


Figure 33: Number of submitted tasks within the corneal transplantation workflow for each form over time.

The evaluation of the artefact was conducted in collaboration with healthcare professionals from the ULSdSA to assess its usability, functionality, and impact on healthcare documentation practices. The evaluation process consisted of three main stages: a usability test, a survey, and a SWOT analysis.

4.2.5.1 Usability Test

The usability test involved healthcare professionals interacting with the artefact to create user-interface forms from openEHR templates in a simulated clinical scenario. The participants in the usability test included doctors and nurses to ensure representation across different clinical roles and perspectives. Initially, it was intended to involve a eight healthcare professionals in the usability test, in order to be able to observe more directly the interactions and gather more detailed feedback. However, due to the constraints regarding the availability of clinical staff with knowledge of openEHR, only three healthcare professionals were able to participate in the usability test. Despite the reduced number of participants, the feedback provided by the participants was valuable and provided insights into the usability and functionality of the artefact.

The test began with an orientation session where participants were provided with an overview of the artefact's functionality. The usability test was conducted in a controlled environment, where participants were asked to perform a series of tasks using the artefact, while their interactions were observed:

- Selecting a specific openEHR template for form creation;
- Configuring the layout and fields of the form to match clinical requirements. This involved hidding, configuring, and rearranging form elements according to a pdf document representing the desired paper form;
- Adding validation rules and logic to the form elements;

- Testing the form by filling in sample data and submitting it.

During the test, participants were encouraged to think aloud and provide feedback on their interactions with the artefact. All these interactions were recorded for later analysis. After completing the tasks, a debriefing interview allowed to provide their overall impressions of the artefact, including its ease of use, usefulness, and any areas for improvement.

Overall the results were positive, with participants expressing satisfaction with the artefact's functionality and usability. The lengthiest part of the test was the configuration of the form layout, as participants needed to familiarise themselves with the interface and the available options. However, once they became accustomed to the interface, they were able to complete the tasks efficiently and accurately. The participants appreciated the real-time preview of the form, which allowed them to visualise the impact of their changes immediately. Although, the participants have found the validation rules and logic features to be useful for ensuring data accuracy and consistency, they suggested some improvements to enhance the usability and flexibility of these features. The most complex part of the test was in fact the addition of refsets and functions to the form elements, as it required a deeper understanding of the openEHR specifications and the artefact's capabilities. However, with guidance and support, the participants were able to successfully complete the tasks and create the desired form. Finally, since the participants were not acquainted with the validation features, consequently, one participant found it difficult to submit the form because he/she was entering data that did not meet the validation rules, which prevented the form from being submitted. However, upon receiving guidance on what was causing the error, the participant was able to correct the data just by reading the validation error message and then successfully submitted the form. This issue highlighted the importance of thorough testing and validation of the form to ensure that it functions as intended and that users are able to complete the form without encountering errors.

Despite the positive feedback, it is important to acknowledge that the usability test had some limitations, including the small number of participants and the controlled environment in which the test was conducted. These limitations may have influenced the results and prevented a comprehensive evaluation of the artefact. Thus, future evaluations should involve a larger and more diverse group of participants, including healthcare professionals from different specialties and backgrounds, to ensure that the artefact meets the needs and expectations of a wider range of users. Moreover, the use of other evaluation methods can complement the usability test results obtained and provide additional insights into the usability and functionality of the artefact.

4.2.5.2 TAM questionnaire

Individual survey questionnaires were distributed to a group of users to gather feedback on their experience with the artefact, including their perceptions of its usability, functionality, and impact on healthcare documentation practices. The survey responses were analyzed to identify common themes and patterns, which were used to evaluate the artefact's effectiveness and inform recommendations for future development. The survey questionnaire was the strategy chosen to collect data from the participants, as it is a

direct, economic, and fast execution method. In addition, this is a practical approach that allows easy distribution through Google Forms.

Since the success of the artefact ultimately hinges on its acceptance and usability by end-users, TAM3 proves invaluable in evaluating the artefact's potential impact on healthcare documentation practices. The inclusion of perceived enjoyment aligns well with the user experience considerations inherent in web application development. Given React's focus on creating engaging and responsive user interfaces, assessing users' perceived enjoyment provides insights into the extent to which the artefact elicits positive emotions and satisfaction during interaction. In addition, subjective norms, for instance, offer insights into the social influences and norms that shape users' attitudes towards the application. Understanding these contextual factors can help identify potential barriers to adoption and inform strategies to foster a supportive and conducive environment. Overall, most of the constructs of the TAM3 model are relevant to the artefact and can provide valuable insights. Hence, the survey questionnaire was designed based on the constructs of the TAM3 model, which includes the original constructs of perceived usefulness and perceived ease of use, as well as additional constructs of perceived enjoyment, perceived trust, and subjective norms.

Thus, the survey is divided into precise sections according to the constructs of the adopted model. It is important to point out that since not all the constructs of the TAM3 model were relevant or could be applied to the artefact, some of them were not included in the questionnaire.

The questionnaire consisted of a series of Likert-scale questions, where participants were asked to rate their agreement with statements related to each construct. The responses were then analyzed to determine the participants' perceptions of the artefact and its potential impact on healthcare documentation practices. The Likert scale was chosen to define the possible options of response, with answers ranging from 1 to 5:

1. Not satisfied / Full disagreement
2. Unsatisfied / Some level of disagreement
3. Satisfied / Indifferent
4. Fairly satisfied / Partially agree
5. Completely satisfied / Strongly agree

The survey questionnaire was distributed to a group of healthcare professionals at the ULSdSA who had experience with openEHR and were familiar with the artefact. The participants were asked to complete the questionnaire anonymously and provide honest and constructive feedback on their experience with the artefact. The structure and content of the questionnaire can be found in Appendix A.

The analysis of the survey responses did not yield conclusive results, as the number of participants was insufficient to derive meaningful conclusions. However, the feedback provided by the participants was generally positive, with many expressing satisfaction with the artefact's usability, functionality, and impact on healthcare documentation practices.

4.2.5.3 SWOT Analysis

A SWOT analysis was conducted to identify the artefact's strengths, weaknesses, opportunities, and threats in the context of healthcare documentation practices and EHR systems. Therefore, it is intended to identify factors that influence the functioning of the openEHR-based artefact, as well as to understand the internal and external environment in which it operates. The SWOT analysis provides a comprehensive overview of the artefact's current state and its potential for future development and integration in healthcare settings. As mentioned in Chapter 3 - Research Methodologies in Section 3.3.2, the SWOT analysis is mainly divided in two parts:

- Internal Factors: It concerns the strengths and weaknesses of the artefact, which are internal to the artefact and can be controlled or influenced by the development team.
- External Factors: It concerns the opportunities and threats that are influenced by the external environment in which the artefact operates and therefore outside the control of the development team.

The SWOT analysis was conducted based on the feedback received from the usability test, survey, and interviews with healthcare professionals, researchers with knowledge on the topic, members of the IT department, and common individuals to simulate as patients. In addition, the observations made during the artefact's development and implementation were also taken into account. Therefore, a qualitative analysis was performed with different data collection and analysis techniques, such as participant observation, focus groups, and semi-structured interviews.

The data collection was carried out in a non-participatory manner, with the researcher observing the participants' interactions with the artefact and taking notes on their behaviour, reactions, and feedback. The focus groups and semi-structured interviews were conducted to gather additional insights into the artefact's usability, functionality, and impact on healthcare documentation practices. The questionnaires were used to collect quantitative data on the artefact's perceived usefulness, ease of use, enjoyment, trust, and subjective norms. The data collected through these methods were then analyzed to identify the artefact's strengths, weaknesses, opportunities, and threats, providing a comprehensive overview of its current state.

Accordingly, the results of the SWOT analysis allowed to identify the following factors:

STRENGTHS:

- The artefact is based on openEHR specifications, which are comprehensive, solid, and increasingly becoming widely adopted in the healthcare industry;
- User-friendly interface that simplifies the creation of UI forms from openEHR templates;
- Real-time preview of form changes, enabling users to visualise the impact of their modifications;

- High degree of customization and flexibility in form layout and field configuration;
- Different data visualization options, such as lists or table views, to facilitate data analysis and interpretation;
- Validation rules and logic features that ensure data accuracy and consistency;
- Integration with the DLM for dynamic form updates based on predefined rules and conditions as users fill in the form;
- Standardization of data entry according to openEHR templates, ensuring data accuracy and compliance with clinical standards;
- No need for programming skills to create user-interface forms from openEHR templates, making it accessible to a wide range of users
- Easy to integrate with other clinical applications, being already integrated in four clinical tools at the ULSdSA;
- Integration with TPlan, allowing the completion of tasks based on the form data;
- Ability to incorporate AI and ML technologies to enhance data analysis and decision-making capabilities;
- Easy accommodation of different clinical scenarios and healthcare settings, making it versatile and adaptable to various use cases;
- Fast to deploy and easy to scale, enabling rapid implementation in different healthcare settings.

WEAKNESSES:

- Limited support for validation rules and logic, with room for improvement;
- The artefact may be overwhelming for users unfamiliar with openEHR specifications. Although it does not require programming skills, a deeper understanding of openEHR specifications is needed;
- No integration with mobile devices, limiting its accessibility and usability on ios and Android platforms;
- There is no built-in support for different languages, limiting its usability in multilingual healthcare settings.

OPPORTUNITIES:

- Need for lifelong EHRs to support patient care throughout their entire care journey across different healthcare settings and providers;
- Growing demand for standardised EHRs to improve data accuracy, consistency, and interoperability;

- Expansion of openEHR adoption and implementation in healthcare organizations worldwide, particularly in the EU and UK;
- Rise of the number of guides, tutorials, and training resources on openEHR specifications;
- Strong openEHR community support and collaboration for developing and sharing archetypes, templates, and tools;
- Ability to integrate with ML and AI technologies to enhance data analysis and decision-making capabilities;
- Pressure for government regulations and incentives to promote openEHR adoption and implementation in healthcare systems.

THREATS:

- Absence of regulatory guidelines for openEHR forms implementation;
- Resistance to change among healthcare professionals, hindering the adoption and implementation of the artefact;
- Lack of awareness and understanding of openEHR specifications, impeding the artefact's full potential;
- Competition from other EHR systems and software solutions;
- No government support, funding, or incentives for openEHR adoption and implementation.

4.2.6 Results and Discussion

The openEHR based solution detailed in this section was developed and implemented at the ULSdSA to address the challenges associated with healthcare documentation practices and lack of interoperability and standardization within EHR systems. The artefact was designed to be user-friendly, intuitive, and adaptable to the clinical needs of healthcare professionals, facilitating the creation of UI forms from openEHR templates tailored to their specific requirements. It enables users to configure form layouts, add validation rules and logic, and customise form elements to capture, store, and exchange structured health data in a standardised format, ensuring data accuracy and consistency across different healthcare settings. By enhancing the quality and interoperability of health data, the proposed artefact seeks to empower clinicians to make informed decisions, improve patient outcomes, and ensure the delivery of high-quality, ethical healthcare services.

The case study approach provided a nuanced examination of the artefact's deployment and operationalization across different healthcare settings, each with unique data requirements and clinical demands. Notably, the artefact showcased its versatility and adaptability by seamlessly integrating into the infrastructure of the ULSdSA. Moreover, its integration with other openEHR artefacts, such as the DLM and TPlan, further enhances its interoperability and functionality. The scaling of the artefact across multiple

clinical tools at the ULSdSA without the need for cumbersome manual processes and technical adjustments further demonstrated its versatility and adaptability to different healthcare settings, showcasing its potential to streamline healthcare documentation practices and improve data management capabilities. The demonstration yielded promising outcomes, highlighting the potential of the artefact to improve the quality of care and patient outcomes in healthcare.

After the demonstration, the artefact was evaluated through a usability test, a survey, and a SWOT analysis to assess its effectiveness and impact on healthcare documentation practices. The results of the evaluation process provided valuable insights into the usability, functionality, and significance of the artefact on healthcare.

Most healthcare professionals who contacted with the artefact were enthusiastic about its potential to streamline the creation of user-interface forms from openEHR templates, making it easier to manage patient data and improve the quality of care. The artefact's user-friendly interface, real-time preview of form changes, and high degree of customization and flexibility were particularly well-received by participants, who found these features to be beneficial in creating forms that meet their clinical needs. The validation rules and logic features were also appreciated for ensuring data accuracy and consistency, although some participants suggested improvements to enhance their creation.

Some ideas are already being discussed for future development, such as the refinement of logic functions available in the artefact, the possibility of enabling users to create costum functions, and the governance and categorization of these functions in the artefact's interface. In addition, the integration with mobile devices and the support for different languages are also being considered for future development. As such, from all the weaknesses identified, the one left to be addressed is related to the artefact being overwhelming for users unfamiliar with openEHR specifications. This issue can be mitigated by providing users with more training and resources on openEHR specifications, which is already mentioned in the opportunities identified in the SWOT analysis. In fact, given the expansion of openEHR adoption and implementation in healthcare organizations worldwide, there is a growing concern with the matter of training clinicians to be acquainted with openEHR modelling process and specifications.

Therefore, the artefact can play a key role in the near future, where healthcare professionals will be able to create their own forms without the need for programming skills, which is a significant advantage in the context of healthcare settings, bringing the power of data management to the hands of the users themselves.

Furthermore, as emphasised earlier in this thesis, addressing the problems identified in the data layer, such as the lack of standardization and consistency, is imperative to uphold accuracy and ethical integrity in healthcare decision-making. Minimizing the occurrence of missing values, structuring fields with appropriate clinical terminologies, and reducing the predominance of free-text fields are essential steps in this process. By doing so, it is possible to glean real semantic insights from the data, fostering a deeper understanding of patient conditions and treatment options. Ultimately, rectifying data layer deficiencies lays the groundwork for informed and ethical decision-making, underpinning the delivery of high-quality healthcare services.

4.3 From openEHR Data to Ethical Decision-Making in the ICU using Artificial Intelligence

4.3.1 Introduction and Problem Contextualization

ICU is a critical environment where patients with severe illnesses or injuries often teeter on the brink of life and death, requiring vigilant monitoring and timely interventions. Within the confines of the ICU, patients often present severe and rapidly evolving conditions. Every decision made within this environment can have far-reaching consequences, not only for the patients themselves but also for their loved ones and the healthcare professionals entrusted with their care. Given its intricacies, the ICU represents a pivotal locus of decision-making, where healthcare professionals are pressured to make rapid and accurate decisions that are determinant for patient outcomes.

Nonethless, the abundance of data being generated on a daily basis in the ICU coupled with the heterogeneity and lack of standardization in data storage poses significant challenges for healthcare professionals, which are often exacerbated by the complexity and unpredictability of patient conditions. Moreover, ethical considerations regarding patient privacy, autonomy, and beneficence further complicate decision-making processes in ICUs.

The lack of implementation of clinical standards and the vendor lock-in of proprietary EHR systems hinder the aggregation of patient data from different sources, which is essential to provide a comprehensive view of the patient's health status and facilitate informed decision-making. The absence of interoperability and data standardization in ICUs not only impedes the exchange of information between healthcare professionals but also limits the potential for leveraging advanced technologies, such as AI, to support clinical decision-making processes.

Hence, the complexity and urgency of patient care in ICUs necessitate the integration of advanced technologies and data-driven approaches to support effective information processing and interpretation. Amidst the intricacy of ICU workflows, the adoption of openEHR standards has emerged as a promising solution for mitigating data fragmentation and enhancing interoperability, through the standardization and structuring of clinical data, ultimately, leading to the delivery of more efficient and comprehensive care to patients.

Despite the strides made in leveraging openEHR data within the ICU, there persists a critical gap between data availability and its translation into actionable insights. Therefore, there is a critical need to explore how openEHR data can be effectively used to support decision-making using AI, especially in the context of the ICU, while ensuring the protection of patient rights. This chasm underscores the need for innovative approaches that harness the potential of AI to navigate the structured clinical data and distill meaningful insights.

This problem contextualization sets the stage for investigating the challenges inherent in the intersection of openEHR data, AI algorithms, and ICU decision-making, with a focus on their synergies, opportunities, and implications. By exploring the transformative potential of openEHR structured data in the

performance of AI algorithms, it is possible to uncover opportunities to enhance the accuracy, reliability, and timeliness of predictions in the ICU. Through a multifaceted examination, this section endeavors to shed light on the challenges, opportunities, and imperatives inherent in leveraging AI to harness the full potential of openEHR data for clinical decision-making in the ICU setting.

4.3.2 Research Purposes

The problems identified in the previous section underscore the need for innovative solutions that bridge the gap between openEHR data and its translation into actionable insights for clinical decision-making, particularly within the intricacies of the ICU environment. To this end, the research purposes of this contribution are as follows:

- **AI-Driven Clinical Decision-Making:** To explore the potential of AI-driven algorithms in leveraging openEHR data to enhance clinical decision-making in the ICU;
- **Generability and Demonstrability:** Assess the generability and demonstrability of ML models trained on public data for predicting patient LOS;
- **Performance Evaluation:** Identify the strengths and limitations of different ML algorithms in predicting patient LOS in ICUs, considering factors such as accuracy, precision, and recall.
- **Feasibility and Effectiveness:** Investigate the feasibility and effectiveness of leveraging openEHR-based solutions for predicting COVID-19 diagnosis based on clinical symptomatology data.
- **Performance and Robustness:** Evaluate the performance and robustness of ML models trained on openEHR artefacts in predicting COVID-19 diagnosis, considering metrics such as accuracy, sensitivity, and specificity.
- **Bias Examination:** Examine the potential biases and inaccuracies introduced by relying on self-reported data, particularly regarding clinical measurements such as body temperature;
- **Speed/Accuracy Trade-Off:** Explore the trade-off between speed and accuracy in developing predictive models for COVID-19 diagnosis during the pandemic, considering the implications for data quality and completeness;
- **Ethical Implications:** To provide recommendations for the ethical deployment of AI algorithms in the ICU based on the findings of the research.
- **Interoperability Enhancement:** Investigate the role of openEHR in improving the interoperability and utility of healthcare data for predictive analytics and clinical decision support, focusing on separating demographic data from clinical data and using standardised terminologies and coding systems.

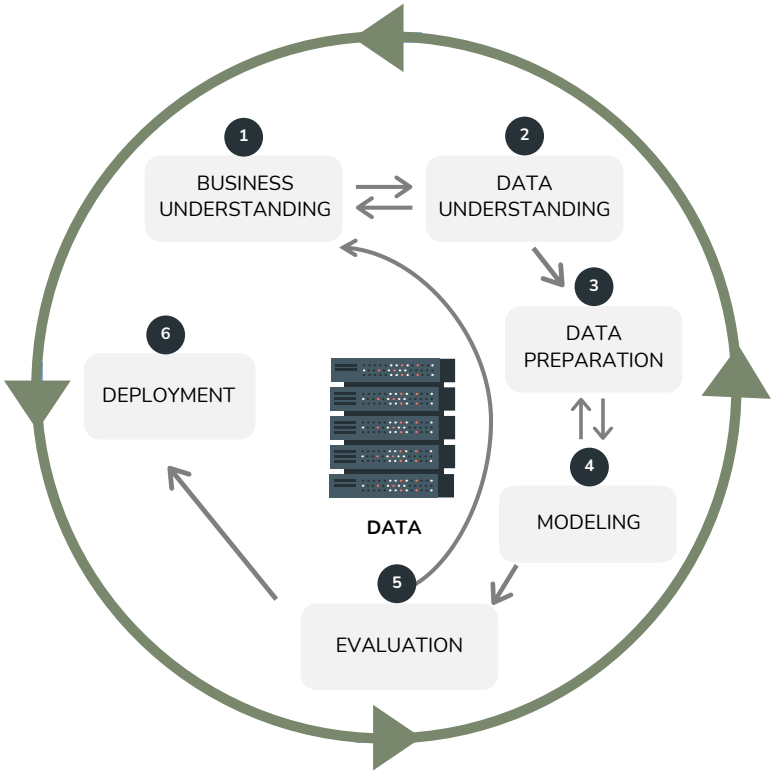


Figure 34: CRISP-DM Methodology

- **Recommendations:** Provide recommendations for the ethical deployment of AI algorithms in the ICU based on the findings of the research.

4.3.3 Design and Development

The design and development of the contribution based on AI-driven algorithms was guided by the CRISP-DM methodology, which provides a structured framework for the development of data-driven solutions. CRISP-DM is a popular framework used worldwide for easy project replication, planning, and management. It is a cyclical process comprised of six stages: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment (Martins et al., 2021). Figure 34 illustrates the CRISP-DM methodology and its sequential stages.

The CRISP-DM methodology was chosen for its flexibility and adaptability to different project requirements, allowing for iterative development and refinement (Ferreira et al., 2020). The following sections provide an overview of the CRISP-DM stages and their implementation within the context of the thesis.

As mentioned in the previous section, given the constraints imposed by the COVID-19 pandemic, the development of AI algorithms on top of openEHR data referring to the gIsICU context was compromised, given the lack of availability of these units and clinicians to provide the necessary contributions. Therefore, the development of the AI in this thesis focused on the development of predictive models based on two

case studies:

Case Study I: Predicting Patient Length of Stay in Intensive Care Units (Neto et al., 2021)

Case Study II: Prediction of COVID-19 diagnosis based on openEHR artefacts (Oliveira et al., 2022)

The development of these predictive models was based on the CRISP-DM methodology, which guided the project from the initial stages of business understanding and data preparation to the final stages of model evaluation and deployment. The following sections provide a detailed overview of each case study, including the business understanding, data preparation, modelling, evaluation, and results. The deployment stage was not included in this thesis due to the need for further validation and refinement of the predictive models before they can be deployed in a clinical setting.

4.3.3.1 Case Study I: Predicting Patient Length of Stay in Intensive Care Units

Business Understanding

The CRISP-DM methodology begins with the Business Understanding phase, which aims to acquire a comprehensive understanding of the business objectives, requirements, and constraints surrounding the prediction task. In the context of the ICU, predicting patient LOS is a critical task that can inform resource allocation, bed management, and treatment planning. The LOS refers to the duration a patient spends within the hospital, encompassing the time from admission to discharge. During the COVID-19 pandemic, the pressure on hospital resources has intensified, with surges in patient admissions straining capacities to their limits. In such circumstances, accurately predicting and managing LOS becomes indispensable for optimising patient flow, reduce overcrowding, and enhance the quality of care delivered in the ICU. By minimizing unnecessary delays in patient discharge and streamlining care pathways, hospitals can alleviate the burden on their resources, ensuring that critical beds are available for patients who require them most urgently.

Moreover, effective LOS management enhances patient flow within the hospital, fostering smoother transitions between care settings and reducing bottlenecks in the healthcare delivery process. This not only enhances the overall efficiency of healthcare operations but also contributes to better patient outcomes by minimizing the risk of exposure to nosocomial infections and facilitating timely access to appropriate levels of care.

Ideally, the engagement of stakeholders would be meaningful to gather insights into the key factors such as clinical variables, treatment protocols, and institutional factors that influence patient LOS in the ICU. However, the focus on containing the spread of COVID-19 pandemic and the security restrictions imposed on healthcare facilities limited the availability of stakeholders for direct engagement. As a result, the business understanding phase was informed by the existing literature and clinical guidelines.

The focus of this case study resides on the challenge of scheduling hospital admissions, with a particular emphasis on ensuring an adequate allocation of beds for patients requiring critical care while concurrently reducing the substantial financial burdens incurred by the institution due to inefficient resource management. The LOS in hospitals is a critical metric, particularly amid the new coronavirus pandemic as healthcare facilities struggle with strained resources, notably in terms of bed availability. Prolonged stays in the ICU tie up beds and other essential resources, hindering the facility's ability to admit new patients promptly. On the contrary, shorter LOS facilitates a more efficient turnover of beds, allowing hospitals to accommodate a larger volume of patients within the same infrastructure. With hospitals grappling with limited capacities to accommodate patients, optimizing LOS becomes paramount for effective planning and resource management.

In order to provide ICUs with valuable informational resources, this case study will include several facets, such as the identification of the most strongly correlated features to the LOS, the investigation of the optimal parameter values for the given ML algorithms, and the prediction of LOS through the construction of different DMM. Afterwards, a comprehensive one-to-one comparative analysis will be conducted for each DMM, considering both optimised and non-optimised parameters as well as strongly correlated features. The study will entail the comparison of different techniques in terms of their performance across various selected metrics. The objective is therefore to discern the most suitable approach to address the specific challenges inherent in the problem domain and the dataset under examination.

Data Understanding

During the Data Understanding stage, the focus shifts to acquiring, exploring, and familiarizing with the dataset to gain insights into the underlying patterns, relationships, and characteristics of the collected data. A better perception of the existing data is crucial for the development of accurate predictive models.

The dataset used in this case study comprises approximately 60,000 records from a clinical computer system utilised between 2001 and 2008 at the Beth Israel Deaconess Medical Center, a medical care unit located in Boston. The attributes stored in the dataset are essential for determining the LOS in the ICU, encompassing patient details, reasons for admission, discharge details, laboratory test results, and other medical features. The dataset is characterised by a mix of numerical and categorical attributes, with the target variable being the LOS in days. Each record is described by a total of 28 attributes, as presented in Table 18.

Table 18: Characterization of the dataset attributes for ICU length of stay prediction

Attribute	Description	Attribute	Description
hadm_id	Hospital admission identifier	NumOutput	Number of outputs performed

(Table 18 continued on next page)

(Table 18 continued from previous page)

Attribute	Description	Attribute	Description
gender	Patient's gender	NumCPTevents	Number of performed current procedural terminology (CPT) events
age	Patient's age (in years)	NumInput	Number of performed inputs
insurance	Patient's insurance type	NumLabs	Number of performed laboratory tests
religion	Patient's religion	NumMicroLabs	Number of performed microbiology tests
marital_status	Status Patient's Marital status	NumNotes	Number of written notes and summaries
ethnicity	Patient's ethnicity	NumRx	Number of performed x-rays
admit_type	Admission type	NumProcEvents	Number of performed procedure events
admit_location	Admission location	NumTransfers	Number of performed transfers
AdmitDiagnosis	Diagnosis Admission Diagnosis	NumChartEvents	Number of performed chart events
NumDiagnosis	Number of diagnoses entered to the system	ExpiredHospital	Indicates if the patient has died or not
NumCallouts	Number of performed callouts	TotalNumInteract	Total number of performed interactions
NumProcs	Number of performed procedures	LOSdays	Length of Stay in Intensive Care Unit
AdmitProcedure	Performed procedure on admission	LOSgroupNum	Length of Stay in Intensive Care Unit[1]

[1] 0: LOS 0-4 days; 1: LOS 4-8 days; 2: LOS 8-12 days; 3: LOS > 12 days}

A detailed examination of Table 18 reveals the resemblance between the attributes *LOSdays* and *LOSgroupNum* While the former represents the number of days each patient stayed in the ICU, the latter discretises the *LOSdays* attribute into four distinct bins, delineating a four-day range for each category, with the exception of the final label (as elucidated in the table footnote). Given the evident interdependence between these attributes, deliberation on which feature to retain will be undertaken in the next stage, aiming to prevent any inconsistencies in the models.

Furthermore, it was noticed that all numerical features required additional processing to address inconsistencies manifested in high standard deviation values relative to the features' mean values. The

data analysis carried out aims to ensure data consistency and reliability to fortify the integrity of subsequent analyses and model developments.

Data Preparation

The Data Preparation stage is a critical phase in the CRISP-DM methodology, involving the preprocessing of the dataset to ensure its quality, consistency, and readiness for model development.

The first step in the data preparation was to ensure that the raw data did not contain any duplicate, inconsistent, or incomplete information. A verification process was conducted to identify and remove any duplicate records, ensuring that each patient's data was unique and accurate. In addition, the dataset was examined for missing values and inconsistencies that could potentially impact the performance of the predictive models.

It was essential to verify the presence of any personal information that could compromise patient privacy and confidentiality. Hence, all nominal personal information was transformed into numerical features with the purpose of allowing the use of ML algorithms like Neural Network, which only run over numerical features.

The next step involved the detection and removal of outliers from the dataset, as outliers can significantly impact the performance of predictive models. Outliers were identified using the Interquartile Range (IQR) method, a simple algorithm that calculates the range between the first and third quartiles of the data and identifies any values that fall outside this range. Equation 4.1 illustrates the calculation of the IQR value, where Q_1 and Q_3 represent the first and third quartiles, respectively.

$$IQR = Q_3 - Q_1 \tag{4.1}$$

After calculating the IQR values, each value was then multiplied by a constant known as the Interquartile Range Multiplier (IQRM), which is typically set to 1.5 to identify outliers. Equation 4.2 illustrates the calculation of the constant IQRM, where C represents the constant value.

$$C = IQR * 1.5 \tag{4.2}$$

Lastly, the constant value C was then subtracted from Q_1 and added to Q_3 to establish a range of non-outlier records. Any record located outside this range was labeled as an outlier, as shown in Equations 4.3 and 4.4, where x represents the value of the record and Q_1 and Q_3 denote the first and third quartiles, respectively. The records identified as outliers were subsequently removed from the dataset to ensure the integrity and reliability of the data for model development.

$$x < Q_1 - C \rightarrow Outlier \tag{4.3}$$

$$x > Q_3 + C \rightarrow Outlier \tag{4.4}$$

After detecting and removing all outliers, a discretization process was then applied to the continuous features of the dataset to enhance the performance of the predictive models during the modelling phase. Discretization is a technique that transforms continuous features into categorical features by dividing the data into distinct intervals or bins. This process facilitates the identification of patterns and relationships within the data, enabling the models to make more accurate predictions. In this case study, the *age* feature was discretised into ten-year intervals to simplify the data and enhance the interpretability of the models.

The generation of new clinical features was another crucial step in the data preparation process, aimed at enriching the dataset with additional information that could enhance the predictive power of the models. Two new features were created based on the number of procedures and diagnoses performed on each patient per day. The *NumProcsPerDay* feature was calculated by dividing the total number of procedures performed on a patient by the length of stay in days, as shown in Equation 4.5. Similarly, the *NumDiagsPerDay* feature was calculated by dividing the total number of diagnoses made on a patient by the length of stay in days, as illustrated in Equation 4.6. These new features were added to the dataset to provide additional insights into the patient's condition and treatment trajectory.

$$NumProcsPerDay = \frac{NumProcs}{LOSdays} \tag{4.5}$$

$$NumDiagsPerDay = \frac{NumDiagnosis}{LOSdays} \tag{4.6}$$

Afterwards, an overview of each feature was conducted to understand its relevance and impact on the predictive models. During this analysis, it was found that *ExpiredHospital*, a binary feature that indicates if a patient died or not in the ICU, only presents records of patients who have not died (value = 0). Accordingly, the approach was to remove this feature due to its lack of ability to influence the DMMs.

On the other hand, the aforementioned dependency between the *LOSdays* and *LOSgroupNum* attributes was also addressed. The strong correlation between these two features could potentially introduce bias into the predictive models and, in worst-case scenarios, return corrupted and invalid results, as the use of one feature to forecast another could lead to overfitting. Hence, the decision was made to remove the *LOSdays* attribute from the dataset to prevent complications. The *hadm_id* feature was also removed as it serves as an identifier and does not contribute to the predictive power of the DMMs.

The final step in the data preparation phase involved normalizing the features to ensure that all attributes were on the same scale. Normalization is a crucial preprocessing step that standardises the range of values for each feature, preventing certain features from dominating the model due to their larger scale. In this case study, the Min-Max normalization technique was applied to scale all features between 0 and 1, as shown in Equation 4.7, where x represents the original value of the feature, x_{min} denotes the minimum value of the feature, and x_{max} represents the maximum value of the feature.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4.7}$$

Modelling

The modelling phase of the CRISP-DM methodology involves the selection, construction, and evaluation of predictive models based on the preprocessed dataset.

As expressed in Equation 4.8, each DMM can be described as belonging to:

- an Approach (A);
- being composed by a Scenario (S);
- a Missing Value Approach (MVA);
- a DMT;
- a Sampling Method (SM);
- a Data Approach (DA);
- and a Target (T).

$$DMM = \{A, S, MVA, DMT, SM, DA, T\} \tag{4.8}$$

In this case study, there was only one T, which was the *LOSgroupNum* variable. Since Classification was the chosen A, seven different classifiers were selected to be used as DMTs, namely Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB), Gradient Boosted Tree (GBT), Artificial Neural Network (ANN), k Nearest Neighbours (kNN), and DL. It is important to mention that the DL algorithm is a Rapidminer operator implementation that is based on a multi-layer feed-forward ANN trained with stochastic gradient descent using back-propagation.

Each algorithm relies on a different set of parameters that can be adjusted to enhance the model's performance. Hence, different parameter values determine different results, and the selection of the optimal parameters is crucial for the model's success. Therefore, a model tuning was performed before the model's execution, using the RapidMiner *Optimise Parameters* operator.

Afterwards, a feature correlation analysis was conducted to identify the most relevant features for the prediction task. The correlation analysis was performed using the *Weight by Correlation* operator, which calculates the correlation weight between each feature and the target attribute. The higher the weight of an attribute, the more relevant it is considered. The weighting scheme is based upon correlation and it returns the absolute or squared value of correlation as attribute weight. There were two features that presented a strong negative correlation with the target attribute, *LOSgroupNum*, specifically *NumTransfers* and *NumDiagsPerDay*. These features were evaluated with values of -0.71 and -0.69, respectively.

According to the results of the correlation analysis and the parameter optimization, the DMMs were constructed using a total of four scenarios:

- S1 - Without correlated attributes and the default parameters;
- S2 - With correlated attributes and the default parameters;
- S3 - Without correlated attributes and the optimised parameters;
- S4 - With correlated attributes and the optimised parameters;

Cross Validation with 10 folds was the sampling method chosen. Since the dataset presented a balanced distribution of the target attribute, there was no need to perform any sampling technique. Similarly, the absence of missing values eliminated the need for any data imputation method.

As a result, the following DMMs were defined:

- A = {Classification}
- S = {S1, S2, S3, S4}
- MVA = {N/A}
- DMT = {DT, RF, NB, GBT, kNN, ANN, DL}
- SM = {Cross Validation (10 folds)}
- DA = {N/A}
- T = {LOSgroupNum}

In the end, a total of 28 DMMs were created according to Equation 4.9.

$$DMM = 1(A) \times 4(S) \times 1(MVA) \times 7(DMT) \times 1(SM) \times 1(DA) \times 1(T) \tag{4.9}$$

Evaluation

The quality assessment of the DMMs was conducted using several statistical metrics to provide a comprehensive evaluation of their performance. The confusion matrix is a valuable tool for evaluating classification models, as it provides a detailed breakdown of the model's predictions. The confusion matrix consists of four key metrics:

- True Positive (TP) - The number of positive instances correctly classified by the model;
- True Negative (TN) - The number of negative instances correctly classified by the model;
- False Positive (FP) - The number of negative instances incorrectly classified as positive by the model;
- False Negative (FN) - The number of positive instances incorrectly classified as negative by the model.

As the target attribute *LOSgroupNum* encompasses four discrete potential outcomes, the diagonal of the confusion matrix encompasses all the TP entries corresponding to accurately forecasted instances for each outcome. In turn, the rows and columns associated with each predicted outcome denote the FN and FP entries, respectively, signifying instances of erroneous predictions. In this study, particular attention is drawn to the FN and FP entries, as they represent instances where the model incorrectly forecasts the target feature. These misclassifications are of significant concern, as they can potentially lead to erroneous clinical decisions or interventions. Hence, minimizing these errors is crucial to ensure the model's reliability. The remaining values in the confusion matrix pertain to TN records, indicating instances correctly classified as having a different length of stay. These entries reflect instances where the model successfully identifies cases not belonging to the predicted class. In summary, the confusion matrix provides a comprehensive breakdown of the model's predictive performance, highlighting areas of both success and concern in accurately predicting the target feature across multiple classes.

A variety of performance metrics can be derived from the confusion matrix. This case study will use accuracy, precision, recall, kappa coefficient, and Root Mean Squared Error (RMSE) to assess the predictive power of the models.

The description of each of these measures, as well as the calculation behind them, is detailed below.

- **Accuracy:** This indicator calculates the ratio between the instances correctly classified by the predictive model and all classified instances for the correctly TP classified instances, which answers the question:

  How many patients' LOS were accurately classified out of the total?

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{4.10}$$

- **Precision:** This parameter measures the proportion of positive occurrences correctly classified by the model to the total number of positive instances, which answers the question:

  How many of patients who were classified with a certain LOS category actually had a LOS that matched that timing?

$$Precision = \frac{TP}{TP + FP} \tag{4.11}$$

- **Sensitivity/Recall:** This metric is considered an integrator indicator and measures the ratio of positive instances correctly classified by the model to the total positive instances, which answers the question:

  How many patients' LOS were successfully predicted out of all of them?

$$Sensitivity = \frac{TP}{TP + FN} \tag{4.12}$$

Stepping away from the evaluation metrics derived from the confusion matrix, the kappa coefficient also known as Cohen's kappa was selected given its adequacy for multi-class and imbalanced problems or when measures like accuracy, precision, or recall are not able to provide a complete performance overview of a classifier. It takes into account the possibility of agreement occurring by chance and provides a more robust measure of agreement than simple accuracy. Its values vary between 0 and 1, where 0 indicates that the classifier's predictions are no better than random, and 1 signifies perfect predictions. The kappa coefficient is calculated as follows:

$$Kappa = \frac{p_o - p_e}{1 - p_e} \tag{4.13}$$

where p_o represents the observed agreement and p_e denotes the expected agreement. The observed agreement is the proportion of instances correctly classified by the model, while the expected agreement is the probability of chance agreement occurring. The kappa coefficient provides a more nuanced evaluation of the model's performance, accounting for the possibility of random agreement and offering a more comprehensive assessment of the classifier's predictive power.

Results

The evaluation of the predictive capabilities of the DMMs was conducted using the performance metrics described in the previous section. The results of the evaluation provide insights into the predictive power of the models and their ability to accurately forecast patient LOS in the ICU across four distinct scenarios. The performance metrics were calculated for each DMM and compared to determine the most effective model for predicting patient LOS.

The results of the evaluation are presented in Table 19, which displays the performance metrics for each DMM across the different classifiers and scenarios.

Table 19: Algorithm's performance evaluation in the ICU LOS case study

Metric	Algorithm	S1	S2	S3	S4
Accuracy (%)	NB	62.44	67.39	69.10	78.69
	kNN	55.50	59.15	57.08	61.54
	ANN	91.46	93.59	91.46	95.09
	DL	90.10	95.60	92.45	96.87
	DT	70.94	72.20	87.90	96.82
	RF	67.42	75.07	90.90	94.42
	GBT	92.07	96.85	97.47	99.19

(Table 19 continued on next page)

(Table 19 continued from previous page)

Metric	Algorithm	S1	S2	S3	S4
Precision (%)	NB	62.65	68.34	82.01	85.22
	kNN	57.30	61.31	61.58	65.98
	ANN	91.57	93.75	91.68	95.06
	DL	90.42	95.85	92.31	96.60
	DT	75.37	76.38	87.37	96.62
	RF	74.68	81.18	91.50	94.80
	GBT	92.69	96.92	97.53	99.25
Recall (%)	NB	61.23	67.30	56.52	70.54
	kNN	51.87	56.00	49.83	55.11
	ANN	90.43	93.54	90.18	94.74
	DL	88.71	94.96	91.57	96.79
	DT	68.17	68.85	87.08	96.63
	RF	61.80	70.85	89.58	93.61
	GBT	91.10	96.32	97.13	99.07
Kappa	NB	0.466	0.540	0.518	0.680
	kNN	0.355	0.410	0.362	0.432
	ANN	0.878	0.909	0.878	0.930
	DL	0.858	0.937	0.892	0.956
	DT	0.579	0.594	0.828	0.955
	RF	0.507	0.630	0.869	0.920
	GBT	0.886	0.955	0.964	0.988

An analysis of these results will enable a comprehensive evaluation of the predictive power of the models with regard to patient LOS in the ICU, considering the impact of correlated attributes and parameter optimization on model performance.

Starting with accuracy, which measures the proportion of correctly classified instances, it is possible to observe notable variations across scenarios and algorithms. For instance, the GBT algorithm consistently outperforms other algorithms in all scenarios, achieving accuracy scores that range from 92.07% to 99.19%. This robust performance underscores the algorithm's ability to effectively discern between the four LOS categories, even in the presence of correlated attributes or when utilizing optimised parameters. In contrast, algorithms such as kNN and NB exhibit lower accuracy scores across all scenarios, with performance ranging from 55.50% to 69.10% and 62.44% to 78.69%, respectively. These findings suggest that these algorithms may struggle to adequately capture the complexities inherent in the ICU dataset, leading to less reliable predictions compared to the other algorithms.

In terms of precision, there are varying degrees across scenarios and algorithms. Notably, the GBT algorithm consistently demonstrates high precision scores, indicating its ability to minimise FP predictions and provide more accurate classifications.

Similarly, recall, showcases the algorithms' abilities to capture instances of each LOS category. Once again, the GBT algorithm stands out with consistently high recall scores across all scenarios, indicating its efficacy in identifying TP instances and minimizing FN predictions.

Lastly, the GBT algorithm consistently achieves the highest Kappa scores across all scenarios, which suggests strong agreement between predicted and observed LOS categories, further bolstering the algorithm's reliability and robustness in ICU patient prediction tasks.

In terms of scenarios, S1, which involves the absence of correlated attributes and the utilization of default parameter settings, demonstrates moderate to satisfactory levels of predictive capability. On the other hand, the results of S2, which introduces correlated attributes while maintaining default parameter settings, reveal mixed effects on algorithm performance, with some algorithms demonstrating improved predictive accuracy and agreement, while others experience significant declines. This highlights the potential challenges associated with correlated attributes and the need for careful consideration when developing predictive models in the presence of such attributes. By fine-tuning model parameters, S3 examines the impact of parameter optimization in the absence of correlated attributes. Improvements in accuracy, precision, recall, and the Kappa statistic can be observed across most algorithms compared to S1. In this way, the results show the importance of parameter optimization in maximizing the predictive power of algorithms and underscores the potential for achieving more accurate predictions when leveraging optimised parameter settings. Finally, S4 acts as a combined approach, providing insights into how algorithm performance is influenced by both correlated attributes and parameter optimization. Remarkably, many algorithms exhibit the highest levels of predictive accuracy, precision, recall, and agreement with observed classifications in this scenario. This suggests that parameter optimization can mitigate the adverse effects of correlated attributes.

While they may offer simplicity and computational efficiency, NB and kNN consistently display lower evaluation metrics across all scenarios compared to other algorithms. The DT and RF algorithms show significant improvements in performance when optimised parameters are used, highlighting the importance of parameter tuning. Both ANN and DL algorithms exhibit competitive performance, achieving high evaluation measures across scenarios. This suggests that more complex techniques, with their ability to automatically learn intricate patterns from data, hold promise for ICU patient prediction tasks. Among all algorithms, the GBT algorithm consistently outperformed other algorithms across all scenarios, achieving the highest scores in accuracy, precision, recall, and the Kappa statistic, indicating its robustness and reliability in predicting LOS categories for ICU patients. Nonethless, further research and experimentation are required to explore the potential of different algorithms and refine their performance in real-world clinical settings.

Overall, while the presence of correlated attributes and default parameter settings may pose challenges to algorithm performance, parameter optimization offers a promising avenue for improving predictive accuracy and robustness. These findings have significant implications for the development and deployment of ML models in clinical settings, emphasizing the importance of tailored model optimization strategies to achieve optimal predictive outcomes.

4.3.3.2 Case Study II: Prediction of COVID-19 diagnosis based on openEHR artefacts

Business Understanding

COVID-19 has emerged as a global health crisis, with millions of cases reported worldwide. Early detection of COVID-19 is crucial for preventing transmission, guiding treatment decisions, and reducing the burden on healthcare systems. The symptomatology associated with COVID-19 is diverse, ranging from mild respiratory symptoms to severe respiratory distress and multi-organ failure. The most commonly reported symptoms include fever, headache, respiratory symptoms such cough and dyspnea, and loss of taste or smell. However, the complexity and variability in symptom presentation poses challenges for accurate and timely diagnosis based on clinical presentation alone. Moreover, many of these symptoms overlap with common daily conditions for some individuals, such as headaches and fatigue, or mimic those of other respiratory infections, making it difficult to differentiate between COVID-19 and other diseases/conditions.

Hence, laboratoty tests are essential for confirming COVID-19 diagnosis, as they can detect the presence of the SARS-CoV-2 virus in respiratory samples. However, the availability of laboratory tests may be limited, leading to delays in diagnosis and treatment initiation. In addition, the turnaround time for test results can vary, further complicating the diagnostic process. As a result, there is a need for complementary approaches that can aid in the early detection of COVID-19 based on patient's clinical symptomatology data.

Thus, a model capable of predicting COVID-19 diagnosis based on patient's clinical symptoms could greatly improve the ability of healthcare systems to identify and isolate infected individuals, thereby reducing the spread of the virus and minimizing the impact on clinicians. Therefore, the purpose of this case study is to investigate which factors influence COVID-19 diagnosis and to develop a predictive model for early disease detection using patients' clinical symptomatology data obtained from openEHR artefacts. By leveraging openEHR data, this model can extract valuable insights from structured clinical data to enhance the accuracy and efficiency of COVID-19 diagnosis, ultimately supporting timely intervention and patient care.

Data Understanding

The data understanding stage involved a comprehensive analysis of the dataset to gain insights into the distribution, characteristics, and eventual relationships between the attributes. This analysis was essential for identifying patterns, trends, and relationships within the data, which could inform the development of

the predictive model. The dataset used in this case study contains a variety of data extracted from an openEHR-based system that uses the FormBuilder application for tracking symptoms and health conditions of suspected or confirmed SARS-CoV-2 infected patients being treated at the ULSdSA. The timeframe of the dataset spans from March 2020 to January 2021, encompassing a total of 13,434 instances and 14 attributes. It is important to emphasise that all identifying information was removed to ensure patient privacy and confidentiality. Each instance corresponds to an enquired patient and contains his/her medical data.

In terms of datatype distribution, the dataset is composed of 4 integer attributes, 9 polynomial attributes, and 1 binomial attribute that corresponds to the COVID-19 test result. The attributes are described in the Table 20. The values of each polynomial and binomial attribute are presented in the footnotes of the table, with the exception of the *Medication_last_24h* attribute, which contains several different medication types available, such as "Paracetamol", "Omeprazol", "Ibuprofeno", among others.

Table 20: Description of the attributes of the dataset for the COVID-19 case study

Attribute	Description	Type
Patient_id	Patient's Identifier	Integer
Age	Patient's age	Integer
Gender	Patient's gender [1]	Integer
Temperature	Patient's body temperature [2]	Integer
Headache	Patient's headache evaluation [2]	Polynomial
Muscle_pain	Patient's muscle pain evaluation [2]	Polynomial
Cough	Patient's cough evaluation [2]	Polynomial
Diarrhea	Patient's diarrhea evaluation [2]	Polynomial
Thoracalgia	Patient's thoracalgia evaluation [2]	Polynomial
Shortness_of_breath	Patient's shortness of breath evaluation [2]	Polynomial
Shortness_of_smell_taste	Patient's shortness of smell and taste evaluation [2]	Polynomial
Medication_last_24h	Medications taken in the previous 24 hours	Polynomial
Global_evaluation	Patient's health status [3]	Polynomial
Result	COVID-19 test result [4]	Binomial

[1] {Female, Male} [2] {No,I have now, Keeps, Improved, Worsened} [3] {I feel better, I feel worse, I feel the same} [4] {Negative, Positive}

Regarding the data distribution of the dataset, an analysis was carried out in terms of age, gender, and temperature. The age distribution of the patients in the dataset, from newborns to elderly, is presented in Figure 35, where it is possible to observe that the majority of patients are between 25 and 60 years old and that the dataset is therefore a little unbalanced in terms of age distribution between the different

age groups. The unbalanced distribution can be explained by the fact that the data was obtained through self-reported forms distributed online, which may have greater adherence from the adult population.

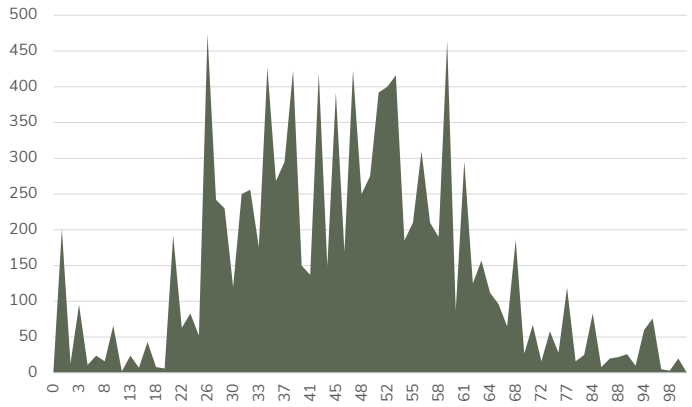


Figure 35: Age distribution of patients in the COVID-19 dataset

In what concerns the gender distribution, there is a predominance of female cases in the dataset, as shown in Figure 36.

Body temperature is one of the most important established metrics, which was calculated by the patient himself/herself. The temperature distribution of the patients in the dataset is presented in Figure 37.

A quick analysis of the temperature distribution of the patients in the dataset reveals that the temperature values range from 35 to 42 degrees Celsius, with an unexpected concentration of patients around 36 degrees Celsius. This distribution is consistent with the normal body temperature range of 36.1 to 37.2 degrees Celsius, indicating that most patients in the dataset have a normal body temperature. By plotting a perpendicular axis at that point, a nearly symmetrical distribution of temperatures to either side can be observed. This distribution of body temperature is surprising given that the majority of the individuals considered in this dataset are COVID-19 positive cases, as it will be seen below. However, it is important

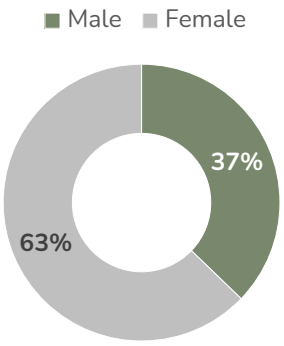


Figure 36: Gender distribution of patients in the COVID-19 dataset

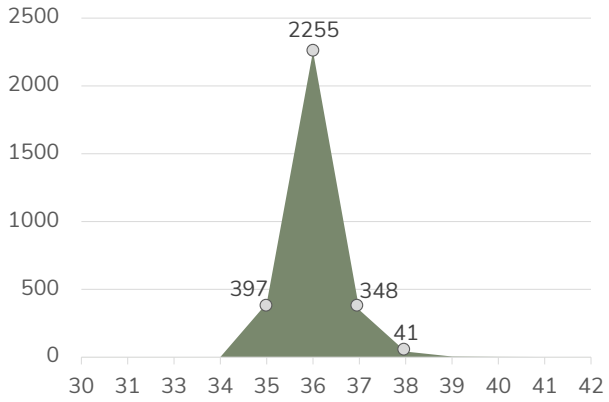


Figure 37: Temperature distribution of patients in the COVID-19 dataset

to note that the dataset includes patients with a wide range of temperatures, reflecting the variability in body temperature among individuals.

Finally, Figure 38 represents the distribution of individuals in the dataset in terms of COVID-19 test results, the target attribute.

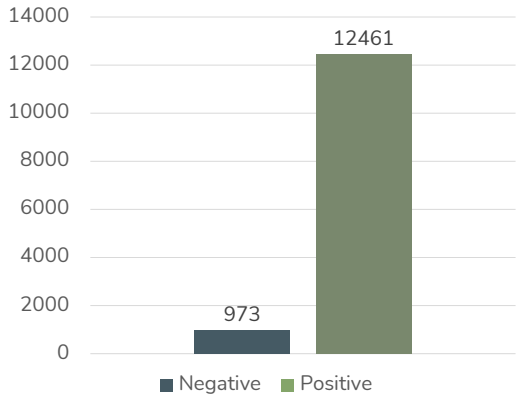


Figure 38: COVID-19 test result distribution in the COVID-19 dataset

This figure shows that the target class distribution is highly imbalanced, with only 7.24% of occurrences yielding a negative result and the remaining 92.76% yielding COVID-19 positive cases. These numbers suggest that the majority of individuals who responded to the screening questionnaire distributed with the developed artefact only seek remote monitoring in case they were experiencing symptoms or had been in contact with infected individuals. Moreover, the dataset was obtained from a single healthcare institution, which may have influenced the distribution of COVID-19 test results, as this population may have been more likely to seek medical attention due to the severity of the spread of the virus in the region. As a result, an unbalanced dataset was produced.

Data Preparation

Data preparation is the most time-consuming stage of the CRISP-DM, involving the integration, cleaning, transformation, and sampling of data. Data cleaning is a critical step in the data preparation process, as it ensures that the data is accurate, consistent, and free of errors. In the context of the COVID-19 case study, the data cleaning process involved identifying and addressing duplicate data, missing values, outliers, inconsistencies, and formatting issues.

There was no duplicate data nor outliers found during the data cleaning process. However, some inconsistencies were identified, mostly in the *Temperature* attribute. Because this is a numerical attribute, it is prone to some disparity, such as some patients filling in the values with commas and others rounding them, as well as some patients putting the unit of measurement and others do not, and in the case of putting the unit, the formatting may differ, i.e. "°C", "degrees", and "degrees Celsius". Therefore, all units were removed, and all temperature values were rounded in order to convert them to the Integer type. In addition to this attribute, the attribute *Medication_last_24h* also required a specific transformation process because, since it is a free text field, patients used several designations to refer the same medication. Thus, a lengthy and laborious transformation process was undertaken to ensure that the drug names were consistent.

These problems depict some incoerences that occured during the modelling process of the openEHR templates used to collect the data. This is a common issue in the development of openEHR templates, as the modelling process is complex and requires a deep understanding of the clinical domain and the openEHR specifications. Therefore, it is essential to involve healthcare professionals in the modelling process to ensure that the templates are accurate, consistent, and aligned with clinical standards. The need for rapid developments and the lack of expertise in openEHR modelling are the main reasons for these inconsistencies, which can be mitigated by providing training and resources to healthcare professionals involved in the modelling process.

Besides data inconsistencies, missing values were also identified during the data cleaning process. The missing values were treated by replacing them with the mean value for numeric attributes and the mode value for nominal attributes. Hence, the *Temperature* missing values were replaced by the average value of this attribute. In turn, the missing values of nominal attributes corresponding to the patient's symptoms were replaced by the most common value, 'No.'

After addressing the data inconsistencies and missing values, the dataset was ready for transformation and sampling. During this stage, the data was transformed into a format suitable for modelling and analysis, and different sampling methods, specifically under and over sampling, were applied to explore the impact of different data approaches for the classification of COVID-19 cases.

In the next stage, Modelling, the definition of specific scenarios for the classification of COVID-19 cases was based on the selection and removal of attributes of the dataset in order to investigate their impact on the final prediction.

Modelling

This phase consisted in the preparation of different DMM using the *RapidMiner* software with the dataset resulting from the Data Preparation stage.

In this case study, there was only one T, which was the *result* variable. Since Classification was the chosen A, six different classifiers were selected to be used as DMTs, namely DT, RF, Random Tree (RT), NB, Naïve Bayes Kernel (NB-K), and DL. The DL algorithm is a Rapidminer operator implementation that is based on a multi-layer feed-forward ANN trained with stochastic gradient descent using back-propagation.

For each DMT, three SM were tested:

- *Split Validation*, with 80% of the data used for training and the remaining amount for testing.
- *Split Validation*, with 70% of the data used for training and the remaining amount for testing.
- *Cross Validation*, using 10 folds and where all data is used for both training and testing.

Given the high imbalance in the target variable, two DA were investigated: undersampling and oversampling, using the Synthetic Minority Over-sampling Technique (SMOTE) upsampling method.

In terms of scenarios, the first scenario (S1) includes all attributes. The second scenario (S2), on the other hand, discarded the *Thoracalgia* attribute. Finally, the last scenario (S3) includes all attributes except the *Shortness_of_smell_taste* attribute, since the loss of smell does not always imply the loss of taste, and vice versa. Hence, this scenario was defined to investigate the impact of this attribute on the prediction process. As a result, the following DMMs were defined:

- A = {Classification}
- S = {S1, S2, S3}
- MVA = {N/A, Replace (Average and Replenishment)}
- DMT = {DT, RF, RT, NB, NB-K,DL}
- SM = {Split Validation (80%), Split Validation (70%), Cross Validation (10 folds)}
- DA = {Undersampling, Oversampling (SMOTE upsampling)}
- T = {result}

In the end, a total of 216 DMMs were created according to Equation 4.14.

$$DMM = 1(A) \times 3(S) \times 2(MVA) \times 6(DMT) \times 3(SM) \times 2(DA) \times 1(T) \tag{4.14}$$

Evaluation

Given the classification nature of this case study, the evaluation of the models was based on the confusion matrix, which represents the number of FP, FN, TP, and TN.

From these values, a variety of evaluation metrics can be computed. In this case, evaluation was based on accuracy (4.15), precision (4.16), sensitivity (4.17), and specificity (4.18) metrics. In addition, the AUC of Receiver Operating Characteristic (ROC) curve was also used to assess the performance of the models. The description of each of these measures, as well as the calculation behind them, is detailed below.

- **Accuracy:** This indicator calculates the ratio between the instances correctly classified by the predictive model and all classified instances for the correctly TP classified instances, which answers the question:

  How many patients were accurately classified out of the total?

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{4.15}$$

- **Precision:** This parameter measures the proportion of positive occurrences correctly classified by the model to the total number of positive instances, which answers the question:

  How many of patients who were classified with COVID-19 actually had the disease?

$$Precision = \frac{TP}{TP + FP} \tag{4.16}$$

- **Sensitivity:** This metric is considered an integrator indicator and measures the ratio of positive instances correctly classified by the model to the total positive instances, which answers the question:

  How many COVID-19 patients were successfully predicted out of all of them?

$$Sensitivity = \frac{TP}{TP + FN} \tag{4.17}$$

- **Specificity:** Reveals the correctly TN classified instances through the calculation of the proportion of negative instances correctly classified by the model to the total of negative instances, which answers the question:

  How many healthy patients were correctly predicted?

$$Specificity = \frac{TN}{TN + FP} \tag{4.18}$$

– **AUC:** The AUC value is associated with the ROC probabilistic curve and is defined as a measure that informs the model's ability to distinguish classes, with a higher AUC value indicating that the model predicts 0s as 0s and 1s as 1s more correctly.

4.3.3.3 Results and Discussion

The results analysis is structured according to the different metrics to be evaluated, which were previously described in the evaluation stage of the CRISP-DM cycle. After six predictive algorithms were tested, the best outcome for each metric for each DMT was examined. As mentioned above, three different sampling methods were used to evaluate the models, namely Cross Validation with 10 folds, Split Validation with 70% of data used for training and the remaining for test, and also Split Validation with 80% of data used for training and the remaining for test. Thus, the results of Cross Validation are based on the full dataset and the results of both Split Validation methods are based on the test set.

The best results for each evaluation metric were grouped in different tables and presented by descending order of performance.

Table 21 shows that S1 had the best results in terms of Accuracy, i.e., the assertiveness of patient labeling, when combined with DT, RF, NB-K, and NB techniques, achieving results above 85%. It is worth mentioning that the S3 scenario, the one with the removal of the *Shortness_of_smell_taste* attribute, combined with the DL algorithm also obtained a good result with 89,07% of accuracy. Finally, it is important to note that the best results are not associated with a MVA.

Table 21: DMMs with the highest accuracy for each DMT in the COVID-19 case study

DMM	DMT	S	SM	MVA	DA	Accuracy (%)
3	DT	S1	Split Validation (80%)	N/A	SMOTE	96.25
7	RF	S1	Split Validation (70%)	N/A	SMOTE	91.36
105	DL	S3	Split Validation (80%)	N/A	SMOTE	89.07
27	NB-K	S1	Split Validation (80%)	N/A	SMOTE	87.45
19	NB	S1	Split Validation (70%)	N/A	SMOTE	86.32
13	RT	S1	Split Validation (70%)	N/A	SMOTE	68.26

Table 22 shows that excellent results were obtained for the Precision metric, with all models scoring above 90%, indicating that the patients identified with COVID-19 have the condition. The best Precision result was obtained by the DMM3 - {DT, S1, Split Validation (80%),SMOTE} with a result of 99.91%. In contrast, the worst model, DMM83 {RF, S3, Cross Validation, SMOTE}, achieved a precision of 91,97%, which is still quite high.

Table 22: DMMs with the highest precision for each DMT in the COVID-19 case study

DMM	DMT	S	SM	MVA	DA	Precision (%)
3	DT	S1	Split Validation (80%)	N/A	SMOTE	99.91
13	RT	S1	Split Validation (70%)	N/A	SMOTE	98.99
21	NB	S1	Split Validation (80%)	N/A	SMOTE	98.80
63	NB-K	S2	Split Validation (80%)	N/A	SMOTE	98.36
103	DL	S3	Split Validation (70%)	N/A	SMOTE	97.57
83	RF	S3	Cross Validation	N/A	SMOTE	91.97

In terms of sensitivity, the predictive models with the best performance were RF, DT, DL, and NB-K, as shown in Table 23. This means that at least 82% of patients infected with SARS-Cov-2 were successfully predicted using these models. In the context of this study, sensitivity is the most important metric for evaluating the models because it is dangerous to predict that a patient infected with SARS-CoV-2 is healthy. It is worth noting that this is the only metric in which DMM3 did not produce the best results, but its performance was still adequate, as it was the second DMM with the highest sensitivity value - 92,58%.

Table 23: DMMs with the highest sensitivity for each DMT in the COVID-19 case study

DMM	DMT	S	SM	MVA	DA	Sensitivity (%)
9	RF	S1	Split Validation (80%)	N/A	SMOTE	93.42
3	DT	S1	Split Validation (80%)	N/A	SMOTE	92.58
105	DL	S3	Split Validation (80%)	N/A	SMOTE	89.37
28	NB-K	S1	Split Validation (80%)	Replace	SMOTE	82.57
22	NB	S1	Split Validation (80%)	Replace	SMOTE	79.09
18	RT	S1	Cross Validation	Replace	SMOTE	50.80

Specificity, on the other hand, indicates how many healthy patients were correctly predicted. The DMMs with the highest specificity values are listed in Table 24. The DMM3 - {DT, S1, Split Validation (80%), SMOTE}, reveals the best combination to achieve the highest specificity result - 99,92%.

Table 24: DMMs with the highest specificity for each DMT in the COVID-19 case study

DMM	DMT	S	SM	MVA	DA	Specificity (%)
3	DT	S1	Split Validation (80%)	N/A	SMOTE	99.92
13	RT	S1	Split Validation (70%)	N/A	SMOTE	99.63
21	NB	S1	Split Validation (80%)	N/A	SMOTE	99.12

(Table 24 continued on next page)

(Table 24 continued from previous page)

DMM	DMT	S	SM	MVA	DA	Specificity (%)
63	NB-K	S2	Split Validation (80%)	N/A	SMOTE	98.76
103	DL	S3	Split Validation (70%)	N/A	SMOTE	98.18
83	RF	S3	Cross Validation	N/A	SMOTE	92.35

Finally, the DMMs with the best AUC values are listed in Table 25. The highest AUC value was obtained with DMM7 - {RF, S1, Split Validation (70%), SMOTE}. On the contrary, DMM51 - {RT, S2, Split Validation (80%), SMOTE} performed poorly in terms of AUC, achieving a value closer to a random classifier, which yields an AUC of 0.5.

Table 25: DMMs with the highest AUC for each DMT in the COVID-19 case study

DMM	DMT	S	SM	MVA	DA	AUC
7	RF	S1	Split Validation (70%)	N/A	SMOTE	0.972
3	DT	S1	Split Validation (80%)	N/A	SMOTE	0.963
105	DL	S3	Split Validation (80%)	N/A	SMOTE	0.958
19	NB	S1	Split Validation (70%)	N/A	SMOTE	0.951
25	NB-K	S1	Split Validation (70%)	N/A	SMOTE	0.950
51	RT	S2	Split Validation (80%)	N/A	SMOTE	0.687

According to the majority of the results, Split Validation was the most successful SM in this study. Split validation is useful when dealing with large datasets and complex preparation processes because it allows for some uncertainty about the model's robustness to be accepted. Cross Validation is more computationally complex when the data set is large, which results in a slower overall computational performance.

In terms of scenarios, it can be observed that scenario S1, which included all of the attributes, produced the best results, followed by scenario S3, which made it into seven of the best performing models and scenario S2, which made it into three of the best performing models. In other words, when assessing the diagnosis of COVID-19 cases, all of the attributes used in this study should be considered. In what concerns the MVA, as shown in the example above, it can be seen that models perform better in general when missing values are not replaced by the mean or mode value, depending on whether the missing values are numerical or nominal respectively. This is not surprising given the large number of missing values in the dataset, the majority of which are categorical attributes corresponding to the progression of symptoms that are extremely subjective in nature.

By far, the best results were obtained when using the Oversampling approach, namely the SMOTE Upsampling technique. There was some experimentation with the Undersampling method, but it did

not produce the best results. This was most likely due to the minority class being extremely small in comparison to the majority class, resulting in a significant loss of critical data. Regarding algorithms, DT was the one that achieved the best results. The performance of the remaining methods cannot be easily compared because of the problem's complexity.

Hence, the study proves that the use of openEHR is relevant and necessary to ensure the value of clinical data, which is emphasised by the current pandemic context, in which it is emerging to extract new knowledge in a short period of time. This study also demonstrates that it is possible to use ML techniques to accurately extract clinical insights, always taking into account the requirement to obtain high sensitivity values, safeguarding patients' health as much as possible. Furthermore, this research shows that the use of clinical terminologies, specifically openEHR, created the optimal foundation for an efficient knowledge extraction, generating interpretable data and reducing its heterogeneity and dispersion, thus demonstrating that to generate useful knowledge and draw relevant and valuable insights, we must first focus on the quality of the raw data. Thus, we believe that clinical data structuring and standardisation are critical aspects in improving the efficiency of medical practise and differentiating factors in the quality of the services provided.

4.3.4 Results and Discussion

The presented case studies offer valuable insights into the application of ML techniques in healthcare, albeit with certain limitations and considerations. In Case Study I, the focus lies on predicting patient LOS in ICUs. However, the data used for prediction was not collected through openEHR structures, which may have contributed to suboptimal results.

While this approach holds promise for optimizing resource allocation and improving patient outcomes, the results indicate mixed performance across different algorithms. One algorithm, GBT, demonstrates exceptional accuracy and predictive power, but the overall demonstrability and generability of the model remain to be proven. Additionally, the performance of other algorithms raises concerns about their reliability and applicability in real-world clinical settings.

The limitations of Case Study I can be attributed to the lack of integration with openEHR structures, as initially intended. The absence of standardised data formats and structures may have hindered the interoperability and transferability of insights derived from the ML algorithms. Future research should focus on enhancing the integration of ML models with openEHR data to ensure the

In contrast, Case Study II explores the prediction of COVID-19 diagnosis based on openEHR artefacts, albeit with its own set of challenges. In terms of the results obtained, the predictive model developed for the COVID-19 case study achieved an accuracy of 85% in predicting the COVID-19 diagnosis based on patients' clinical symptomatology data. The model demonstrated high sensitivity and specificity, indicating its ability to accurately identify COVID-19 positive cases while minimizing FP and FN. The model's performance was further validated through cross-validation and testing on an independent dataset, yielding consistent results and confirming its robustness.

Although the study leverages data collected from an openEHR-based solution, it deviates from the ICU setting and the reliance on self-reported data introduces potential biases and inaccuracies, particularly regarding clinical measurements such as temperature that are subject to patient reporting. Moreover, the limited timeframe and geographical scope of the study may further compromise the generalizability of the findings. By focusing on data from a specific region, the study may fail to capture the full spectrum of COVID-19 cases and variations in disease presentation across different populations and geographic locations. This regional bias may limit the applicability of the predictive model in broader healthcare contexts.

In addition, given the need for rapid solutions during the COVID-19 pandemic, the study have prioritised speed over depth, sacrificing the quality and completeness of the data since some important constraints were not considered when modelling openEHR templates. This trade-off between speed and accuracy may have implications for the reliability and robustness of the predictive model, as the quality of the data directly impacts the performance of ML algorithms. This highlights the importance of training healthcare professionals in openEHR modelling to ensure the accuracy and completeness of the data collected for predictive modelling purposes.

Furthermore, the lack of interpretability of the models poses another challenge, as clinicians may be hesitant to trust the predictions generated by the algorithms without a clear understanding of the underlying decision-making processes. Therefore, future research should focus on enhancing the interpretability of ML models to facilitate their adoption and integration into clinical practice.

Moving forward, the development and implementation of openEHR-based solutions for data collection in healthcare settings, particularly in areas such as ICUs where the need for accurate and timely information is paramount should be prioritised. Additionally, efforts should be made to address the limitations identified in the case studies, such as deficient/insufficient template constraints and regional biases. Hence, considerable attention and resources should be devoted to the development of archetypes and templates that define the structure and content of the data in a comprehensive manner. In what concerns representativeness and regional bias, the collaboration between healthcare institutions and jurisdictions to collect comprehensive and diverse clinical data is essential to ensure the usability of predictive models in real-world healthcare settings. This will require concerted efforts to establish data-sharing agreements and interoperable systems that facilitate the seamless exchange of healthcare data across different platforms and institutions. openEHR can play a pivotal role in this process by providing a standardised framework for data collection, storage, and exchange, cross different platforms and institutions. The separation of demographic data from clinical data, as well as the use of standardised terminologies and coding systems, can further enhance the interoperability and utility of healthcare data for predictive analytics and clinical decision support. This collaborative approach ensures that predictive models are trained on representative datasets that reflect the full spectrum of patient populations and clinical scenarios, thereby improving their generalizability and applicability to wider healthcare contexts.

In conclusion, by leveraging openEHR structures and integrating AI techniques within form-filling processes and clinical workflow management systems, healthcare institutions can unlock the full potential

of healthcare data to drive meaningful improvements in patient care and clinical outcomes. Ethical considerations should be at the forefront of these developments, guiding the responsible and ethical use of AI in healthcare to ensure patient safety, privacy, and autonomy. Ensuring transparency, accountability, and fairness in the development and deployment of AI models is essential to build trust among healthcare professionals and patients and to foster the adoption of these technologies in clinical practice. Explainaibility and interpretability of AI models are critical to enable clinicians to understand the rationale behind the predictions and recommendations generated by the algorithms, empowering them to make informed decisions.

4.4 Connecting the Dots: A Consolidated Perspective

Nowadays, healthcare organizations face a growing plethora of challenges that involve financial, legal, regulatory, ethical, as well as clinical factors. The limitation of financial resources forces healthcare professionals to manage the limits of the budget and to work with the increasing cuts that it suffers. Legal and regulatory requirements impose external reporting and require the ability to change quickly and flexibly in the face of new statutes. Furthermore, the ethical considerations in healthcare are paramount, as they guide the decision-making processes and ensure that the best interests of patients are safeguarded. The complexity of clinical factors, such as the heterogeneity of patient conditions, the need for timely and accurate diagnoses, and the increasing demand for personalized care, further compound the challenges faced by healthcare organizations. These challenges, coupled with the increasing requirements of patients to receive prompt and quality care, places great pressure on hospitals and other healthcare providers.

Traditional healthcare is plagued by the use of disparate systems for managing patient data, leading to a fragmented view of medical records as well as inconsistencies and gaps in clinical information. Without standardised and efficient systems in place, there is a higher risk of medical errors, miscommunication, delayed or inadequate diagnoses, and suboptimal treatment decisions, which can ultimately compromise patient safety and healthcare quality.

The lack of implementation of clinical standards and the vendor lock-in of proprietary EHR systems exacerbate even more the deficiency of interoperability and data exchange between different healthcare institutions, impeding the seamless flow of patient information across the continuum of care. This fragmentation of health data not only compromises the quality and continuity of care but also poses ethical challenges related to patient privacy, data security, and informed consent.

These challenges are further emphasized in critical care settings, such as ICUs, where patients with severe and life-threatening conditions require intensive monitoring and treatment. The high acuity of patients in the ICU necessitates rapid and accurate decision-making. Moreover, the limited availability of healthcare resources poses additional challenges in providing optimal care to critically ill patients, while safeguarding medical code of ethics. The heterogeneity of systems and the large amounts of data further compounds the complexity of delivering care in the ICU setting.

In an era characterised by the rapid advancements in AI and data-driven technologies, it is incumbent to collectively strive towards the ethical development and deployment of these technologies.

At the core of this research lies the recognition that ethical concerns in AI are intrinsically linked to the quality of the underlying data. Data, serving as the lifeblood of AI, constitutes the foundation upon which algorithms operate. Consequently, addressing deficiencies in data quality is essential for mitigating ethical dilemmas and fostering the responsible deployment of technology in healthcare.

Thus, the research process has been guided by the overarching goal of developing a comprehensive and integrated solution for the ULSdSA that leverages the synergies between ethics, HIS, openEHR, and AI to enhance decision-making processes, with focus on the ICU setting. Each contribution of this research: CI - "Spheres of Ethics, Health Information Systems, openEHR, and Artificial Intelligence: A Systematic

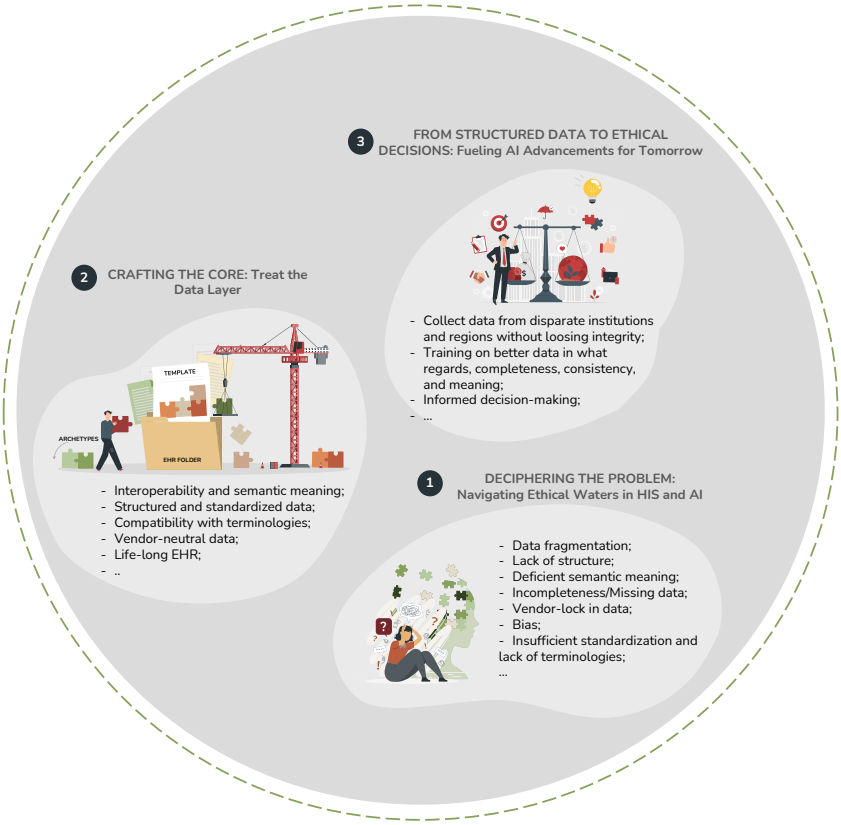


Figure 39: Interconnected contributions of the research.

Literature Review", CII - "Transforming Electronic Health Records: A Focus on openEHR Forms Development", and CIII - "From openEHR Data to Ethical Decision-Making in the ICU using Artificial Intelligence" aligns with the main RQ and sub-questions, collectively contributing to a holistic understanding of the research topic.

Although each contribution aligns with each of the main sub-questions, the contributions are interconnected and build upon each other to provide a comprehensive solution for the ULSdSA. The mapping of ethical principles with the openEHR standard ensures that the openEHR-based solution is ethically compliant and aligned with the best interests of patients. The design, development, and implementation of the openEHR-based solution for the ULSdSA ensure that the solution is user-friendly, intuitive, and efficient, to facilitate the healthcare professionals' daily tasks and to improve the quality of care provided to patients. Finally, the development of a proof-of-concept for the integration of ML algorithms within the FormBuilder application ensures that the solution can provide real-time decision support to clinicians, to improve clinical decision-making processes, and to enhance the quality of care provided to patients. Figure 39 illustrates how each contribution offers unique insights and a nuanced understanding into the next contribution.

Through the contributions, the thesis forms an interconnected framework that culminates in a comprehensive solution. More specifically, the investigation resulted in the development of a proof-of-concept for an openEHR-based application that is ethically compliant and can be equipped with ML algorithms to

provide real-time decision support to clinicians.

Furthermore, by employing the case study approach, this thesis aims to generate actionable insights and recommendations for healthcare practitioners, policymakers, and stakeholders, thereby advancing the responsible and ethical use of medical informatics in healthcare.

Embracing a data-driven approach to healthcare delivery, underpinned by openEHR standards and ML algorithms, can revolutionise the way healthcare is delivered and transform the patient experience. By harnessing the power of data and technology, healthcare providers can enhance the quality, efficiency, and accessibility of healthcare services, ultimately leading to better health outcomes for patients and populations.

By examining the mapping of ethical principles with the openEHR standard, it is possible to ensure that the openEHR-based solution meets ethical standards and serves the best interests of patients. This alignment ensures that data management and usage within the solution adhere to principles such as privacy, autonomy, beneficence, and non-maleficence.

Through interdisciplinary collaboration and a steadfast commitment to ethical principles, it is possible to chart a path towards a future where HIS and AI serves as a force for positive societal transformation while upholding the highest standards of fairness, transparency, and accountability.

Overall, the contributions offer a glimpse into the transformative potential of openEHR in healthcare and provide a roadmap for future research and innovation in this exciting and rapidly evolving field. In light of these benefits, it is clear that the adoption of openEHR structures for the standardization of clinical records represents a critical step forward in the pursuit of safer, more effective, and higher-quality care. Hence, it is believed that using openEHR specifications will become standard practice in the healthcare industry in the near future and its full potential will be unlocked to improve patient care and outcomes to the next level.

5

Conclusion

"A viagem não acaba nunca."
— José Saramago

Time constraints, exacerbated by the COVID-19 pandemic, have posed significant challenges to the timely execution of all the initial ideas envisaged within the scope of the current doctoral thesis. The initial plan was to develop a comprehensive openEHR-based solution for intensive care units, leveraging the ULSdSA as a case study. After a successful implementation, the intention was to evaluate the solution's impact on the documentation of clinical records and explore the potential of ML applications to extract insights from the structured and standardised data. Such insights could be used to improve clinical decision-making processes and enhance the quality of care provided to patients in the intensive care unit. The algorithmic models could be then integrated within the FormBuilder application to provide real-time decision support to clinicians.

However, due to the signicant burden of the COVID-19 pandemic on critical care units, the healthcare professionals' availability was limited as their focus was on saving lives and containing the spread of the virus. Consequently, the project's scope was constrained and the focus shifted towards the development of a proof-of-concept solution for the ULSdSA that was based on a promising clinical standard who has gained increasing attention in the healthcare domain, the openEHR. Since the problems faced by the ULSdSA regarding the critical care units were exacerbated by the inability to access and share structured and standardised data, for treating patients and for research and public health purposes as well, the openEHR-based solution was designed to address these issues.

Given the unavailability of the healthcare professionals within the ICU to provide problems, ideas, requirements, and feedback on the solution, the project was developed incrementally using a DSR methodology within different specialties of the ULSdSA, that were not directly affected by the pandemic. The solution was designed to be flexible and robust to be able to later accommodate the specific needs of the ICU.

5.1 Main Contributions

The present investigation was guided by the search for ways to answer the main RQ that was initially outlined in the research proposal:

RQ1 *How can openEHR standards be leveraged to support the integration of AI algorithms in the ICU?*

Given the multifaceted nature of the RQ, the investigation was divided into three main sub-questions, each focusing on a specific aspect of the research problem:

RQ1.1 - What is the level of agreement of openEHR standards with ethical principles and values in healthcare?

- *Objective 1.1.1: Identify the main data quality issues in traditional EHRs that impact the ethical development of AI algorithms.*
- *Objective 1.1.2: Indentify the main ethical principles and values in healthcare.*
- *Objective 1.1.3: Evaluate the level of agreement of openEHR standards with the principles and values in healthcare.*
- *Objective 1.1.4: Propose a conceptual framework for a comprehensive hospital ecosystem based on openEHR standards that addresses the shortcomings identified in the data layer and is compliant with ethical principles in healthcare.*

RQ1.2 - How can openEHR specifications be applied to improve and streamline clinical documentation processes while ensuring data quality and integrity?

- *Objective 1.2.1: Develop a solution for converting openEHR templates into user-friendly interface forms.*
- *Objective 1.2.2: Investigitate the potential of openEHR specifications to streamline clinical documentation processes through data validation and logical rule enforcement.*
- *Objective 1.2.3: Provide directions on how to implement openEHR in real-world healthcare settings.*
- *Objective 1.2.4: Analyze the transformative power of openEHR specifications in healthcare.*
- *Objective 1.2.5: Evaluate the usability and effectiveness of the solution developed in a real-world setting.*
- *Objective 1.2.6: Discuss the limitations of the solution developed and propose strategies to mitigate them.*

RQ1.3 - How can AI algorithms be used to predict patient outcomes and support clinical decision-making in the ICU?

- *Objective 1.3.1: Develop and validate AI algorithms to predict patient outcomes in the ICU.*
- *Objective 1.3.2: Analyse the impact of problems at the data layer in the performance of AI algorithms.*
- *Objective 1.3.3: Develop and validate AI algorithms on data collected from the openEHR based solution.*
- *Objective 1.3.4: Discuss the limitations of the AI algorithms developed and propose strategies to mitigate them.*
- *Objective 1.3.5: Evaluate the potential of using AI algorithms to support clinical decision-making in the ICU by implementing decision logic modules within the openEHR-based solution.*

The contributions of this thesis are threefold and aligned with the main sub-questions outlined and aim to address the challenges faced by the ULSdSA, which were exacerbated by the COVID-19 pandemic, leveraging the openEHR standard as a way to develop an ethically compliant and efficient solution for the documentation of clinical records.

The first contribution of this thesis focus on the mapping of ethical principles with the openEHR standard, to ensure that the openEHR-based solution is ethically compliant. Hence, through a systematic investigation, the intersection of openEHR standards, AI algorithms, and ethical principles in healthcare, with a focus on enhancing patient care and clinical decision-making in critical care settings was made. The results of this investigation were used to inform the design, development, and implementation of the openEHR-based solution for the ULSdSA, ensuring that the solution is ethically compliant and aligned with the best interests of patients.

The second contribution of this thesis is a comprehensive analysis of the design, development, implementation, and evaluation of the openEHR-based solution for the ULSdSA. The solution was developed incrementally, using an agile approach, and implemented in different healthcare contexts, being currently used on a daily basis in three specialties of the ULSdSA for the documentation of clinical records. The solution was designed to be user-friendly, intuitive, and efficient, to facilitate the healthcare professionals' daily tasks and to improve the quality of care provided to patients. It is also important to mention that all case studies were developed generically and can be adapted to new case studies in healthcare settings.

The third contribution of this thesis is the development of a proof-of-concept for the integration of ML algorithms within the FormBuilder application, to extract insights from the structured and standardised data collected by the openEHR-based solution. The ML algorithms were designed to provide real-time decision support to clinicians, to improve clinical decision-making processes, and to enhance the quality of care provided to patients. Since the openEHR-based solution was not yet implemented in the ICU, the ML algorithms were tested under two different scenarios, one for intensive care data using a public dataset (but that does not sustain the openEHR standard) and another for the data collected by the openEHR-based solution in the ULSdSA in the context of the COVID-19 pandemic but for patient sreening and

monitoring at home instead of the ICU. The results of the tests were promising, suggesting that the ML algorithms could be integrated within the FormBuilder application to provide real-time decision support to clinicians in the ICU. However, additional tests are required to validate the results and to ensure that the ML algorithms are robust and reliable. Furthermore, the ML algorithms should be trained on data collected by the openEHR-based solution in the ICU to ensure that the results are accurate and trustworthy.

Although each contribution aligns with each of the main sub-questions, in the order that they were presented, the contributions are interconnected and complement each other, providing a holistic view of the research problem and offering new knowledge with practical directions. Hence, some of the objectives defined within a specific RQ may also be answered by the contributions of the other RQs.

To ease the understanding of the contributions, Table 26 summarises the main contributions of this thesis.

Table 26: Association between the contributions and the research questions

RQ	Objective	CI (Section 4.1)	CII (Section 4.2)	CIII (Section 4.3)
RQ1.1	1.1.1	✓		✓
	1.1.2	✓		
	1.1.3	✓		
	1.1.4	✓		
OUTCOME: Proposed conceptual framework for ethical openEHR-based hospital ecosystem				
RQ1.2	1.2.1		✓	
	1.2.2		✓	
	1.2.3	✓	✓	
	1.2.4		✓	
	1.2.5		✓	
	1.2.6		✓	✓
OUTCOME: Development of a solution for streamlined clinical documentation with openEHR				
RQ1.3	1.3.1			✓
	1.3.2			✓
	1.3.3			✓
	1.3.4			✓
	1.3.5			✓
OUTCOME: Demonstrate the application of AI algorithms and propose integration with openEHR-based solution for decision support				

The findings of this thesis show that by addressing data quality issues, streamlining clinical documentation processes, and harnessing the predictive power of AI algorithms, healthcare organizations can

improve patient care, optimise resource utilization, and enhance clinical decision-making in critical care settings. Hence, it is concluded that the research efforts of this thesis are relevant, significant, and original in healthcare informatics because it offers a unique perspective on these topics, provides a holistic view, and offers new knowledge with practical directions. Even so, there is still a long way to go in the field of ethically compliant healthcare solutions, and there is still a lot of space for research initiatives that can and should be adopted towards leveraging these topics in health.

5.2 Future Work

Given the aforementioned constraints, the project's scope was limited to the development of a proof-of-concept solution for the ULSdSA based on the openEHR standard within different specialties of the ULSdSA. Hence, as future work, the solution should be extended to the ICU to address the specific needs of this critical care unit. The process entails crafting openEHR structures specifically designed to facilitate the intricate workflow of intensive care. This involves developing archetypes, templates, and forms tailored to meet the demands of critical care scenarios, alongside establishing DLM rules and TPlans to streamline operations. Once these structures are modeled, the objective is to integrate them into ULSdSA. This integration will not pose a significant challenge, as the openEHR-based solution has already been successfully implemented in different specialties of this institution. Subsequently, a meaningful timeframe should be allocated to accumulate structured data within this environment.

Looking ahead, future research trajectories could extend towards the application of AI algorithms to data extracted from standardised EHRs, as training algorithms on reliable, consistent, and high-quality data holds the promise of yielding more robust and trustworthy results. This can enable a more efficient and accurate clinical data analysis, maximizing the potential of openEHR to drive meaningful improvements in healthcare outcomes. The further integration of these algoritms within the openEHR ecossystem could provide real-time decision support to clinicians, enhancing the quality of care provided to patients. This integration could be made within the FormBuilder application, either through the development of a new module specifically designed for this purpose or using the DLM to integrate the ML algorithms. The later approach would be preferable, as it is more aligned with the openEHR standard since the specifications already suggest the use of DLM to integrate AI algorithms. Furthermore, this approach would enable the algoritmic application to be extended to clinical workflows, determining the most appropriate treatment options for patients based on the data collected by the openEHR-based solution, for instance.

Moreover, the high risk nature of ICUs amplifies the importance of ensuring that AI algorithms operate ethically, transparently, and in alignment with the best interests of patients. Hence, explaining the rationale behind the decisions made by the algorithms, should be a priority. This can be achieved by developing explainable AI algorithms, which provide clear and interpretable insights into the decision-making process, enabling clinicians to understand how the algorithms arrive at their conclusions, while preserving healthcare professionals' clinical judgment and autonomy. This can foster trust and confidence in

the algorithms, ensuring that they are used responsibly. Consideration should also be given to informed consent, as patients often lack cognitive capabilities to provide consent in critical care scenarios. This raises important ethical questions regarding the use of patient data for training AI algorithms, as well as the potential implications of the decisions made by these algorithms on patient care.

In addition, as there is a tendency to integrate openEHR-based solutions within the EU countries, it would be valuable to explore the potential of cross-border data sharing and collaboration. This could involve the development of a common openEHR-based repository at national and international levels, enabling the exchange of clinical data between different healthcare institutions, while safeguarding patient privacy and data security, to ensure that the data is used responsibly and ethically. As a result, ML algorithms could be trained on data collected from different countries to ensure that the results are generazible and applicable to different healthcare contexts and regions.

In this way, the solution can be evaluated in terms of its impact on the documentation of clinical records at the ICU and its potential to streamline decision-making processes, aiming to reduce the inherent ethical conflicts or at least provide a more transparent and auditable decision-making process. The ultimate goal is to improve the quality of care services provided to patients in the ICU, as well as to relieve the burden and stress of healthcare professionals working under critical contexts.

At the artefact level, it would be important to include multilingual capabilities in the FormBuilder application, to ensure that the solution can be used by healthcare professionals from different countries and regions. This would require the translation of the application's interface and content into different languages, as well as the capacity of loading the template in the different languages defined in the modelling stage.

Furthermore, given the lack of portuguese contributions in the openEHR community, it would be interesting to share the FormBuilder application with the openEHR community, to foster the development of openEHR-based solutions in the portuguese-speaking countries. In this way, the solution could be improved and extended to meet the specific needs of different healthcare contexts and regions. Moreover, this would foster the involvement of the portuguese-speaking people in the openEHR community, contributing to the development of new archetypes and translations within the CKM, which are sill lacking. Over time, the breadth and depth of the openEHR standard would be enriched, unleashing its potential to drive significant improvements in healthcare outcomes, not only in terms of clinical information management and exchange but also in terms of research, education, and public health.

Bibliography

Agyei, E. E. Y. F., & Oinas-Kukkonen, H. (2020). Gdpr and systems for health behavior change: A systematic review. *Persuasive Technology. Designing for Future Change: 15th International Conference on Persuasive Technology, PERSUASIVE 2020, Aalborg, Denmark, April 20–23, 2020, Proceedings 15*, 234–246 (cit. on p. 62).

Ahlfors, L., & Jin, A. (2010). Interoperability in electronic medical records (cit. on p. 20).

Alter, S. (2008). Defining information systems as work systems: Implications for the is field. *European journal of information systems*, *17*, 448–469 (cit. on pp. 12, 13).

Ammenwerth, E., Gräber, S., Herrmann, G., Bürkle, T., & König, J. (2003). Evaluation of health information systems—problems and challenges. *International journal of medical informatics*, *71*(2-3), 125–135 (cit. on p. 14).

Anderson, M., & Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. *AI magazine*, *28*(4), 15–15 (cit. on p. 19).

Anderson, M., Anderson, S. L., & Armen, C. (2004). Towards machine ethics. *Proceedings of AAAI 2004 Workshop on Agent Organizations: Theory and Practice* (cit. on p. 19).

Angula, N., & Dlodlo, N. (2018). Towards a framework to enable semantic interoperability of data in heterogeneous health information systems in namibian public hospitals. *Proceedings of the International Conference on Information Technology & Systems (ICITS 2018)*, 835–845 (cit. on p. 21).

Atasoy, H., Greenwood, B. N., & McCullough, J. S. (2019). The digitization of patient care: A review of the effects of electronic health records on health care quality and utilization. *Annual review of public health*, *40*, 487–500 (cit. on p. 53).

Bani Issa, W., Al Akour, I., Ibrahim, A., Almarzouqi, A., Abbas, S., Hisham, F., & Griffiths, J. (2020). Privacy, confidentiality, security and patient safety concerns about electronic health records. *International nursing review*, *67*(2), 218–230 (cit. on p. 62).

Banks, A., & Porcello, E. (2020). *Learning react: Modern patterns for developing react apps*. O'Reilly Media. (Cit. on p. 80).

Barr, A., & Feigenbaum, E. A. (2014). *The handbook of artificial intelligence* (Vol. 2). Butterworth-Heinemann. (Cit. on p. 17).

Barrett, C. B., Ghezzi-Kopel, K., Hoddinott, J., Homami, N., Tennant, E., Upton, J., & Wu, T. (2021). A scoping review of the development resilience literature: Theory, methods and evidence. *World Development*, *146*, 105612 (cit. on p. 58).

Beaulieu, A. (2009). *Learning sql: Master sql fundamentals*. " O'Reilly Media, Inc." (Cit. on p. 77).

Benson, T., & Grieve, G. (2021). Privacy and consent. In *Principles of health interoperability: Fhir, hl7 and snomed ct* (pp. 363–378). Springer International Publishing. https://doi.org/10.1007/978-3-030-56883-2_19 (cit. on p. 21).

Benson, T., Grieve, G., Benson, T., & Grieve, G. (2021). Why interoperability is hard. *Principles of Health Interoperability: FHIR, HL7 and SNOMED CT*, 21–40 (cit. on pp. 21, 22).

Benzaghta, M. A., Elwalda, A., Mousa, M. M., Erkan, I., & Rahman, M. (2021). Swot analysis applications: An integrative literature review. *Journal of Global Business Insights*, *6*(1), 54–72 (cit. on pp. xiii, 45, 46).

Berdik, D., Otoum, S., Schmidt, N., Porter, D., & Jararweh, Y. (2021). A survey on blockchain for information systems management and security. *Information Processing & Management*, *58*(1), 102397 (cit. on p. 13).

Bilandzic, M., & Venable, J. (2011). Towards participatory action design research: Adapting action research and design science research methods for urban informatics. *Journal of Community Informatics*, *7*(3) (cit. on p. 41).

Bolloju, N., Alter, S., Gupta, A., Gupta, S., & Jain, S. (2017). Improving scrum user stories and product backlog using work system snapshots. *AMCIS* (cit. on pp. 74–76).

Bönisch, C., Sargeant, A., Wulff, A., Parciak, M., Bauer, C. R., & Sax, U. (2019). Fairness of openehr archetypes and templates. *SWAT4HCLS*, 102–111 (cit. on p. 63).

Bonney, W. (2013). Applicability of business intelligence in electronic health record. *Procedia-Social and Behavioral Sciences*, *73*, 257–262 (cit. on p. 13).

Bourne, R., Steinmetz, J. D., Flaxman, S., Briant, P. S., Taylor, H. R., Resnikoff, S., Casson, R. J., Abdoli, A., Abu-Gharbieh, E., Afshin, A., et al. (2021). Trends in prevalence of blindness and distance and near vision impairment over 30 years: An analysis for the global burden of disease study. *The Lancet global health*, *9*(2), e130–e143 (cit. on pp. 104, 105).

Bourne, R. R., Flaxman, S. R., Braithwaite, T., Cicinelli, M. V., Das, A., Jonas, J. B., Keeffe, J., Kempen, J. H., Leasher, J., Limburg, H., et al. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: A systematic review and meta-analysis. *The Lancet Global Health*, *5*(9), e888–e897 (cit. on p. 105).

Brall, C., Schröder-Bäck, P., & Maeckelberghe, E. (2019). Ethical aspects of digital health from a justice point of view. *European journal of public health*, *29*(Supplement_3), 18–22 (cit. on p. 62).

Brown, R. L., & Barrett, A. E. (2011). Visual impairment and quality of life among older adults: An examination of explanations for the relationship. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *66*(3), 364–373 (cit. on pp. 104, 105).

Buna, S. (2017). Yes, react is taking over front-end development. *The question is why* (cit. on pp. 78, 79).

Butler, C. E., Noel, S., Hibbs, S. P., Miles, D., Staves, J., Mohaghegh, P., Altmann, P., Curnow, E., & Murphy, M. F. (2015). Implementation of a clinical decision support system improves compliance with restrictive transfusion policies in hematology patients. *Transfusion*, *55*(8), 1964–1971 (cit. on p. 16).

Canova-Barrios, C., & Machuca-Contreras, F. (2022). Interoperability standards in health information systems: Systematic review. *Seminars in Medical Writing and Education*, *1*, 7–7 (cit. on pp. 21, 25).

Cardoso, L., Marins, F., Quintas, C., Portela, F., Santos, M., Abelha, A., & Machado, J. (2018). Interoperability in healthcare. In *Health care delivery and clinical science: Concepts, methodologies, tools, and applications* (pp. 689–714). IGI Global. (Cit. on pp. 20, 22).

Carvalho, M., Bandiera-Paiva, P., Marques, E., & Machado, J. M. (2021). Health information systems (his) privacy restrictions for gdpr: Assessing initial impacts perceived by patients and healthcare professionals. *International Journal of Reliable and Quality E-Healthcare (IJRQEH)*, *10*(2), 4–16 (cit. on p. 62).

Castaneda, C., Nalley, K., Mannion, C., Bhattacharyya, P., Blake, P., Pecora, A., Goy, A., & Suh, K. S. (2015). Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. *Journal of clinical bioinformatics*, *5*, 1–16 (cit. on p. 16).

Connolly, T. M., & Begg, C. E. (2005). *Database systems: A practical approach to design, implementation, and management*. Pearson Education. (Cit. on p. 77).

Coronel, C., & Morris, S. (2016). *Database systems: Design, implementation, & management*. Cengage Learning. (Cit. on p. 77).

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future healthcare journal*, *6*(2), 94 (cit. on p. 53).

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319–340 (cit. on p. 47).

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management science*, *35*(8), 982–1003 (cit. on p. 47).

Davis, G. B., & Olson, M. H. (1984). *Management information systems: Conceptual foundations, structure, and development*. McGraw-Hill, Inc. (Cit. on p. 12).

De Vries, N., Lavreysen, O., Boone, A., Bouman, J., Szemik, S., Baranski, K., Godderis, L., & De Winter, P. (2023). Retaining healthcare workers: A systematic review of strategies for sustaining power in the workplace. *Healthcare*, *11*(13), 1887 (cit. on p. 25).

Degoulet, P. (2014). Hospital information systems. In *Medical informatics, e-health* (pp. 289–313). Springer. (Cit. on p. 14).

de Mello, B. H., Rigo, S. J., da Costa, C. A., da Rosa Righi, R., Donida, B., Bez, M. R., & Schunke, L. C. (2022). Semantic interoperability in health records standards: A systematic literature review. *Health and technology*, *12*(2), 255–272 (cit. on pp. 21, 23, 27).

de Moraes, J. L. C., de Souza, W. L., Pires, L. F., & do Prado, A. F. (2016). A methodology based on openehr archetypes and software agents for developing e-health applications reusing legacy systems. *Computer methods and programs in biomedicine*, *134*, 267–287 (cit. on p. 108).

Dobrow, M. J., Bytautas, J. P., Tharmalingam, S., & Hagens, S. (2019). Interoperable electronic health records and health information exchanges: Systematic review. *JMIR medical informatics*, *7*(2), e12607 (cit. on pp. 105, 106).

Duarte, J., Salazar, M., Quintas, C., Santos, M., Neves, J., Abelha, A., & Machado, J. (2010). Data quality evaluation of electronic health records in the hospital admission process. *Computer and Information Science (ICIS), 2010 IEEE/ACIS 9th International Conference on*, 201–206 (cit. on p. 15).

Feeney, O., Werner-Felmayer, G., Siipi, H., Frischhut, M., Zullo, S., Barteczko, U., Øystein Ursin, L., Linn, S., Felzmann, H., Krajnović, D., et al. (2020). European electronic personal health records initiatives and vulnerable migrants: A need for greater ethical, legal and social safeguards. *Developing World Bioethics*, *20*(1), 27–37 (cit. on p. 62).

Ferreira, D., Silva, S., Abelha, A., & Machado, J. (2020). Recommendation system using autoencoders. *Applied Sciences*, *10*(16), 5510 (cit. on p. 128).

Finlay, P. N. (1989). *Introducing decision support systems*. NCC Blackwell. (Cit. on p. 16).

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. (Cit. on p. 47).

Fisher, B., & Thompson, M. (2022). From raw information to strategic insights: A multi-faceted investigation into the challenges and opportunities of big data analytics in contemporary organizations. *International Journal of Electrical and Electronics Engineering Studies*, *8*(1), 88–95 (cit. on p. 12).

Flaxman, S. R., Bourne, R. R., Resnikoff, S., Ackland, P., Braithwaite, T., Cicinelli, M. V., Das, A., Jonas, J. B., Keeffe, J., Kempen, J. H., et al. (2017). Global causes of blindness and distance vision impairment 1990–2020: A systematic review and meta-analysis. *The Lancet Global Health*, *5*(12), e1221–e1234 (cit. on p. 105).

Floridi, L. (2008). *The blackwell guide to the philosophy of computing and information*. John Wiley & Sons. (Cit. on pp. 17, 18).

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., et al. (2018). Ai4people—an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707 (cit. on p. 19).

Flyvbjerg, B. (2011). Case study. *The Sage handbook of qualitative research*, *4*, 301–316 (cit. on p. 43).

Foshay, N., & Kuziemsky, C. (2014). Towards an implementation framework for business intelligence in healthcare. *International Journal of Information Management*, *34*(1), 20–27 (cit. on pp. 16, 17).

Frade, S., Beale, T., & Cruz-Correia, R. J. (2022). Openehr implementation guide: Towards standard low-code healthcare systems. In *Medinfo 2021: One world, one health–global partnership for digital innovation* (pp. 52–55). IOS Press. (Cit. on p. 108).

Frexia, F., Mascia, C., Lianas, L., Delussu, G., Sulis, A., Meloni, V., Del Rio, M., & Zanetti, G. (2021). Openehr is fair-enabling by design. In *Public health and informatics* (pp. 113–117). IOS Press. (Cit. on p. 63).

Garcia, L. A., OliveiraJr, E., & Morandini, M. (2022). Tailoring the scrum framework for software development: Literature mapping and feature-based support. *Information and Software Technology*, *146*, 106814 (cit. on pp. xv, 74–76).

Genesereth, M. R., & Nilsson, N. J. (2012). *Logical foundations of artificial intelligence*. Morgan Kaufmann. (Cit. on p. 17).

Gillenson, M. L., Sherrell, D. L., et al. (2002). Enticing online consumers: An extended technology acceptance perspective. *Information & management*, *39*(8), 705–719 (cit. on p. 47).

Gonçalves-Ferreira, D., Sousa, M., Bacelar-Silva, G. M., Frade, S., Antunes, L. F., Beale, T., Cruz-Correia, R., et al. (2019). Openehr and general data protection regulation: Evaluation of principles and requirements. *JMIR medical informatics*, *7*(1), e9845 (cit. on p. 63).

Goodman, K. W. (2020). Ethics in health informatics. *Yearbook of medical informatics*, *29*(01), 026–031 (cit. on p. 64).

Greenwald, R., Stackowiak, R., & Stern, J. (2013). *Oracle essentials: Oracle database 12c.* " O'Reilly Media, Inc." (Cit. on p. 77).

Hasman, A., Haux, R., & Albert, A. (1996). A systematic view on medical informatics. *Computer methods and programs in biomedicine*, *51*(3), 131–139 (cit. on p. 14).

Haux, R. (2006a). Health information systems–past, present, future. *International journal of medical informatics*, *75*(3-4), 268–281 (cit. on pp. 14, 15).

Haux, R. (2006b). Individualization, globalization and health–about sustainable information technologies and the aim of medical informatics. *International journal of medical informatics*, *75*(12), 795–808 (cit. on p. 14).

Haux, R. (2010). Medical informatics: Past, present, future. *International journal of medical informatics*, *79*(9), 599–610 (cit. on p. 14).

Haux, R., Ammenwerth, E., Winter, A., & Brigl, B. (2004). *Strategic information management in hospitals: An introduction to hospital information systems*. Springer Science & Business Media. (Cit. on p. 14).

Hawken, S., Han, H., & Pettit, C. (2020). *Open cities| open data: Collaborative cities in the information era*. Springer. (Cit. on p. 11).

Heale, R., & Twycross, A. (2018). What is a case study? (Cit. on p. 43).

Hevner, A., & Chatterjee, S. (2010). Design science research in information systems. In *Design research in information systems* (pp. 9–22). Springer. (Cit. on pp. 41, 42).

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2008). Design science in information systems research. *Management Information Systems Quarterly*, *28*(1), 6 (cit. on p. 41).

Holden, R. J., & Karsh, B.-T. (2010). The technology acceptance model: Its past and its future in health care. *Journal of biomedical informatics*, *43*(1), 159–172 (cit. on pp. xiii, 47, 48, 51).

Hollins, M. (2022). *Understanding blindness: An integrative approach*. Routledge. (Cit. on pp. 104, 105).

Ide, N., & Pustejovsky, J. (2010). What does interoperability mean, anyway? toward an operational definition of interoperability for language technology. *Proceedings of the Second International Conference on Global Interoperability for Language Resources. Hong Kong, China* (cit. on p. 20).

Ingram, D. (2016). Origins of openehr. Retrieved 2024-03-03, from https://www.openehr.org/about/origins (cit. on p. 28).

Jaleel, A., Mahmood, T., Hassan, M. A., Bano, G., & Khurshid, S. K. (2020). Towards medical data interoperability through collaboration of healthcare devices. *IEEE Access*, *8*, 132302–132319 (cit. on p. 20).

Kalra, D., Beale, T., & Heard, S. (2005). The openehr foundation. *Studies in health technology and informatics*, *115*, 153–173 (cit. on pp. 28, 54).

Keen, P. G. (1978). Decision support systems: The next decade. *Decision Support Systems*, *3*, 253–265 (cit. on p. 16).

Kim, J. (2018). *Philosophy of mind*. Routledge. (Cit. on p. 18).

King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & management*, *43*(6), 740–755 (cit. on p. 47).

Kirch, W. (2008). *Encyclopedia of public health: Volume 1: A-h volume 2: I-z*. Springer Science & Business Media. (Cit. on p. 20).

Lai, P. (2017). The literature review of technology adoption models and theories for the novelty technology. *JISTEM-Journal of Information Systems and Technology Management*, *14*(1), 21–38 (cit. on p. 48).

Laudon, K. C., & Laudon, J. P. (2004). *Management information systems: Managing the digital firm*. Pearson Educación. (Cit. on p. 12).

Legris, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? a critical review of the technology acceptance model. *Information & management*, *40*(3), 191–204 (cit. on p. 47).

Lenz, R., & Reichert, M. (2007). It support for healthcare processes–premises, challenges, perspectives. *Data & Knowledge Engineering*, *61*(1), 39–58 (cit. on pp. 14–16).

Leslie, H. (2020). Openehr archetype use and reuse within multilingual clinical data sets: Case study. *Journal of medical Internet research*, *22*(11), e23361 (cit. on p. 30).

Li, Y., Duan, Y., Maamar, Z., Che, H., Spulber, A.-B., & Fuentes, S. (2021). Swarm differential privacy for purpose-driven data-information-knowledge-wisdom architecture. *Mobile Information Systems*, *2021*, 1–15 (cit. on p. 12).

Lippeveld, T., Sauerborn, R., Bodart, C., et al. (2000). *Design and implementation of health information systems*. Citeseer. (Cit. on p. 15).

Machado, J., Oliveira, L., Barreiro, L., Pinto, S., & Coimbra, A. (2016). Applying soft computing to clinical decision support. In *Applying business intelligence to clinical and healthcare organizations* (pp. 256–271). IGI Global. (Cit. on p. 16).

Maddox, T. M., Rumsfeld, J. S., & Payne, P. R. (2019). Questions for artificial intelligence in health care. *Jama*, *321*(1), 31–32 (cit. on p. 53).

Mansoor, M. E., & Majeed, R. (2010). Achieving interoperability among healthcare organizations. *Blekinge Institute of Technology* (cit. on p. 20).

March, S. T., & Storey, V. C. (2008). Design science in the information systems discipline: An introduction to the special issue on design science research. *MIS quarterly*, 725–730 (cit. on pp. 41, 42).

Martins, B., Ferreira, D., Neto, C., Abelha, A., & Machado, J. (2021). Data mining for cardiovascular disease prediction. *Journal of medical systems*, *45*, 1–8 (cit. on p. 128).

Mattila, J., Koikkalainen, J., Virkki, A., van Gils, M., Lötjönen, J., Initiative, A. D. N., et al. (2012). Design and application of a generic clinical decision support system for multiscale data. *IEEE Transactions on Biomedical Engineering*, *59*(1), 234–240 (cit. on p. 16).

Meredith, J., Whitehead, N., & Dacey, M. (2023). Aligning semantic interoperability frameworks with the foxs stack for fair health data. *Methods of Information in Medicine*, *62*(S 01), e39–e46 (cit. on p. 62).

Milosevic, Z. (2019). Ethics in digital health: A deontic accountability framework. *2019 IEEE 23rd International Enterprise Distributed Object Computing Conference (EDOC)*, 105–111 (cit. on p. 62).

Min, L., Atalag, K., Tian, Q., Chen, Y., & Lu, X. (2021). Verifying the feasibility of implementing semantic interoperability in different countries based on the openehr approach: Comparative study of acute coronary syndrome registries. *JMIR Medical Informatics*, *9*(10), e31288 (cit. on pp. 28, 30, 106).

Minnick, C. (2022). *Beginning reactjs foundations building user interfaces with reactjs: An approachable guide*. John Wiley & Sons. (Cit. on pp. 78–80).

Miranda, M., Salazar, M., Portela, F., Santos, M., Abelha, A., Neves, J., & Machado, J. (2012). Multi-agent systems for hl7 interoperability services. *Procedia Technology*, *5*, 725–733 (cit. on p. 15).

Moner, D., Maldonado, J. A., & Robles, M. (2018). Archetype modeling methodology. *Journal of biomedical informatics*, *79*, 71–81 (cit. on p. 108).

Moor, J. (2006). The dartmouth college artificial intelligence conference: The next fifty years. *Ai Magazine*, *27*(4), 87–87 (cit. on p. 17).

Moor, J. H. (2005). Why we need better ethics for emerging technologies. *Ethics and information technology*, *7*(3), 111–119 (cit. on p. 18).

Muguira, J. A., & Tolk, A. (2006). Applying a methodology to identify structural variances in interoperations. *The Journal of Defense Modeling and Simulation*, *3*(2), 77–93 (cit. on pp. xiii, xv, 22, 23).

Müller, S. (2022). Is there a civic duty to support medical ai development by sharing electronic health records? *BMC Medical Ethics*, *23*(1), 134 (cit. on p. 64).

Musen, M. A., Middleton, B., & Greenes, R. A. (2014). Clinical decision-support systems. In *Biomedical informatics* (pp. 643–674). Springer. (Cit. on p. 16).

Neto, C., Pontes, G., Domente, A., Reinolds, F., Costa, J., Ferreira, D., & Machado, J. (2021). Step towards predicting patient length of stay in intensive care units. *World Conference on Information Systems and Technologies*, 287–297 (cit. on p. 129).

Oliveira, D., Ferreira, D., Abreu, N., Leuschner, P., Abelha, A., & Machado, J. (2022). Prediction of covid-19 diagnosis based on openehr artefacts. *Scientific Reports*, *12*(1), 12549 (cit. on p. 129).

openEHR Foundation. (n.d.). Openehr specifications. Retrieved 2024-03-03, from https://specifications.openehr.org/ (cit. on pp. 28, 30).

openEHR Specifications. (latest). Openehr reference model. Retrieved 2024-03-03, from https://specifications.openehr.org/releases/RM/latest/ehr.html (cit. on pp. xiii, 29).

Organization, W. H., et al. (2022). *Package of eye care interventions*. World Health Organization. (Cit. on pp. 104, 105).

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., et al. (2021). The prisma 2020 statement: An updated guideline for reporting systematic reviews. *Bmj*, *372* (cit. on p. 55).

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., et al. (2021). Prisma 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *bmj*, *372* (cit. on p. 56).

Pannu, A. (2015). Artificial intelligence and its application in different areas. *Artificial Intelligence*, *4*(10), 79–84 (cit. on p. 17).

Parciak, M., Suhr, M., Schmidt, C., Bönisch, C., Löhnhardt, B., Kesztyüs, D., & Kesztyüs, T. (2023). Fairness through automation: Development of an automated medical data integration infrastructure for fair health data in a maximum care university hospital. *BMC Medical Informatics and Decision Making*, *23*(1), 94 (cit. on p. 62).

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), 45–77 (cit. on pp. xiii, 42).

Peixoto, H., Santos, M., Abelha, A., & Machado, J. (2012). Intelligence in interoperability with aida. *International Symposium on Methodologies for Intelligent Systems*, 264–273 (cit. on p. 15).

Piasecki, J., Walkiewicz-arek, E., Figas-Skrzypulec, J., Kordecka, A., & Dranseika, V. (2021). Ethical issues in biomedical research using electronic health records: A systematic review. *Medicine, Health Care and Philosophy*, *24*(4), 633–658 (cit. on p. 62).

Pikkarainen, T., Pikkarainen, K., Karjaluoto, H., & Pahnila, S. (2004). Consumer acceptance of online banking: An extension of the technology acceptance model. *Internet research*, *14*(3), 224–235 (cit. on p. 47).

Pineda, R. (2020). World corneal blindness. *Foundations of Corneal Disease: Past, Present and Future*, 299–305 (cit. on p. 105).

Plug, R., Liang, Y., Basajja, M., Aktau, A., Jati, P. H. P., Amare, S. Y., Taye, G. T., Mpezamihigo, M., Oladipo, F., & van Reisen, M. (2022). Fair and gdpr compliant population health data generation, processing and analytics. *SWAT4HCLS*, 54–63 (cit. on p. 62).

Power, D. J. (2008). Understanding data-driven decision support systems. *Information Systems Management*, *25*(2), 149–154 (cit. on p. 16).

Prins, N. W., Monteleone, R., Soldado-Magraner, J., Nash, J. E., Young, M. J., & Cabrera, L. (2022). Reexamining the ethical, legal, social, and cultural implications for cochlear implants through a novel neuroethics framework. *2022 IEEE International Symposium on Technology and Society (ISTAS)*, *1*, 1–6 (cit. on p. 11).

Puyt, R., Lie, F. B., De Graaf, F. J., & Wilderom, C. P. (2020). Origins of swot analysis. *Academy of management proceedings*, *2020*(1), 17416 (cit. on p. 45).

Puyt, R. W., Lie, F. B., & Wilderom, C. P. (2023). The origins of swot analysis. *Long Range Planning*, *56*(3), 102304 (cit. on p. 45).

Ramakrishnan, R., & Gehrke, J. (2000). *Database management systems*. McGraw Hill. (Cit. on p. 77).

Rawat, P., & Mahajan, A. N. (2020). Reactjs: A modern web development framework. *International Journal of Innovative Science and Research Technology*, *5*(11), 698–702 (cit. on p. 80).

Razzak, M. I., Imran, M., & Xu, G. (2020). Big data analytics for preventive medicine. *Neural Computing and Applications*, *32*(9), 4417–4451 (cit. on p. 53).

Ruotsalainen, P., & Blobel, B. (2019). How does gdpr support healthcare transformation to 5p medicine? *MEDINFO 2019: Health and Wellbeing e-Networks for All-Proceedings of the 17th World Congress on Medical and Health Informatics* (cit. on p. 62).

S. Rubí, J. N., & L. Gondim, P. R. (2019). Iomt platform for pervasive healthcare data aggregation, processing, and sharing based on onem2m and openehr. *Sensors*, *19*(19), 4283 (cit. on p. 108).

Schmidt, B. (2006). Proof of principle studies. *Epilepsy research*, *68*(1), 48–52 (cit. on pp. 44, 45).

Schwaber, K., & Sutherland, J. (2011). The scrum guide. *Scrum Alliance*, *21*(1), 1–38 (cit. on pp. xiii, 76).

Sensmeier, J. (2022). Interoperability: There is no digital health without health it standards. In *Nursing informatics: A health informatics, interprofessional and global perspective* (pp. 285–302). Springer. (Cit. on p. 21).

Sergey, A. B., Alexandr, D. B., & Sergey, A. T. (2015). Proof of concept center—a promising tool for innovative development at entrepreneurial universities. *Procedia-Social and Behavioral Sciences*, *166*, 240–245 (cit. on p. 44).

Shakya, S., & Nepal, L. (2020). Computational enhancements of wearable healthcare devices on pervasive computing system. *Journal of Ubiquitous Computing and Communication Technologies (UCCT)*, *2*(02), 98–108 (cit. on p. 19).

Sharma, R., & Mishra, R. (2014). A review of evolution of theories and models of technology adoption. *Indore Manag. J*, *6*(2), 17–29 (cit. on p. 48).

Shirabad, J. S., Wilk, S., Michalowski, W., & Farion, K. (2012). Implementing an integrative multi-agent clinical decision support system with open source software. *Journal of medical systems*, *36*(1), 123–137 (cit. on p. 16).

Shortliffe, E. H., Shortliffe, E. H., Cimino, J. J., & Cimino, J. J. (2014). *Biomedical informatics: Computer applications in health care and biomedicine*. Springer. (Cit. on p. 3).

Singh, R., Gupta, N., Vanathi, M., & Tandon, R. (2019). Corneal transplantation in the modern era. *Indian Journal of Medical Research*, *150*(1), 7–22 (cit. on p. 105).

Sousa, M., Ferreira, D. N. G., Pereira, C. S., Bacelar, G., Frade, S., Pestana, O., & Correia, R. C. (2018). Openehr based systems and the general data protection regulation (gdpr). *Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth* (cit. on p. 63).

Steinmetz, J. D., Bourne, R. R., Briant, P. S., Flaxman, S. R., Taylor, H. R., Jonas, J. B., Abdoli, A. A., Abrha, W. A., Abualhasan, A., Abu-Gharbieh, E. G., et al. (2021). Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to vision 2020: The right to sight: An analysis for the global burden of disease study. *The Lancet Global Health*, *9*(2), e144–e160 (cit. on p. 105).

Takahashi, A. R. W., & Araujo, L. (2020). Case study research: Opening up research opportunities. *RAUSP Management Journal*, *55*, 100–111 (cit. on p. 43).

Thiel, R., Lupiáñez-Villanueva, F., Deimel, L., Gunderson, L., & Sokolyanskaya, A. (2021). Ehealth, interoperability of health data and artificial intelligence for health and care in the eu: Interoperability of electronic health records in the eu. *The European Commission* (cit. on p. 68).

Tolk, A., Diallo, S. Y., & Turnitsa, C. D. (2007). Applying the levels of conceptual interoperability model in support of integratability, interoperability, and composability for system-of-systems engineering. *Journal of Systems, Cybernetics, and Informatics*, *5*(5) (cit. on pp. xv, 22, 23).

Torab-Miandoab, A., Samad-Soltani, T., Jodati, A., & Rezaei-Hachesu, P. (2023). Interoperability of heterogeneous health information systems: A systematic literature review. *BMC Medical Informatics and Decision Making*, *23*(1), 18 (cit. on pp. 21, 22, 24, 25).

Trowbridge, R., & Weingarten, S. (2001). Clinical decision support systems. *Making health care safer: a critical analysis of patient safety practices. Evidence Report/Technology Assessment*, (43) (cit. on p. 16).

Turban, E., Aronson, J., & Liang, T. (2004). Decision support systems and intelligent systems, 7th editionprentice hall inc. *New Jersey* (cit. on p. 16).

Turban, E. (1993). *Decision support and expert systems: Management support systems*. Prentice Hall PTR. (Cit. on p. 16).

Uslu, A., Stausberg, J., et al. (2021). Value of the electronic medical record for hospital care: Update from the literature. *Journal of medical Internet research*, *23*(12), e26323 (cit. on p. 105).

Vaishnavi, V. K., & Kuechler, W. (2015). *Design science research methods and patterns: Innovating information and communication technology*. Crc Press. (Cit. on p. 42).

Van Meter, H. J. (2020). Revising the dikw pyramid and the real relationship between data, information, knowledge, and wisdom. *Law, Technology and Humans*, *2*(2), 69–80 (cit. on p. 12).

Vassilakopoulou, P., & Hustad, E. (2023). Bridging digital divides: A literature review and research agenda for information systems research. *Information Systems Frontiers*, *25*(3), 955–969 (cit. on p. 13).

Vayena, E., Haeusermann, T., Adjekum, A., & Blasimme, A. (2018). Digital health: Meeting the ethical and policy challenges. *Swiss medical weekly*, *148*, w14571 (cit. on p. 62).

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information systems research*, *11*(4), 342–365 (cit. on p. 48).

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, *39*(2), 273–315 (cit. on pp. xiii, 48, 51).

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, *46*(2), 186–204 (cit. on p. 47).

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425–478 (cit. on p. 48).

Wager, K. A., Lee, F. W., & Glaser, J. P. (2021). *Health care information systems: A practical approach for health care management*. John Wiley & Sons. (Cit. on p. 53).

Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37 (cit. on p. 53).

Wang, W., Tolk, A., & Wang, W. (2009). The levels of conceptual interoperability model: Applying systems engineering principles to m&s. *arXiv preprint arXiv:0908.0191* (cit. on p. 22).

Wang, Y., Liu, C., Zhao, Y., Huang, W., You, B., & Lin, J. (2021). The new theory of learning in the era of educational information 2.0—connected constructivism. *Intelligent Computing and Block Chain: First BenchCouncil International Federated Conferences, FICC 2020, Qingdao, China, October 30–November 3, 2020, Revised Selected Papers 1*, 417–428 (cit. on p. 12).

Weber-Jahnke, J., Peyton, L., & Topaloglou, T. (2012). Ehealth system interoperability. *Information Systems Frontiers*, *14*(1), 1–3 (cit. on pp. 14, 15).

Winter, A., Ammenwerth, E., Haux, R., Marschollek, M., Steiner, B., & Jahn, F. (2023). *Health information systems: Technological and management perspectives*. Springer Nature. (Cit. on p. 12).

Wohlin, C. (2021). Case study research in software engineering—it is a case, and it is a study, but is it a case study? *Information and Software Technology*, *133*, 106514 (cit. on p. 43).

World Health Organization. (2019). *World report on vision*. (Cit. on pp. 104, 105).

Yang, L., Huang, X., & Li, J. (2019a). Discovering clinical information models online to promote interoperability of electronic health records: A feasibility study of openehr. *Journal of medical Internet research*, *21*(5), e13504 (cit. on pp. 28, 30).

Yang, L., Huang, X., & Li, J. (2019b). Discovering clinical information models online to promote interoperability of electronic health records: A feasibility study of openehr. *Journal of medical Internet research*, *21*(5), e13504 (cit. on pp. 106, 108).

Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). sage. (Cit. on p. 43).

A

Questionnaire for Assessing the Acceptance of the FormBuilder Artefact

This appendix chapter presents the TAM3 questionnaire developed to assess the acceptance of the proposed solution. The questionnaire was developed based on the TAM3 model, which is an extension of the original TAM model. The questionnaire was developed and distributed in Portuguese and was translated into English for the purpose of this thesis. It is important to remember that all items of the questionnaires are measured on a 5-point Likert scale, where:

1. Not satisfied / Full disagreement
2. Unsatisfied / Some level of disagreement
3. Satisfied / Indifferent
4. Fairly satisfied / Partially agree
5. Completely satisfied / Strongly agree

Table 27 presents the questionnaire structure based on the TAM3 constructors for the artefact. The questionnaire is divided into ten sections, each corresponding to a TAM3 constructor. Each section contains a set of questions that assess the user's perception of the artefact.

Table 27: Structure of the survey questionnaire based on TAM3 constructors for assessing the FormBuilder

Question	Description	1	2	3	4	5
	1. Perceived Ease Of Use (PEOU)					
1.1	The application is easy to use.	○	○	○	○	○
1.2	The application is rigid and inflexible to interact with.	○	○	○	○	○

(Table 27 continued on next page)

(Table 27 continued from previous page)

Question	Description	1	2	3	4	5
1.3	The system provides helpful guidance in creating forms based on openEHR templates and filling them in.	○	○	○	○	○
1.4	The application makes it easy to recognise key information.	○	○	○	○	○
1.5	The application displays visually pleasing design.	○	○	○	○	○
1.6	Each display page focuses on a single topic.	○	○	○	○	○
1.7	I can determine my position within the application.	○	○	○	○	○
1.8	The application uses consistent and understandable graphics.	○	○	○	○	○
1.9	The display pages within the application are easy to read.	○	○	○	○	○
1.10	The application uses understandable terms.	○	○	○	○	○
1.11	The information I need is easy to find within the application.	○	○	○	○	○
1.12	The application loads quickly.	○	○	○	○	○
	2. Perceived Usefulness (PU)					
2.1	Using this application gives me greater control over my work.	○	○	○	○	○
2.2	Using this application makes it easier to do my job.	○	○	○	○	○
2.3	I find the system to be useful in my job.	○	○	○	○	○
2.4	The application enables me to accomplish tasks more quickly.	○	○	○	○	○
2.5	The application supports critical aspects of my job.	○	○	○	○	○
2.6	The application addresses my job-related needs.	○	○	○	○	○
2.7	The application enables decisions based on better evidence.	○	○	○	○	○
	3. Behavioural Intention to Use (BIU)					
3.1	I identify this tool as a standard everyday application.	○	○	○	○	○
	4. Experience (EXP)					
4.1	I have had experience with systems similar to this.	○	○	○	○	○
	5. Job Relevance (REL)					
5.1	The use of the system is pertinent to my various job-related tasks.	○	○	○	○	○
	6. Output Quality (OUT)					
6.1	The quality of the output I get from the system is high.	○	○	○	○	○
6.2	The presentation of the system information is adequate.	○	○	○	○	○

(Table 27 continued on next page)

(Table 27 continued from previous page)

Question	Description	1	2	3	4	5
6.3	The amount of information presented in the system is adequate.	○	○	○	○	○
	7. Result Demonstrability (RES)					
7.1	The results of using the system are important to my job.	○	○	○	○	○
7.2	Using the system will simplify decision making in my work.	○	○	○	○	○
	8. Computer Self-Efficacy (CSE)					
8.1	I could complete the job using the application if there was no one around to tell me what to do as I go.	○	○	○	○	○
8.2	I could complete the job using the application if someone showed me how to do it first.	○	○	○	○	○
8.3	I could complete the job using the application if I had used similar applications before this one to do the same job.	○	○	○	○	○
8.4	I could complete the job using the application if I had just the built-in help facility for assistance.	○	○	○	○	○
	9. Perceptions of External Constrol (PEC)					
9.1	I have control over using the system.	○	○	○	○	○
9.2	The system is not compatible with other systems I use.	○	○	○	○	○
	10. Perceived Enjoyment (ENJ)					
10.1	The use of the application is enjoyable / beneficial.	○	○	○	○	○

B

Scientific Contributions

Throughout the course of this doctoral thesis, the author has contributed to several scientific publications, both in conferences and journals. It is important to note that some of these publications are still under review or awainting publication. Nonetheless, the status of each publication is clearly stated. The following sections detail the papers where the author has contributed, either as the main author or as a co-author.

B.1 Journal Publications

B.1.1 Standardizing Corneal Transplantation Records with openEHR: Case Study

Authors: ***Diana Ferreira***, Cristiana Neto, Francini Hak, António Abelha, Manuel Santos, José Machado.

Title: Standardizing Corneal Transplantation Records with openEHR: Case Study

Abstract: **Background:** Corneal transplantation, also known as Keratoplasty, is a widely performed surgical procedure that aims to restore vision in patients with corneal damage. The success of corneal transplantation relies on the accurate and timely management of patient information, which can be enhanced with Electronic Health Records (EHRs). However, conventional EHRs are often fragmented and lack standardization, leading to difficulties in information access and sharing, increased medical errors, and decreased patient safety. In the wake of these problems, there is a growing demand for standardized EHRs that can ensure the accuracy and consistency of patient data across healthcare organizations.
Objective: This paper proposes the use of openEHR structures for standardizing corneal transplantation records. The main objective of this research is to improve the quality and interoperability of EHRs in corneal transplantation, making it easier for healthcare providers to capture, share, and analyze clinical information.
Methods: A series of sequential steps were carried out in this study to implement standardized clinical records using openEHR specifications. These specifications furnish a methodical approach that ascertains

the development of high-quality clinical records. In broad terms, the methodology followed encompasses the conduction of meetings with healthcare professionals and the modeling of archetypes, templates, forms, decision rules, and work plans.

Results: This research resulted in a tailored solution that streamlines healthcare delivery and meets the needs of medical professionals involved in the corneal transplantation process, while seamlessly aligning with contemporary clinical practices. The proposed solution culminated in the successful integration within a Portuguese hospital of three key components of openEHR specifications: forms, Decision Logic Modules (DLMs), and Work Plans (WPs). A statistical analysis of data collected from May 1st, 2022 to March 31st, 2023 allowed to perceive the usage of the new technologies within the corneal transplantation workflow. Despite the completion rate being only 64%, which can be explained by external factors such as patient health and availability of donor organs, there was an overall improvement in terms of task control and follow-up of the patients' clinical process.

Conclusions: This study shows that the adoption of openEHR structures represents a significant step forward in the standardization and optimization of corneal transplantation records. It offers a detailed demonstration of how to implement openEHR specifications and highlights the different advantages of standardizing EHRs in the field of corneal transplantation. Furthermore, it serves as a valuable reference for researchers and practitioners who are interested in advancing and improving the exploitation of EHRs in healthcare.

Keywords: Eletronic Health Record; Corneal Transplantation; Keratoplasty; openEHR; Data Representation; Data Exchange; Templates; Archetypes; Forms.

Submission Date: April, 2023

Journal: Journal of Medical Internet Research (JMIR)
SJR 2023: 2.02 (Q1 - Health Informatics)

Status of Publication: 2nd round of reviews

DOI: N/A

B.1.2 Automation of Hospital Workflows Using International Standards: a Case Study With Imaging Exams

Authors: Cristiana Neto, Francini Hak, ***Diana Ferreira***, António Abelha, José Machado.

Title: Automation of Hospital Workflows Using International Standards: a Case Study With Imaging Exams

Abstract: **Background:** Hospital workflow automation holds paramount importance in modern healthcare systems. By streamlining complex processes, it is possible to enhance operational efficiency and patient care. Automated workflows aligned with open standards, such as openEHR, enable accurate data capture, seamless communication, and timely task execution among healthcare professionals. This leads to reduced errors, improved patient outcomes, and optimized resource utilization.
Objective: The main goal of this study is to improve and automate the medical imaging exams workflow of a Portuguese hospital through the implementation of clinical standards.
Methods: This work followed the openEHR specifications using a case study research design for the implementation of a workflow automation system in a real healthcare environment. Although the main focus is the task planning specification, the document also approaches other openEHR components, such as forms and decision rules.
Results: After implementing the proposed solution, information about the screening process of more than 6800 requests for imaging exams was gathered, revealing the willingness of the health professionals to use the tools provided. The relevance of the solution was sustained by the statistical study carried out, which showed a significant decrease in waiting times for outsourcing management, with approximately 30,857 fewer waiting days on average in 2022 compared to 2019.
Conclusions: This study shows the development of a new approach to the imaging exams workflow using the openEHR specifications achieving an automated process with structured data generation. The authors believe that the developed system represents an important step towards efficient knowledge discovery from the registry of quality and structured data.

Keywords: Health Information Systems; Hospital Workflow; openEHR; Task Planning.

Submission Date: August, 2023

Journal: International Journal of Medical Informatics
SJR 2023: 1.11 (Q1 - Health Informatics)

Status of Publication: In review

DOI: N/A

B.1.3 An Interoperable Approach to Support Clinical Pathway Decisions for Nutritional Screening using openEHR

Authors: Francini Hak, ***Diana Ferreira***, Cristiana Neto, Ricardo Marinho, António Abelha, José Machado, Manuel Santos.

Title: An Interoperable Approach to Support Clinical Pathway Decisions for Nutritional Screening using openEHR

Abstract: In today's healthcare landscape, an increasing number of hospitalized patients are being recognized as being at risk of malnutrition. To deal with this problematic, Clinical Decision Support Systems (CDSS) are being adopted to assist the decision-making process. In this context, this article explores an interoperable approach to support clinical pathway decisions for malnutrition screening by utilizing the openEHR standard. The study presents an architecture that integrates different openEHR components designed to define workflow tasks, create decision rules, and visualize forms through archetypes and templates. To facilitate the malnutrition pathway implementation within a Portuguese hospital center, the clinical guidelines Nutritional Risk Screening (NRS 2002) and Global Leadership Initiative on Malnutrition (GLIM) were employed. As a result, three forms and three decision modules were developed following the proposed architecture. This approach streamlines the integration of data from various sources and holds the potential to enhance coordination among healthcare professionals involved in a patient's care.

Keywords: Clinical Decision Support Systems; Interoperability; Workflow; Nutritional Screening; openEHR.

Submission Date: January, 2024

Journal: Scientific Reports - Nature Portfolio
SJR 2023: 0.9 (Q1 - Multidisciplinary)

Status of Publication: In review

DOI: N/A

B.1.4 Prediction of COVID-19 diagnosis based on openEHR artefacts

Authors: Daniela Oliveira, ***Diana Ferreira***, Nuno Abreu, Pedro Leuschner, António Abelha, José Machado.

Title: Prediction of COVID-19 diagnosis based on openEHR artefacts

Abstract: Nowadays, we are facing the worldwide pandemic caused by COVID-19. The complexity and momentum of monitoring patients infected with this virus calls for the usage of agile and scalable data structure methodologies. OpenEHR is a healthcare standard that is attracting a lot of attention in recent years due to its comprehensive and robust architecture. The importance of an open, standardized and adaptable approach to clinical data lies in extracting value to generate useful knowledge that really can help healthcare professionals make an assertive decision. This importance is even more accentuated when

facing a pandemic context. Thus, in this study, a system for tracking symptoms and health conditions of suspected or confirmed SARS-CoV-2 patients from a Portuguese hospital was developed using openEHR. All data on the evolutionary status of patients in home care as well as the results of their COVID-19 test were used to train different ML algorithms, with the aim of developing a predictive model capable of identifying COVID-19 infections according to the severity of symptoms identified by patients. The CRISP-DM methodology was used to conduct this research. The results obtained were promising, with the best model achieving an accuracy of 96.25%, a precision of 99.91%, a sensitivity of 92.58%, a specificity of 99.92%, and an AUC of 0.963, using the Decision Tree algorithm and the Split Validation method. Hence, in the future, after further testing, the predictive model could be implemented in clinical decision support systems.

Keywords: COVID-19; openEHR; Data Mining; Predictive Models; Decision Tree.

Publication Date: July, 2022

Journal: Scientific Reports - Nature Portfolio
JCR IF 2022: 4.6 (Q2-MULTIDISCIPLINARY SCIENCES)
SJR 2022: 0.97 (Q1-Multidisciplinary)

Status of Publication: Published

DOI: 10.1038/s41598-022-15968-z

B.1.5 Data Mining for Cardiovascular Disease Prediction

Authors: Bárbara Martins, ***Diana Ferreira***, Cristiana Neto, António Abelha, José Machado.

Title: Data Mining for Cardiovascular Disease Prediction

Abstract: Cardiovascular diseases (CVDs) aredisorders of the heart and blood vessels and are a major cause of disability and premature death worldwide. Individuals at higher risk of developing CVD must be noticed at an early stage to prevent premature deaths. Advances in the field of computational intelligence, together with the vast amount of data produced daily in clinical settings, have made it possible to create recognition systems capable of identifying hidden patterns and useful information. This paper focuses on the application of Data Mining Techniques (DMTs) to clinical data collected during the medical examination in an attempt to predict whether or not an individual has a CVD. To this end, the CRossIndustry Standard Process for Data Mining (CRISP-DM) methodology was followed, in which five classifiers were applied, namely DT, Optimized DT, RI, RF, and DL. The models were mainly developed using the RapidMiner software with the assist of the WEKA tool and were analyzed based on accuracy, precision, sensitivity,

and specificity. The results obtained were considered promising on the basis of the research for effective means of diagnosing CVD, with the best model being Optimized DT, which achieved the highest values for all the evaluation metrics, 73.54%, 75.82%, 68.89%, 78.16% and 0.788 for accuracy, precision, sensitivity, specificity, and AUC, respectively.

Keywords: Cardiovascular disease; Health information systems; Decision support systems; Data mining; CRISP-DM; Classification.

Publication Date: January, 2021

Journal: Journal of Medical Systems
JCR IF 2021: 4.92 (Q2-Medical Informatics; Q1-Healthcare Sciences and Services)
SJR 2021: 1.13 (Q1-Information Systems)

Status of Publication: Published

DOI: 10.1007/s10916-020-01682-8

B.1.6 Different Scenarios for the Prediction of Hospital Readmission of Diabetic Patients

Authors: Cristiana Neto, Fábio Senra, Jaime Leite, Nuno Rei, Rui Rodrigues, ***Diana Ferreira***, José Machado.

Title: Different Scenarios for the Prediction of Hospital Readmission of Diabetic Patients

Abstract: Hospitals generate large amounts of data on a daily basis, but most of the time that data is just an overwhelming amount of information which never transitions to knowledge. Through the application of Data Mining techniques it is possible to find hidden relations or patterns among the data and convert those into knowledge that can further be used to aid in the decision-making of hospital professionals. This study aims to use information about patients with diabetes, which is a chronic (long-term) condition that occurs when the body does not produce enough or any insulin. The main purpose is to help hospitals improve their care with diabetic patients and consequently reduce readmission costs. An hospital readmission is an episode in which a patient discharged from a hospital is admitted again within a specified period of time (usually a 30 day period). This period allows hospitals to verify that their services are being performed correctly and also to verify the costs of these re-admissions. The goal of the study is to predict if a patient who suffers from diabetes will be readmitted, after being discharged, using Machine Leaning algorithms. The final results revealed that the most efficient algorithm was Random Forest with 0.898 of accuracy.

Keywords: Data mining; Diabetes; Weka; RapidMiner studio; Prediction; Readmission.

Publication Date: January, 2021

Journal: Journal of Medical Systems
JCR IF 2021: 4.92 (Q2-Medical Informatics; Q1-Healthcare Sciences and Services)
SJR 2021: 1.13 (Q1-Information Systems)

Status of Publication: Published

DOI: 10.1007/s10916-020-01686-4

B.1.7 Predicting the Survival of Primary Biliary Cholangitis Patients

Authors: ***Diana Ferreira***, Cristiana Neto, José Lopes, Júlio Duarte, António Abelha, José Machado.

Title: Predicting the Survival of Primary Biliary Cholangitis Patients

Abstract: Primary Biliary Cholangitis, which is thought to be caused by a combination of genetic and environmental factors, is a slow-growing chronic autoimmune disease in which the human body's immune system attacks healthy cells and tissues and gradually destroys the bile ducts in the liver. A reliable diagnosis of this clinical condition, followed by appropriate intervention measures, can slow the damage to the liver and prevent further complications, especially in the early stages. Hence, the focus of this study is to compare different classification Data Mining techniques, using clinical and demographic data, in an attempt to predict whether or not a Primary Biliary Cholangitis patient will survive. Data from 418 patients with Primary Biliary Cholangitis, following the Mayo Clinic's research between 1974 and 1984, were used to predict patient survival or non-survival using the Cross Industry Standard Process for Data Mining methodology. Different classification techniques were applied during this process, more specifically, Decision Tree, Random Tree, Random Forest, and Naïve Bayes. The model with the best performance used the Random Forest classifier and Split Validation with a ratio of 0.8, yielding values greater than 93% in all evaluation metrics. With further testing, this model may provide benefits in terms of medical decision support.

Keywords: Classification; Data Mining; Predictive Models; Primary Biliary Cholangitis.

Publication Date: August, 2022

Journal: Applied Sciences
JCR IF 2022: 2.7 (Q2-ENGINEERING, MULTIDISCIPLINARY)
SJR 2022: 0.49 (Q2-Engineering (Miscellaneous); Q3-Computer Science Applications)

Status of Publication: Published

DOI: 10.3390/app12168043

B.1.8 Integrating a New Generation of Interoperability Agents into the AIDA Platform

Authors: Francisca Nogueira, ***Diana Ferreira***, Regina Sousa, António Abelha, José Machado.

Title: Integrating a New Generation of Interoperability Agents into the AIDA Platform

Abstract: Health is an information rich and complex environment, which makes it essential to implement interoperability in different health organizations and the consequent homogeneity among Health Information Systems (HIS). The Agency for the Integration, Dissemination and Archiving of Medical and Clinical Information (AIDA) is a consistent agent monitoring platform capable of guaranteeing the automation of information as well as the interoperability and integration of HIS. This platform was designed as a solution to the information islands that are commonly found in hospital systems, and it is currently being used in several hospitals throughout Portugal. However, like any technological innovation, the solution requires a constant health technology assessment (HTA) to ensure the absence of obsolescence and a continued efficiency and security of the platform. Hence, this article focuses on the relevance and the need for vigilance, culminating in the restructuring of certain intelligent agents that make up the AIDA platform.

Keywords: Health Information Systems, Interoperability, Obsolescence, Health Technology Assessment.

Publication Date: June, 2021

Journal: Journal of Digital Science

Status of Publication: Published

DOI: 10.33847/2686-8296.3.1_5.

B.1.9 Recommendation System Using Autoencoders

Authors: ***Diana Ferreira***, Sofia Silva, António Abelha, José Machado.

Title: Recommendation System Using Autoencoders

Abstract: The magnitude of the daily explosion of high volumes of data has led to the emergence of the Big Data paradigm. The ever-increasing amount of information available on the Internet makes it increasingly difficult for individuals to find what they need quickly and easily. Recommendation systems have appeared as a solution to overcome this problem. Collaborative filtering is widely used in this type of systems, but high dimensions and data sparsity are always a main problem. With the idea of deep learning gaining more importance, several works have emerged to improve this type of filtering. In this article, a product recommendation system is proposed where an autoencoder based on a collaborative filtering method is employed. A comparison of this model with the Singular Value Decomposition is made and presented in the results section. Our experiment shows a very low Root Mean Squared Error (RMSE) value, considering that the recommendations presented to the users are in line with their interests and are not affected by the data sparsity problem as the datasets are very sparse, 0.996. The results are quite promising achieving an RMSE value of 0.029 in the first dataset and 0.010 in the second one.

Keywords: Big Data; Recommendation Systems; Collaborative Filtering; Autoencoders.

Publication Date: August, 2020

Journal: Applied Sciences
IF 2020: 2.68 (Q2-ENGINEERING, MULTIDISCIPLINAR)
SJR 2020: 0.44 (Q2-Computer Science Applications)

Status of Publication: Published

DOI: 10.3390/app10165510.

B.1.10 Monitoring and Maintenance of Web Service Processes in Health Units

Authors: ***Diana Ferreira***, Cristiana Neto, António Abelha.

Title: Monitoring and Maintenance of Web Service Processes in Health Units

Abstract: The delivery of quality health care is a demanding information technology challenge. The last years have witnessed profound changes in healthcare institutions to improve patient experience. Those changes focus on the needs of the health professionals that directly contact with the patients. However, it is equally important to aid the professionals that work in the background of the health units. This study proposes the monitoring and maintenance of web service processes that occur in the Centro Hospitalar do Porto (CHP) through an intuitive and user-friendly platform developed in ReactJS. These processes are responsible for critical tasks within the hospital and thus, the daily control of its activities ensure that there are no unnecessary failures, supporting an efficient health care delivery. The development of the application followed the Design Science Research (DSR) methodology and was submitted to a Strengths Weaknesses Opportunities and Threats (SWOT) analysis which results were considered optimistic.

Keywords: Centro Hospitalar do Porto; Control; Design Science Research; Healthcare; Information Technology; ReactJS; Web Application; Web Service Activities.

Publication Date: January, 2020

Journal: International Journal of Reliable and Quality E-Healthcare (IJRQEH) - Igi Global
SJR 2021: 0.215 (Q3- Medical Laboratory Technology; Q4-Health Informatics)

Status of Publication: Published

DOI: 10.4018/IJRQEH.2020010103

B.1.11 Improving the Decision-Making Process in a Hospital Environment With New Interactive Visualization Methods

Authors: Cristiana Neto, ***Diana Ferreira***, António Abelha.

Title: Improving the Decision-Making Process in a Hospital Environment With New Interactive Visualization Methods

Abstract: A large number of medical errors occurring in hospitals happen in the diagnostic process, such as the use of outmoded tests and the failure to perform on test results, leading to delays in the diagnosis. Thus, fast and efficient access to the patients' exams and test results is essential to provide them adequate treatment. This need demands solving problems that exist during this process of consulting the results in order to reduce clinical risk. Therefore, an analysis of the problems in the visualization process of the exams in Centro Hospitalar do Porto was carried out, leading to the development of a new platform - AIDA-MCDT, whose main objective was to fill the gaps verified, making the access to

exams' results more efficient and intuitive. This platform has significantly improved the system's usability, including several features designed to help healthcare professionals make decisions more quickly and securely.

Keywords: Centro Hospitalar do Porto; Clinical Risk; Electronic Health Records; Health Information System; Healthcare; Information Technology; Medical Errors; Medical Exams; ReactJS.

Publication Date: January, 2020

Journal: International Journal of Reliable and Quality E-Healthcare (IJRQEH) - IGI Global
SJR 2021: 0.215 (Q3- Medical Laboratory Technology; Q4-Health Informatics)

Status of Publication: Published

DOI: 10.4018/IJRQEH.2020010102

B.2 Conference Publications

B.2.1 Recommendation of Medical Exams to Support Clinical Diagnosis Based on Patient's Symptoms

Authors: Cristiana Neto; ***Diana Ferreira***; Hugo Cunha; Maria Pires; Susana Marques; Regina Sousa; José Machado.

Title: Recommendation of Medical Exams to Support Clinical Diagnosis Based on Patient's Symptoms

Abstract: Nowadays, it is essential that the error in the decisions made by health professionals is as small as possible. This applies to any medical area, including the recommendation of medical exams based on certain symptoms for the diagnosis of diseases. This study aims to explore the use of different Machine Learning techniques to increase the confidence of the medical exams prescribed by healthcare professionals. A successful implementation of this proposal could reduce the probability of medical errors in what concerns the prescription of medical exams and, consequently, the diagnosis of medical conditions. Thus, in this paper, six Machine Learning models were applied and optimized, namely, RF, DT, k-NN, NB, SVM and RNN, in order to find the most suitable model for the problem at hand. The results obtained with this study were promising, achieving high accuracy values with RF, DT and k-NN.

Keywords: Recommender system, Medical exams, Crisp-dm, Classification

Publication date: July, 2023

Conference: International Conference on AI-assisted Solutions for COVID-19 and Biomedical Applications in Smart Cities

Status of Publication: Published

DOI: 10.1007/978-3-031-38204-8_8

B.2.2 OpenEHR based bariatric surgery follow-up

Authors: Ana Afonso, César Alvaréz, ***Diana Ferreira***, Daniela Oliveira, Hugo Peixoto, António Abelha, José Machado.

Title: OpenEHR based bariatric surgery follow-up

Abstract: According to the World Health Organization (WHO), more than one billion people in the world are obese, and this number is still increasing. When this disease becomes a high risk to the individual's health and the non-invasive approach does not result in weight loss, it is usual to resort to an invasive intervention, bariatric surgery. Due to all the specifications, as well as the multidisciplinary treatment inherent to this procedure, the need for a specialized environment emerges. In this context, the main objective of this topic focuses on the development of a platform for registration and monitoring of bariatric surgery. For this purpose, a web platform was created, which integrates openEHR forms for the registration of appointments regarding this intervention, and uses openEHR modeling to build the interoperable template. The development of this project aims to provide greater ease and speed in patient care, assisting health professionals in their daily lives.

Keywords: Bariatric surgery; Obesity; Health Information Systems; Electronic Health Records; Interoperability; OpenEHR.

Publication Date: November, 2022

Journal: Procedia Computer Science

Status of Publication: Published

DOI: 10.1016/j.procs.2022.10.149.

B.2.3 Classification of Dementia in Adults

Authors: Cristiana Neto, ***Diana Ferreira***, João Nunes, Luís Braga, Luís Martins, Luís Cunha, José Machado.

Title: Classification of Dementia in Adults

Abstract: Dementia is a broad term for a large number of conditions, and it is often associated with Alzheimer's disease. A reliable diagnosis of this disease, especially in the early stages, may prevent further complications. As such, machine learning algorithms can be applied in order to validate and correctly classify cases of dementia or non dementia in adults, assisting physicians in the diagnosis and management of this clinical condition. In this study, a dataset containing magnetic resonance imaging comparisons of demented/non demented adults was used to conduct a Data Mining process, following the Cross Industry Standard Process for Data Mining methodology, with the main goal of classifying instances of dementia. Different machine learning algorithms were applied during this process, more specifically Support Vector Machines, Decision Trees, Logistic Regression, Neural Networks, Naïve Bayes and Random Forest. The maximum accuracy of 95.41% was achieved with the Naïve Bayes algorithm using Split Validation.

Keywords: Dementia, Alzheimer, Data Mining, Classification;

Publication Date: October, 2021

Conference: Developments and Advances in Defense and Security

Status of Publication: Published

DOI: 10.1007/978-981-16-4884-7_23

B.2.4 Detecting Autism Spectrum Disorder Using Data Mining

Authors: Ana C. Guisasola, ***Diana Ferreira***, Cristiana Neto, António Abelha, José Machado

Title: Detecting Autism Spectrum Disorder Using Data Mining

Abstract: Autism spectrum disorder (ASD) is a set of neurodevelopmental disorders that affect cognitive development, social and communication skills, and behavior of affected individuals. The faster traces of ASD are identified, the faster the stimulation will begin and the more effective the gains in neuropsychomotor development will be. That being said, the earlier the diagnosis of ASD, the easier it is

to control the disorder. Therefore, this study aims to classify the cases of ASD as "yes" if a patient has been diagnosed with ASD or "no" if a patient has not, using data mining (DM) models with classification techniques. The methodology of cross-industry standard process for data mining (CRISP-DM) was followed, and to induce the data mining models, the Rapidminer software was used. The results were quite promising reaching a level of accuracy of 97%, specificity of 95.45%, sensitivity of 100%, and precision of 95.65%.

Keywords: Autism Spectrum Disorder; Data Mining; CRISP-DM; Classification

Publication Date: October, 2021

Conference: Developments and Advances in Defense and Security: Proceedings of MICRADS 2021

Status of Publication: Published

DOI: 10.1007/978-981-16-4884-7_22

B.2.5 Prediction Models for Coronary Heart Disease

Authors: Cristiana Neto, ***Diana Ferreira***, José Ramos, Sandro Cruz, Joaquim Oliveira, António Abelha, José Machado.

Title: Prediction Models for Coronary Heart Disease

Abstract: In the current days, it is known that a great amount of effort is being applied to improving healthcare with the use of Artificial Intelligence technologies in order to assist healthcare professionals in the decision-making process. One of the most important field in healthcare diagnoses is the identification of Coronary Heart Disease since it has a high mortality rate worldwide. This disease occurs when the heart's arteries are incapable of providing enough oxygen-rich blood to the heart. Thus, this study attempts to develop Data Mining models, using Machine Learning algorithms, capable of predicting, based on patients' data, if a patient is at risk of developing any kind of Coronary Heart Disease within the next 10 years. To achieve this goal, the study was conducted by the CRISP-DM methodology and using the RapidMiner software. The best model was obtained using the Decision Tree algorithm and with Cross-Validation as the sampling method, obtaining an accuracy of 0.884, an AUC value of 0.942 and an F1-Score of 0.881.

Publication Date: September, 2021

Conference: Distributed Computing and Artificial Intelligence, Volume 1: 18th International Conference 18)

Status of Publication: Published

DOI: 10.1007/978-3-030-86261-9_12

B.2.6 A CRISP-DM Approach for Predicting Liver Failure Cases: An Indian Case Study

Authors: António F. Cunha, ***Diana Ferreira***, Cristiana Neto, António Abelha, José Machado.

Title: A CRISP-DM Approach for Predicting Liver Failure Cases: An Indian Case Study

Abstract: We are currently living a period in which data processing and analysis are increasingly relevant and the health sector is no exception. In this way, through data mining processes, it is possible to make a number of predictions in the medical field, such as predicting medical conditions and disease progression. Acute Liver Failure (ALF) is a rare but critical disorder associated with high mortality. The aim of this work is to predict cases of acute hepatic insufficiency based on clinical data through data mining techniques. To this end, the CRISP-DM methodology was followed, in which five classifiers were applied, namely, Decision Tree, k-Nearest Neighbor, Random Forest, Rule Induction, and Naïve Bayes. Throughout this work, the RapidMiner software was used and the different models were analyzed based on Accuracy, Precision, Recall, Kappa Statistic, and Specificity. The best data mining model achieved an Accuracy of 0.925, a Precision of 0.869, a Recall of 1.000, a Kappa of 0.849, and a Specificity of 0.849, using split validation and the k-Nearest Neighbor algorithm.

Keywords: Acute Liver Failure; Data Mining; Classification Algorithms; CRISP-DM.

Publication Date: July, 2021

Conference: Advances in Artificial Intelligence, Software and Systems Engineering: Proceedings of the AHFE 2021

Status of Publication: Published

DOI: 10.1007/978-3-030-80624-8_20

B.2.7 Data Mining Approach to Classify Cases of Lung Cancer

Authors: Eduarda Vieira, ***Diana Ferreira***, Cristiana Neto, António Abelha, José Machado.

Title: Data Mining Approach to Classify Cases of Lung Cancer

Abstract: According to the World Cancer Research Fund, a leading authority on cancer prevention research, lung cancer is the most commonly occurring cancer in men and the third most commonly occurring cancer in women, with the 5-year relative survival percentage being significantly low. Smoking is the major risk factor for lung cancer and the symptoms associated with it include cough, fatigue, shortness of breath, chest pain, weight loss, and loss of appetite. In an attempt to build a model capable of identifying individuals with lung cancer, this study aims to build a data mining classification model to predict whether or not a patient has lung cancer based on crucial features such as the above mentioned symptoms. Through the CRISP-DM methodology and the RapidMiner software, different models were built, using different scenarios, algorithms, sampling methods, and data approaches. The best data mining model achieved an accuracy of 93%, a sensitivity of 96%, a specificity of 90% and a precision of 91%, using the Artificial Neural Network algorithm.

Keywords: Healthcare; Lung cancer; Data mining; Classification; CRISP-DM.

Publication Date: April, 2021

Conference: Trends and Applications in Information Systems and Technologies (WorldCIST 2021)

Status of Publication: Published

DOI: 10.1007/978-3-030-72657-7_49

B.2.8 Step Towards Predicting Patient Length of Stay in Intensive Care Units

Authors: Cristiana Neto, Gabriel Pontes, Alexandru Domente, Francisco Reinolds, José Costa, ***Diana Ferreira***, José Machado.

Title: Step Towards Predicting Patient Length of Stay in Intensive Care Units

Abstract: Increasingly, hospitals are collecting huge amounts of data through new storage methods. These data can be use to extract hidden knowledge, which can be crucial to estimate the length of stay of admitted patients in order to improve the management of hospital resources. Hence, this article portrays the performance analysis of different data mining techniques through the application of learning algorithms in order to predict patients' length of stay when admitted to an Intensive Care Unit. The data used in this study contains about 60,000 records and 28 features with personal and medical information. A full analysis of the results obtained with different Machine Learning algorithms showed that the model trained with the Gradient Boosted Trees algorithm and using only the features that were strongly correlated to the patient's length of stay, achieved the best performance with 99,19% of accuracy. In this sense, an accurate understanding of the factors associated with the length of stay in intensive care units was achieved.

Keywords: Length of Stay; Hospitalization Stay; Intensive Care Unit; Data Mining; Classification; CRISP-DM.

Publication Date: March, 2021

Conference: Trends and Applications in Information Systems and Technologies (WorldCIST 2021)

Status of Publication: Published

DOI: 10.1007/978-3-030-72654-6_28

B.2.9 Diagnosis of Diabetic Retinopathy Using Data Mining Classification Techniques

Authors: Ana Abreu, ***Diana Ferreira***, Cristiana Neto, António Abelha, José Machado.

Title: Diagnosis of Diabetic Retinopathy Using Data Mining Classification Techniques

Abstract: Diabetic retinopathy is one of the complications of diabetes that affects the small vessels of the retina, being the main cause of blindness in adults. An early detection of this disease is essential, as it can prevent blindness as well as other irreversible harmful outcomes. This article attempts to develop a data mining model capable of identifying diabetic retinopathy in patients based on features extracted from eye fundus images. The data mining process was carried out in the RapidMiner software and followed the CRISP-DM methodology. In particular, classification models were built by combining different scenarios, algorithms, and sampling methods. The data mining model which performed best achieved an accuracy of 76.90%, a precision of 85.92%, and a sensitivity of 67.40%, using the Logistic Regression algorithm and Split Validation as the sampling method.

Keywords: Healthcare; Diabetic Retinopathy; Data Mining; CRISP-DM; Classification.

Publication Date: March, 2021

Conference: Advances in Digital Science (ICADS 2021)

Status of Publication: Published

DOI: 10.1007/978-3-030-71782-7_18

B.2.10 Prediction Models for Polycystic Ovary Syndrome Using Data Mining

Authors: Cristiana Neto, Mateus Silva, Mariana Fernandes, ***Diana Ferreira***, José Machado.

Title: Prediction Models for Polycystic Ovary Syndrome Using Data Mining

Abstract: Polycystic Ovary Syndrome is an endocrine abnormality that occurs in the female reproductive system and is considered a heterogeneous disorder because of the different criteria used for its diagnosis. Early detection and treatment are critical factors to reduce the risk of long-term complications, such as type 2 diabetes and heart disease. With the vast amount of data being collected daily in healthcare environments, it is possible to build Decision Support Systems using Data Mining and Machine Learning. Currently, healthcare systems have advanced skills like Artificial Intelligence, Machine Learning and Data Mining to offer intelligent and expert healthcare services. The use of efficient Data Mining techniques is able to reveal and extract hidden information from clinical and laboratory patient data, which can be helpful to assist doctors in maximizing the accuracy of the diagnosis. In this sense, this paper aims to predict, using the classification techniques and the CRISP-DM methodology, the presence of Polycystic Ovary Syndrome. This paper compares the performance of multiple algorithms, namely, Support Vector Machines, Multilayer Perceptron Neural Network, Random Forest, Logistic Regression and Gaussian Naïve Bayes. In the end, it was found that Random Forest provides the best classification, and the use of data sampling techniques also improves the results, allowing to achieve a sensitivity of 0.94, an accuracy of 0.95, a precision of 0.96 and a specificity of 0.96.

Keywords: Data Mining; Classification; CRISP-DM; Decision Support Systems; Health Information System; Polycystic Ovary Syndrome.

Publication Date: March, 2021

Conference: Advances in Digital Science (ICADS 2021)

Status of Publication: Published

DOI: 10.1007/978-3-030-71782-7_19

B.2.11 Data Mining Approach to Understand the Association Between Mental Disorders and Unemployment

Authors: Cristiana Neto, Caroline Rodrigues, Emely Mendonça, Laercio Sartori, Rafaela de Pinho, ***Diana Ferreira***, António Abelha, José Machado.

Title: Data Mining Approach to Understand the Association Between Mental Disorders and Unemployment

Abstract: Over the years, mental illness has affected the life of numerous human beings and nowadays is a matter of great concern. The problems that arise with this clinical condition, such as social isolation, unemployment, and others, have been a subject of study. The purpose of this study is to use a survey that aims to assess the situation of unemployment among individuals with mental illness. Hence, this article focuses on using the result of this research to identify if there is a connection between having mental illness and being in a situation of unemployment, as well as, which factors can be determinant for such a relationship and also if there is any way to anticipate them. In this context, this research attempts to develop an accurate prediction mechanism, using Data Mining, capable of predicting, based on the answers of a similar questionnaire, if an individual will be in risk of unemployment. Throughout this research, the CRISP-DM methodology was adopted and the RapidMiner Studio software was the tool used for the learning process. The best percentages of accuracy were between 0.79 and 0.86, of sensitivity between 0.75 and 0.88, and of specificity between 0.66 and 0.93.

Keywords: Mental illness; Unemployment; Data mining; Classification; RapidMiner; CRISP-DM.

Publication Date: January, 2021

Conference: Information Technology and Systems (ICITS 2021)

Status of Publication: Published

DOI: 10.1007/978-3-030-68418-1_8

B.2.12 Data Mining for the Prediction of Fetal Malformation Through Cardiotocography Data

Authors: Maria Nogueira, ***Diana Ferreira***, Cristiana Neto, António Abelha, José Machado.

Title: Data Mining for the Prediction of Fetal Malformation Through Cardiotocography Data

Abstract: Despite advances in technology and health, the number of maternal and fetal deaths during and after pregnancy and childbirth remains significant. Most of these deaths could be avoided if there was prenatal care before and during pregnancy, which could assist in monitoring the fetal heart rate (FHR). Thus, medical methods have been developed for assisting fetal monitoring, such as cardiotocography (CTG). To collaborate with the methods developed, advances in the field of machine learning and computational intelligence made it possible to increase the effectiveness of classification and recognition systems and, thus, to predict possible maternal or fetal death. To this end, this paper tries to predict fetal well-being, through the classification of data resulting from fetal CTGs using two different types of classification, fetal state and morphological pattern. The classification by fetal state, using methods such as Decision Tree (DT) and k-Nearest Neighbors (kNN), presented high accuracy values, achieving values that range from 93% to 98%. However, although not expected, the classification by morphological standards also showed high accuracy values, achieving the best model a value of 93% of accuracy with the kNN. Therefore, the complementary between both classifications may guarantee success in predicting fetal well-being.

Keywords: Cardiotocography(CTG); Fetal Heart Rate (FHR); Machine Learning; CRISP-DM; Data Mining.

Publication Date: January, 2021

Conference: Information Technology and Systems (ICITS 2021)

Status of Publication: Published

DOI: 10.1007/978-3-030-68418-1_7

B.2.13 Review of Trends in Automatic Human Activity Recognition in Vehicle Based in Synthetic Data

Authors: Ana Coimbra, Cristiana Neto, ***Diana Ferreira***, Júlio Duarte, Daniela Oliveira, Francini Hak, Filipe Gonçalves, Joaquim Fonseca, Nicolas Lori, António Abelha, José Machado .

Title: Review of Trends in Automatic Human Activity Recognition in Vehicle Based in Synthetic Data

Abstract: Driverless vehicles are more and more becoming a reality. However, people still have some concerns in using them, the main concern is fear, hence the importance of creating a surveillance system inside those vehicles. For the detection and classification of human movements to be possible it is necessary to train the system with data representative enough for all kinds of possibilities. Although the production of large quantities of data becomes an expensive process and adds the problem of data protection, the use of synthetic data once they are artificially generated allows lower costs and eliminates the problem of data protection. A bibliographic study was carried out in this paper with articles from 2017 or later on the use of synthetic data. In these studies, it is noted that synthetic data is widely used with good results. As far as image capture is concerned, they show that 3D cameras have better results, but they are more expensive, so 2D cameras are more often used with later conversion to 3D images. The stitched puppet (SP) model is capable of adapting to the most difficult poses having obtained good results in its application in the FAUST dataset.

Kewords: Autonomous Car; Multisensory Fusion; Audio-Visual Synthetic Data

Publication Date: October, 2020

Conference: Intelligent Data Engineering and Automated Learning – IDEAL 2020 (IDEAL 2020)

Status of Publication: Published

DOI: 10.1007/978-3-030-62365-4_35.

B.2.14 Review of Trends in Automatic Human Activity Recognition Using Synthetic Audio-Visual Data

Authors: Tiago Jesus, Júlio Duarte, ***Diana Ferreira***, Dalila Durães, Francisco Marcondes, Flávio Santos, Marco Gomes, Paulo Novais, Filipe Gonçalves, Joaquim Fonseca, Nicolas Lori, António Abelha, José Machado .

Title: Review of Trends in Automatic Human Activity Recognition Using Synthetic Audio-Visual Data

Abstract: An in-depth study of knowledge and technologies was made related to the various scientific, technical, and industrial domains necessary for the acquisition of skills and capabilities for the design and development of a multisensory fusion system for vehicle cockpits. After an extensive literature review, it was possible to determine the baselines of the solution to be developed and obtain a pipeline prototype.

Keywords: Autonomous Car; Multisensory Fusion; Audio-Visual Synthetic Data

Publication Date: October, 2020

Conference: Intelligent Data Engineering and Automated Learning – IDEAL 2020 (IDEAL 2020)

Status of Publication: Published

DOI: 10.1007/978-3-030-62365-4_53.

B.2.15 Bridging the Gap of Neuroscience, Philosophy, and Evolutionary Biology to Propose an Approach to Machine Learning of Human-Like Ethics

Authors: Nicolas Lori, ***Diana Ferreira***, Victor Alves, José Machado.

Title: Bridging the Gap of Neuroscience, Philosophy, and Evolutionary Biology to Propose an Approach to Machine Learning of Human-Like Ethics

Abstract: The growing explosion of ideas such as Artificial Intelligence (AI), smart environments and ubiquitous computing has led to the creation of the Ambient Intelligence (AmI) paradigm. As AmI begins to take place, moving from a futuristic idea to a reality, we are gradually witnessing the creation of an omnipresent, responsive, and intelligent atmosphere in which thousands of tiny sensors and natural user interfaces will be embedded in our natural movements and in our social and physical interactions. Hence, a key challenge in this multi-disciplinary approach is to get machines to act according to ethical priorities that make sense to human beings. In this study, we improve the capacity for machine ethics to approach human ethics by assessing the computation of transaction values and we argue that it is possible to perform such a computation using recent work that describes the effects of human decision-making using an axiomatic framework. This paper clarifies the relationship between the brain's 3-axes of neuroscience, the 3 Plato's Transcendentals of philosophy and the biological evolution's 3-components, as well as the top-down vs. bottom-up approaches to machine ethics.

Keywords: Artificial Intelligence; Ambient Intelligence; Machine Ethics; Transaction Value; Aesthetics Evolution; Plato's Transcendentals; Axiomatic Systems

Publication Date: October, 2020

Conference: Intelligent Data Engineering and Automated Learning – IDEAL 2020 (IDEAL 2020)

Status of Publication: Published

DOI: 10.1007/978-3-030-62365-4_30.

B.2.16 Step Towards Monitoring Intelligent Agents in Healthcare Information Systems

Authors: Regina Sousa, ***Diana Ferreira***, António Abelha, José Machado.

Title: Step Towards Monitoring Intelligent Agents in Healthcare Information Systems

Abstract: A platform for establishing interoperability between heterogeneous information systems implemented in a hospital environment is more a requirement than an option. The Agency for the Integration, Diffusion and Archiving of Medical and Clinical Information (AIDA) is an interoperability platform designed specifically to address the problem of integrating information from multiple systems and addressing interoperability, confidentiality, integrity and data availability. This article focuses on the relevance and need for such vigilance, finding and designing effective new ways to establish them. This study culminated in the creation of AIDAMonit, a surveillance platform developed and tested by ALGORITMI Center researchers, which has shown promise and is extremely beneficial for the well-functioning of the health facilities currently using the AIDA platform.

Keywords: Health Information Systems (HIS) · Intelligent agents · AIDA · Interoperability · Monitoring · AIDAMonit.

Publication Date: May, 2020

Conference: Trends and Innovations in Information Systems and Technologies (WorldCIST 2020)

Status of Publication: Published

DOI: 10.1007/978-3-030-45697-9_50

B.2.17 Prediction of Mental Illness Associated with Unemployment Using Data Mining

Authors: Carina Gonçalves, ***Diana Ferreira***, Cristiana Neto, António Abelha, José Machado.

Title: Prediction of Mental Illness Associated with Unemployment Using Data Mining

Abstract: Mental illness is a concern these days, affecting people worldwide and across all kinds of ages. This article aims to predict mental illness and discover its association with unemployment as well as other possible causes behind the illness. In order to accomplish this goal, a Data Mining (DM) process was performed using the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology and the RapidMiner Studio software. In the end, the results obtained were considered promising since all the evaluation metrics, namely accuracy, sensitivity, and specificity, obtained values above 90%. The study also allowed, in the end, to identify the factors associated with the prediction of mental illness.

Keywords: Mental Illness; Unemployment; Data Mining; CRISP-DM; Classification.

Publication Date: January, 2020

Conference Proceedings: Procedia Computer Science

Status of Publication: Published

DOI: 10.1016/j.procs.2020.10.078.

B.2.18 Data Mining to Predict Early Stage Chronic Kidney Disease

Authors: Ana Pinto, ***Diana Ferreira***, Cristiana Neto, António Abelha, José Machado.

Title: Data Mining to Predict Early Stage Chronic Kidney Disease

Abstract: Chronic Kidney Disease (CKD) is a condition characterized by a gradual loss of kidney function over time. In national and international guidelines, CKD is organized into different degrees of risk stratification using commonly available markers. It is usually asymptomatic in its early stages, and early detection is important to reduce future risks. This study used the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology and the WEKA software to build a system that can classify the chronic condition of kidney disease based on accuracy, sensitivity, specificity and precision. The results obtained were considered satisfactory, achieving the most suitable result of 97.66% of accuracy, 96.13% of sensitivity, 98.78% of specificity and 98.31% of precision with the J48 algorithm.

Keywords: Chronic Kidney Disease; Data Mining; Classification; CRISP-DM; Weka.

Publication Date: January, 2020

Conference Proceedings: Procedia Computer Science

Status of Publication: Published

DOI: 10.1016/j.procs.2020.10.079.

B.2.19 Intelligent Nutrition in Healthcare and Continuous Care

Authors: Rui Miranda; ***Diana Ferreira***; António Abelha; José Machado.

Title: Intelligent Nutrition in Healthcare and Continuous Care

Abstract: In the healthcare industry, the patient's nutrition is a key factor in their treatment process. Every user has their own specific nutritional needs and requirements. An appropriate nutrition policy can therefore help the patient's recovery process and alleviate possible symptoms. Food recommender systems are platforms that offer personalised suggestions of recipes to users. However, there is a lack of usage of recipe recommendation systems in the healthcare sector. Multiple challenges in representing the domain of food and the patient's needs make it complicated to implement these systems. The present project aims to develop a platform for an intelligent planning of the user's meals, based on their clinical conditions. The application of machine learning algorithms on nutrition, in healthcare services and continuous care is thus a key topic of research. This platform will be tested and deployed at the Social Cafeteria of Vila Verde (Cantina Social da Santa Casa da Misericórdia de Vila Verde). The development of this project will use the Design Science Research (DSR) investigation methodology, ensuring that the solution to the problem accomplishes all needs and requirements of the professionals, while elucidating new knowledge both for the institution and the scientific community.

Keywords: machine learning, recommender systems, meal planning, decision support systems.

Publication Date: July, 2019

Conference: IEEE-International Conference in Engineering Applications (ICEA)

Status of Publication: Published

DOI: 10.1109/CEAP.2019.8883496

B.2.20 Prediction of Neoadjuvant Chemotherapy Outcome in Breast Cancer Patients

Authors: José Neves, Almeida Dias, Cristiana Silva, ***Diana Ferreira***, Luís Costa, Filipa Ferraz, Victor Alves, João Neves, Jorge Ribeiro, Henrique Vicente.

Title: Prediction of Neoadjuvant Chemotherapy Outcome in Breast Cancer Patients

Abstract: Breast Cancer is the most common invasive cancer in women worldwide. Indeed, it is imperative to investigate which factors influence the development of this disease in order to improve the efficiency of the treatment and to allow for a balanced follow-up. In fact, this article has in mind an original Case Based Reasoning (CBR) approach to problem solving, complemented with a novel approach to Knowledge Representation and Reasoning that takes into consideration the data items entropic states. It works towards cancer's assessment after Neoadjuvant Chemotherapy in terms of its size, shape, and texture.

Keywords: Breast Cancer; Neoadjuvant Chemotherapy; Knowledge Representation and Reasoning; Case Based Reasoning; Entropy; Logic Programming.

Publication Date: June, 2019

Conference: Applied Physics, System Science and Computers III (APSAC 2018)

Status of Publication: Published

DOI: 10.1007/978-3-030-21507-1_45.

B.2.21 Steps Towards Online Monitoring Systems and Interoperability

Authors: ***Diana Ferreira***, Cristiana Neto, José Machado, António Abelha .

Title: Steps Towards Online Monitoring Systems and Interoperability

Abstract: In the health area, there is, on a daily basis, an enormous amount of data being produced and disseminated. The fast-growing amount of collected data and the rich knowledge, possibly life-saving, that could be extracted from these data has demanded the search of new ways to ensure the reliability and availability of the information with an emphasis on the efficient use of information technology tools. Although the main focus of the information systems is the health professionals who contact directly with the patient, it is also imperative to have tools for the background of the health units (information services, managers of systems, etc.). The main purpose of this work is the development of an innovative and interactive web platform for the daily monitoring of the web services activities of a Portuguese hospital, Centro Hospitalar do Porto (CHP). This platform is a web application developed in React that aims to ensure the correct functioning of the web services, that are responsible for numerous tasks within the hospital environment, and which failure could result in disastrous consequences, both for the health institution and for the patients. The development of the web application followed the six stages of the Design Science

Research (DSR) methodology and was submitted to the Strengths Weaknesses Opportunities and Threats (SWOT) analysis, which results were considered optimistic.

Keywords: Information technology; Monitoring; Web services; Web application.

Publication Date: March, 2019

Conference: New Knowledge in Information Systems and Technologies

Status of Publication: Published

DOI: 10.1007/978-3-030-16187-3_51.

B.2.22 Improving Healthcare Delivery with New Interactive Visualization Methods

Authors: Cristiana Neto, ***Diana Ferreira***, António Abelha, José Machado.

Title: Improving Healthcare Delivery with New Interactive Visualization Methods

Abstract: Over the last years, the implementation and evolution of computer resources in hospital institutions has been improving both the financial and temporal efficiency of clinical processes, as well as the security in the transmission and maintenance of their data, also ensuring the reduction of clinical risk. Diagnosis, treatment and prevention of human illness are some of the most information-intensive of all intellectual tasks. Health providers often do not have or cannot find the information they need to respond quickly and appropriately to patient's medical problems. Failure to review and follow up on patient's test results in a timely manner, for example, represents a patient's safety and malpractice concern. Therefore, it was sought to identify problems in a medical exams results management system and possible ways to improve this system in order to reduce both clinical risks and hospital costs. In this sense, a new medical exams visualization platform (AIDA-MCDT) was developed, specifically in the Hospital Center of Porto (CHP), with several new functionalities in order to make this process faster, intuitive and efficient, always guaranteeing the confidentiality and protection of patients' personal data and significantly improving the usability of the system, leading to a better health care delivery.

Keywords: Information technology; Health information system; Electronic Health Records; Medical exams.

Publication Date: March, 2019

Conference: New Knowledge in Information Systems and Technologies

Status of Publication: Published

DOI: 10.1007/978-3-030-16187-3_52.

B.2.23 Predictive Data Mining in Nutrition Therapy

Authors: ***Diana Ferreira***; Hugo Peixoto; José Machado; António Abelha.

Title: Predictive Data Mining in Nutrition Therapy

Abstract: The assessment and measurement of health status in communities throughput the world is a massive information technology challenge. Data mining, plays a vital role in health care industry since it really has the potential to generate a knowledge-rich environment that reduces medical errors, decreases costs by increasing efficiency, improves the quality of clinical decisions and significantly enhances patient's outcomes and quality of life. This study falls within the context of nutrition evaluation and its main goal is to apply classification algorithms in order to predict if a patient needs to be followed by a nutrition specialist. One of the tools resorted in this study was the Waikato Environment for Knowledge Analysis (Weka in advance) Workbench since it allows to quickly try out and compare different machine learning solutions. The tasks involved in the development of this project included data preparation, data preprocessing, data transformation and cleaning, application of several classifiers and its respective evaluation through performance measures that include the confusion matrix, accuracy, error rate, and others. The accomplished results showed to be quite optimistic presenting promising values of performance measures. specifically an accuracy around 91

Keywords: information technology; data mining; healthcare; clinical decisions; nutrition evaluation; machine learning; classification algorithms; performance measures.

Publication Date: March, 2019

Conference: IEEE-APCA International Conference on Automatic Control and Soft Computing (CONTROLO)

Status of Publication: Published

DOI: 10.1109/CONTROLO.2018.8516413

B.3 Book Chapters

B.3.1 Classification of Polycystic Ovary Syndrome Based on Correlation Weight Using Machine Learning

Authors: Marcelo Marreiros, ***Diana Ferreira***, Cristiana Neto, Deden Witarsyah, José Machado.

Title: Classification of Polycystic Ovary Syndrome Based on Correlation Weight Using Machine Learning

Abstract: Polycystic ovarian syndrome (PCOS) is the most common endocrine pathology in reproductive-age women worldwide. Research has shown that the application of machine learning (ML) and data mining (DM) can have a positive impact in this condition's diagnosis. This study aims to develop a model to identify patients with PCOS using different scenarios based on correlation weights. Five DM techniques were applied, namely random forest (RF), decision tree (DT), naive bayes (NB), logistic regression (LR), and artificial neural network (ANN), to determine the best model, which was the RF classifier. Additionally, the results show that the model was able to predict PCOS with 93.06% of accuracy, 92.66% of precision, 93.52% of sensitivity, and 92.59% of specificity. Compared with a previous work conducted by the authors, the feature selection-based solo on the correlation weight decreased the accuracy values by 1.9%, precision by 3.7%, sensitivity by 0.3%, and specificity by 3.6%.

Keywords: Data Mining, Machine Learning, Classification, Polycystic Ovary Syndrome, Correlation, Feature Selection, Random Forest.

Publication Date: January, 2022

Book: Big Data Analytics and Artificial Intelligence in the Healthcare Industry - Igi Global

Status of Publication: Published

DOI: 10.4018/978-1-7998-9172-7.ch006.

B.3.2 Business Intelligence for Nutrition Therapy

Authors: Rita Reis, Ana Mendonça, Diana Lisandra Azevedo Ferreira, Hugo Peixoto, José Machado.

Title: Business Intelligence for Nutrition Therapy

Abstract: The assessment of health status in communities throughout the world is a massive information technology challenge. Data warehousing provides a flexible environment to support the business management and serve as an integrated repository for data. With the addition of models and analytic tools that have the potential to provide actionable information resources and support effective problem identification, critical decision-making, and strategy formulation, implementation, and evaluation. Of particular interest are the factors of influence like the patient's height or weight and its impact on processes and results. A multidimensional process is a way to discover health care processes according to certain factors of influence. This study aims to implement a data warehousing environment for decision support, in the context of nutrition evaluation, to integrate data obtained from a health care facility. This paper highlights the implementation of Business Intelligence in health care settings allows searching and interpreting stored information to support decisions concerning people's life.

Keywords: Information Technology, Decision Support System, Data Warehouse, Business Intelligence, Multidimensional Process, Health Care, Nutrition Evaluation.

Publication Date: January, 2018

Book: Healthcare Policy and Reform: Concepts, Methodologies, Tools, and Applications.

Status of Publication: Published

DOI: 10.4018/978-1-5225-6915-2.ch022.