UQAC
Université du Québec
à Chicoutimi

Context-Aware Ontology-Guided Healthcare Learning Model for Performance and Interpretability; Asthma disease as a case Study

By Batoul Msheik

THESIS PRESENTED TO L'UNIVERSITE DU QUEBEC A CHICOUTIMI IN PARTIAL FULFILLMENT OF THE REQUIERMENTS FOR THE DEGREE OF PHILOSOPHIA DOCTOR (PH.D.) IN THE SUBJECT OF INFORMATIQUE

Québec, Canada

ABSTRACT

Context-aware medical systems have become an important advancement in healthcare, enabling a transformation in how medical services interact with patients and clinical environments. These systems rely on the concept of context-awareness to understand patient conditions, environmental factors, and medical history to support more informed decisions. However, context medical recommendation systems face three major challenges: the complexity of the medical context (its definition and representation) and the reliability of healthcare systems. The first challenge relates to the proposed structures and understanding of "chronic diseases context", and these structures have not been able to meet the needs of the medical field, either because the structure of context is too general, or because it is too specific. The second challenge is the representation of the context, which encompasses all the information in a comprehensive knowledge base of the medical field. The third challenge relates to the reliability of reasoning based on the medical context during intervention. This reliability raises issues for the development of contextual systems, especially the requirements (performance and interpretability), needed by the healthcare domain and medical staff (physicians and nurses).

This research addresses these challenges in general terms, and in the case of patients with chronic diseases requiring continuous monitoring without hospitalization. As case study, we consider chronic disease (asthma), for three reasons: (i) scientific, due to the complexity of the biological systems, the control requires a combination of modeling and empirical and dynamic observations; (ii) social, because 3.8 million Canadians suffer from asthma; and (iii) medical, because the causes of allergies and exacerbations are complex and poorly understood by medical science.

These chronic diseases are long-term conditions that progress gradually, do not resolve spontaneously, and usually cannot be completely treated. Asthma demands ongoing monitoring, timely intervention, and individualized care. Although various systems have been proposed to support asthma management, most lack the ability to provide reliable, context-aware responses tailored to real-world patient environments.

Modern healthcare systems require flexible, reliable, and interpretable learning processes that support medical staff in tasks such as diagnostic reasoning and patient monitoring. However, these systems still have many issues such as context understanding, context representation, and interpretability of the decision by healthcare staff.

To manage the complexity of the healthcare domain and ensure a clear progression from raw inputs to interpretable outputs, we develop a five-layer context-aware ontology-guided Healthcare architectural model (COHM), composed of data acquisition, context categorization, context representation, operation, and application layers. This model supports the methodological workflow of the study, including data collection, preprocessing, contextual labeling, ontology-guided structuring, extraction rules method, and its evaluation.

First, in this COHM model we build a fifteen-category context model to enriche it with additional dimensions such as social factors and Service Level Agreements (SLAs), strengthening its relevance for clinical settings. By leveraging knowledge representation models, the designed general contextual ontology enables structured organization and semantic understanding of healthcare data and increasing interpretability of the system.

Second, we develop a structured ontology representation for sharing contextual medical knowledge and for integrating heterogeneous healthcare data. In this work, a processing pipeline is developed to organize the different stages of the study, from contextual categorization to data preparation and analysis. This pipeline also supports the extension of the ontology to the asthma domain using real-world datasets.

Thirdly, we develop and validate a rule extraction approach within the COHM model to generate interpretable rules. This approach considers two inputs to extract rules: based on data raw and based on ontology-enhanced data. We classify the rule-extraction methods into four categories to analyze their interpretability in healthcare decision-support systems. These include expert rules, direct rule-learning models (Apriori), ML based models (XGBoost with TE2Rules), and Large Language Models (LLMs). Based on these categories, we extract the rules by designing rule extraction approach using datasets obtained from the Kaggle platform under two scenarios: with raw data and with ontology-enhanced data. The results demonstrate that ontology-guided feature structuring improves rule

interpretability by producing rules that are more concise, semantically consistent, and clinically meaningful.

RÉSUMÉ

L'informatique ubiquitaire sensible au contexte constitue l'une des avancées majeures des dernières décennies, ayant profondément transformé les interactions des utilisateurs finaux grâce au concept de sensibilité au contexte. Elle offre de nouvelles possibilités pour repenser les solutions conventionnelles en proposant des services personnalisés fondés sur le contexte propre à chaque environnement. Toutefois, les systèmes de recommandation médicale sensibles au contexte sont confrontés à trois défis principaux : premièrement, la complexité du contexte médical, tant dans sa définition que dans sa représentation ; deuxièmement, la fiabilité des décisions, souvent limitée par un manque de confiance dans leurs résultats.

Le premier défi concerne les définitions proposées du « contexte », qui ne répondent pas pleinement aux besoins du domaine médical, car elles sont soit trop vagues ou trop générales, soit excessivement spécifiques. Le deuxième défi porte sur la représentation du contexte, laquelle doit intégrer l'ensemble des informations au sein d'une base de connaissances complète dans le domaine médical. Le troisième défi concerne l'interprétabilité des règles en santé. Ce travail se concentre sur l'extraction de règles et sur l'évaluation systématique de leur interprétabilité dans les systèmes d'aide à la décision médicale.

Cette recherche aborde ces défis de manière générale, ainsi que dans le cas des patients atteints de maladies chroniques nécessitant un suivi continu sans hospitalisation. L'asthme est retenu comme étude de cas pour trois raisons : (i) scientifique, en raison de la complexité des systèmes biologiques, dont le contrôle requiert une combinaison de modélisation et d'observations empiriques et dynamiques ; (ii) sociale, puisque 3,8 millions de Canadiens souffrent d'asthme ; et (iii) médicale, car les causes des allergies et des exacerbations demeurent complexes et encore insuffisamment comprises par la science médicale.

Les maladies chroniques sont des affections de longue durée qui évoluent progressivement, ne disparaissent pas spontanément et ne peuvent généralement pas être complètement guéries. Des maladies telles que l'asthme devraient représenter près des trois quarts des décès dans le monde. L'asthme exige un suivi continu, des interventions rapides et des soins individualisés. Bien que divers systèmes aient été proposés pour soutenir la gestion de l'asthme, la plupart ne sont pas en mesure de fournir des réponses fiables et sensibles au contexte, adaptées aux environnements réels des patients.

Pour répondre à ces enjeux, les systèmes de santé modernes nécessitent des processus d'apprentissage flexibles, fiables et transparents, capables d'assister le personnel médical dans des tâches telles que le raisonnement diagnostique et le suivi des patients. Ces processus prennent en compte la compréhension du contexte, sa représentation et l'interprétabilité des méthodes d'extraction de règles.

Cette thèse conçoit un modèle d'apprentissage enrichi sémantiquement pour les systèmes d'aide à la décision en santé. Ce modèle intègre trois composantes principales : (i) l'identification du contexte, (ii) la représentation des connaissances à travers une ontologie contextuelle spécifique au domaine, et (iii) un modèle d'évaluation indépendant à double voie qui évalue séparément la précision prédictive et l'interprétabilité. Le modèle s'appuie sur un cadre de catégorisation du contexte composé de quinze catégories et l'enrichit par l'ajout de dimensions supplémentaires telles que les facteurs sociaux et les accords de niveau de service (SLA), renforçant ainsi sa pertinence en milieu clinique. Grâce à des techniques avancées de représentation des connaissances, l'ontologie permet une organisation structurée et une compréhension sémantique des données de santé, favorisant l'interopérabilité entre systèmes hétérogènes.

Le modèle ontologique constitue une base formelle pour le partage des connaissances contextuelles et intègre des sources de données médicales hétérogènes. Le pipeline soutient également des stratégies analytiques visant à identifier les paramètres pertinents et à étendre l'ontologie spécifiquement au cas de l'asthme à partir de jeux de données réels. Le système traite les informations des patients en temps réel et utilise des méthodes de raisonnement fondées sur des règles ainsi que des approches d'apprentissage automatique afin de produire des prédictions et des recommandations personnalisées.

Dans la continuité de ce travail, la méthode d'extraction de règles étend le cadre fondé sur l'ontologie en mettant l'accent sur l'interprétabilité dans l'aide à la décision médicale. Cette recherche examine comment différentes méthodes d'extraction de règles, telles qu'Apriori, les arbres de décision, la distillation (XG-Boost avec TE2Rules) et les grands modèles de langage (LLM), se comportent selon deux conditions : avec et sans intégration de l'ontologie. Ces méthodes sont classées en quatre catégories, puis comparées à l'aide de plusieurs métriques. En combinant une représentation des connaissances guidée par l'ontologie avec des techniques d'intelligence artificielle fondées sur les données, notre approche d'extraction de règles génère des règles cliniquement pertinentes, transparentes et fiables, renforçant ainsi la confiance dans les systèmes décisionnels automatisés.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI — Artificial Intelligence
ML — Machine Learning
AMIA — American Medical Informatics Association
BMI — Body Mass Index
COHM — Context-Aware Ontology-Guided Healthcare Learning Model
Conf_min — Minimum Confidence Threshold
COPD — Chronic Obstructive Pulmonary Disease
CORELS — Certifiably Optimal Rule Lists
CVD — cardiovascular disease
DNN — Deep Neural Network
ECG — Electrocardiogram
ECLAT — Equivalence Class Transformation
EHR — Electronic Health Record
FEV1 — Forced Expiratory Volume in 1 Second
F1-Score — Harmonic Mean of Precision and Recall
FP — False Positive
FP-Growth — Frequent Pattern Growth Algorithm
FVC — Forced Vital Capacity
GDPR — General Data Protection Regulation
HL7 — Health Level Seven
HIPAA — Health Insurance Portability and Accountability Act
ICU — Intensive Care Unit
IEEE — Institute of Electrical and Electronics Engineers
IoT — Internet of Things
KDD — Knowledge Discovery and Data Mining
KRMs — Knowledge Representation Models
LLM — Large Language Model
LIME — Local Interpretable Model-Agnostic Explanations
NB — Naïve Bayes
NCD — Noncommunicable Disease
OWL — Web Ontology Language
PCA — Principal Component Analysis
RDF — Resource Description Framework
RF — Random Forest
RIPPER — Repeated Incremental Pruning to Produce Error Reduction
ROC — Receiver Operating Characteristic
SENSUS — Ontology Development Methodology
SLA — Service Level Agreement
SNOMED CT — Systematized Nomenclature of Medicine – Clinical Terms
SPARQL — SPARQL Protocol and RDF Query Language
SVM — Support Vector Machine
Sup_min — Minimum Support Threshold
SWRL — Semantic Web Rule Language
TE2Rules — Tree Ensemble to Rules Extraction Method
TN — True Negative
TP — True Positive
TPR — True Positive Rate
TNR — True Negative Rate
TSR — Total Satisfaction ratio
UML — Unified Modeling Language
XAI — Explainable Artificial Intelligence
XGBoost — Extreme Gradient Boosting

LIST OF SYMBOLS

∧ — Logical AND

∨ — Logical OR

¬ — Logical NOT

→ — Logical implication (if–then)

↔ — Logical equivalence

∀ — Universal quantifier (“for all”)

∃ — Existential quantifier (“there exists”)

≤ — Less than or equal to

≥ — Greater than or equal to

< — Less than

— Greater than

= — Equality

≠ — Not equal to

∈ — Element of / belongs to

∉ — Does not belong to

⊆ — Subset of (or equal to)

⊂ — Proper subset of

⊇ — Superset of (or equal to)

∪ — Union of sets

∩ — Intersection of sets

⊤ — Universal class (top concept in ontology)

⊥ — Empty class / inconsistency (bottom concept)

Σ — Summation

Π — Product

/ — Division

× — Multiplication

( ) — Grouping operator

{ } — Set notation

[ ] — Interval or indexing notation

? — Variable indicator (used in SPARQL queries)

DEDICATION

To my beloved parents,
To my lovely husband, my little princess
To my dear sister and brothers,
To my dear nieces and nephews,
To my cherished friends,
Your endless love, unwavering support, and constant encouragement have been my greatest strengths.
Thank you from the bottom of my heart.

ACKNOWLEDGEMENTS

I first and foremost thank God Almighty, the source of all knowledge, for granting me the perseverance, strength, and determination needed to complete this doctoral journey.

I am deeply grateful to my supervisors, Professor Hamid Mcheick, Professor Mehdi Adda, and Professor Mohamed Dbouk for their invaluable guidance, insightful feedback, and continuous support throughout my studies. Their advice greatly shaped this work while allowing me the intellectual freedom to explore my own ideas and research directions.

I would also like to express my sincere appreciation to Université du Québec à Chicoutimi for providing the academic environment in which this research was conducted. Special thanks are extended to the Lebanese University, Lebanon, my cotutelle institution, for their support and collaboration during this doctoral program.

Finally, I owe my deepest gratitude to my family. To my parents, for their endless prayers, sacrifices, and unwavering belief in me, and to my husband, for his constant encouragement, patience, and support. This achievement would not have been possible without their love and understanding throughout these years.

CHAPTER 1
INTRODUCTION

Context medical systems are used in several healthcare situations because they allow systems to monitor patient conditions, process medical data, and support decision-making [1]. In home healthcare, they help monitor patient status through connected medical devices [1,2]. In hospitals, they support clinical activities by integrating information from different medical systems [2]. In general, context medical systems organize contextual medical information to support healthcare services and patient monitoring [3].

In healthcare, everyday computing helps create smart environments for continuous patient monitoring, allowing for quick actions and personalized care. Wearable sensors and medical devices can track important health data, like heart rate, breathing, and activity levels in real time [4]. These smart environments combine clinical information with factors such as location and movement to help identify health risks and disease flare-ups early [5]. Context aware medical systems are ubiquitous computing technologies that support remote patient monitoring and telemedicine by keeping a constant connection between patients and healthcare providers [5]. By integrating technology into daily activities, healthcare systems can become more proactive and better at delivering support to patients outside of traditional doctor offices.

As these embedded technologies gather information about where people are, what they are doing, the time, and their surroundings, they help systems adjust their actions based on the user's situation [3]. Context-aware systems are a key part of everyday computing because they turn the data collected from the environment into meaningful actions, personalized services, and smart decision support [4]. In this way, everyday computing provides the necessary framework, and context-awareness gives systems the ability to respond and adapt to real-world conditions.

In today's understanding, context is very important because it helps systems respond appropriately to their surroundings. According to [3], when applications are aware of their environment, they can better interpret information and adjust to the user’s needs. This is particularly vital in systems that need to adapt to changing situations [4]. In healthcare, patient conditions can change due to various factors, while structuring context is essential for effective monitoring and decision-making [6,7]. This focus on using medical context is the main topic of this thesis.

At the same time, Artificial Intelligence-based models for illness diagnosis (chronical diseases, Asthma, CVD, diabetes, etc.) have recently emerged granting a high level of interpretability and detection [8]. Improving decision' interpretability in diseases detection is always considerable, especially in the healthcare field where sensitive data are used. The integration of these AI models into healthcare has revolutionized diagnostic capabilities, particularly in disease prediction and explainability. However, the widespread adoption of AI in clinical settings is significantly impeded by challenges related to context representation and interpretability in AI-driven decision-making processes. This thesis aims to handle these critical challenges by introducing ontology-based model that enhances the interpretability, and usability, of AI systems in healthcare, with a particular focus on improving chronic disease management through more transparent and actionable insights.

Context aware system is defined as a system that collects and uses contextual information to deliver relevant information and services to users, adjusting their operations according to the user's activities and the surrounding environment.

A context medical system uses contextual information to understand patient conditions and support medical services. These systems collect information from different sources such as medical records, sensors, and healthcare devices [5]. The collected information describes the patient's situation and allows the system to adjust its responses according to the medical context [6]. The main goal of a context medical system is to support healthcare activities by organizing and using contextual medical information [6].

1.1. Context Scenario

Let us consider a patient arriving at a hospital emergency room. The triage process involves determining the priority level for examination by the emergency physician based on the patient's symptoms and medical history. This decision is made according to two main factors: (i) the likelihood of an underlying medical condition and (ii) its severity. The reasoning behind triage considers constraints such as time (a lack of time or the high cost of performing a full blood panel) and limited resources. This initial triage helps select the first set of tests that are most likely to assist the physician. The emergency doctor, working with limited resources (including time, equipment, and staff), focuses on identifying and treating life-threatening conditions. Subsequently, a second round of tests may be requested to confirm or rule out a diagnosis.

To illustrate this idea, let’s take an example from the scenario (healthcare) of measuring heart rate in a patient arriving at the emergency room. The data might show a heart rate of 110 beats per minute, while this number doesn’t tell us a lot of issues, it could mean different things depending on the situation. That is where context becomes essential. For example, if the patient is young, 110 bpm might be perfectly normal. But if the patient is an elderly person who is sitting calmly and has a history of heart disease, the same heart rate could be a sign of a serious condition. Context helps to understand the relations between patient’s age, medical history, and even the time and place where the data was collected [6]. It helps doctors interpret the data correctly, prioritize the patient’s care, and decide which initial tests are most relevant.

As the process continues to the second stage of diagnosis, context remains very efficient. For instance, the context will guide additional tests to perform such as an ECG, blood enzymes, or stress test based on the patient’s profile, which can help distinguish between a temporary elevation in heart rate and a deeper cardiac issue like arrhythmia or early signs of heart failure.

Both scenarios involve similar stages:

1. Collecting initial measurements and assessing risk based on symptoms and history,
2. Making a preliminary diagnosis that leads to a second round of assessments,
3. Generating measurement data while awaiting a formal diagnosis,
4. Formulating a final diagnosis.

If the second round of data fails to produce a clear diagnosis, steps 2 to 4 are repeated either to reach a diagnosis or at least to mitigate the immediate risk.

We are aiming to systematize the first three pre-diagnostic phases (1, 2, and 3) using an effective context-aware model and provide explanation of the final decisions.

Several limitations remain in the development of context medical systems. The first limitation concerns context identification, where existing medical systems often lack a clear and standardized structuring of contextual information related to the patient's condition [11]. The second limitation relates to context representation, as contextual medical information is heterogeneous, and difficult to organize in a formal structure that can be processed by computational systems, particularly in chronic diseases such as asthma [23]. The third limitation concerns context-aware reasoning, where current systems have limited ability to effectively use contextual information to support decisions while maintaining interpretability in the reasoning process [23]. These limitations reduce the effective use of contextual information in healthcare systems and highlight the need for structured approaches to support context-aware medical systems.

1.2. Motivations

To address this problem in the general case, and more specifically for chronic disease, patients require continuous monitoring without hospitalization using context-aware system (CA). The designer of a CA does not know what constitutes a context, how to take it into account, and how to represent it in the CA, and how to reason efficiently based on that context. As a case study, we consider Asthma for three reasons:

Scientific: Asthma represents a complex biological and physiological disease that involves numerous interacting factors such as genetic predisposition, environmental triggers, and immune responses. Understanding and managing these interactions requires both advanced modeling techniques and continuous empirical observation to capture the dynamic nature of disease progression and patient variability [8].

Social: From a societal perspective, asthma has a significant public health impact. In Canada, more than 3.8 million people are affected by asthma, influencing not only individual quality of life but also healthcare costs, productivity, and community well-being. The social challenge lies in ensuring equitable access to preventive care, monitoring technologies, and personalized management for diverse populations [6, 9].

Medical: Medically, the causes and triggers of asthma exacerbations remain multifactorial and not fully understood, involving complex interactions between allergens, pollutants, psychological stress, and lifestyle factors. Current medical approaches often rely on reactive treatments rather than predictive and preventive models, highlighting the need for data-driven, context-aware, and ontology-supported systems that can assist healthcare professionals in anticipating exacerbations and tailoring interventions [10].

In addition, context refers to the set of information and relationships that allow us to understand a patient's current situation and predict potential changes. Beyond improving patient quality of life by enabling at home monitoring, also reducing the cost of patient follow-up.

1.3. Research Questions

The main research questions can be summarized as follows:

Q1: How can medical context be categorized and structured to support more effective clinical decision-making for chronic disease management?

Q2: How can we build a model that effectively represents medical context and knowledge in a structured and visual way to facilitate sharing, organization, and reasoning?

Q3: How can ontology enhance the interpretability of extracted rules in the healthcare domain?

1.4. Problem statement

Asthma progresses gradually, but certain risk factors like air pollution, smoking, temperature changes, humidity, vapor pressure, air pressure, and rainfall can make symptoms worse and accelerate the condition [21].

In healthcare, context encompasses the data describing a patient's condition, such as medical history, symptoms, risk factors, and treatments. However, the lack of a standardized definition and structured representation of context makes it difficult to model, share, and interpret this information effectively. This ambiguity limits the system's ability to reason patient situations and deliver personalized, context-aware medical decisions. Therefore, defining and structuring the dimensions of context is an essential problem in designing an effective and intelligent healthcare domain model.

While defining and structuring context is a fundamental step, representing it in a precise and comprehensible manner remains a major challenge in healthcare systems. Achieving a through understanding of medical situations requires models capable of expressing complex relationships between patient data, symptoms, and environmental factors in a way that is both interpretable and computationally useful. Current approaches often lack intuitive and standardized structures for visualizing and reasoning context. Therefore, there is a critical need to develop knowledge representation models (KRMs) that can organize, visualize, and formalize contextual information to enhance clarity, reasoning, and knowledge sharing within intelligent healthcare systems.

Even when contextual information defined, structured, and represented, the interpretation of the decisions produced by medical systems remains limited. Many models used in healthcare generate results without clearly explaining how different contextual elements influence the final decision. This lack of interpretability makes it difficult for healthcare professionals to understand how patient data, symptoms, environmental factors, and risk conditions contribute to the system's decision. As a result, the reasoning process of these systems can remain unclear, which limits their transparency and their acceptance in medical practice.

1.5. Goals and contributions

The main goal of this thesis is to help medical staff perform their tasks in monitoring and diagnosis processes.

To overcome the problems described in previous section, we propose a context medical ontology-based architectural model that includes many layers: i) design context definition to identify, structure, and interpret complex medical information, ii) a knowledge representation model and iii) interpretable reasoning model. This model formalizes context, integrates diverse data sources, supports risk reduction, improves patient self-management, and helps healthcare professionals in diagnosis and follow-up.

The specific goals can be summarized as follows:

G1: Develop a structural context model that categorizes medical, environmental, behavioral, social, and temporal factors relevant to chronic disease monitoring.

G2: Build a domain-specific contextual ontology that formalizes the identified context, integrates heterogeneous data sources, and supports structured semantic representation of healthcare.

G3: Design and evaluate a semantic-enhanced learning model where ontology representation influences rule interpretability, through two separate evaluation paths.

The primary contribution of this thesis is the design of a unified, context-aware ontology that guides healthcare systems.

The thesis makes the following key contributions:

i. Survey and analysis are done to categorize the existing solutions of context definition, context representation, and context reasoning in general and especially in healthcare systems. Then, we identify their limitations.The thesis proposes a fifteen-categories context model that organizes medical, environmental, behavioral, social, and temporal information essential for chronic disease management.

ii. This research introduces a domain-specific contextual ontology that formalizes this model and provides a structured semantic representation capable of integrating heterogeneous patient data.

iii. This thesis develops a semantic-enhanced learning model in which ontology-guided feature structuring supports an interpretability path (using four categories of rule extraction methods, including Apriori, ML based models, and LLMs) [24, 25].

iv. Finally, the thesis accomplishes a comparative analysis demonstrating that ontology affects rule structure improving interpretability by simplifying rule sets.

Together, these contributions offer a unified, context-aware foundation for building transparent, structured, and clinically aligned decision-support systems.

1.6. Methodology

The methodology of this thesis follows a structured four-steps process designed to operationalize the main objective of this work: context-aware semantic enhancement to improve the interpretability of rule-extraction methods in healthcare. The proposed methodology ensures logical progression from raw contextual information to clinically meaningful and interpretable rules.

1.6.1 Step 1 - Context Identification & Categorization

Step 1 begins by identifying the parameters of patient, environmental, social, and medical variables that describe the real-world asthma context. The fifteen-category contextual model introduced in our context-categorization provides the structure for this step, ensuring that all dimensions, including temporal, environmental, organizational, technological, physiological, and social factors, are systematically captured. To support this step, thirty relevant articles were reviewed, information was collected from healthcare professionals and organizations such as the WHO, and the identified parameters were organized into fifteen contextual categories while grouping similar healthcare entities to ensure alignment with medical context requirements such as accuracy, interpretability, explainability, and performance.

By organizing raw variables into meaningful semantic groups, Step 1 creates a clinically coherent view of the patient's state and lays the foundation for transforming complex, multiscale information into structured knowledge, as emphasized in both the context model and the knowledge-representation paper.

1.6.2 Step 2 - Ontology Construction & Semantic Formalization

Step 2 translates the contextual categories into a formal ontology based on the Sanchez methodology [26]. This methodology constitutes four steps: i) The ontology scope is defined for chronic diseases and aims to support healthcare data representation and analysis; ii) Key domain concepts (symptoms, triggers, biological indicators, patient, treatment) are identified, along with their attributes and relationships, and organized into hierarchical classes reflecting contextual dimensions such as temporal, environmental, organizational, and medical contexts; iii) The ontology integrates heterogeneous data sources including questionnaires, environmental parameters, biomarkers, and patient history, while using standard medical terminologies such as SNOMED CT to ensure semantic clarity and consistency; iv) The ontology is evaluated to ensure that it correctly represents the domain knowledge and supports healthcare data analysis and interpretation.

The asthma extension of the contextual ontology (General Contextual Ontology Extended to Asthma) demonstrates how context dimensions are formalized into an interconnected semantic graph that supports interpretation of patient state. This ontology is not used for automated reasoning; instead, it provides the semantic backbone needed to unify data, aggregate related variables, and expose the clinical relationships that raw data does not explicitly reveal.

1.6.3 Step 3 - Rule Extraction Method using Ontology-Guided

In this step, we distinguish four categories of rule-extraction methods to enable a structured analysis of rule interpretability within the proposed model. These include expert rules derived from physicians and medical standards, direct rule-learning such as association-rule mining that generate rules directly from data, ML based models approaches that extract rules from learning models, and Large Language Models (LLMs) capable of generating rules from textual knowledge. This categorization provides a systematic view of how rule-extraction methods behave when combined with semantic enhancement, while highlighting key challenges related to interpretability.

Following this distinction, we design a semantic-enhanced learning architectural model composed of five layers, where each layer provides specific services to develop our system and we applied the model directly in our methodological setup, supporting the interpretability evaluation settings of our study. We designed a new rule extraction approach that integrates ontology into rule extraction methods. The semantic enrichment produced by the contextual ontology enables the aggregation of clinically related variables into higher-level concepts, providing more meaningful inputs to the rule extraction methods.

Through this integration, the extracted rules are expected to become more concise, semantically coherent, and clinically interpretable.

The resulting rule sets constitute the basis for the interpretability evaluation performed in the subsequent step.

1.6.4 Step 4 – Interpretability Model Evaluation

Step 4 evaluates the impact of ontology-guided structuring on rule interpretability. Each rule-extraction method applied to two datasets obtained from the Kaggle [28]: the original dataset and on the other hand a rules extraction method is illustrated in for extracting rules where we use ontology as model to represent this data.

This comparison allows the evaluation of the impact of semantic enrichment on rule interpretability. Interpretability assessed using rule quantity metrics such as:

i. Number of extracted rules

ii. Average antecedent length

iii. Rule coverage

The goal is to determine that ontology integration can enhance the interpretability of extracted rules in healthcare contexts.

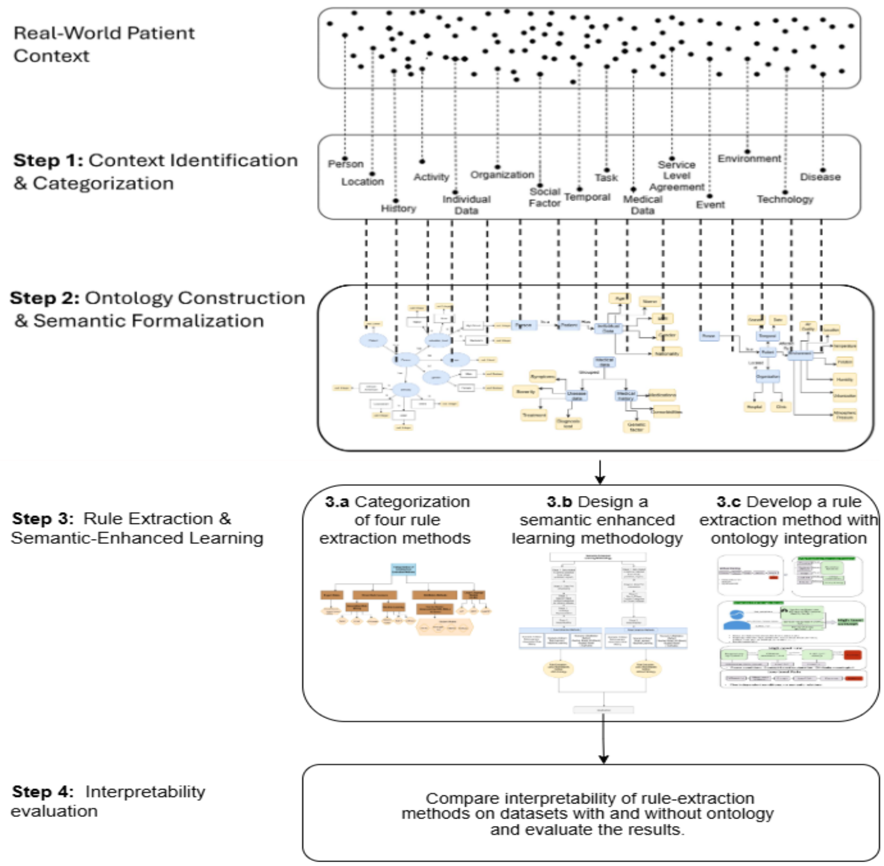

Figure 1.1 Context-Aware Ontology-Guided Learning Methodology

1.7. Organization of this thesis

This thesis is organized into six chapters. Chapter 2 defines medical context in detail and presents a fifteen-category contextual model, explaining how such categorization clarifies clinical information and strengthens contextual understanding. Chapter 3 surveys and analyzes existing knowledge-representation models in the healthcare domain and highlights their role in structuring and formalizing medical data. Chapter 4 presents the design of the general contextual ontology developed in this work, describing how contextual categories and medical concepts are modeled and semantically organized. Chapter 5 introduces COHM, the Context-Aware Ontology-Guided Healthcare Learning Model, and explains how ontology-structured features are leveraged to support rule interpretability. Finally, Chapter 6 concludes the thesis by summarizing the main findings, discussing the implications of the proposed approach, and outlining directions for future research.

CHAPTER 2
FROM DATA TO MEDICAL CONTEXT: THE POWER OF CATEGORIZATION IN HEALTHCARE

Résumé

Dans le domaine de la santé en constante évolution, la capacité à structurer et interpréter les données médicales contextuelles est essentielle pour fournir des soins personnalisés et efficaces aux patients. Bien que de nombreuses études existantes tentent de définir le contexte médical à travers diverses catégorisations, elles manquent souvent de complétude ou d'applicabilité dans le domaine réel des soins de santé. Cet article propose un modèle novateur et complet de catégorisation du contexte, composé de quinze catégories clairement définies, comblant l'écart entre les modèles théoriques et les exigences pratiques des systèmes de télésurveillance pour la gestion des maladies chroniques. En intégrant des composantes importantes mais souvent négligées, telles que les déterminants sociaux, les accords de niveau de service (SLA) et les facteurs environnementaux, notre modèle améliore la clarté et renforce la prise de décision en milieu clinique. Nous validons l'applicabilité de ce cadre à travers des études de cas détaillées portant sur l'asthme, la MPOC et les maladies cardiovasculaires, démontrant son utilité pour l'amélioration des solutions de télésanté et le soutien aux stratégies d'intervention précoce**.**

Abstract

In the rapidly evolving healthcare domain, the ability to structure and interpret contextual medical data is crucial for delivering personalized and efficient patient care. While many existing studies attempt to define medical context through diverse categorizations, they often lack completeness or applicability in the real-world healthcare domain. This paper introduces a novel and comprehensive context categorization model composed of fifteen well-defined categories, bridging the gap between theoretical models and practical requirements in telemonitoring systems for chronic disease management.

By incorporating important but often overlooked components such as social determinants, Service Level Agreements (SLAs), and environmental factors our model enhances clarity and strengthens decision-making in clinical settings. We validate the applicability of this framework through detailed case studies on asthma, COPD, and cardiovascular diseases, demonstrating its utility in enhancing telehealth solutions and aiding early intervention strategies.

Keywords: context categorization, context of healthcare domain, telemedicine, chronic disease, clinical support

2.1 Introduction

The development of technologies and the management of knowledge have improved many human life routines and practices in education, finance, industry, and healthcare [29] Healthcare incorporates such technologies as information and communication tools, artificial intelligence and its successors, such as machine learning and deep learning, computing devices, and sensors [29. 30]. In other words, combining technology and healthcare improves services for individuals and society [31, 32]. Electronic health records (EHRs) [31], smart wearables [32], mobile health applications [32], decision support systems [33], and a variety of emerging technologies that can collect, store, and even analyze health data in real time are part of the integration between technology and healthcare [34]. The healthcare domain has grown rapidly over the last ten years and can now support the provision of high-quality care to patients, satisfying both patients and healthcare professionals and reducing healthcare consumption and costs [35]. Contextual characteristics play a central role in the effectiveness of healthcare, emphasizing the necessity of thinking about the definitions and dimensions of context in designing a healthcare domain model. [36] Context is the key element of successful computing in the telemedicine domain. In medical domains, context refers to the relevant data that relate to a patient, such as medical history, symptoms, risk factors, and medications. It also allows more precise diagnoses and establishes more powerful treatment plans [37].

The existing categories of context in healthcare organizations are not very efficient for retrieving information that the user's needs and demands require [38]. Moreover, structured medical terminologies such as SNOMED CT provide standardized clinical vocabularies [39], they primarily focus on clinical documentation, and functional status. These systems are not designed to organize contextual information dynamically for use in real-time decision-making, and AI-driven healthcare system [40]. Our proposed categorization may fills this gap by offering a structure that integrates social, environmental, technological, and temporal dimensions elements that are often missing or scattered across current frameworks. In addition, the context may vary from one field to another, making a standardized approach to understanding and applying context in a specific domain [41]. Accordingly, capturing the needs and preferences that provide a complete and clear picture of the patient's information demand is crucial [42, 43]. Making a more meaningful contribution in this research area requires a better understanding of the most critical aspects of context in the healthcare domain. This approach aims to utilize the most relevant data that may improve an individual's health outcomes and quality of life [44, 45]. Its main purpose is to understand and identify the context and provide a generic model for monitoring daily life activity and handling emergency situations for a patient. This paper focuses on context categorization in healthcare, proposing a fifteen-category framework and applying it to three chronic diseases (COPD, asthma, and cardiovascular diseases). In the medical field, context is key for getting diagnoses right, predicting outcomes accurately, and treating correctly.

2.1.1 Research gap and novelty

Existing studies on medical context categorization have laid the groundwork for intelligent health monitoring systems, yet they fall short in several areas. Some existing Models lack specific healthcare parameters or crucial dimensions like social determinants of health, Service Level Agreements (SLA), or real-time medical data. Therefore their applicability in developing robust, responsive, and personalized telemonitoring systems remains limited.

2.1.2 Motivation and contribution

The need for a unified and detailed context categorization framework in healthcare remains unmet. Existing models often lack medical specificity, adaptability, or completeness. This study makes three key contributions:

- Introduce a fifteen-category context model tailored for chronic disease management.

- It extends existing frameworks by incorporating critical dimensions such as Service Level Agreements (SLA) and social context, while also refining other categories through the inclusion of additional contextual entities.
- Applies the model to three prevalent chronic diseases demonstrating their practical relevance and adaptability.

The article consists of five parts. The first presents the existing context categorizations. The second part is an overview of noncommunicable diseases. The third describes our perception of medical concepts. The fourth part provides a case study of designing healthcare systems for chronic diseases (COPD, asthma, and cardiovascular diseases). The final part discusses this research work and offers the authors' perspectives.

2.2 Medical context in healthcare

2.2.1 State of the art

Weiser defines context as "all the information that should be taken into consideration for an adjustment." While this is a helpful starting point, it is too abstract to apply directly in healthcare. Our approach builds on this idea by identifying and organizing the key context elements that matter most for healthcare delivery, monitoring, and personalized care. To address this issue, researchers have proposed various definitions of "context," some of which focus on listing specific contextual information, such as location, time, and environment. Lacking standard definitions for this term, some useful characteristics can emerge from some of its most common features. While everyone has a general idea of what context is, finding a precise definition is not easy (see table 2.1).

Table 2.1 Dimensions context

Authors	Dimension
Gwizdka (2000) [46]	Internal and external
Gross and Specht (2001) [47]	Location, identity, time, environment, or activity

Antti Aaltonen (2002) [48]	Location, target, calendar, address book, users nearby, history, profile, direction, and speed
Prekop et al. (2003) [49]	Physical and logical
Mayrhofer (2004) [50]	Geographical, physical, organizational, social, emotional, user, task, action, technological, and time
Bunningen et al. (2005) [51]	Operational and conceptual
Miao et al. (2006) [52]	Sensed, profiled, and derived
Chong et al. (2007) [53]	Computing, physical, history, identity, and time
Miraoui and Tadj (2008) [54]	Trigger information and quality-changing information
Arianti Kurti (2009) [55]	User's profile, activity, and location/environment
Soylu (2009) [56]	User and environment
Zhong (2009) [57]	User, system, environment, social, and time
Tamine et al. (2010) [58]	User, platform, and environment.
Rizou et al. (2010) [59]	Low level and high level
Nageba E. (2011) [60]	Physical and abstract
Kim et al. (2012) [61]	5W1H (Who, When, Where, What, Why, and How)
Bin Guo (2013) [62]	Individual, social, and urban context
Boughareb et al. (2014) [63]	Device, task, user, document, spatiotemporal, environmental, and event
Meshali et al. (2015) [64]	Unusual behaviour
Banhato et al. (2015) [65]	Social-demographic factors, such as physical, emotional, and cognitive
Zhang et al. (2016) [66]	Vital signs, medical symptoms, risk factors, activities, and environment
Ameyed (2016) [67]	Time, space, and purpose. Psychological Identity, location, status (physical parameter, medical device), and time
Ahmed et al. (2017) [68]	Individual, medical, auxiliary, location, device, activity, and environmental
Neumuth et al. (2018) [69]	Quality of patient life
A.S. Almobarak (2019) [70]	Patient profile, time, location, environment, and social factors
Ajami et al. (2018) [71]	Medical history, disease, treatment, person, technology, task, and event Organization and policies
Taiwo et al. (2020) [72]	Physical factors
Kayes et al. (2020) [73]	Patient information in smart space
Kouamé et al. (2022) [74]	User, resource, and environment

2.2.2 Review state of the art

Many researchers have proposed different ways to categorize medical and general context information. For example, Gwizdka introduced a basic distinction between internal context (related to the user, like emotions or thoughts) and external context (environmental factors) [138]. Gross and Specht suggested five key context categories: location, identity, time, environment, and activity [139]. Aaltonen extended this by including calendar data, user profiles, nearby users, direction, and speed [140]. Prekop et al. proposed two types of context: physical context, which refers to concrete and observable factors, and logical (or abstract) context, which includes internal and mental states[141]. Mayrhofer expanded the scope further by including dimensions such as organizational, social, emotional, user, task, action, technological, and time [142]. Bunningen et al. organized context into two layers: operational (raw, collected data) and conceptual (interpreted or abstracted information) [143]. Miao et al. introduced three types of context: sensed (captured directly), profiled (based on past data), and derived (inferred from combinations) [144]. Chong et al. categorized context into computing, physical, history, identity, and time [145]. Miraoui and Tadj divided contextual information into trigger information (which initiates automatic services) and quality- changing information (which affects the format or quality of service delivery) [146]. Kurti focused on three key areas: the user profile (age, gender), the user's activity, and the location or environment [147]. Soylu grouped context into two main dimensions: user factors (preferences, physical/mental traits) and environmental factors (location, time, lighting) [148]. Zhong proposed five dimensions: user, system, environment, social, and time [149]. Tamine et al. classified context into user, platform, and environment [150]. Rizou et al. distinguished between low-level context (basic sensor data) and high-level context (interpreted information) [151]. Nageba presented a simple classification: physical and abstract context [152]. Kim et al. introduced the well-known 5W1H modelWho, What, When, Where, Why, and How to represent minimal, yet complete, context dimensions [153]. Guo considered three levels: individual, social, and urban context [154]. Boughareb et al. proposed a more extensive taxonomy including device, task, user, document, spatiotemporal data, environment, and event [155]. Meshali introduced unusual behavior as a contextual element [156]. Banhato emphasized sociodemographic context, covering physical, emotional, and cognitive factors that affect health [157]. Zhang et al. introduced a healthcare-focused model using vital signs, symptoms, risk factors, activities, and environment [158]. Ameyed described context in terms of time, space, purpose, and psychological aspects [159]. Ahmed et al. identified individual, medical, auxiliary, location, device, activity, and environmental components [160]. Neumuth et al. added quality of patient life as an important category [161], while Almobarak presented context as patient profile, time, location, environment, and social factors [162]. Ajami et al. proposed a comprehensive model covering medical history, disease, treatment, person, task, technology, event, organization, and policies [163]. More recent models include Taiwo, who focused on physical context like location and orientation [164]; Kayes, who identified user, resource, and environment as the core dimensions [165]; and Kouamé, who connected computing context with service-level agreements (SLAs) for healthcare systems [166].

2.2.3 Limitation state of the art

Despite the numerous categorizations proposed in the literature, may not provide a comprehensive, balanced, and practically applicable structure for medical context in healthcare. Many existing models are either too generic, lacking the granularity needed for real-world systems, or too narrow, focused on isolated factors such as time, location, or user preferences.

For example, studies such as those by Gross et al. [138], Soylu [148], and Kim et al. [153] focus on fundamental context components but do not consider disease-specific parameters, technological constraints, or organizational infrastructure. On the other hand, taxonomies like those of Boughareb et al. [155] and Ajami et al. [163] attempt broader frameworks but still lack critical dimensions such as social determinants of health, Service Level Agreements (SLAs), or real-time medical data integration.

Moreover, no existing framework has been thoroughly validated through detailed case studies across multiple chronic diseases. This limits their effectiveness in building context-aware telemonitoring systems, Our work addresses these gaps by:

Proposing a novel and structured categorization of 15 context categories tailored for medical applications.

Introducing new categories such as social factors, medical data, and SLA factors rarely or insufficiently addressed in prior models.

Applying and validating this categorization through three practical chronic disease case studies (asthma, COPD, and cardiovascular diseases), showcasing its adaptability and completeness.

A comparison of previous categorization frameworks highlights several important limitations. Table 2.2 summarizes the key models, their areas of focus, and the contextual dimensions they overlook. Our proposed framework builds upon these existing models while addressing their gaps by introducing essential categories such as social factors, Service Level Agreements (SLA), and real-time medical data and by refining and expanding others to better support healthcare system.

Moreover, our contribution lies not only in refining existing taxonomies but also in introducing a categorization model that may enhance telemonitoring systems, particularly within telemedicine contexts.

Table 2.2 Comparison of context categorization models.

Author(s)	Number of categories	Domains covered	Key categories included	Missing categories compared to our model
Ajami et al. (2018) [163]	14	Healthcare	Medical history, disease, person, event, organization	Social factors, SLA, medical data
Gross and Specht (2001) [139]	5	General computing	Location, identity, time, environment, activity	Medical-specific, organizational, SLA

Kim et al. (2012) [153]	6	General modeling	Who, What, When, Where, Why, How	Social and policy categories
Zhang et al. (2016) [158]	5	Vital sign monitoring	Vital signs, symptoms, risk factors, activities, environment	Organizational, task, SLA
Ahmed et al. (2017) [160]	7	Mobile healthcare	Medical, auxiliary, location, device, activity	Social, organizational, SLA
Boughareb et al. (2014) [155]	7	Smart environments	Device, task, user, document, environment, spatiotemporal	Social, SLA, policy
Almobarak and Jaziri (2019) [162]	6	Patient profile and environment	Time, location, environment, social factors	SLA, technology, task
Kouamé et al. (2022) [166]	3	Computing context	User, resource, environment	Medical-specific dimensions
Our model	15	Chronic disease healthcare	Person, location, history, activity, environment, *temporal*, *individual data*, *social factors*, *SLA*, *medical data*, disease, task, event, organization	None

2.3 Categorization of medical context in healthcare

The process of developing the medical context requires a deep understanding of the healthcare domain and its dimensions. Due to the complexity of context, the creation of a customized context is a feasible solution. Therefore, the medical context can comprise several categories, a division that allows understanding, standardizing, and centralizing the management of all diseases. It connotes representing a disease by investigating a patient's signs, medical history, and demographic information. In the healthcare field, context-aware systems can play a role in enabling the self-management of a disease by collecting data and dividing it into categories that help to build the structure of a medical context [167]. In addition, using this information will enable personalized care designed and tailored to each patient, to identify symptoms, risk factors, and effective self-management strategies for managing a particular disease and to provide more targeted care and the ability to make a suitable decision. Moreover, the medical context's efficiency in the healthcare domain appears when working in a real-time system that may help provide and assure services to the patient, especially in an emergency [168]. The structure of context depends on analyzing health problems to generalize context categorization applicable in the medical domain. The set of context information presented above clearly demonstrates that what context is depends on what requires description.

Several studies categorize the medical context with similar entities. For example, Zhang et al. define the medical context as the information for a patient's medical situation, roughly divisible into the following categories: patient's vital parameters, medical symptoms, risk factors, activities (standing, walking), and surrounding environment (room temperature) [158]. Neumuth et al. mentioned that when dealing with context, the categories can form four entities: identity, location, status (e.g., physical parameters, processes running currently on a device), and time [161]. In addition, Almobarak et al. mentioned that typical user contexts may include time, location, gender, age, weather, culture, financial level, educational level, and demography [162]. Ajami et al. constructed a classification of fourteen categories that built the structure of a medical context for further use in the management of various diseases [163]. They became more specific in the medical domain context by adding some categories for example, "medical history," the disease, and the treatment that are important because they help to illuminate the patient's medical situation. In addition, the "person" category includes all individuals involved in building the context, such as the patient, the healthcare provider, and the family member. "Technology" is also an important category; it includes devices for monitoring the patient's health status. They also added "tasks, ""events," "organizations," and "policies" as categories. By identifying and analyzing these dimensions, researchers can develop a more comprehensive and effective approach to contextual management in healthcare environments, thereby improving patient outcomes and enhancing the overall quality of healthcare services.

This chapter builds a context categorization in the healthcare domain by enhancing the structure of the context categorization that Ajami et al. proposed [163], adding other new categories (“social factors,” “medical data,” and “SLA”), eliminating some categories, and transferring others. Studies have shown that social factors impact population health outcomes. Each person’s behavior affects his or her health, which, in turn, is associated with his or her social or economic status and the corresponding environmental conditions [169]. For example, people with higher incomes generally live longer, with better overall health, associating higher incomes with supporting better healthcare, especially for high-cost diseases. Higher education leads to higher income, as well as the ability to understand healthy habits. Social relationships and environmental conditions may affect health status; people living in poor neighbourhoods are more susceptible to different types of diseases [169–172]. In fact, the medical data that include symptom data, disease data, and health data have an important role and are in a separate category with different entities. Moreover, following its introduction in 1980, various fields, such as telecommunications, call centres, security, cloud computing, and health systems, have utilized SLAs. However, remote medical monitoring platforms manipulate large volumes of patient data, and the risks of data loss or poor data quality are real. The context changes dynamically, and meeting quality of service (QoS) requirements is a challenge. Moreover, a virtual dynamic SLA monitors the patient context and updates the violation control interface, the main document that guarantees (by SLA) the QoS a computer system provides and defines an SLA between the IT service provider and the consumer of its services [173, 174].

The context categorization occurred by completing three main steps:

Review and analyze thirty relevant, pertinent, and helpful articles, to explore existing parameters, acquire information from healthcare professionals, and gather data from the WHO and such health standards as

Identify the parameters and organize them into fifteen categories; the classification that the previous section discusses provides a broad foundation for determining the context categorization of the medical field, as the list below describes:

1. The "person” context contains three subcategories: a patient, the main element of the medical context: a physician who specializes in the diagnosis and treatment of a disease;

a family member, since genetic factors may play a role in the development of many diseases [175].

2. “Individual data” consists of a set of basic characteristics essential to providing suitable healthcare customized to patient needs. Demographic factors can be an example of individual data. They include four subcategories: age, gender, BMI, and occupation [176].

3. The “social factors” context includes factors that relate to patient social status:

    - Educational level: Educated people are more aware of what keeps them healthy; they experience better health that high levels of self-reported health and low levels of morbidity, mortality, and disability reflect [177].
    - Income: Economic or financial status; some chronic diseases require high-cost healthcare [178].
    - Social relations that affect how people behave can, in turn, affect their health [179].

    The WHO defines “quality of life” as “an individual's perception of his position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns” [180, 181].

4. “Temporal context” relates to date and seasons [182, 183]. Some diseases may be more common during certain seasons or times of the year. Understanding temporal context is important in healthcare, where the change in season or time can be critical and affect patients.

5. “Location”: With the availability of location awareness technology, tracking and transporting patients to hospitals or medical centres when they require urgent medical intervention become easier. This highlights the importance of determining the location of patients in pervasive healthcare system [184]. Location information provides a more detailed, meaningful, and identifiable description of the physical characteristics of a place, which can impact patient health status. For example, the altitude of a location may be harmful to certain types of patients. Representing location in healthcare system takes three forms:

geographic coordinates, named spaces (such as rooms), and relative location descriptions that describe the position of an object in relation to surrounding objects [185].

6. “Activity context” consists of everyday activities, such as exercising and kind of work. Physical activities may hold the greatest risk of triggering or exacerbating a patient's health condition. In the healthcare domain, automated recognition of human activities is an increasing need. By tracking the current situation of users in smart spaces, new features that provide more accurate and consistent results can enhance healthcare applications. Additionally, utilizing the activity context can help warn users if they engage in excessive levels of exercise, to prevent exacerbations or serious complications [186].

7. “Technology” encompasses both hardware and software components that humans create. It includes computing devices, mobile phones, personal digital assistants (PDAs), and sensors [187,188]. The technology context not only refers to computing resources but also such factors as network connectivity and platform characteristics. Additionally, this category includes biomedical equipment, which can comprise three types: fixed infrastructure equipment (heating, ventilation, and air conditioning), support equipment (laboratory equipment), and medical equipment [189].

8. The “environment” context consists of the environmental risk factors that lead to the development of certain diseases, such as urbanization, pollution, allergens, radiation, and weather conditions [186]. Ubiquitous healthcare systems should acknowledge that even small changes in the indoor or outdoor environment can significantly influence patient behaviour [190]. Many environmental factors have been associated with disease progression.

9. “Medical data” consists of signals that biomedical equipment, sensors, or smart wearable devices capture. Treatment of a disease often requires this data [132].

10. “History” consists of the patient’s medical information and his/her long-term follow-up. This part of the context should contain sufficient information about the physical examination,

diagnostic test results, family diseases, comorbidities, and medications [191]. This can help in providing proper patient care.

11. “Disease category” refers to the causal relationship between symptoms, causes, and treatments that accompany a particular medical condition. Providing fully customizable services responsive to the patient's condition requires having a classification system for human diseases. This classification system will enhance the existing taxonomy of medical context, by linking diseases with appropriate medications, allowing for efficient treatment administration [192, 193].
12. “Task” is an essential category where healthcare providers utilize the context to carry out tasks and interact with patients to manage critical or uncertain situations. Therefore, action plays a crucial role in the proposed categorization, as Lasierra et al. described [194]. Medical tasks can comprise four types: monitoring, analysis, planning, and execution tasks. A set of conditions that rules express determines the relationship between them.
13. Service Level Agreement (SLA): Many contexts can qualify for classification into the computing context category [138]:

    The percentage of time the system must be available.

    The number of users it can serve simultaneously.

    The delivery of the message and the latency time.

    Message security, the notification schedule, and available dial-in.

    Help-desk is always available to ensure appropriate user service.

    Quality of service (QoS) .
14. “Event”: This category can help to detect occurences of expected conditions, i.e., indicating abnormal physical situations [195].

15. “Organization” refers to the healthcare situation, such as a hospital or a clinic. A healthcare organization is responsible for tasks that include defining and monitoring the delivery process of service care, assigning a care provider and medical equipment when a patient

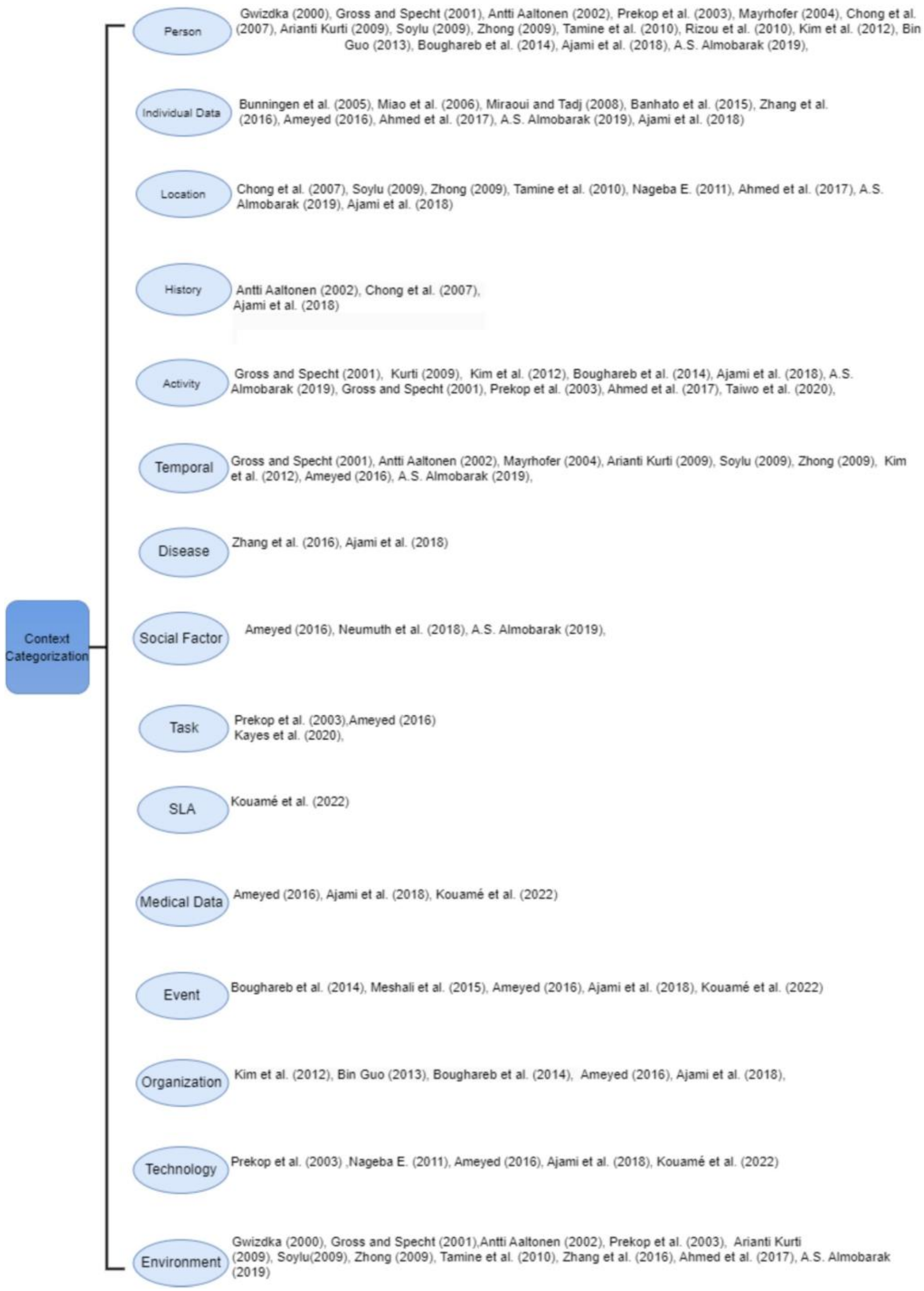

Figure 2.1 Existing context categories in medical domain

requires it, managing human and physical resources between services, and collaborating with other centres [196].

Group together the same entities within the healthcare environment, ensuring that they align with medical context requirements (accuracy, interpretability, explainability, performance).

Figure 2.1 shows the existing categorization proposed by several researchers that can easily help in interpretation of the context in healthcare domain. It illustrates that most of the existing categorizations missed out the essential entities in medical context. Only a few of these researchers partially address the contextual parameters of the medical domain; some researchers focus on presenting basic context components like location, environment, and activity, while others shed light on the technology, SLA and social factors. The structure of this figure highlights fifteen categories by combining the proposed entities used to satisfy medical requirements when working over context.

2.4 Overview of noncommunicable diseases

To define and understand the context, we propose to analyze the context categorization of three important chronic diseases, namely, asthma, COPD, and cardiovascular diseases. Noncommunicable diseases (NCDs), commonly referred to as chronic diseases, have a prolonged duration and arise from the interplay of genetic, physiological, environmental, and behavioral factors. Examples of NCDs include cardiovascular diseases (such as heart attacks and stroke) and chronic respiratory diseases (such as chronic obstructive pulmonary disease and asthma) [197]. NCDs are a major cause of death worldwide claiming the lives of 41 million people each year, about 74% of all deaths globally. Cardiovascular diseases are responsible for 17.9 million deaths each year, while asthma and COPD cause 4.1 million deaths annually [198]. Nowadays, telemonitoring can play an important role in achieving the management of chronic diseases and building a useful healthcare system to engage anytime and anywhere in a relevant context.

2.4.1 Asthma

Asthma is a chronic respiratory disease whose symptoms include wheezing, shortness of breath, chest tightness, and coughing [199]. These symptoms vary in severity and frequency among asthmatic patients [200]. Asthma often develops during childhood, but some individuals show late onset, sometimes after the age of 40 years [201]. In 2019, estimates showed that asthma affected more than 260 million people worldwide [202]. In Canada, more than 3.8 million individuals [203] over one year old are asthmatics. Epidemiological studies have shown that asthma is more prevalent among children, especially boys, and it affects women more than men [204].

2.4.2 Cardiovascular Diseases

Cardiovascular diseases (CVDs) are the leading causes of death in the world. Recent World Health Organization (WHO) statistics show an increase in the number of CVD patients worldwide, affecting 523 million patients, with an increase in the number of deaths this disease caused, reaching 18.6 million (32% of global mortality) in 2019 [205]. For example, in the United States, a person dies from CVD at least every 34 seconds, and in Canada, a death from CVD occurs at least every 5 min. Moreover, CVDs are among the costliest diseases; the total cost of CVD in the United States reached 378 billion USD between 2017 and 2018, according to the Medical Expenditure Panel Survey [205].

2.4.3 COPD

COPD is justly regarded as one of those dangerous maladies, likely to become the third leading cause of death worldwide by 2030 [206]. The main goal of COPD management is to keep the disease under control for as long as possible. This involves selecting the right medications, minimizing side effects and other health issues, preventing severe lung damage, and avoiding the worsening of symptoms [207]. The WHO reported 3.23 million people dying from COPD in 2019, the third most common cause of death worldwide [208].

2.5 Medical Context Case Studies

Management in the healthcare system includes detecting and treating diseases and avoiding the associated risk factors. Asthma, cardiovascular disease, and COPD are three chronic diseases. Telemonitoring can play an important role in chronic disease management; it helps in detecting symptoms, avoiding risk factors, and achieving self-management of the disease, reducing hospitalizations and long stays at high costs.

2.5.1 Asthma Context Categorization

Inhalation of specific allergens can trigger allergic asthma that other allergic diseases, such as food allergy, allergic rhinitis and atopic eczema, usually accompany, and which early onset in life characterizes[52]. While onset late in life characterizes nonallergic asthma, allergic reaction does not trigger it [77]. Environmental factors, including indoor allergens, outdoor allergens, air pollutants, respiratory viruses, tobacco smoke, irritants in the workplace, and weather conditions, can contribute to asthma pathogenesis. Other factors include psychological factors, physical activity, and obesity, in addition to certain medications [48,50,78]. On the other hand, genetic factors have an important effect on the inception, severity, and treatment of asthma [19]. Studies have reported the prevalence of allergic disease in first-degree relatives of affected individuals. Children of asthmatic parents are more likely to develop asthma than those whose parents have no history of allergic diseases [78]. The context categorization above (section III) can apply

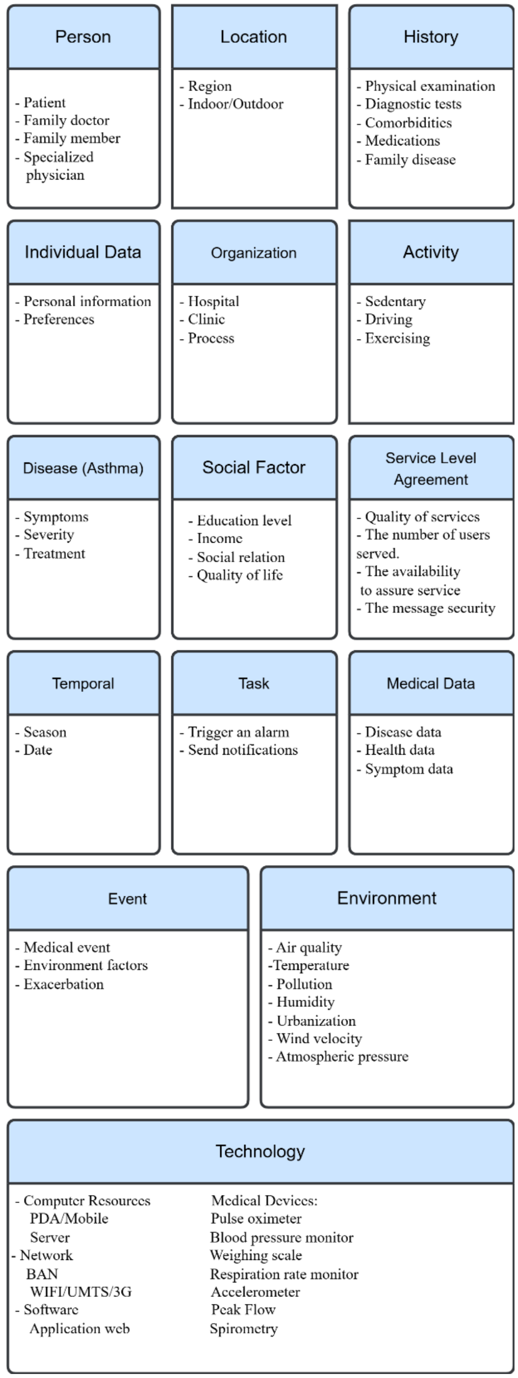

Figure 2.2 Asthma Context Categorization

in a case study of asthma. These basic categories with their entities appear below in Figure 2.2.

2.5.2 Cardiovascular Diseases Context Categorization

Cardiovascular diseases affect the heart and blood vessels. These diseases can range from very minor, such as varicose veins, to potentially fatal, such as a heart attack or stroke. Cardiovascular disease is the leading cause of death worldwide, and effective treatment of risk factors could prevent many such deaths. Diseases of the heart and blood vessels (CVDs) are often the most serious diseases worldwide. Reducing the likelihood of developing cardiovascular disease requires first identifying and then controlling risk factors. Examples of these risk factors include smoking, genetic factors, pollution, medical history, BMI, and stress. Knowing the risk factors associated with cardiovascular disease and acting accordingly increases the chances of maintaining excellent cardiovascular health and reducing the risk of potentially fatal disease [79,80,81]. The context categorization in section III can also apply to the cardiovascular diseases case study, with some modifications (Figure 2.3).

2.5.3 COPD Context Categorization

COPD is a respiratory condition that persistent airway obstruction not fully reversible (incurable) characterizes. COPD has symptoms similar to asthma: coughing, wheezing, chest tightness, and shortness of breath. Its cause is an inflammatory response to certain irritants, leading to the narrowing of the airways and limiting the airflow in the lungs [82,83]. COPD has several risk factors; active smoking is the cause of most COPD cases. Other risk factors of COPD include indoor fumes (tobacco smoke and biomass fuel combustion), outdoor pollution, meteorological factors (cold weather, atmospheric pressure, wind speed, and humidity), and occupational exposure to certain chemicals, gases, and organic and inorganic substances [84,85,86].

Furthermore, the development of COPD is associated with demographic features, such as age and gender, as well as social factors, including educational level and income

[87,88]. We can use the context categorization described above (section III) as a case study for COPD. Following (Figure 2.4) are the fundamental categories with their corresponding entities.

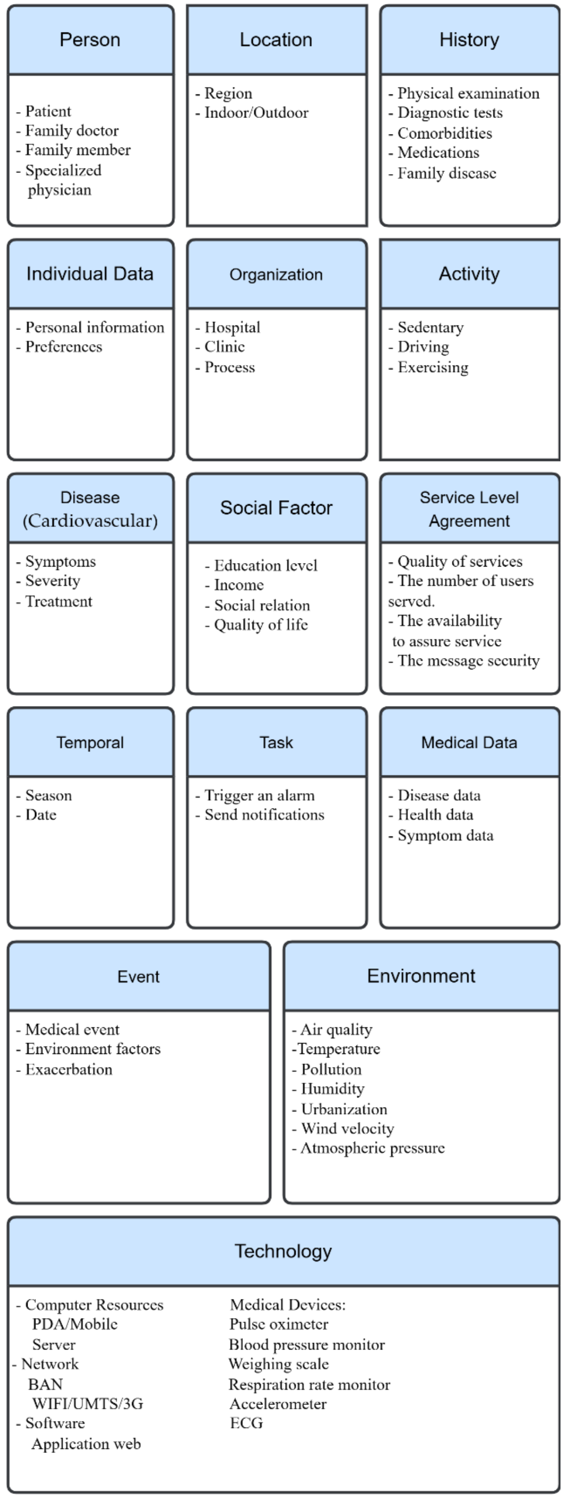

Figure 2.3 Cardiovascular Context Categorization

In summary, the case studies demonstrate how the fifteen categories model can adapt to disease-specific characteristics while maintaining a unified structure, highlighting its real-world feasibility.

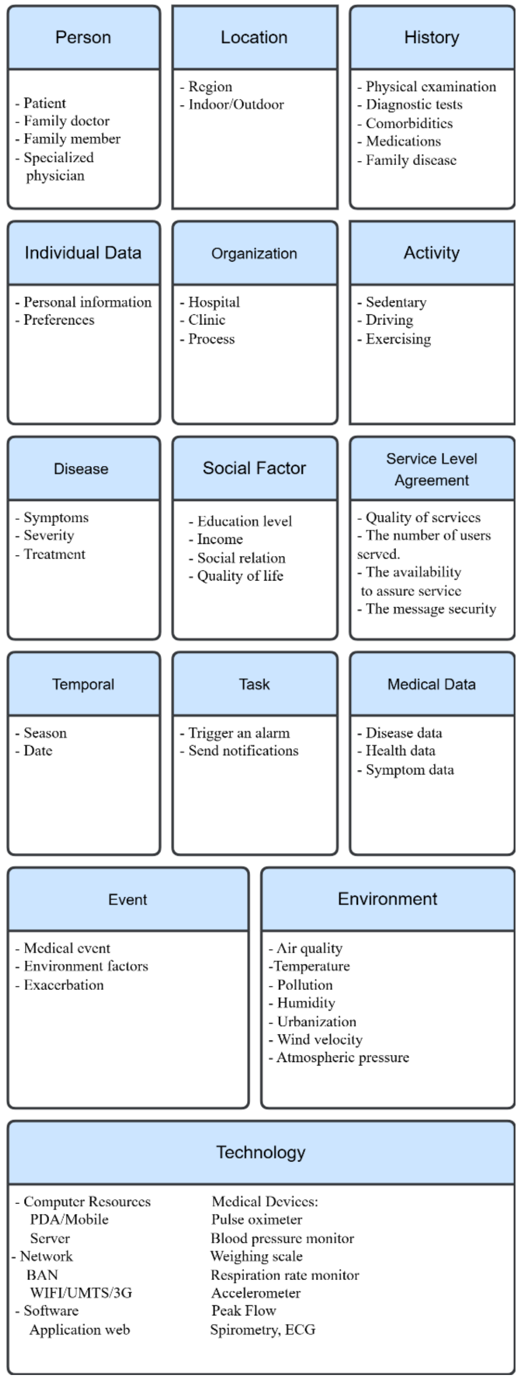

Figure 2.4 Context Categorization of the Medical Domain

2.6 Result and Discussion

A comparative analysis between our proposed fifteen context categories model and the taxonomy developed by Ajami et al. [36] reveals several significant enhancements. Ajami's framework, while comprehensive, does not account for key real-world parameters like social inequality, device-level service expectations (SLA), or dynamic medical data streams all of which are now central to modern healthcare delivery. Our model builds upon and extends this foundation by:

Introducing social factors (education, income, social relationships) as standalone elements influencing disease outcomes aligned with findings in public health and chronic care literature [60, 62].

Defining SLAs within medical systems to reflect service expectations, data transmission, and latency vital in telemonitoring platforms [44].

Highlighting medical data as a distinct category to represent real-time sensor readings, vital signs, and symptom tracking crucial for AI-based alerts and interventions [36].

The proposed context categorization model can be automatically populated using real-time inputs from IoT-enabled devices, including wearable sensors [63]. For instance, physiological data captured by smartwatches or respiratory monitors can be used to detect abnormal trends and trigger early warnings in COPD patients [64].

In addition, these well-structured context categories serve as standardized inputs for AI-based decision support systems, enhancing their ability to deliver timely and personalized care [65]. Table 2.4 presents examples of how specific context categories can be leveraged to support intelligent healthcare systems.

Table 2.3 Role of Context Categories in Enabling AI-Based Health Monitoring.

Context Category	AI Use Case

Medical Data	Used as input for diagnostic algorithms (e.g., ECG + HR to detect arrhythmias)
Activity	Tracked via wearables to detect abnormal behavior or exercise intolerance
History	Combined with real-time data to estimate risk scores (e.g., cardiovascular events)
Environment	Adjusts alerts for asthma or COPD based on pollution or allergens
SLA	Ensures alerts are sent within agreed response times for critical cases

Through case studies involving asthma, COPD, and cardiovascular disease, we demonstrate that the proposed model as shown in figure 2.5 below adapts flexibly across conditions. Each disease exhibits distinct contextual entity profiles, confirming the model's adaptability while preserving a unified structure. The context categorization reduces the risk of incomplete context modeling, which is often a source of false diagnosis or ineffective intervention [66]. Overall, this refined framework not only improves the clarity and usability of medical context but also enhances diagnostic precision, self-management, and healthcare personalization.

2.6.1 Benefits of context categorization in medical domain

Context categorization not only supports accurate diagnoses but also allows developers and healthcare systems to tailor services to individual patients [67]. The ability to isolate and update context entities directly within categories improves system performance and supports healthcare application design [68]. The proposed context categorization framework can be integrated into existing healthcare systems by aligning each context category with data structures and resource definitions used in HL7as illustrated in table 2.5 [69]. The model has the ability to grow depends on its modular structure, allowing healthcare organizations to adopt relevant categories based on their needs and gradually expand [33]. While this study is conceptual, future work includes building prototype systems and collaborating with healthcare providers to test deployment and refine integration strategies [36].

Table 2.4 Context Category Integration into Healthcare Systems via HL7 mapping

Framework Category	**HL7 Resource(s)**
Medical Data	Vitals, symptoms, test results
Individual Info	Demographics, family context
Task & Activity	Scheduling, tracking, interventions
Time & Events	Timing, acute events, session records
SLA & Organization	Governance, provider

In the healthcare domain, context categorization has proven their efficiency, not only by defining and organizing context in the medical domain, but also by helping doctors and medical experts in diagnosing disease and taking precluding procedures to avoid worsening of symptoms. Moreover this classification helps in self management tasks which aim to avoid risk factors and reduce hospitalization. There are many diseases that are comorbid, in other words, they share similar symptoms but they are diagnosed in various methods [92]. For example, Asthma and COPD have common signs and symptoms, but they are assessed in different ways [92, 93]; many researchers made use of the different risk factors and symptoms in order to take the final decision, in other words they used many context categorization to achieve the right diagnosis. As a result, any missing entity or category can lead to wrong prediction which is unacceptable in the medical domain [90]. Context categorization can be useful for developers to improve system performance and understand different situations by making them simple to handle [92, 93]. For instance, changing a certain entity can be resolved by going directly to the category, where the context belongs, and making the appropriate change, which can be valuable for the performance [92]. The table below (table 2.6) shows that different researchers have mentioned many entities that we used to built the context categorization. This table can demonstrate the important role played by the context categorization in the medical domain, particularly in the accurate diagnosis of diseases.

In addition the proposed context model would manage sensitive patient information applied in real-world systems. Ensuring privacy and data security is essential where the model supports context-aware handling of sensitive information, allowing non-critical personal data (location or social factors) to be selectively processed or restricted according to defined SLA rule.

Table 2.5 Recent evaluation of diseases using contextual categorization

Context Categorisation	References
Person	Toskala & Kennedy, 2015 [94]; Poplin et al., 2018 [95]
Individual Data	Moore et al., 2010 [96]; Vaghefi et al., 2019 [97]
Social Factors	Braveman & Gottlieb, 2014 [98]; Abohelwa et al., 2023 [99]
Temporal Context	Schatz & Rosenwasser, 2014 [100]; Abohelwa et al., 2023 [101]
Location	Zhang et al., 2020 [102]; Baggett et al., 2018 [103]
Activity Context	Moshawrab et al., 2023 [104]; Cillekens et al., 2023 [105]

Technology	van Vliet et al., 2017 [106]; Arnould et al., 2021 [107]
Environment	WHO Asthma Fact Sheet, 2020 [108]; Vaghefi et al., 2019 [109]
Medical Data	Huffaker et al., 2018 [110]; Poplin et al., 2018 [111]
History	Moore et al., 2010 [112]; Abohelwa et al., 2023 [113]
Disease	Agache & Akdis, 2012 [114]; Zhang et al., 2023 [115]
Task	Baker et al., 2023 [116]
Service Level Agreement (SLA)	Kouame et al., 2018 [117]
Event	Sanchez-Morillo et al., 2016 [118]
Organization	Finkelstein et al., 2017 [119]; Tsiligianni et al., 2020 [120]

2.7 Conclusion and future works

Examining context categorization in the healthcare domain reveals that this method can apply to different medical conditions. Visualizing context categorization for chronic diseases, such as asthma, COPD, and cardiovascular disease, has yielded valuable insights for managing these conditions. This visualization offers a promising framework for advancing our understanding and management of chronic illnesses within healthcare systems. In addition, in this context with all risk factors, symptoms within this categorization are important because of their strong influence on the accuracy of the medical domain.

Although this study focused primarily on chronic conditions, the proposed context model is also applicable to emergency and acute care scenarios. For instance, the “event” category supports detection of incidents such as trauma, stroke, or cardiac attack. Moreover location data facilitates faster intervention through GPS or hospital-based positioning systems. In addition medical data can be streamed in real-time from emergency monitoring devices, while categories such as “SLA,” “task,” and “organization” enable the coordination of emergency workflows and enforcement of care quality standards. These categories demonstrate its potential use across a broader range of healthcare context.

However, acknowledging the approach's limitations and understanding the drawbacks of this method are critical. Providing a useful structure in the healthcare field requires continuous updates to guarantee precise categorization. Nonetheless, we must remember that this way of organizing context might not fit all healthcare conditions. While our framework has been validated conceptually and across multiple case studies, future work will involve empirical testing through integration with a live telemonitoring platform and urgent care. We plan to evaluate context recognition accuracy, system performance, and its effect on patient outcomes using real-world datasets.

CHAPTER 3
KNOWLEDGE REPRESENTATION MODELS IN HEALTHCARE : A SURVEY

Résumé

Les modèles de représentation des connaissances visant à présenter les données de manière structurée et compréhensible ont gagné en popularité en tant qu'axe de recherche dans la quête de l'intelligence de niveau humain. Les êtres humains possèdent la capacité de comprendre, de raisonner et d'interpréter les connaissances. Ils acquièrent ces connaissances à travers leurs expériences et les utilisent pour accomplir diverses actions dans le monde réel. De manière similaire, les machines peuvent également effectuer ces tâches, un processus connu sous le nom de représentation et raisonnement des connaissances. Dans cette étude, nous présentons une analyse approfondie des modèles de représentation des connaissances et de leur rôle essentiel dans la gestion de l'information au sein du domaine de la santé. Nous proposons une vue d'ensemble de différents modèles, notamment les ontologies, la logique du premier ordre et les systèmes à base de règles. Nous classons quatre modèles de représentation des connaissances selon leur type, tels que les modèles graphiques, mathématiques et autres types. Nous comparons ces modèles selon quatre critères : l'hétérogénéité, l'interprétabilité, la scalabilité et le raisonnement, afin de déterminer le modèle le plus approprié pour répondre aux défis du domaine de la santé et atteindre un niveau élevé de satisfaction.

Abstract

Knowledge representation models that aim to present data in a structured and comprehensible manner have gained popularity as a research focus in the pursuit of achieving human-level intelligence. Humans possess the ability to understand, reason and interpret knowledge. They ac- quire knowledge through their experiences and utilize it to carry out various actions in the real world. Similarly, machines can also perform these tasks, a process known as knowledge representation and reasoning. In this

survey, we present a thorough analysis of knowledge representation models and their crucial role in information management within the healthcare domain. We provide an overview of various models, including ontologies, first-order logic and rule-based systems. We classify four knowledge representation models based on their type, such as graphical, mathematical and other types. We compare these models based on four criteria: heterogeneity, interpretability, scalability and reasoning in order to determine the most suitable model that addresses healthcare challenges and achieves a high level of satisfaction.

Keywords: knowledge representation model; healthcare requirements; heterogeneity; interpretability; scalability and reasoning

3.1. Introduction

As the volume and diversity of data rapidly increase, the need to effectively represent and process knowledge becomes more critical [163]. It is not feasible to extract knowledge directly from large, unstructured data repositories without structured models. As Stafford explains, “data” refers to raw facts, “information” is organized data, and “knowledge” is the meaningful interpretation of that information [164].

A precise understanding of any domain relies on accurate contextual representation. This highlights the need for methods to express knowledge via visual (e.g., trees and frames) or logical (e.g., first-order logic) structures [165]. Knowledge representation models (KRMs) serve this purpose by enabling structured sharing, organization, and inference across scientific domains [166]. They are widely adopted in expert systems, AI, robotics, and natural language processing, with recent studies estimating that 80% of applications in these domains use KRMs for reasoning and decision support [167].

KRMs also facilitate communication and knowledge delivery across sectors such as healthcare, education, and finance [168]. Specifically in healthcare, KRMs enhance the integration of electronic health records (EHRs), clinical guidelines, and decision support systems to assist in treatment planning and accurate diagnostics [169].

Despite their value, KRMs in healthcare face ongoing challenges including heterogeneity, interpretability, scalability, and reasoning complexity [170]. Healthcare data is particularly diverse and merely aggregating it without modeling contextual relationships may obscure crucial insights. Interpretability is especially vital in clinical settings, where end-users must understand and trust model outputs to make informed decisions [171]. Scalability is another pressing requirement given the exponential growth of healthcare data. Models must handle this volume efficiently without sacrificing performance [172].

This review emphasizes the strategic importance of KRMs in healthcare, outlines specific sector demands, and compares model efficiency using satisfaction metrics to determine the most suitable approaches [173]. This chapter comprehensively reviews KRMs in general, with a specific focus on their applications in healthcare. It includes an overview of recent surveys and chapters on the topic, emphasizing the significance of these models. This review process is critical for determining the functional relevance of each model and understanding its potential applications. The chapter also highlights the specific requirements of the healthcare sector and compares KRMs to assess their efficiency using a satisfaction ratio formula. The formula, calculated for each model, aids in identifying the most suitable one based on healthcare needs.

3.2. Background Of Knowledge Concepts And Types

3.2.1 Data, Information, Knowledge and Wisdom

Data are fundamental facts used for transactions and decision-making. However, raw data are not useful until they are organized, structured, and interpreted to produce meaningful information [174]. While all data can technically be part of information, not all information is merely raw data contextualization is key to extracting meaning and insight [175]. For example, recording that "Sami has a body temperature of 38 °C" is a raw data point. But if Sami’s normal body temperature is usually 36.5 °C, then the added context transforms it into information: Sami’s temperature is elevated.

Understanding occurs when one comprehends this pattern, knowing that 38 °C generally signifies a mild fever and recognizing that Sami may be ill. Wisdom, in turn, is the intelligent application of that knowledge to act: for instance, monitoring Sami's condition and consulting a doctor if symptoms persist. This progression from data to decision illustrates the DIKW (Data, Information Knowledge and Wisdom) model, represented in Figure 3.1, and highlights how structured representation improves health decision-making.

Figure 3.1 Data, information, knowledge and wisdom chain

3.2.2 Knowledge Types

Understanding the different types of knowledge is crucial because it greatly influences how models are represented. Additionally, the specific domain of the knowledge also affects the type of model that is used [176]. Knowledge can be categorized into three main categories: explicit knowledge, tacit knowledge and implicit knowledge, as shown in Table 3.1. Regardless of the type, all knowledge needs to be represented and expressed in a suitable format and language to be processed by a computer.

Table 3.1 Types of knowledge.

Knowledge Types	**Description**	**Examples**
Explicit knowledge	A form of knowledge that can be expressed in various formats such as plain text, documents, spreadsheets,	Scientific formulas; recipes for a chef; books.For example: The capital of France is Paris.

	databases, images, etc., and can exist in structured or unstructured forms [177,178].	
Implicit knowledge	Knowledge gained through incidental activities, or without awareness that learning is occurring [178].	How to walk, run, ride a bicycle or swim.
Tacit knowledge	This type of knowledge is subjective, based on personal experience, and often difficult to express. It resides in the human brain in an inexpressible form and is of an intellectual nature. This original form of knowledge is primarily obtained through learning and experiences [179].	Doctor's diagnosis; musician's improvisation.

Tacit knowledge is acquired naturally through learning and experience, and it becomes explicit knowledge when we document or represent it in various ways. In contrast, implicit knowledge, which is a component of this innate knowledge, can be acquired incidentally.

3.3. Knowledge Representation Models (KRMs)

According to Shrobe et al., “A knowledge representation is a fragmentary theory of intelligent reasoning; it should provide the set of inferences the representation allows or recommends” [180]. Knowledge representation involves a formal description of knowledge that emphasizes computer processing as a means of representation. KRMs are used to organize and describe data in a standardized model, facilitating access and handling by other components [181]. There are various models, each with its own strengths and weaknesses. This survey categorizes KRMs into five main categories based on types of representation. Each category offers distinct advantages and disadvantages depending on the specific problem domain, highlighting the importance of selecting the appropriate model for each use case.

3.3.1 Graphical Representation Models

The graphical representation model focuses on visually depicting relationships between variables and components using graphs [182]. These models typically utilize nodes and edges to represent entities and their interconnections, visually illustrating knowledge relationships. Below are brief descriptions of nine graphical models.

A Bayesian network is a powerful model that combines probabilistic and graph-based representations [158]. It finds extensive application across domains such as machine learning, data mining and diagnosis due to its ability to perform evidence-based inference that aligns with human intuition [183]. To illustrate, consider the example mentioned earlier represented in a Bayesian network graph (Figure 3.2). In this graph, tables are populated with arbitrary probability values. Each table represents the probabilities based on whether specific conditions are satisfied or not. For example, the probability of a patient having a disease might be 0.9, while the probability of having both a high temperature and cough could be 0.8. Conversely, the probability of not having a disease might be 0.6, even if the probability of having a high temperature alone (without a cough) is 0.9.

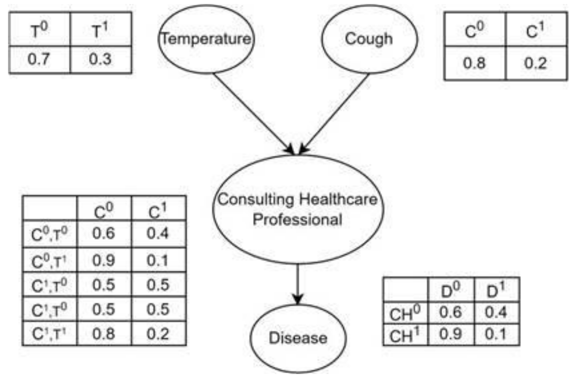

Figure 3.2 Bayesian network representation

Ontology is a structured method of representing knowledge by defining concepts and their relationships, using two different languages: RDF and OWL [185]. Figure 3.3 displays the ontology representation for the previously mentioned example. Ontologies have a wide range of applications in various fields, such as biology, medicine, cultural heritage, accounting and social media [186].

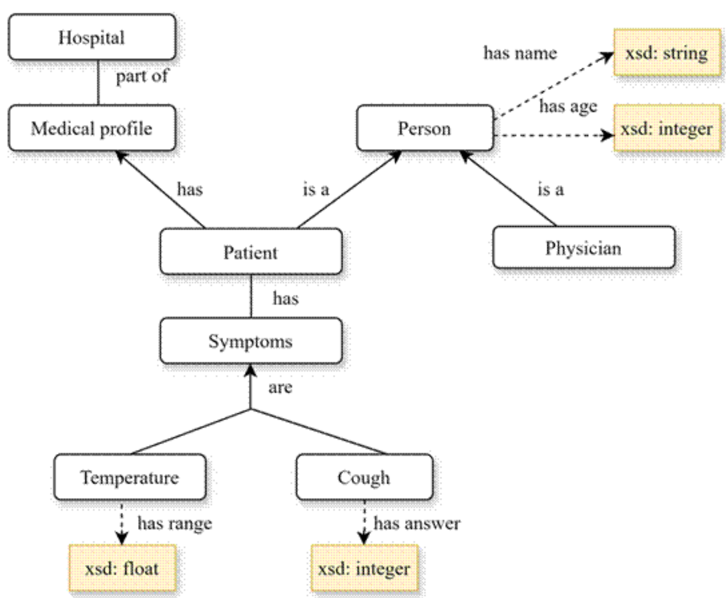

Figure 3.3 Ontology representation

A decision tree is a graphical model used to represent knowledge, where each node corresponds to a test or decision based on an attribute or feature [187], as represented in Figure 3.4. Decision trees find applications in various fields, including engineering, education, law, business, healthcare and finance [188].

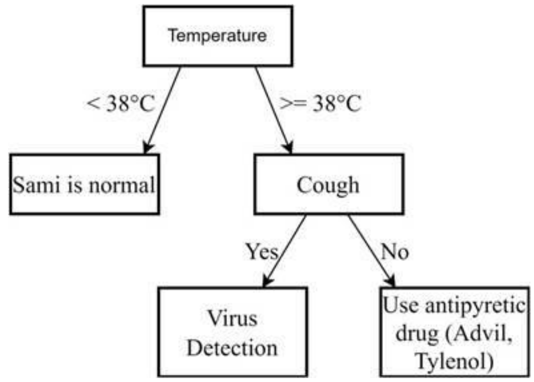

Figure 3.4 Decision tree representation

A neural network is a machine learning model that mimics the structure and functionality of the human brain, comprising interconnected layers of nodes or neurons [189] as shown in Figure 3.5. Neural networks find applications across diverse domains including finance, healthcare and education [190].

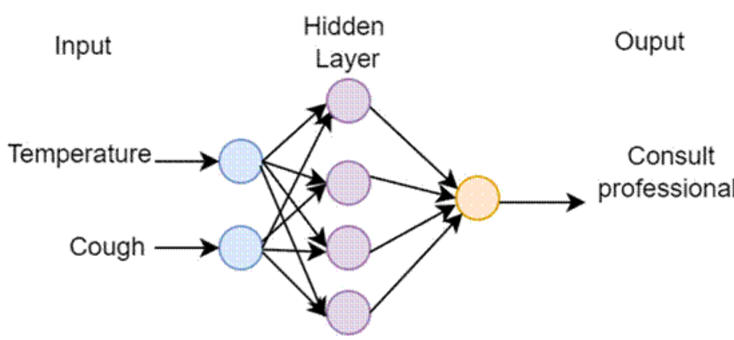

Figure 3.5 Neural network graph

Unified Modeling Language (UML) is rooted in the object-oriented paradigm, providing a framework for knowledge representation where relationships are defined using associations, aggregations and generalizations [191]. Figure 3.6 illustrates the previously mentioned example. The UML class diagram, a key component of UML, finds versatile applications across various domains including banking, finance, internet, aerospace and healthcare [192].

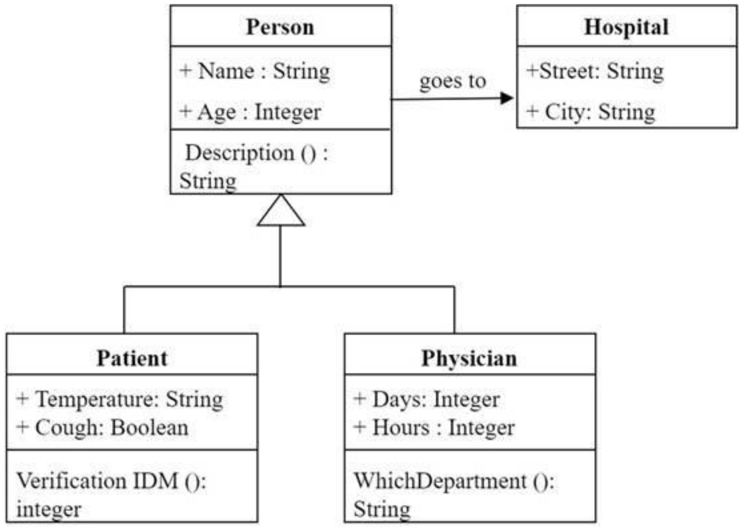

Figure 3.6 Unified modeling language representation

A frame is a model that represents concepts as structured sets of attributes and values [193] as illustrated in Figure 3.7. It finds application in diverse fields such as education, finance and others. Frames are designed primarily for straightforward one-to-one relationships, which can sometimes limit their ability to represent intricate or nuanced relationships between entities.

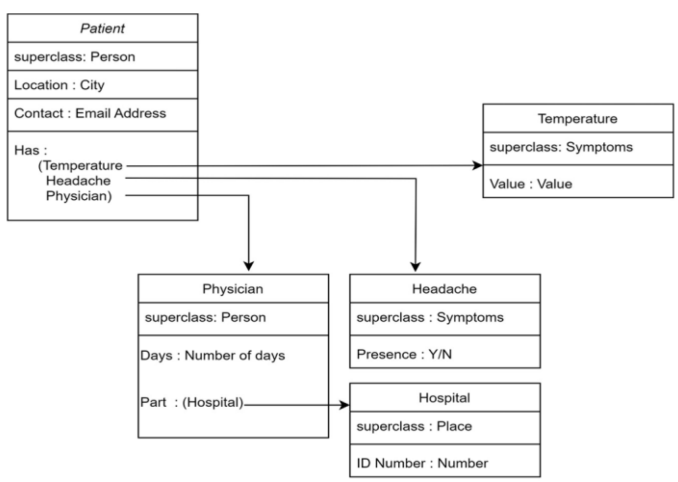

Figure 3.7 Frame representation

Tuple-based representation is a model that uses sets of data, called tuples, to represent knowledge. In relational databases, a tuple corresponds to a record, where each tuple represents a specific item. For example, in a database table, each row is a tuple associated with a particular entity or object [194]. This model is straightforward and easy to use, as data can be represented in various forms such as tables or lists. In Table 3.2, a tuple would represent the record or row related to Sami in the context of the database.

Table 3.2 Tuple representation

Name	Age	Sex	Description	Temperature	Cough
Sami	32	Male	Patient	38 °C	Yes

A semantic network uses nodes and edges to visually represent information [195]. For instance, within the medical field, the nodes depicted in Figure 3.8 represent various aspects such as diseases, symptoms, or treatments. Where the connections (edges) between these nodes highlight dependencies and relationships between different concepts, providing an effective graphical representation model for understanding complex data structures.

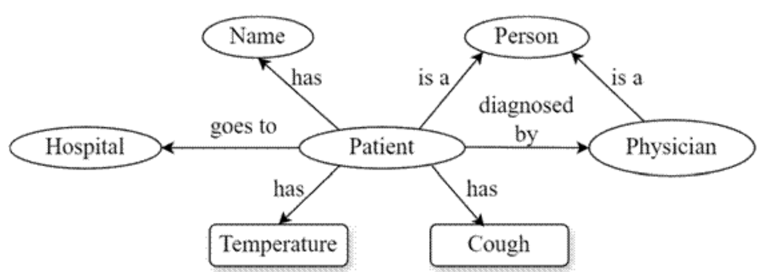

Figure 3.8 Semantic network graph

3.3.2 Learning-Based Models

Learning-based models encompass machine learning algorithms capable of learning relationships and dependencies from datasets, enabling them to perform predictions and classifications.

Random forest, an ensemble of decision trees, is used for classification or regression tasks, with each tree trained on a distinct subset of the data [196]. A key advantage of random forest is its ability to manage high-dimensional datasets effectively [197], making it widely applicable across various industries [197]. In our example, a random forest could represent multiple patients in the same graph. Other models such as neural networks, decision trees and Bayesian networks are commonly employed in machine learning [197, 198]. Naïve Bayes, a supervised machine learning model, represents data using probability theory, allowing for the calculation of feature probabilities belonging to a class [199].

3.3.3 Rule-Based Models

Rule-based models operate on a set of rules that derive conclusions from expressed knowledge. In these systems, information is structured as a collection of “if–then” statements, where each rule consists of a condition (“if”) and an action (“then”) [200]. For example, in a medical context, when a new case is presented, symptoms are matched against conditions in the rules, and relevant rules are applied to make a diagnosis. Fuzzy logic, a mathematical approach that addresses uncertainty and imprecision in data, is one method that allows for partial truth values ranging from 0 to 1 [201]. While traditional rule-based systems provide binary decisions (disease present or absent), fuzzy logic enables a more nuanced assessment, indicating the level or degree of disease presence [202,203]. Rulebased systems find applications in diagnosing medical conditions, evaluating financial investments, optimizing manufacturing processes and more. For instance:

Rule 1: If the patient has medium fever and joint pain, then they take Advil.

Rule 2: If the patient has high fever and severe cough, then they visit the physician.

3.3.4 Mathematical Models

Mathematical models represent terms and rules using logical and probabilistic notation, allowing for the generation of new knowledge. This category encompasses several subcategories, including logic models and probabilistic models, some of which are also considered graphical models and have been previously explained.

Propositional logic is a declarative statement that can be classified as either true or false, serving as a method to represent knowledge in a logical or mathematical format [204]. However, propositional logic has limitations in representing complex sentences or natural language statements [205], and it cannot describe statements in terms of their properties or logical relationships. An example can be expressed as:

$$(T \wedge H) \rightarrow C$$

This statement means that if Sami has a temperature above 38 ◦C (T) and also has a headache (H), then Sami should consult a healthcare professional (C).

First-order logic, also known as predicate logic, offers a more sophisticated way to describe information about objects and their relationships [206]. Unlike propositional logic, which deals with simple true or false statements, first-order logic recognizes that the world is more complex and operates similarly to natural language, allowing for nuanced expressions. In the example below, for all individuals (people) X and all times T:

If the patient x has a temperature greater than 38 ◦C (denoted as 1 if true and 0 if false) and also has a cough (similarly treated as 1 if true and 0 if false), then both conditions must be true for their sum to equal 2:

(Temperature(x) > 38) + Cough(x) = 2

For each patient x belonging to X at time t, the following statement holds true:

$\forall x \forall t$: ((Temperature(x,t) > 38) + Cough(x,t) = 2)

(VirusDetection(x,t) = 1 ∧ ConsultHealthcare(x,t) = 1)

This indicates that if a patient x at time t meets the criteria of having both a temperature greater than 38 ◦C and a cough (sum equals 2), it implies positive virus detection (VirusDetection(x,t) = 1) and recommends healthcare consultation (ConsultHealthcare(x,t) = 1).

Second-order logic extends first-order logic by allowing quantification over sets or collections of objects. While first-order logic quantifies over individual objects like people, animals or numbers, second-order logic permits quantification over collections such as properties, relations and functions [207]. In second-order logic, we use P to quantify over properties (such as temperature, cough and consulting healthcare), allowing us to directly describe these properties and their relationships. For example, if the patient x has a temperature greater than 38 ◦C (denoted as 1 if true and 0 if false) and also has a cough (similarly treated as 1 if true and 0 if false), then both conditions must be true for their sum to equal 2:

P(Temperature (x) > 38) + P(Cough (x)) = 2

For each patient x, property P and time t, the following statement holds true:

$\forall x \forall P \forall t$: (P(Temperature(x,t) > 38) + P(Cough(x,t)) = 2) →

(P(VirusDetection(x,t) = 1,t) ∧ P(ConsultHealthcare(x,t) = 1))

3.3.5 Hybrid Models

Hybrid models combine two or more KRMs to leverage the strengths of each and mitigate their individual limitations [208]. Combining ontologies with rule-based systems, for instance, overcomes many limitations by providing a more comprehensive and adaptable representation of knowledge. This hybrid approach excels in handling complex and uncertain information while offering easier management and maintenance over time [209]. Another effective combination is between ontologies and Bayesian networks, which enhances knowledge representation by structuring domain knowledge and modelling uncertainty and probabilistic relationships effectively [209]. Moreover, limitations, such as the challenge of accurately capturing probabilities in complex systems, can emerge when using either model alone. By integrating both models, ontology can offer a structured representation of domain knowledge, while Bayesian networks can effectively model uncertainty and probabilistic relationships among concepts and entities [209, 210], thereby mitigating some of these challenges. Additionally, combining neural networks with semantic networks harnesses their respective strengths to develop a more adaptable and robust knowledge representation system [210]. While neural networks excel in processing data but often lack interpretability, semantic networks are effective at representing concepts and their relationships but typically do not learn directly from data [211]. Integrating these two approaches allows us to develop a system that combines the data processing capabilities of neural networks with the structured representation strengths of semantic networks, resulting in a more comprehensive and meaningful knowledge representation framework.

Figure 3.9 illustrates the basic categories and their subcategories for KRMs. The five main categories are distinguished by the beige colour, while models belonging to more than one category are marked in blue. Models that belong exclusively to one category are listed in regular text. Additionally, the mathematical category includes two subcategories highlighted in pink. Table 3.3 provides a detailed overview of the advantages and disadvantages of each model.

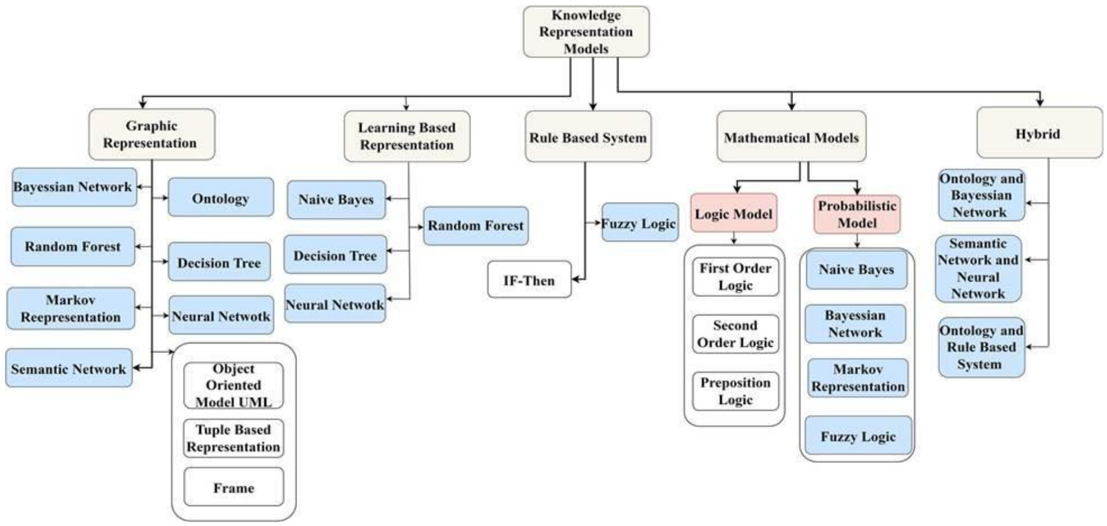

Figure 3.9 Knowledge representation model categorization

Table 3.3 A comprehensive overview: advantages and disadvantages of knowledge models.

Model	Advantages	Disadvantages
Naïve Bayes [212]	Simple; manages noisy data and missing values.	Inability to handle continuous features.
Decision tree [213]	Intuitive interpretation; manages missing values.	Difficulty in designing large trees; low ability to manage noisy data.
Random forest [214]	Ability to manage noisy data; suitable for large and heterogeneous data.	Requires more memory; difficult implementation and complicated in some cases; low ability to manage missing values.
Bayesian network [215]	Handles missing or incomplete data; graphical structure facilitates interoperability.	Memory requirements increase with the number of variables. Large and complex networks may face scalability issues.

Markov representation [216]	Ability to represent large data; robustness in handling noisy data.	Future transitions depend only on the current state, not on the entire history. This memoryless property can be limiting, especially when considering long-term effects or complex dependencies.
UML [217]	Graphical representation for software design.	Becomes complex for large systems.
Frame [218]	Hierarchical organization of knowledge; easy to construct.	Difficulty in representing complex relationships.
Semantic network [219]	Effective when representing relationships between entities.	Faces challenges with large graphs.
Ontology [220]	Structured representation of data; allows navigation of complex processes. Enables integration of domain knowledge for high-level reasoning.	Creating an ontology model can be time-consuming.
First-order logic [221]	Expressive; handles complex relationships.	Faces challenges in representing certain types of data.

Second-order logic [222]	Adds expressiveness over first-order logic.	Increased complexity in reasoning.
Propositional logic [223]	Simple and easy to understand.	Limited expressiveness for complex relationships.

In the previous section, we classified different models utilized for data representation, such as mathematical, graphical and logical models, and examined their respective strengths and weaknesses. Now, we will investigate specific applications of KRMs in healthcare during the past decade.

3.4. Importance of Knowledge Representation Models in the Medical Domain

In the medical field, healthcare is increasingly embracing tele-monitoring systems, which give patients self-management tools [223]. The effective management and interpretation of vast, complex data are crucial for accurate diagnoses, optimal treatment planning and improved patient outcomes. KRMs play a crucial role in providing a structured framework for organizing and representing medical data. This allows healthcare professionals to store and utilize valuable information effectively, including risk factors, treatments, symptoms and other patient-related data [224]. Various methods are used to represent different types of medical data, each with unique strengths and weaknesses. For example, Bayesian networks can model risk factors, decision trees are effective for treatment pathways, fuzzy logic-based models manage symptoms, while UML or frame notations are useful for demographic variables [225, 226].

Over the past decade, the International Workshop on Knowledge Representation for Health Care (KR4HC) has highlighted the prevalence of several generic models in healthcare. Ontologies constitute 31%, semantic web-related formalisms account for 26%, decision trees and rules represent 19%, logic covers 14%, and probabilistic models encompass 10% of the represented knowledge models. These contributions primarily aim to formalize and represent medical knowledge [227].

In the healthcare field, KRMs provide benefits to various stakeholders such as doctors, patients, hospitals and research institutions across multiple dimensions including disease diagnosis, monitoring and treatment [225]. The monitoring process involves analyzing relevant parameters, measurements and activities over time to maintain control over a patient's health, monitor treatment progress and provide better care by detecting changes [228]. Organizing medical knowledge through KRMs establishes rules and relationships across different contexts, facilitating decision support during monitoring. This enables medical staff to interpret data and make informed decisions for improved patient care [229]. Furthermore, disease diagnosis involves identifying specific illnesses based on established classifications, which enables accurate diagnosis [198]. Knowledge representation and reasoning techniques assist doctors in decision-making and inferring new information about diseases from previously represented data [229, 230]. For example, Shi et al. [231] developed a knowledge graph based on medical texts and applied semantic reasoning to enhance disease diagnosis. Table 3.4 provides an overview of the utilization of KRMs in the healthcare domain.

Table 3.4 Utilization of knowledge representation models (KRMs) in the healthcare domain.

Ref.	KRMs	Health Domain
[232]	Mathematical representation models	Disease diagnosis.
[233]	Random forest	Treatment for rheumatoid arthritis.
[234]	Random forest	Detection of inflammatory bowel disease.
[235]	Decision tree	Patient and disease diagnosis.
[236]	Ontology	Data representation in medical domain.
[237]	K-nearest neighbours	Diagnosis and data in X-ray image.
[238]	Neural network	Finding signs of diabetic retinopathy.

[239]	Naïve Bayes	Data representation in congenital heart surgery.
[240]	Decision tree	Evaluation of the risk of amputation in individuals with diabetic foot.
[241]	Random forest	Antibody data representation in kidney transplantation.
[242]	Naïve Bayes	Data representation from ultrasound images.
[243]	Bayesian network	Used in treatment process.
[244]	Decision tree	Specific details in the diagnosis process.
[245]	K-nearest neighbours	Forecasting the risk of retinopathy.
[246]	Random forest	Monitoring for school children.
[247]	Decision tree	Detection of spread patterns in recent ischemic stroke.
[248]	K-nearest neighbours	Detection of thyroid disease.
[249]	Neural network	Used in cancer treatment using advanced therapy techniques.
[250]	Markov model	Disease diagnosis using ECG.
[251]	Decision tree	Patient monitoring.
[252]	Neural network	Identification of coronary calcium.
[253]	Fuzzy logic	Medical record representation for diagnosis prediction.
[254]	Random forest	Diagnosing CKD disease.

The next section outlines the specific requirements and challenges within healthcare, demonstrating how KRMs can address these needs.

3.5. Requirements in the Medical Domain

In healthcare and medicine, knowledge plays a crucial role. Healthcare professionals depend on understanding the workings of the human body and follow established methods outlined in clinical guidelines for diagnosis, monitoring and treatment [181]. Understanding healthcare organization is also vital for effective patient management. Currently, we are evaluating different KRMs based on specific requirements. Our objective is to choose a model that can meet most of our needs. However, before selecting a model, it is important to define and fully understand the significance of each characteristic.

Heterogeneity: KRMs must manage various types of information sources in healthcare, such as data from sensors, like GPS sensors, which require interpretation, and user profiles that are updated less frequently and generally do not need additional interpretation. An effective model in the medical domain should be capable of managing these diverse information types to ensure accurate decisions and comprehensive representations [255].

Interpretability: Interpretability is crucial as it enables medical staff and patients to understand results easily and comprehend how inputs are processed and how outputs are derived. Therefore, when developing KRMs for the medical field, prioritizing interpretability is essential to ensure practical usability and acceptance by healthcare professionals [256].

Reasoning: Reasoning refers to the ability to draw deductions and inferences from the information stored in the knowledge base, even when that information is incomplete or uncertain. This capability is vital in medical domain models as it empowers healthcare providers to make informed decisions based on available information [257].

Scalability: Scalability is the model’s ability to manage large volumes of data or knowledge efficiently. It is a critical factor in design as it directly impacts performance and efficiency. Models that are not scalable may struggle with larger datasets or more complex problems, resulting in slower or less effective performance [258].

The evaluation of KRMs in Table 3.5 is based on the four essential requirements explained earlier. We use the term "satisfy" to indicate supported characteristics, "not satisfy" for characteristics that are not supported and "partial satisfaction" for characteristics where support is unclear or not fully met.

To effectively represent knowledge, it is crucial to utilize models with clear representations. Our analysis, based on approximately 28 articles, examines and evaluates these models against four key criteria using KRMs.

Table 3.5 Assessing knowledge representation models (KRMs) in achieving medical domain needs.

KRMs	Heterogeneity	Interpretability	Reasoning	Scalability
Bayesian etwork	Satisfy [259]	Satisfy [260]	Satisfy [259]	Partial satisfaction [260]
Markov representation	Satisfy [181]	Partial satisfaction [261]	Satisfy [261]	Satisfy [262]
Ontology	Satisfy [263]	Satisfy [263]	Satisfy [263]	Satisfy [264]
UML	Satisfy [265]	Satisfy [265]	Partial satisfaction [266]	Partial satisfaction [265]
Frame	Satisfy [266]	Satisfy [265]	Not satisfy [266]	Partial satisfaction [266]

Tuple-based representation	Satisfy [267]	Satisfy [267]	Partial satisfaction [267]	Partial satisfaction [267]
Semantic network	Satisfy [268]	Satisfy [268]	Partial satisfaction [269]	Partial satisfaction [269]
First-order logic	Partial satisfaction [270]	Satisfy [270]	Satisfy [270]	Partial satisfaction [270]
Second-order logic	Satisfy [271]	Satisfy [272]	Satisfy [273]	Partial satisfaction [272]
Propositional logic	Partial satisfaction [272]	Satisfy [272]	Partial satisfaction [272]	Not satisfy [273]
Rule-based system	Satisfy [274]	Satisfy [274]	Satisfy [274]	Partial satisfaction [274]
Naïve Bayes	Partial satisfaction [275]	Partial satisfaction [275]	Satisfy [275]	Satisfy [275]
Random forest	Satisfy [276]	Partial satisfaction [276]	Satisfy [276]	Satisfy [276]
Decision tree	Satisfy [276]	Satisfy [276]	Satisfy [276]	Partial satisfaction [276]

Neural network	Satisfy [277]	Not satisfy [277]	Satisfy [278]	Satisfy [278]

As previously discussed, there are five primary categories of representations, each encompassing multiple models. Table 3.5 assesses these models according to specific medical needs within this domain, as previously outlined. It is important to note that certain models have inherent limitations. For instance, Naïve Bayes struggles with certain data types like numerical data [210]. Additionally, propositional logic lacks essential quantifiers such as "for all" (∀) and "there exists" (∃), which are necessary for expressing statements about entire classes of objects or specific instances [211].

In terms of interpretability, models like Markov models are considered somewhat non-understandable to a degree. The complexity of the data limits the level of detail and explanation that can be derived from these models [274]. Similarly, a random forest, which comprises multiple decision trees, makes it challenging to comprehend the relationships among them, hindering clear and intuitive interpretation of the entire model [276]. Regarding reasoning, several models such as propositional logic, frames, UML and semantic networks have limitations in making inferences based on data representation [266,268].

Lastly, concerning scalability, both frames and semantic networks may struggle to scale when managing large volumes of interconnected information [266, 268]. Naïve Bayes, on the other hand, faces challenges in capturing complex relationships, and its scalability may be impacted as the number of features increases [275].

We can derive a general equation regardless of the number of requirements and their respective weights in the medical domain. Consider R, a set of n requirements:

R = {r1, r2, r3, . . ., rn}

And consider W, a set of n different weights corresponding to these requirements:

W = {w1, w2, w3, . . ., wn}

Therefore, a general formula to calculate the total satisfaction ratio (TSR) can be described as shown in Formula (1):

$$TSR = \frac{value\ of\ Rk * wk}{wk}$$

Depending on the requirements in the healthcare domain—heterogeneity, interpretability, reasoning and scalability we select the model with the highest satisfaction ratio using Formula (2):

$$TSR = \frac{(\text{heterogeneity} * \text{W1} + \text{interpretability} * \text{W2} + \text{reasoning} * \text{W3} + \text{scalability} * \text{W4})}{\sum_{k=1}^{4} \text{wk}}$$

To calculate the overall ratio for the set of requirements, we assign a weight factor “w” to each requirement based on its importance. While adjusting these weights depending on the context remains a future challenge we aim to address, our survey currently assumes equal priority among the four requirements. Therefore, each weight factor is set to 1, enabling us to apply Formula (3) for calculating the overall ratio (TSR):

$$TSR = \frac{\text{ValueOf(heterogeneity} + \text{interpretability} + \text{reasoning} + \text{scalability)}}{4}$$

We used Table 3.5, replacing “satisfy”, “partial satisfaction” and “not satisfy” with 1, 0.5 and 0, respectively, in each column. Using Formula (2), we calculated the satisfaction ratio for each model listed in the table by dividing it by 4. The satisfaction ratios of all models are presented in Figure 3.10, which reflects their performance against the healthcare requirements mentioned earlier. Figure 3.11 displays citation numbers indicating the usage of each model in the medical field, with higher numbers indicating greater adoption. According to our survey, ontology achieved the highest ratio and citation number, indicating its suitability for meeting healthcare requirements.

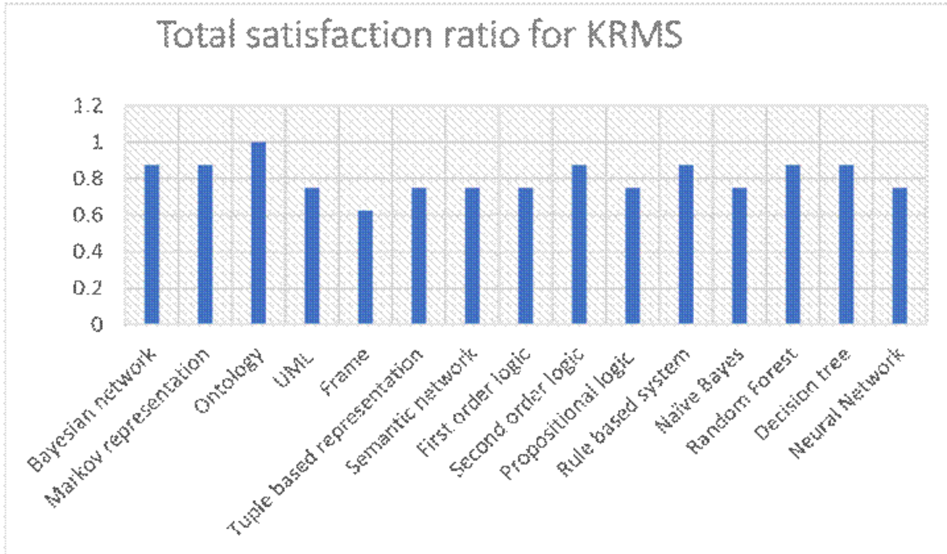

Figure 3.10 Satisfaction ratio for knowledge representation models based on medical domain requirements

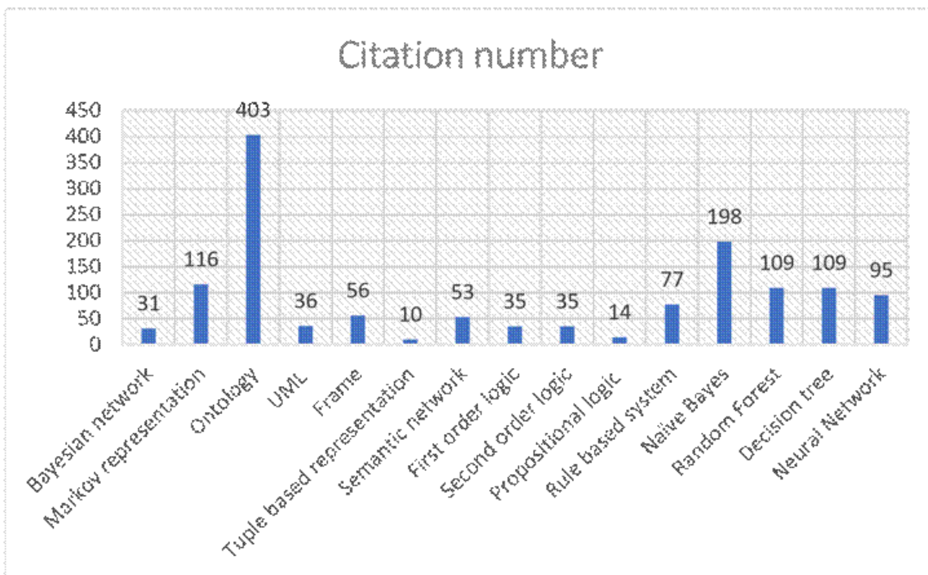

Figure 3.11 Citation number for knowledge representation models (over the last decade)

3.6. Results and Discussion

As discussed in previous sections, each model in the medical domain possesses unique strengths and weaknesses. KRMs are crafted to interpret complex information, making it comprehensible and accessible to machines [278]. According to Figure 4.10, the ontology model demonstrates a higher ratio concerning the healthcare requirements mentioned earlier.

Ontology is recognized as a robust semantic structure that facilitates knowledge representation, integration and reasoning [279]. In the early 1990s, Borst defined ontology as a “formal specification of a shared conceptualization” [281]. Studer later refined this definition, stating that ontology is an “explicit specification of a shared conceptualization” [282]. In this context, a conceptualization is an abstract model that simplifies real-world representation using objects, concepts, entities and their relationships within a specific domain. Explicitness requires precise definitions of all concepts and constraints to prevent misinterpretation and ensure clear understanding of symbols [283]. Formality demands that the ontology be machine-readable. Lastly, “shared” implies that an ontology is consensual knowledge accepted by a group or community [283]. Reusing existing ontologies is beneficial as it saves time, leverages proven effective applications and facilitates interoperability with existing tools [283].

Ontology provides a structured representation of concepts and relationships within a domain, making it well-suited for the complexity and diversity inherent in the medical field [253]. Biomedical ontologies play a critical role in enabling efficient retrieval and reuse of machine-readable information stored in databases [284].

CHAPTER 4
DEVELOPING CONTEXTUAL ONTOLOGY FOR CHRONIC DISEASES: AI-ENHANCED EXTENSION AND PREDICTION IN AN ASTHMA CASE STUDY

Résumé

La complexité croissante et l'interdépendance des données de santé, en particulier pour les maladies chroniques telles que l'asthme, exigent des approches innovantes pour une représentation efficace des connaissances. Cette étude propose un modèle général d'ontologie contextuelle pour les maladies chroniques, étendu spécifiquement au cas de l'asthme. En s'appuyant sur des jeux de données réels, l'ontologie étendue de l'asthme intègre des facteurs clés tels que les symptômes, les déclencheurs, les traitements et les données démographiques des patients, offrant ainsi un cadre complet pour la gestion de la maladie. L'ontologie a été validée à l'aide de métriques intrinsèques telles que la classification, la réutilisabilité et la complétude dans des applications de santé. Pour valider l'ontologie, nous avons utilisé des arbres de décision afin d'extraire des règles après avoir identifié les paramètres les plus pertinents nécessaires à la génération de règles en Semantic Web Rule Language (SWRL). Ces règles facilitent le raisonnement, la validation et la prise de décision au sein de l'ontologie. Les résultats mettent en évidence le potentiel du développement d'une ontologie contextuelle générale et de son extension pour traiter des maladies chroniques spécifiques, telles que l'asthme. Nous avons conçu un cadre général d'ontologie contextuelle en intégrant l'ontologie étendue avec des algorithmes d'intelligence artificielle, en identifiant les paramètres pertinents et en extrayant des règles afin d'améliorer la représentation des connaissances et de soutenir la prise de décision clinique. Ce cadre peut être appliqué à d'autres études de cas de maladies.

Abstract

The growing complexity and interdependence of healthcare data, especially for chronic diseases such as asthma, demand innovative approaches for effective knowledge representation. This study

introduces a general contextual ontology model for chronic diseases, extended specifically to asthma. Leveraging real-world datasets, the extended asthma ontology integrates key factors such as symptoms, triggers, treatments, and patient demographics, providing a comprehensive framework for disease management. The ontology was validated using intrinsic metrics such as classification, reusability, and completeness in healthcare systems. To validate the ontology, we used decision trees to identify the most relevant parameters. We designed a general contextual ontology framework by integrating the extended ontology with artificial intelligence algorithms, identifying relevant parameters, to enhance knowledge representation and support clinical decision-making.

Keywords: ontology development; context-aware systems; ontology reasoning; artificial intelligence in healthcare.

4.1 Introduction

The representation of the medical context requires a deep understanding of the healthcare domain and its various dimensions. Context can significantly enhance the representation of knowledge, particularly in the healthcare domain [285, 286]. Consequently, medical data can encompass several parameters, and the acquisition process must facilitate data representation that enables understanding, standardization, and centralized management of all diseases. Data now plays a crucial role in decision-making across multiple fields. However, datasets in the medical domain are often complex and disorganized, making direct extraction of useful information difficult [287]. Asthma represents a chronic disease that can be investigated through a patient's signs, medical history, demographic information, and other parameters [288].

We distinguish many KRMs, such as frame, UML, and graphical representation. We focus on ontology models that facilitate the management of data by incorporating relevant domain concepts and their associated relations [289]. This shared knowledge system allows for a comprehensive representation of complex information, which is crucial in fields such as health care, where data interdependence is significant [290]. Ontologies help in structuring and managing healthcare data to offer tailored services [291]. Ontologies are used in 50% of applications, emphasizing their crucial role in organizing and representing data. The other 50% comprises alternative modelling approaches, demonstrating a balanced distribution among KRMs [292].

The development and integration of standard ontologies, such as the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), facilitate the exchange of information through medical information systems [293]. The use of standard ontologies ensures consistent and semantically interoperable data management across different healthcare systems [263].

The main contributions of this chapter are:

Designing a general contextual ontology model that includes the key knowledge domains related to chronic disease management.

Extending this general contextual ontology to a domain-specific model for asthma, highlighting the identification of classes, entities and their interrelations.

Supporting the evaluation process and enhancing the inference capabilities (reasoning) of this extended asthma ontology by using artificial intelligence (AI) algorithms.

Evaluating this extended asthma ontology using intrinsic metrics such as classification, reusability and completeness.

This chapter aims to improve the monitoring of healthcare systems for chronic diseases by creating an ontology related to context categorization and extending it specifically to asthma. Additionally, we identify the most relevant parameters using AI algorithms, leading to more accurate, easy-to-understand, and efficient predictions. This approach can be useful in various areas, making it a valuable tool for addressing real-world knowledge representation challenges in disease diagnosis.

4.2 Comparing the General Contextual Ontology with the Existing Ontologies

We review some of the most relevant existing ontologies available in BioPortal and other repositories.

Asthma Ontology (AO): The Asthma Ontology, available on BioPortal, provides a structured representation of asthma-related factors, including symptoms, triggers, and treatments. However, it lacks contextual dimensions such as environmental and organizational factors, which are crucial for real-world asthma management [294].

Disease Ontology (DO): The Disease Ontology offers a broad classification of diseases, including asthma, but does not provide fine-grained relationships between symptoms and external environmental factors [295].

Mondo Disease Ontology: This ontology integrates multiple disease classifications, facilitating cross-referencing. However, it does not explicitly support AI-driven rule generation for clinical decision support [296].

Monarch Initiative: Monarch integrates multiple ontological sources for disease modeling but primarily focuses on genetic and phenotypic aspects, limiting its applicability in real-world asthma management [297].

While these ontologies provide valuable frameworks for disease classification, they lack a comprehensive contextual representation that integrates environmental, temporal, and organizational factors. Additionally, they do not incorporate AI-driven ontology validation methods to enhance reasoning and rule-based decision-making.

To highlight the originality and advantages of our approach, we compare our GCOCD with related existing disease ontologies. Table 4.1 provides a comparative analysis of key features.

Table 4.1 Comparative analysis of contextual classifications within existing disease ontologies

Feature	Asthma Ontology (BioPortal)	Disease Ontology	Mondo Ontology	Monarch Initiative	Our Ontology (GCOCD - Asthma)
Scope	Asthma-specific concepts	Broad disease categories	Integrated disease classification	Focuses on genetic and phenotypic traits	**Context-aware asthma ontology**
Environmental Context	❌ Not included	❌ Not included	❌ Not included	✅ Limited genetic-environment links	✅ **Explicitly modeled (pollution, humidity, allergens)**
Temporal Context	❌ Not included	❌ Not included	❌ Not included	❌ Not included	✅ **Includes seasonal and daily variations**
Medical Standard Integration	✅ Uses SNOMED CT	✅ Uses multiple standards	✅ Uses multiple standards	✅ Uses multiple standards	✅ **Compatible with SNOMED CT**

AI-Driven Rule Generation	❌ Not included	❌ Not included	❌ Not included	❌ Not included	✅ **Decision tree-based rule extraction**
Ontology Reusability	❌ Limited	✅ Broad use	✅ Broad use	✅ Broad use	✅ **Extensible to other chronic diseases**

The GCOCD extends existing models by integrating environmental and temporal factors, which are crucial for asthma monitoring. Additionally, it enhances reasoning capabilities through AI-driven rule extraction, making it a dynamic and adaptive ontology for healthcare systems.

4.3 Methodology for Designing General Contextual Chronic Disease Ontology

Ontology development has introduced many methods over the years, each focusing on different aspects. Ushold and King (1995) focused on creating an ontology with basic guidelines to capture domain knowledge; however, it lacked clear steps for practical implementation [298]. Grüninger and Fox (1995) worked on the TOVE project, which used questions and scenarios to define the purpose of the ontology, but their method was too abstract for broader use [299]. METHONTOLOGY (1997) introduced an organized process with clear steps that define the concepts, formalize them, and evaluate the ontology. This is classified as one of the more detailed methods [300]. Other methods focused on making ontologies reusable, such as Ontolingua (1995), which provided a library of existing ontologies for collaboration but did not explain how to adapt these ontologies for new uses [301].

SENSUS (1996) created a large hierarchy of concepts for natural language processing, but it was not specific enough for certain domains [302]. Additionally, methods such as the 101 Method (2001) provided practical advice for defining classes and relationships in ontologies but omitted important steps, including lifecycle planning [303]. Researchers such as Lopez et al. compared these methods using criteria including reusability and how well they supported collaboration and lifecycle management. They found that no single method met all needs. For example, TERMINAE and Termontography focused on extracting knowledge from text in multiple languages but were limited to English and French or were no longer supported [304]. Efforts to combine methods, such as merging METHONTOLOGY with Cyc 101, have shown potential, especially for designing medical ontologies [305].

Although various ontology development methodologies have been proposed, many lack comprehensive guidance on collaboration, lifecycle management, and adaptation to diverse needs [306]. To create an ontology that is both reusable and extendable, we adopt Sánchez's methodology [307]. This approach (as shown in Figure 5.1) offers clear steps and can accommodate various requirements, making it a reliable choice for long-term and flexible ontology development.

This methodology outlines the process of building an ontology to represent chronic diseases, with a focus on extending and validating it for asthma-specific use cases.

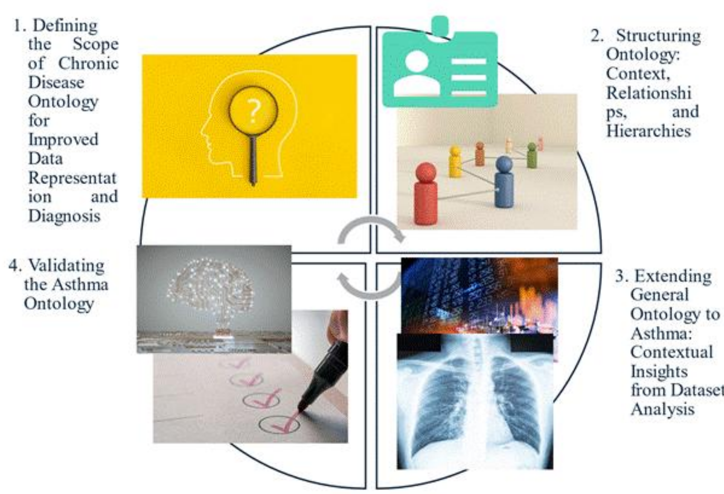

Figure 4.1 Methodology for designing general contextual chronic disease ontology

4.3.1 Designing General Contextual Ontology: Contextual Insights from Dataset Analysis

Define the scope domain of ontology: The ontology focuses on chronic diseases and aims to facilitate data gathering and representation.

The questions guiding the ontology's domain include:

What is the domain the ontology will cover?

Chronic diseases are the focus, addressing conditions such as asthma, diabetes and hypertension.

What is the purpose of this ontology?

To aid in collecting and representing chronic disease data for predictive diagnosis and healthcare research.

Who will use the ontology?

Hospitals, healthcare professionals, researchers and organizations conducting disease-related studies.

What types of questions should the information in the ontology answer? Is this patient diagnosed with a chronic disease?

What are the symptoms of this disease?

The domain scope integrates temporal, environmental, organizational and medical contexts to comprehensively model chronic diseases.

Developing the conceptual model following Sánchez's methodology, the ontology is structured using context categorization. Identify key terms in the ontology:

Nouns (Concepts): Patient, Physician, Symptom, Treatment, Diagnosis, Allergy.

Attributes (Attributes of Concepts): Age, Gender, Diagnosis Date, Environmental Factors (pollution levels).

Verbs (Relationships): “hasSymptom” (Patient → Symptom), “isTreatedBy” (Disease → Treatment), “isTriggeredBy” (Symptom → Environmental Factor).

Standardization: SNOMED CT is used for consistent representation of symptoms and medical terms.

Classes and their hierarchy: Top-Level Classes: Temporal Context, Environmental Context, Medical Data, Organizational Context, Events.

Subclasses: Under medical data: Symptoms, Treatments, Diagnoses, Allergies. Under environmental Context: Pollution, Temperature, Humidity.

Define class properties:

Object Properties: hasSymptom: Domain = Patient, Range = Symptom. isTreatedBy: Domain = Symptom, Range = Treatment.

Datatype Properties: age: Integer, diagnosisDate: Date.

Define Slot Facets: Assign value types, permitted values properties:

Age: Integer (1 to 100)

PollutionLevel: Float (0.0 to 500.0)

Create Instances: Populate slots with real-world values from datasets: A patient named “Sami” with: Age = 45, Gender = Male, Symptom = “Chest Tightness”.

Sánchez’s Methodology with a Comparative Perspective

Although several ontology development methodologies exist, such as METHONTOLOGY Ontolingua, and the 101 Method, many fall short in addressing long-term reusability and modular extension. Sánchez’s methodology was selected for its ability to integrate structured development steps with adaptability across medical domains [306, 307]. Sánchez’s method combines conceptual modeling, class–property definition, and real-world dataset mapping, making it a strong fit for developing a general ontology that can later be extended to specific diseases like asthma. Moreover, Sánchez's method supports ontology lifecycle planning and enables context categorization, which is essential for chronic diseases influenced by environmental, organizational, and temporal dimensions.

4.3.2 Extending the General Contextual Ontology to Asthma

The GCOCD was developed to provide a flexible and reusable framework for modeling chronic diseases, but it lacks specificity when applied to disease-specific contexts such as asthma. The key missing information includes:

Asthma-specific symptoms and triggers, such as wheezing, airway inflammation, and pollen exposure, are absent from the general ontology.

Environmental and temporal factors, including pollution, humidity, and seasonal variations, which influence asthma severity but are not fully captured in GCOCD.

Granular treatment pathways, as the ontology does not distinguish between general chronic disease management and targeted asthma interventions, such as bronchodilators and corticosteroids.

To address these limitations, the extension process focuses on context-based adaptation and hierarchical refinement, ensuring that the ontology is compatible with GCOCD while accurately modeling asthma-specific knowledge. The detailed construction process is outlined in the next section, where targeted modifications will be introduced to specialize symptoms, treatment pathways, and environmental factors relevant to asthma.

4.3.2.1 Ontology Construction from Dataset Schema

To refine the ontology structure, dataset-driven approaches were used:

a) Feature-Based Extension Using Relevant Parameters
   Key features were prioritized based on their impact on asthma severity and treatment. Feature selection was performed based on AI algorithms to be further discussed in Section 4.4.

b) Context-Specific Data Integration, most of the parameters present in the general ontology are reused in the extended asthma ontology: temporal, environmental, medical and organizational.

c) Analyze the dataset to extract key features.

To demonstrate the GCOCD, we conducted a preliminary adaptation to the domain of diabetes. The core structure of GCOCD, including the categories of patient data, environmental factors, temporal context, and organizational context, was reused without modification. New disease-specific classes and properties were added to represent glucose level monitoring, insulin treatment, and dietary control. This quick adaptation illustrates the framework's flexibility and its potential to model diverse chronic conditions. A more comprehensive diabetes-specific ontology will be presented in future work, but this exercise confirms GCOCD's reusability across domains.

4.3.2.2 Validate The Extended Asthma Ontology

Several ontology evaluation methods have been developed, each focusing on specific aspects. Jonathan emphasized the importance of ensuring that ontologies are logical and well-structured to enable the effective organization of knowledge [308]. Pérez et al. highlighted the need to design ontologies that can be applied across various domains [309]. Lovrenčić et al. focused on verifying that ontologies comprehensively capture the necessary domain knowledge, ensuring they include all essential concepts and relationships for their intended use [310]. Together, these approaches provide a robust foundation for ontology reasoning.

4.4 Designing an Ontology Framework: A Case study on Asthma

This section presents the framework (Figure 4.2) for designing the general contextual ontology for chronic diseases through a multi-step approach. It begins by outlining the steps involved in developing the general contextual ontology, followed by data acquisition and preparation. The process then moves to analyzing strategies such as dataset exploration and identifying relevant parameters for developing the extended asthma ontology using specific datasets. Subsequently, rules are derived from the data to enhance ontology reasoning and evaluate the extended ontology.

This approach effectively integrates AI algorithms with ontology-based data representation, organized into clearly defined stages, as illustrated in Figure 4.2.

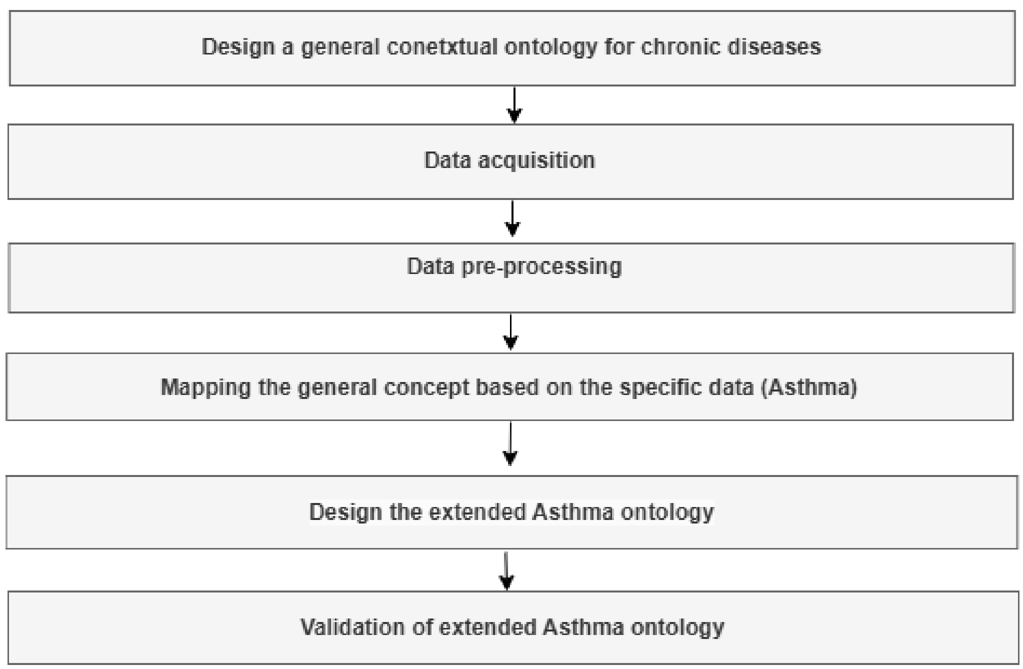

Figure 4.2 Designing a general contextual ontology framework for the extended asthma ontology

4.4.1 Design General Contextual Ontology

An ontology is a formal framework that provides a structured and shared understanding of concepts, properties and their relationships within a particular domain. Ontologies consist of three core components:

- Classes that represent entities or concepts in the domain ("Disease" or "Symptom");
- Properties that describe relationships between these entities ("hasSymptom");

Individuals that serve as specific instances of these classes (e.g., "Asthma" as an individual of the class "Disease"; Horridge, 2005.
We created an ontology using OWL-DL, a specialized language for describing ontologies, and developed it using the Protégé editor from Stanford University, supported by the Pellet reasoner. A general contextual ontology was built to represent chronic diseases, comprising 10 main categories, as illustrated in Figure 4.3. These categories encompass key elements such as patient data, social factors, environmental conditions, and disease-related information. This flexible structure is designed to be adaptable to various chronic diseases by illustrating how these elements are interrelated.

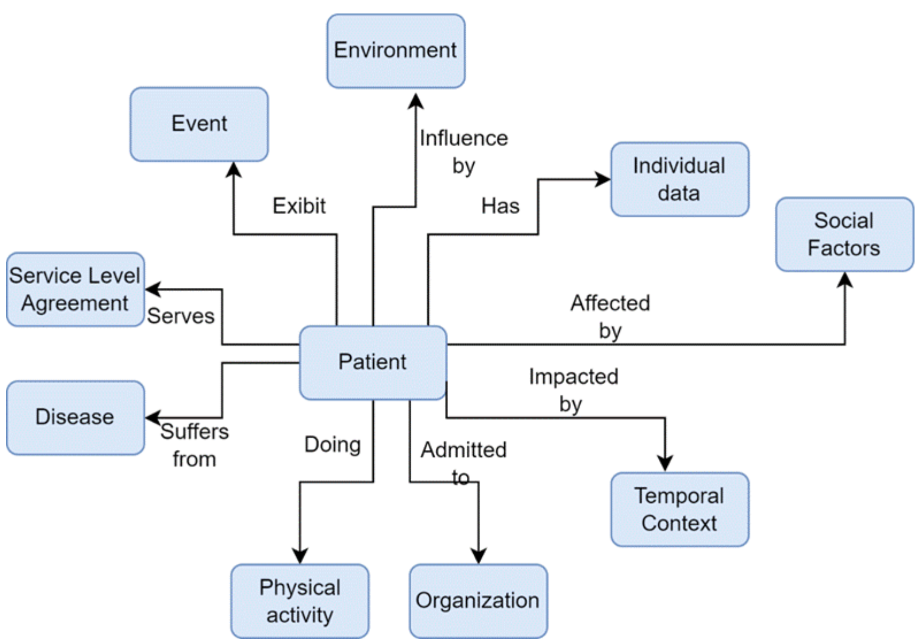

Figure 4.3 General contextual chronic disease ontology

In the general contextual ontology (see Figure 4.3), “Patient” is the central category. It acts as a key point that separates the ontology into nine categories, including individual data, social factors, temporal context, organization and the environmental factors that affect the patient’s context. Additionally, it covers information about physical activity, disease information, health events, relevant organizational contexts and service-level agreements that support healthcare services.

Each of these categories can be further subdivided into subcategories. For example, “Individual Data” can be divided into subcategories addressing the patient’s personal information. Similarly, all other categories can be subdivided based on the “Patient,” creating subcategories that focus on specific aspects of the chronic disease ontology, such as symptoms and treatments. These subdivisions help better organize and represent the interrelations within chronic disease management.

The social and individual data ontology illustrates the factors influencing a patient’s well-being, as shown in Figure 4.4. It captures the relationship between patients; their individual data, including attributes like age, gender, body mass index (BMI), nationality and location; and the social factors impacting their health, such as education level, income, quality of life and social relations. Temporal factors, including seasons and specific dates, are also represented as events associated with the patient.

Figure 4.5 represents the environmental and allergy factors. Ontology demonstrates the interconnected factors influencing a patient’s health. The environmental category includes elements such as air quality, pollution, temperature and humidity, as well as temporal factors like seasons and specific dates. Patients are further linked to healthcare organizations, such as hospitals and clinics, where care is provided. According to studies, these factors play a significant role in chronic diseases and allergies, affecting disease severity and requiring tailored medical attention.

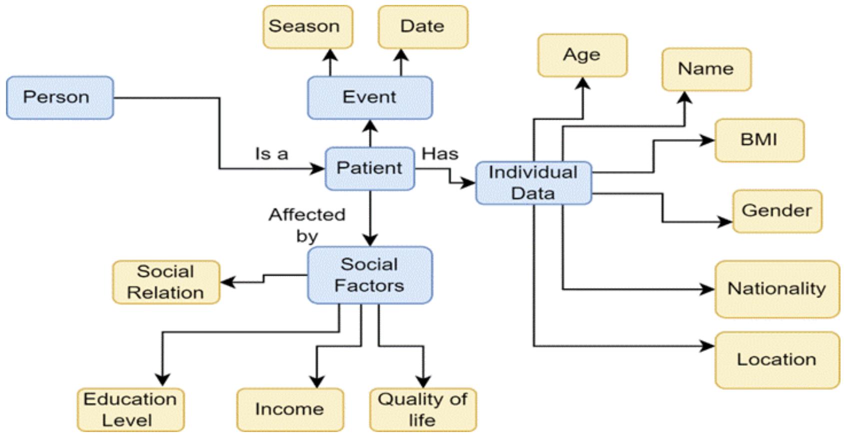

Figure 4.4 Patient ontologies

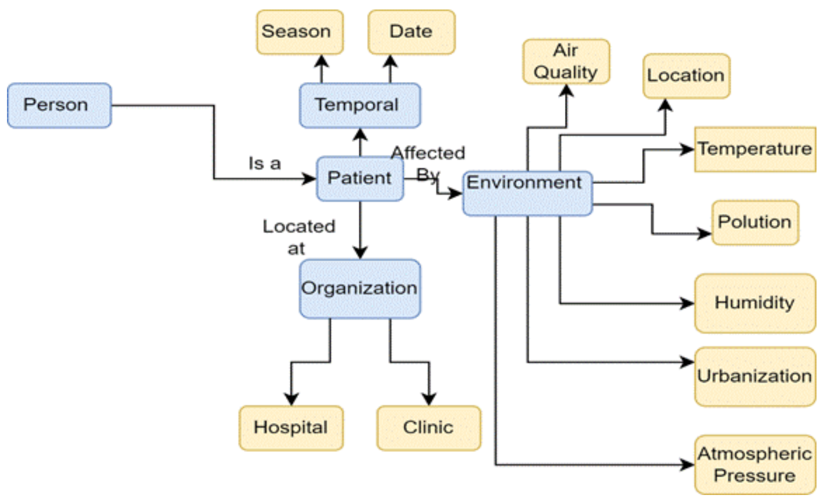

Figure 4.5 Part of the temporal, environmental and organizational ontology

We then extended this general contextual ontology to asthma based on the related dataset by adding specific classes and properties related to asthma, such as pulmonary function test results (FEV1, FVC), depending on the dataset features.

4.4.1.1 Data Acquisition

To support our research objectives, we used the Kaggle platform to find a dataset related to our case studies. We found a dataset focusing specifically on asthma. This dataset included a set of parameters related to symptoms, risk factors and diagnostic test results. It provided detailed information, including demographic parameters (e.g., age, gender, and ethnicity), lifestyle factors (e.g., BMI and smoking status) and clinical symptoms (e.g., chest tightness, coughing, and night-time symptoms). Additionally, it encompassed critical diagnostic indicators such as exercise-induced symptoms and test results. Although a comprehensive asthma ontology should include a broader range of parameters, the limitations of the dataset led us to focus on expanding the ontology based on the data currently available. This approach ensures that the ontology is well-aligned with the dataset, facilitating proper testing and validation.

4.4.1.2 Data Preprocessing

This refers to the initial steps necessary for building robust predictive models for asthma diagnosis. The raw dataset, sourced from an Excel file, contained various medical parameters critical for predicting the presence of asthma. To ensure data integrity, all records with missing values were eliminated, reducing the risk of biased outcomes. Subsequently, a pivotal transformation was performed: converting all categorical data into pure numeric formats.

The dataset used in this chapter consists of 2394 patient records with 25 relevant features, covering demographic information, clinical symptoms, lifestyle factors, and environmental exposures. The patient population includes 1212 males and 1180 females, with ages ranging from children to older adults. The class distribution is imbalanced, with 2268 records labeled as non-asthmatic and 124 as asthmatic. A small number of missing values were present in several features. For example, BMI, FEV1, and FVC had a few missing entries, which were handled using mean imputation, replacing the missing values with the average value of the respective feature. Records with missing values in the target variable (Diagnosis) were excluded from the analysis.

Due to the class imbalance, we applied the Synthetic Minority Over-sampling Technique (SMOTE) to oversample the minority class (asthmatic cases) in the training dataset. Additionally, during feature analysis, potential sources of bias, such as environmental exposure, were monitored. Addressing these factors in more depth will be part of future work during the clinical validation phase.

The dataset was then standardized to a consistent format and saved back to an Excel file for subsequent analysis, as shown in Figure 4.6. This preprocessing ensures that machine learning models operate efficiently and produce accurate results for asthma prediction.

	Age_Label	BMI_Label	Eczema_Label	HistoryOfAllergies_Label	ShortnessOfBreath_Label	ChestTightness_Label	PetAllergy_Label	FEV1_Label	FVC_Label
0	Senior	Underweight	No Eczema	No Allergies	No Shortness of Breath	Has Chest Tightness	Has Pet Allergy	Low	High
1	Young Adult	Normal	No Eczema	Has Allergies	No Shortness of Breath	No Chest Tightness	No Pet Allergy	Moderate	Low
2	Middle-Aged	Underweight	No Eczema	Has Allergies	Has Shortness of Breath	Has Chest Tightness	No Pet Allergy	Low	High
3	Adult	Obese	No Eczema	No Allergies	No Shortness of Breath	Has Chest Tightness	No Pet Allergy	Moderate	Low
4	Senior	Normal	No Eczema	No Allergies	Has Shortness of Breath	Has Chest Tightness	No Pet Allergy	High	Moderate

Figure 4.6 Data pre-processing steps

After the general contextual ontology was implemented in Protégé, the next steps focused on selecting relevant parameters to extract rules.

4.4.1.3 Selection of Relevant Parameters

Relevant parameter selection is used to identify the key features that significantly influence asthma diagnosis. The dataset was divided into input features and a target variable, with “Diagnosis” (0 = no asthma, 1 = asthma) as the target.

4.4.1.4 Rationale Behind the Choice of AI Algorithm

In the context of AI algorithms for parameter selection, a variety of techniques have been explored in the literature, each offering distinct advantages and trade-offs. Commonly used methods include decision trees and logistic regression, which are known for their interpretability and their ability to provide explicit rules. The trade-off between interpretability and accuracy is well-documented in the AI literature, particularly in healthcare applications [312]. In this chapter, we focus specifically on decision trees, logistic regression, acknowledging that while more accurate models exist, their lack of interpretability makes them less ideal for parameter selection in our ontology-based approach. Our goal is not to evaluate all AI algorithms but rather to demonstrate the feasibility of interpretable AI techniques for relevant feature selection and explainable decision-making.

4.4.1.5 Decision Tree

A decision tree is a machine learning algorithm that uses a tree-like structure to represent nodes and their possible outcomes, creating rules based on the selected features used for classification or regression [313]. In this chapter, the decision tree was applied to select relevant parameters due to its ability to prioritize features based on their importance concerning the information gain value calculated for making predictions [314].

4.4.1.6 Logistic Regression

Logistic regression is a machine learning algorithm used for binary classification tasks, predicting the probability of an outcome based on input features [315]. Logistic regression can be used for feature selection because it evaluates the coefficient of each parameter and reflects the strength and the relationship with the target variable. This makes it particularly effective for identifying the most relevant parameters in a dataset [315].

Parameter selection scores were determined based on each parameter's contribution to decision-making, as calculated for each algorithm. The dataset was split into training and testing subsets, with 20% allocated for testing to evaluate the model's performance using the metrics described in the section.

4.4.1.7 Evaluation of AI Algorithms Using Specific Metrics

To evaluate the performance of the algorithms, the author calculated several metrics, including accuracy, precision, recall, and F1-score. Based on the results of these metrics, a suitable algorithm was chosen for identifying the optimal number of relevant parameters.

- Precision

  Precision = True Positives/(True Positives + False Positives) Precision measures how many of the positive predictions made by the model are correct.

- Recall

  Recall = True Positives/(True Positives + False Negatives) Recall measures how many of the actual positive instances the model correctly identifies.

- F1-score

  F1-score = 2 × (Precision × Recall)/(Precision + Recall) The F1-score is the harmonic mean of precision and recall and provides a single metric that balances both concerns.

- Accuracy

  Accuracy = (True Positives + True Negatives)/(True Positives + True Negatives + False Positives + False Negatives) Accuracy measures the proportion of correctly predicted instances out of the total instances.

When comparing decision tree, neural network and logistic regression using the four metrics (accuracy, precision, recall and F1-score), the decision tree consistently achieved higher values. Specifically, the decision tree used 19 relevant parameters to demonstrate the best overall performance across all metrics, making it the most effective choice based on the data provided.

4.4.1.8 Decision Tree Model for Asthma Detection: Implementation Details

A Decision Tree classifier was trained to detect asthma using a structured pipeline where numerical features were standardized, categorical features one-hot encoded, and boolean features binary-encoded. The dataset of 1000 patients were split into 800 for training and 200 for testing to ensure generalizability. Using Information Gain, the top 16 features, such as FEV1 Score, Chest Tightness, and Night-time Symptoms, were selected, excluding less informative ones like Education Level. Decision rules were extracted in an interpretable if-else format, and numerical thresholds were converted back to clinical values (FEV1 ≤ 0.87 → FEV1 ≤ 2.5 L) for usability. Table 4.2 reports the evaluation metrics for the algorithms: Accuracy, Precision, Recall, and F1 Score for Decision Tree, Logistic Regression, and Neural Network. the Decision Tree offered the trade-off between accuracy and interpretability. Logistic Regression, though lower in accuracy, remains valuable for its simplicity. Furthermore, the decision tree thresholds were validated using clinical guidelines such as GINA for FEV1 and the medical literature for BMI and age, ensuring that the extracted rules are not only data-driven but also clinically sound and applicable in real-world medical decision-making [301,302].

Table 4.2 Performance comparison of AI models

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Decision Tree	0.84	0.805	0.81	0.82
Logistic Regression	0.63	0.600	0.590	0.63

4.4.2 Design the Extended Asthma Ontology

Using the asthma disease dataset and its parameters, we found that not all entities and subcategories from the general contextual ontology for chronic diseases were present in the dataset. To address this, we used the general contextual ontology as a guide to develop and extend the asthma ontology, incorporating the entities and subcategories specific to asthma found in the dataset.

The asthma ontology, as illustrated in Figure 4.8, focuses on individual patient information. For example, it captures key attributes such as age, gender, ethnicity and education level. Patients are classified by their gender, represented as either male or female with Boolean values, and their ethnicity, including categories such as African American, Caucasian, Asian and others, each linked to integer values. Education levels, such as high school, bachelor's, higher education or none, are also represented with associated integer values.

The asthma ontology, shown in Figure 4.8, centers on individual patient information. It includes key attributes like age, gender, ethnicity, and education level. Gender is categorized as male or female using Boolean values, while ethnicity is classified into groups such as African American, Caucasian, Asian, and others, each assigned an integer value. Similarly, education levels, such as high school, bachelor's, higher education, or none, are represented with corresponding integer values.

Figure 4.7 Selected individual patient information for the asthma ontology

The crucial next step is building the ontology reasoning to properly evaluate the ontology capabilities.

4.4.3 Ontology Reasoning

Ontology reasoning plays a critical role in both general and medical domains, serving as a vital tool for organizing knowledge, discovering implicit information and maintaining logical consistency. It accomplishes this by analyzing the well-defined concepts, relationships and rules embedded within an ontology [289].

Reasoning tools such as Pellet, Fact++ and Hermit rely on description logic used to define and analyze concepts and their relationships [310]. The steps to use Pellet for reasoning are:

4.4.3.1 Load the Ontology

Import the ontology file (in OWL format) into a reasoning tool such as Protégé [311].

4.4.3.2 Add Rules and Data

Use SWRL to define rules and integrate real-world data, such as symptoms or environmental triggers [312].

4.4.3.3 Run the Pellet Reasoner

Execute reasoning to classify concepts, detect inconsistencies, and derive new insights [292]. These tools are particularly effective in the medical field, where they support clinical decision-making and patient care by managing and interpreting complex healthcare datasets. Ontologies such as SNOMED CT, ICD-10 and specialized medical frameworks enhance reasoning capabilities, enabling better healthcare outcomes [313].

4.4.4 Validation for the Extended Ontology of Asthma Based on Specific Metrics

The evaluation of the extended asthma ontology involves checking its internal qualities to ensure that it is logically consistent and easy to use. Metrics for ontology evaluation include classification, reusability and completeness. Below, we explain the evaluation process used for the extended asthma ontology.

4.4.4.1 Classification

Classification involves determining relationships between classes, validating the hierarchy and ensuring that the ontology's structure is consistent [314]. The reasoner successfully validated the asthma ontology's structure, placing subclasses such as Symptoms, Triggers and Treatments under parent classes such as MedicalData. The changes to the ontology can increase reasoning time. However, the extended asthma ontology was classified using Pellet in approximately 27,000 ms, achieving near-real-time performance. Proper classification ensures that the ontology is well-structured and supports effective reasoning, which is crucial for medical use.

4.4.4.2 Reusability

Reusability assesses whether the ontology can be used in other contexts or integrated with standards such as SNOMED CT [315]. The asthma ontology is modular and aligns with medical standards, enabling its components to be reused in other healthcare ontologies. For example, its structure fits well within SNOMED CT's hierarchy for respiratory conditions. Reusability ensures that the ontology can adapt to different medical datasets, increasing its practical value.

4.4.4.3 Completeness

Completeness refers to an ontology's ability to comprehensively represent essential domain knowledge and infer the necessary information to address competency questions [316]. However, in our case, we refer to relative completeness, as it does not represent general completeness, meaning that the ontology is complete with respect to the available dataset, but its scope remains specific rather than general. The extended asthma ontology incorporates key features from the asthma dataset, such as DietQuality, LungFunction FEV1, PhysicalActivity and Smoking. Additionally, the ontology includes hierarchical structures to represent Symptoms, Triggers, Treatments and Diagnosis, ensuring logical representation of asthma-related data. The extended ontology ensures sufficient completeness to support decision-making and reasoning for asthma. However, gaps in representing rare or less-studied aspects may limit its applicability for edge cases or advanced medical research [317]. These gaps can be addressed by integrating additional datasets or collaborating with domain experts to further refine the ontology.

4.4.4.4 Consistency

Consistency means that the ontology is free from contradictions, allowing for accurate reasoning. In the context of an asthma ontology, maintaining consistency involves accurately modeling relationships among various factors such as environmental triggers, patient activities, and asthma symptoms. For example, if the ontology defines “Smoking” as a trigger for asthma exacerbations, it should not simultaneously represent “Smoking” as a non-trigger without clear contextual differentiation. Ensuring such consistency is essential for the ontology to support effective decision-making and reasoning in asthma management [318].

- The ontology was loaded into Protégé, and the Pellet reasoner was executed.
- No inconsistencies were detected in class hierarchies or property restrictions, confirming logical soundness.

4.4.4.5 Disjointness

In ontology modeling, disjointness is used to specify that certain classes do not share any common instances [319]. For example, declaring classes like MildAsthma and SevereAsthma as disjoint ensures that no individual can simultaneously belong to both categories. To verify the integrity of these constraints, you can execute a SPARQL query to identify any individuals that are erroneously classified under both disjoint classes:

```
SELECT ?individual WHERE {
?individual rdf:type:MildAsthma.
?individual rdf:type:SevereAsthma.
}
```

4.4.4.6 Complexity

Ontology complexity affects how efficiently a system can process and reason with information. When an ontology has many concepts, relationships, and rules, it requires more computing power and time to analyze and infer new knowledge. This can slow down decision-making and make reasoning tasks more complex [320]. Keeping an ontology well-structured and optimized helps improve performance and efficiency. We assessed:

- Total Axioms: 1253
- Class-to-Class Relationships: 320
- Object Properties: 145
- Tools used: Protégé Metrics Plugin and OntoMetrics. The ontology maintains a moderate complexity level, balancing detail and computational efficiency.

To assess the efficiency of the extended asthma ontology, we conducted a benchmark comparison against the SNOMED CT ontology using the Pellet reasoner. The evaluation focused on classification time, consistency checking, and rule inference speed. Our ontology, being more lightweight and domain-specific, achieved significantly faster reasoning times in both classification and consistency checking tasks. For example, classification time averaged 1.8 s for our ontology versus 5.2 s for SNOMED CT under identical conditions (Pellet reasoner, Protégé 5.5). This performance advantage is attributed to the ontology's targeted structure and optimized context modeling, making it suitable for real-time clinical decision support.

4.5 Conclusions

This chapter presents the development of a general contextual ontology for chronic diseases and its extension to an asthma-specific ontology. By incorporating real-world data, the extended asthma ontology effectively captures key elements such as symptoms, triggers, and treatments. Additionally, AI algorithms were employed to enhance ontology reasoning through rule extraction using decision trees and metric-based evaluations. This work can be further expanded by integrating non-interpretable algorithms with AI techniques applied to improve their interpretability.

Although the ontology effectively addressed asthma-related data, this chapter highlights the potential of designing a general contextual ontology and extending it to a specific chronic disease using available datasets. It also emphasizes the importance of integrating AI algorithms with ontology to enhance validation and support decision-making in health care. Future efforts will focus on applying this methodology to other chronic diseases by using larger datasets and incorporating domain expertise, further improving the ontology's completeness and adaptability across diverse healthcare applications.

CHAPTER 5
COHM: CONTEXT-AWARE ONTOLOGY-GUIDED HEALTHCARE LEARNING MODEL FOR INTERPRETABILITY

Résumé

L’extraction de règles a connu des progrès considérables dans l’analyse des données médicales ; toutefois, les applications en santé demeurent confrontées au défi d’atteindre une interprétabilité véritablement significative. De nombreuses approches existantes n’exploitent pas efficacement les connaissances du domaine et ne prennent donc pas en compte des relations cliniques importantes telles que les hiérarchies de symptômes, les groupes de déclencheurs ou les facteurs spécifiques à une maladie. Afin de répondre à ce défi, nous proposons un modèle d’apprentissage enrichi sémantiquement qui intègre l’agrégation de caractéristiques basée sur une ontologie dans deux processus distincts : l’application de techniques d’extraction de règles pour générer des explications interprétables avec ontologie et sans ontologie. Cette séparation permet d’examiner l’influence des connaissances médicales structurées sur l’interprétabilité. Pour évaluer l’interprétabilité, nous distinguons quatre catégories d’approches d’extraction de règles : les règles expertes, les apprenants directs de règles, les méthodes de distillation et les grands modèles de langage (LLM). Afin d’améliorer l’interprétabilité dans ce modèle, nous développons et validons une méthode d’intégration de l’extraction de règles avec et sans ontologie. Les résultats montrent que l’intégration de l’ontologie améliore certaines métriques spécifiques d’interprétabilité, notamment la concision syntaxique et la couverture, à travers plusieurs scénarios. Ces améliorations produisent des ensembles de règles plus concis, réduisant la complexité structurelle, et plus représentatifs d’une population de patients élargie, capturant ainsi une plus grande proportion de cas cliniques.

Ces résultats soulignent la nécessité d’intégrer les ontologies du domaine afin d’améliorer l’interprétabilité, facilitant ainsi la conception de systèmes d’intelligence artificielle fiables, transparents et cliniquement pertinents pour l’aide à la décision en santé.

Abstract

Rule extraction has made considerable progress in analyzing medical data, yet healthcare systems still face the challenge of achieving meaningful interpretability. Many existing approaches do not leverage domain knowledge effectively and therefore fail to account for important clinical relationships such as hierarchies of symptoms, trigger groups, or disease-specific factors. To resolve this challenge, we propose a semantic-enhanced learning model that integrates ontology-based feature aggregation into two distinct processes: applying rule extraction techniques to generate interpretable explanations with ontology and without ontology. This separation allows us to examine how structured medical knowledge influences interpretability. To assess interpretability, we distinguish four categories of rule extraction approaches for interpretability: expert rules, direct rule learners, ML based models, and large language models (LLMs). To enhance the interpretability in this model, we develop and validate a rule extraction integration method with and without ontology. Results show that ontology integration enhances specific interpretability metrics namely syntactic conciseness and coverage using many scenarios. These yield rule sets that are more concise, reducing structural complexity, and representative of a broader patient population, effectively capturing a larger portion of patient cases.

The results underscore the necessity of integrating domain ontologies to interpretability, thereby facilitating the creation of reliable, transparent, and clinically relevant AI systems for healthcare decision support.

Keywords: ontology-guided learning; rule extraction; interpretable models; context-aware healthcare systems.

5.1 Introduction

Healthcare decision-making depends on multifaceted information including clinical histories, symptoms, treatments, behaviors, and environmental factors which must be organized in a way that supports meaningful explanations [321, 322]. Many existing approaches treat these variables in isolation and overlook the role of structured domain knowledge, leading to models that may perform well but fail to articulate the underlying clinical relationships they rely on the research of Peleg and Tu [323]. To handle this gap, rule extraction methods aim to translate complex patterns learned by algorithms into human-readable logic that clinicians can evaluate and understand. When combined with ontology-guided structuring which groups related symptoms, triggers, or risk factors into clinically coherent categories these methods provide clearer and more useful insights into how input factors contribute to an outcome [324]. Together, rule extraction methods with ontology offer a pathway for generating transparent, knowledge-aligned explanations.

5.5.1 Motivation

Healthcare systems are inherently complex, drawing on diverse clinical records, symptoms, patient behaviors, and environmental conditions that must be interpreted together to support safe and reliable decisions [321]. When such information is analyzed without a clear structure, decision-support systems whether rule-based or data-driven often struggle to capture meaningful clinical relationships, leading to outputs that can be difficult for clinicians to understand or trust [322].

From a social perspective, interpretable models help doctors, nurses, and patients better understand each other [323, 324]. Transparently expressing the basis of a prediction improves trust and reduces the risk of communication errors or misunderstandings [325, 326]. Economically, improving interpretability can reduce unnecessary testing and hospital readmissions, contributing to more efficient healthcare management [327, 328].

In addition, ontology-guided structuring of medical knowledge provides a consistent way to organize symptoms, triggers, and clinical factors, which can support more coherent feature representations and clearer rule generation [329, 330]. Overall, applying ontology-guided structuring to rule-extraction methods offers a pathway for developing decision-support tools that are interpretable and aligned with clinical reasoning [331].

Ontology is a formal method to represent knowledge in a certain field. It includes a structured vocabulary of classes (like Patient, Symptom, and clinical_measurments), instances (like individual patients and observations), and relations (like has_symptoms, suffer_from, and hasAllergyStatus) [332, 333]. It also defines data properties and hierarchical groupings (PollenAllergy $\sqsubseteq$ Enironmentfactors). By standardizing names, and conceptual relationships, ontology acts as a knowledge-guided structuring layer, transforming heterogeneous data sources (EHR tables, questionnaires, extracted notes) into a more coherent feature space [344].

However, many existing ontology-based systems face limitations. They often rely on simple hierarchical relations (is-a, part-of), which cannot fully represent the complexity of clinical concepts and may result in less informative patterns [335, 336]. Moreover, clinical language is intricate, and the mapping of unstructured text to specific ontology concepts continues to be difficult, potentially generating noise in subsequent learning [335].

5.5.2 Problem statement

Developing artificial intelligence (AI) models in healthcare faces a major challenge: finding a high prediction accuracy and clear interpretability. AI models such as XGBoost can predict diseases very accurately, but their complex, non-transparent structure makes it difficult for medical staff (physicians, nurses, patients) to understand how predictions are made. In clinical settings, doctors need to understand the justifications and why a model predicts a disease (for example, factors explaining a risk of diabetics, symptoms explaining asthma) to make safe and trustworthy decisions [322, 323]. From this point interpretability is one of the most important requirements in healthcare domain.

Interpretability is also required by laws such as the General Data Protection Regulation (GDPR) in Europe and the Health Insurance Portability and Accountability Act (HIPAA) in the United States, which both stress traceability, accountability, and transparency in automated medical decisions [337, 338].

Therefore, in this study, we aim to design a method that optimize interpretability. We assess how the ontology-guided structuring impacts the rule-extraction methods to enhance the interpretability of a decision. This allows us to analyze the effect of ontology on the interpretability of rule extraction methods as an outcome [324, 331].

Our hypothesis of research can be described as follows: Ontology and context aware can improve the interpretability of a decision, especially in the medical domain.

5.5.3 Goals and contributions

Our goal is to support clinicians by providing rule extraction methods by offering clear, understandable explanations. To achieve this, we develop an ontology-guided learning model that uses domain knowledge to structure input features and improves the clarity of the explanations derived from rule-extraction methods.

We express this objective through two main goals:

G1: Develop a semantic-enhanced learning model that integrates ontology-guided structuring with rule-extraction methods.

G2: Evaluate how ontology-based structuring influences the interpretability of rule extraction methods.

To fulfill these goals G1 and G2, our approach organizes the learning process into a coherent methodology that includes data preparation, feature structuring using ontology concepts, apply rule extraction methods, and explanation generation.

Our contributions are summarized as follows:

C1: Review, understand and classify existing rule-extraction methods and identify their limitations.

C2: Propose a semantic-enhanced learning model that incorporates ontology-guided feature structuring.

C3: Apply and compare this model within a more detailed methodology.

C4: Assess how ontology integration influences the interpretability of extracted rules methods.

C5: publish a research article in a renewed journal with peer review.

Finally, ontology-guided structuring improves interpretability by organizing features through explicit semantic relations (has_Symptom, is_ExposedTo, has_ClinicalMeasurement) and mapping detailed variables into higher-level clinical concepts. This hierarchical representation reduces fragmented conditions and expresses rules using medically meaningful categories, making the explanations more coherent and aligned with clinical reasoning.

5.5.4 Organization of the paper

The remainder of this paper is organized as follows. Section 2 reviews related work on healthcare systems, rule-extraction methods, and ontology-based approaches. Section 3 presents the rule-extraction methods. Section 4 describes the semantic-enhanced learning methodology, integrating ontology-based feature structuring with rule-extraction techniques, and assesses its effects on interpretability utilizing the asthma dataset and the evaluation metrics used in this study. Section 5 presents a comparison between results derived from the dataset without ontology and those obtained using the ontology. Section 6 concludes the paper and outlines directions for future research.

5.2 Related Works

This section examines three essential areas in healthcare decision-support research. First, we review existing healthcare standards such as HL7, openEHR, CDSS systems, and SOA-based which structure clinical data but lack support for context-aware and interpretable AI. Second, we survey the rule extraction methods that aim to make model decisions understandable by producing human-readable rules. Third, we review ontology-based approaches that use structured medical knowledge to enhance data consistency and support more meaningful reasoning. Together, these areas highlight the limitations of current solutions and motivate the layered, ontology-driven approach proposed in this work.

5.2.1 Survey of healthcare frameworks

Several established healthcare standards and system designs such as HL7, openEHR, CDSS rule-based systems, and SOA-based hospital information systems focus primarily on structuring, storing, and exchanging clinical data. HL7 allows standardized communication between systems but does not represent the contextual factors surrounding clinical events, such as asthma symptoms influenced by weather or physical activity [339]. OpenEHR captures structured clinical information but lacks mechanisms for incorporating dynamic environmental or behavioral context [340]. Conventional CDSS rule-based systems can generate recommendations; however, they depend on static logic and fail to represent more complex connections between medical concepts [341]. SOA-based hospital systems support interoperability across services, but they do not combine preprocessing, contextual enrichment, ontology-based structuring, and explainability within one analytical process [342]. Overall, these existing approaches remain effective for data exchange and documentation but provide limited support for context-aware or interpretable rules, which motivates the need for methods that integrate domain knowledge into explanation.

5.2.2 Survey of Rule Extraction Methods

Recently, rule extraction has attracted growing attention, particularly in domains where interpretability and trust are essential, such as healthcare. Rule-based systems have a long history in clinical decision support, offering simple if–then statements that help healthcare providers interpret patient data and assist in diagnosis or treatment selection [343]. With the rise of complex AI algorithms, there has been increasing interest in extracting rules from data-driven systems to ensure transparency [322].

Different methods have been proposed for rule extraction, ranging from symbolic approaches to hybrid methods combining machine learning with domain knowledge. In healthcare, these methods must not only achieve high predictive performance but also maintain clinical relevance and interpretability [323]. Table 5.1 summarizes some of the key previous works in rule extraction, highlighting the method used, the application domain, and the main characteristics.

Table 5.1 Rule Extraction Methods in Healthcare (2015–2024)

Reference	**AI Algorithms**	**Healthcare Application**	**Type of Rules**
[344] Letham *et al.*, 2015	Bayesian Rule	Stroke Risk Prediction	Probabilistic Decision Rules
[345] Yang, Rudin & Seltzer, 2016	Bayesian Rule Lists	Disease Prediction (EHRs)	Probabilistic Decision Rules
[346] Johansson *et al.*, 2016	Causal Inference with Neural Networks	Treatment Effect Estimation (EHRs)	Causal Rules
[347] Wang *et al.*, 2017	Falling Rule Lists	Breast Cancer Risk Prediction	Ordered Rule Lists
[326] Tonekaboni *et al.*, 2019	RuleFit	ICU Mortality Prediction (Pediatrics)	Sparse Decision Rules
[322] Guidotti *et al.*, 2019	LIME, Anchors	Medical Imaging, EHRs	Local Approximate Rules
[348] Lakkaraju *et al.*, 2016	Decision Tree	Diabetes Disease	Decision Lists

[349] Rudin, 2019	CORELS	Sleep Apnea Detection	Compact Rule Lists
[350] Zhang *et al.*, 2020	Interpretable Reinforcement Learning	Sepsis Treatment Policy (ICU)	Decision Lists
[351] Schwab *et al.*, 2020	Personalized Rule Extraction	Heart Failure Risk Stratification	Personalized Decision Rules
[352] Tseng *et al.*, 2021	Decision Tree Pruning for Interpretability	Early Sepsis Detection	Simplified Decision Trees
[353] Sani *et al.*, 2023	Transformer-based Feature Selection + Rules	Oncology: Treatment Pathway Selection	Rule-based Recommendations
[354] Ghassemi *et al.*, 2024	Counterfactual Explanations	ICU Decision Support	Counterfactual Rules
[355] Rochan et al., 2024	Random Forest	Diabetes stages	Decision Lists

5.2.3 Discussion of Rule Extraction Methods

In the studies shown in Table 5.1, we can observe several important trends. Early work focused on probabilistic decision rules and association rules built from electronic health records (EHRs) [344, 345].

Later, researchers introduced causal inference models and ordered rule lists to move beyond simple associations and identify real cause effect relationships, which are critical for healthcare decisions [346], 343].

Since 2019, there has been a shift toward using CORELS [349]. These methods allow machine learning systems to suggest rules while enabling doctors and domain experts to verify their clinical relevance. At the same time, models such as compact rule lists and RuleFit were developed to create smaller and simpler rule sets that are easier for clinicians to interpret [326, 349].

More recent work (2020–2024) focuses on personalized rule extraction and counterfactual explanations [347, 350]. These methods aim to generate patient-specific insights, aligning with the growing focus on precision medicine.

Finally, emerging techniques such as transformer-based models are being used to extract rules from complex medical datasets, including genetics and oncology [349]. These approaches support the management of large-scale, heterogeneous data while maintaining interpretability.

To illustrate how ontologies are applied alongside rule extraction methods, the next section presents selected studies that integrate ontological models with rule extraction techniques in healthcare.

5.2.4 Overview of Ontology-Driven Rule Extraction Approaches in Healthcare

The table below (Table 5.2) outlines representative works where ontologies serve as knowledge representation models integrated with rule association or extraction methods. Each study demonstrates how ontologies help organize and standardize clinical data, while the rule component builds upon this structured knowledge to support various healthcare applications such as diagnosis, prediction, and decision support.

Table 5.2 Existing Works Integrating Ontology and Rule Extraction Methods in Healthcare Systems

Year / Reference	Ontology + Rule Method	Healthcare Application	Integration Summary
2013 – [356] Martínez-Romero *et al.* (iOSC3)	Cardiac Care Ontology (OWL) + SWRL treatment rules	Intensive cardiac care supervision	Ontology models patient conditions: SWRL rules represent therapy decisions.
2017 – [357] Chen *et al.*	Diabetes Ontology + fuzzy logic rules	Type 2 diabetes decision support	Ontology defines patient, lab, and drug relations; fuzzy rules reason on those entities to support individualized medication advice.
2017 – [358] Bytyçi *et al.*	Clinical Context Ontology + Association Rule Mining	Clinical data	Ontology supplies semantic context to Apriori mining.

2018 – [359] El-Sappagh *et al.*	SNOMED-CT Ontology + RuleFit	Interoperability and clinical decision support	Ontology provides semantics; RuleFit built on standardized classes.
2020 – [360] Thamer *et al.*	SNOMED-CT Ontology + Association Rule Mining	Disease diagnosis pattern discovery	Ontology structures dataset attributes; Apriori generates semantically enriched, clinically valid rules.
2023 – [361] Jing *et al.*	Clinical Ontologies (SNOMED, LOINC)	Clinical decision-support rule management	Ontology organizes and governs rule content across CDS systems.
2024 – [362] Ouartani *et al.*	Thyroid Ontology + Random Forest	Thyroid disease prediction	Ontology encodes domain knowledge; random forest rules extracted for explainable prediction.
2024 – [363] El Massari *et al.*	Cardiovascular Ontology + Decision Tree	Cardiovascular disease risk prediction	Ontology models risk factors and lab features; decision-tree path rules are encoded as SWRL, merging semantics with AI output.

5.2.5 Limitations of Existing Works

Many studies have developed data-driven models to create medical rules from clinical data. However, these models often have limitations. For instance, Miotto et al. [364] created rules from Electronic Health Records (EHRs), but these rules depended on local coding systems [365], making them difficult to apply universally in hospitals that do not consistently use standardized vocabularies such as SNOMED CT [366, 367]. Other researchers, including Sappagh et al (2018), Thamer et al. (2020), and Jing et al. (2023), generated understandable rules but frequently overlooked important clinical details such as comorbidities and patient history [369, 370]. Some features were too broadly grouped, risking the loss of clinically relevant details and reducing interpretability [371]. Moreover, large models often produce redundant or conflicting rules, complicating their validation and use in clinical settings [372]. Consequently, many existing rules are fragmented, difficult to explain, and poorly suited to real medical contexts [373].

Ontology-based approaches aim to structure medical knowledge but also face challenges. Many existing healthcare ontologies primarily rely on simple hierarchical relationships such as is-a or part-of, which are insufficient to capture the complex interactions among symptoms, treatments, risk factors, and outcomes reported across multiple studies [374, 375]. Updating ontologies with new data is typically manual and time-consuming [376].

Within this perspective, combining ontology-based representation with data-driven rule generation can offer an effective solution. Ontologies provide a standardized foundation for medical knowledge, while rules make the reasoning process transparent to clinicians by enhancing interpretability in healthcare decision-support systems. This approach ensures that rules remain clinically meaningful and understandable across diverse care environments [376].

5.3 Ontology-Driven Approach and Expected Outcomes

This research presents a semantic-enhanced learning model that integrates ontology-guided structuring into rule-extraction methods stage. In addition, the ontology explicitly models semantic relations between concepts (symptoms disease and risk factors relations), which are preserved during rule extraction [324]. The explicit use of causal relations within the ontology guides the extraction of rules toward clinically coherent reasoning paths, making the derived rules directly interpretable as medical explanations rather than abstract statistical pattern [331].

In this study, we adopt, that ontology is used to improve the interpretability of the extracted rules [330].

This comparison allows us to determine whether ontology-guided aggregation offers benefits beyond standard statistical dimensionality reduction techniques

Finally, because the dataset used in this study is synthetic and obtained from a public repository (Kaggle), the findings should be interpreted as methodological proof-of-concept. The aim is to show how ontology-guided feature structuring and rule extraction methods can be combined to produce more transparent explanations, rather than to make claims about real clinical deployment.

5.4 Categorization of Existing Rule Extraction Methods

This section presents a classification of the main approaches used for rule extraction in the healthcare domain. We distinguished four main categories as illustrated in the figure 5.1 below.

Our aim is to identify an approach for generating rules that allows us to assess improvements in interpretability. Indeed, interpretability is essential for understanding the decision in a critical (medical) domain [342] and meeting regulatory constraints [343]. We distinguish four categories for extracting rules methods:

Category 1: Expert rules, used in our semi-automatic approach [376], which is based on physicians and medical standards [376];

Category 2: Direct rule Learners AI models based directly on data [377], (Association-Rule Mining, ...) [377], but which produce explanations that are too technical for clinicians [377, 378];

Category 3: ML based models [378], which do not simultaneously address these requirements;

Category 4: Large Language Models (LLMs), which generate biased content or fabricate facts [379].

Note that AI models that offer remarkable accuracy (XGBoost) have nonlinear and multilayered structures, making their predictions opaque and difficult to explain to clinicians [349, 350, 351]. This opacity poses regulatory challenges with frameworks (GDPR, HIPAA), which require traceability of decisions [337, 338]. The paradox is compounded by the limitations of explainability methods (examples, SHAP, LIME, TE2Rules), which suffer from instability and produce explanations that are too technical for clinical workflows [347].

Each subsection introduces a specific category and illustrates how rules are generated. To maintain consistency and better comparison, the same example format is used across all methods.

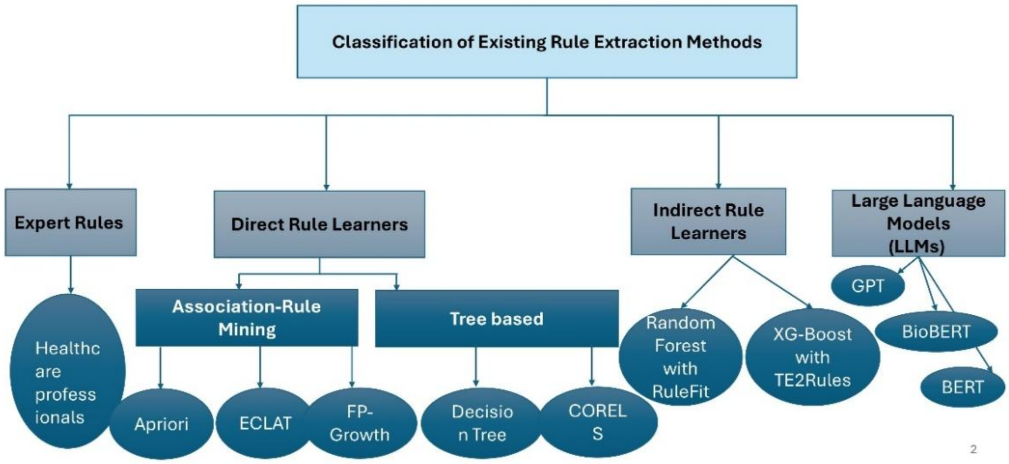

Figure 5.1 Categorization of existing rule extraction methods [366, 367, 368, 369]

5.4.1 Expert Rules Methods

In traditional healthcare, many decision rules are established directly by doctors. These rules stem from their medical education, clinical guidelines, and years of experience with patients. Doctors rely on their knowledge and observations to formulate if-then logic that aids in making consistent and safe decisions [366].

To understand these methods, let us take an example:

IF a patient has chest pain AND elevated troponin levels, THEN suspect a heart attack.

These expert rules are often documented in structured formats that help doctors apply decisions step by step. They are commonly used in training and clinical protocols [367]. Because these rules use familiar medical language and reflect established practices, they are easily understood and followed by healthcare professionals [366, 370].

Expert rules are used for rule extraction for several reasons:

They are understandable and trustworthy because they are based on human reasoning [370].

They are clinically validated, often adhering to official medical standards like practice guidelines [371].

These rule sets are especially valuable in environments that prioritize patient safety and consistency of care, particularly in areas such as emergency medicine, chronic disease management, and diagnostic triage [369, 370].

However, this manual approach has limitations. It often struggles to capture all the nuances of real-world situations, which can lead to poor performance with unseen data [371]. Furthermore, the process of creating these rules can be time-consuming and may vary by region or hospital, depending on how guidelines are interpreted. Maintaining these rules also becomes challenging, as newly created rules can sometimes conflict with existing ones [371].

5.4.2 Direct Rule Learners Methods

Direct rule learners turn patient data into if–then rules that clinicians can read and use. There are two ways to get these rules.

The first is association-rule mining, is the process of finding strong rules that describes the correlations among the items of a certain database. which finds frequent patterns such as “Allergy = True and NighttimeSymptoms = True.” These patterns are valuable for uncovering clinically meaningful combinations but, by themselves, do not provide a complete prediction for every patient [367, 368].

The second is machine-learning rule learners including decision trees, RIPPER, CORELS, and RuleFit [364, 374, 375], which learn a global decision policy where each path or list entry corresponds to an if–then rule [375].

5.4.2.1 Association-Rule Mining

Data mining is a technique used to discover meaningful patterns and relationships in large amounts of data, especially when the dataset is too complex for manual analysis [375, 376]. One of the most widely used methods in data mining is association rule learning, which automatically generates rules [376]. These rules describe how certain features such as symptoms, test results, or treatments are linked together in real patient records [367, 372].

For example:

Eczema = Yes ^ LowFEV1 = Yes → Asthma

These types of rules are not generated manually but are learned automatically from the data using algorithms. The most used algorithms for this rule extraction are Apriori [372], FP-Growth [373], and ECLAT [373]. These methods are especially useful in healthcare because they can analyze electronic health records and extract clinical patterns that may not be immediately obvious to doctors [374]. Such rules help physicians understand hidden relationships, improve diagnosis, and support data driven decisions in clinical care [323,374].

These algorithms are often used in healthcare for tasks such as:

Identifying frequent combinations of symptoms in asthma [375].

Finding which medications are usually prescribed together [376].

Detecting patterns in patient monitoring data [377].

The main strength of this approach is that it can reveal hidden patterns in patient data that clinicians might not immediately recognize [378, 379]. For example, it can show that certain combinations of symptoms and medications are often seen together in patients with a specific diagnosis [379].

However, these methods have limitations. They often assume linear relationships and might fail to capture complex nonlinear patterns present in real medical data [380, 381]. Moreover, association rule mining can generate rules from any combination of items, regardless of their medical relevance, which may lead to spurious or unhelpful rules [372]. A rule might be statistically strong but clinically meaningless unless reviewed and validated by domain experts [372, 373].

5.4.2.2 Tree Based

Machine learning (ML) is a subset of AI algorithms where computers learn from data to provide decisions or predictions [386]. In AI Instead of writing rules manually, the computer finds patterns and creates rules automatically by training on a dataset [385]. Once trained, AI models can generate many "if-then" rules that help in diagnosing diseases or predicting outcomes [323, 340, 345].

For example:

IF a patient has chest pain AND high troponin levels AND high blood pressure, the model predicts cardiovascular disease.

We consider four algorithms represent different ways of generating rules.

Decision Trees (such as CART or C4.5) create rules by splitting the data step by step; each path from root to leaf forms a simple if–then rule, for example: FEV1 ≤ 2.0 ^ Smoking = Yes → Asthma [386].

RuleFit first extracts many simple rules from decision trees and then combines them in a linear model, where each rule receives a weight indicating how strongly it contributes to the prediction [387]. Unlike a single decision tree, RuleFit allows multiple rules to contribute together to the final prediction [387].

Bayesian Rule Lists generate ordered rules associated with probabilities, such as ChestPain ^ HighTroponin → 85% probability of HeartAttack [388].

CORELS uses mathematical optimization to find the most accurate rule list possible, ensuring that the final set of rules is compact and globally optimal [389].

These machine-learned rules are helpful because they can handle large and complex medical datasets more efficiently than traditional methods. They have been successfully used in tasks like predicting disease risk, recommending treatments, and supporting early diagnosis [385, 386].

They can capture complex patterns in the data that would be hard to hand-code, and they produce rules directly from the dataset rather than from manually written logic [388, 389].

5.4.3 ML based models

ML-based models refer to machine learning approaches that learn patterns directly from healthcare data and then transform these learned patterns into interpretable rules. In this approach, a predictive model such as XGBoost or Random Forest is trained on the dataset, and a rule extraction method such as TE2Rules or RuleFit is applied to derive explicit decision rules from the trained model [390, 391]. This allows maintaining strong predictive performance while improving interpretability through rule-based representations [392].

The ML-based models for rule extraction in healthcare are used to identify clinical patterns and convert them into understandable rules that support medical decision-making. For example, an XGBoost model can be trained to classify whether a patient is at risk of asthma exacerbation, and TE2Rules can extract rules that describe which combinations of symptoms, environmental factors, and clinical indicators lead to that prediction. Similarly, a Random Forest model can be used to learn complex clinical relationships, and RuleFit can be applied to extract interpretable rules from the ensemble model [390–392].

The process of ML-based rule extraction is described as follows:

Learning Model a machine learning model such as XGBoost or Random Forest is trained on healthcare data to learn the relationship between input variables and the target outcome [390]. The model captures complex interactions between features such as symptoms, environmental exposures, and patient history. For example, it can learn patterns associated with asthma diagnosis or severity [391].

Rule Extraction, after training, a rule extraction method such as TE2Rules or RuleFit is applied to the model [392]. These methods analyze the behavior of the trained model and convert its internal decision logic into a set of explicit if–then rules.

Generated Rules

- The extracted rules reflect the learned decision patterns of the trained model [390].
- Each rule represents a combination of clinical conditions associated with a specific prediction [391].
- These rules improve transparency and allow clinicians to better understand the reasoning behind predictions [392].

5.4.4 Large Language Models (LLMs)

Large Language Models (LLMs), like BioGPT, ClinicalBERT, and BioBERT, BioMedLM, and Mistral 7B are advanced AI models that can read and medical text., they work within a finite context window, and performance can drop when key facts are buried in long inputs These models are trained on scientific articles, clinical notes, and guidelines [393-395]. They can extract meaningful rules by recognizing patterns in how medical knowledge is written [396].

For example, from a sentence like:

“Chest pain and high troponin suggest a heart attack,”

The model can generate a rule like:

IF chest pain AND high troponin THEN suspect heart attack.

LLMs can be useful in healthcare because they can:

Turn unstructured text (like doctor's notes) into structured rules [396, 397].

Help automate the creation of decision-support rules from large clinical guidelines [396].

However, these models are not perfect. The rules they generate must be carefully reviewed by human experts, as LLMs can sometimes make mistakes [397]. Also, interpretability is still a challenge, and researchers are working on improving explainability and trust in their outputs [323, 326].

5.4 Comparison of Rule Extraction Methods

Table 5.2 compares our rule approaches (expert rules, direct rule learners, ML based models, and LLM rules). It shows where rules come from, how automated they are, typical uses, and key sources. This helps explain how each method supports clinical decisions, from manual guidelines to data-driven rules like machine learning and large language models.

Table 5.2 Comparison of Rule Extraction Methods

Key References	Method	Rule Source	Automation Level	Example Usage
Ordonez 2006 – [78]	Direct Rule learners (Association rule mining)	EHR/Databases	Automatic	Pattern mining in prescriptions
Musen et al. 2014 – [79]	Expert-Rule	Medical Experts	Manual	Clinical guidelines
Rudin 2019 – [80]	Direct Rule learners (machine learning)	Patient Data	Automatic	Risk prediction models
Ghassemi et al. 2024 – [76]	LLMs	Textual Guidelines	Semi-Automatic	Rule generation from medical literature

In addition, rule extraction methods are evaluated by interpretability metrics (number of rules, average antecedent length, and coverage). The next section presents the methodology and the results side by side and shows how ontology affects both metrics.

5.5 Technologies and methods Semantic-Enhanced Learning Architectural Model

To manage the complexity of healthcare domain and ensure a clear progression from raw inputs to interpretable outputs, we build a semantic-enhanced learning architectural model composed of five layers, as illustrated in Figure 5.2. Each layer performs specific services to develop a context-aware healthcare application: Data acquisition, context categorization, context representation, operation and application layer. This organization provides a transparent view of how the dataset is transformed and analyzed throughout the methodology.

Data Acquisition Layer 1: collects medical and contextual data from multiple sources, including open repositories, clinical records, sensors, and wearable devices. The acquired data covers demographic factors, medical profiles, physical activity, and environmental conditions.

Context Categorization Layer 2: Each variable produced in layer 2 is assigned to one of fifteen predefined context categories based on our previously developed context model, which has been applied in three different disease case studies. This step organizes heterogeneous data into coherent groups related to the patient, environment, behavior, and clinical state.

Semantic Representation Layer 3: Using the context-labeled variables from layer 2, this stage aligns them with a domain ontology (and a selected knowledge representation model), yielding shared concept definitions and relationships that support feature representation [324, 401]. It translates this context into machine-understandable and accessible language.

Operation Layer 4: Interpretability: Guided by the developed domain ontology established in layer 3, this layer evaluates the interpretability of the rule-extraction methods identified previously in the section 5.3. Detailed descriptions and supporting studies for each component are provided in Sections 5.4.2 and 5.4.3.

Application Layer 5: Leveraging the validated predictions and interpretable rule sets produced in layer 4, this layer provides services for physicians/nurses and for patients. For example, clinicians can access real-time monitoring, decision support, and treatment evaluation; patients can receive risk assessments of vital signs, external risk factors, and planned activities.

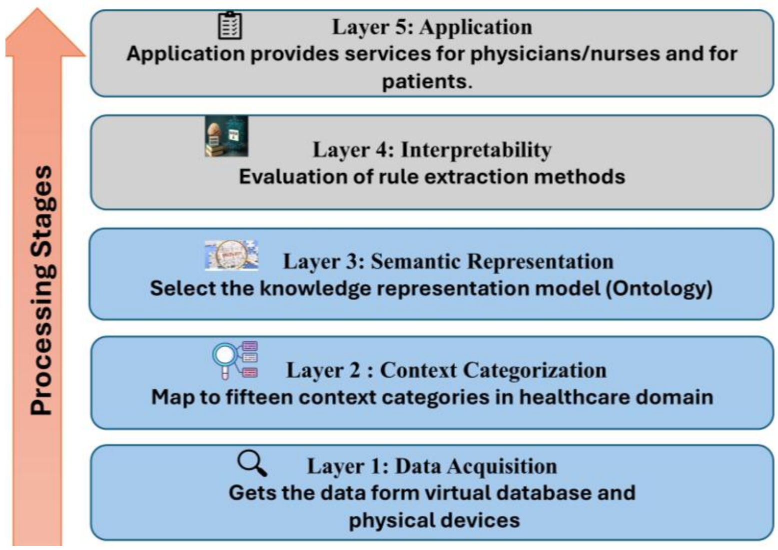

Figure 5.2 Semantic-Enhanced Learning Architectural Model

5.5.1 Semantic-Enhanced Learning Methodology using Four Scenarios

The five-layers semantic-enhanced learning model shown in Figure 5.3 is applied directly in our methodological setup, supporting the evaluation settings of this study: interpretability with raw data, and interpretability with ontology-enhanced data. This methodology focuses on the data-processing workflow data collection, cleaning, contextual labeling, ontology-guided structuring, model evaluation, and output delivery. It is summarized into five steps:

Step 1 provides the input foundation by supplying the raw asthma dataset used across all four experimental configurations. Step 2 prepares this dataset through cleaning, normalization, and transformation, ensuring consistent input regardless of whether ontology is later added. Step 3 enriches the preprocessed data by assigning each feature to one of the fifteen contextual categories defined in our earlier work [401], enabling a structured representation that feeds step 4. Step 4 introduces semantic representation by linking contextualized attributes to our ontology [404]. Step 5 executes the dual evaluation process, where the interpretability path applies the rule extraction methods distinguished before:

Scenario 1: This scenario applies Association Rule Mining (Apriori). Scenario 2: : This scenario applies Machine Learning models, such as Random Forest algorithm with RuleFit

Scenario 3: ML based models: Here, we use a model (like XGBoost) and a simpler model (like TE2Rules) .

We evaluate each scenario using number of rules, average antecedent length, and coverage [402–406]. Finally, we integrate the outputs of evaluation path, allowing us to compare and interpretability across the three applications of the model and quantify the added value of ontology within the overall methodology.

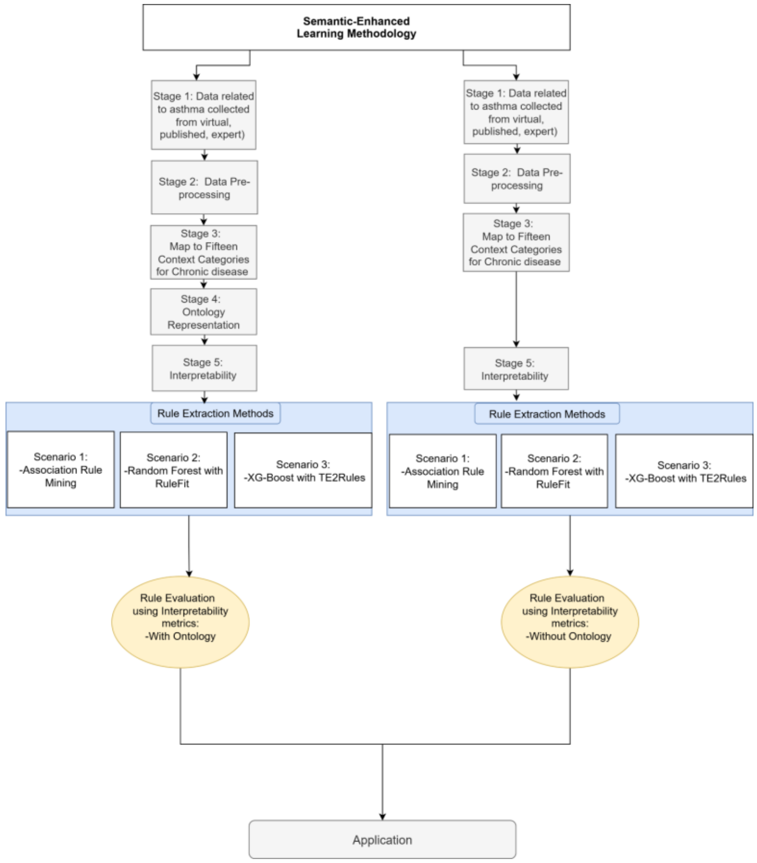

Figure 5.3 Interpretability Rules Extraction Methodology using three scenarios in each pipeline

In the sections below, each stage of the proposed methodology is described in detail, from data acquisition to evaluation. While all stages contribute to the overall methodology, while stage 4 plays a central role by enriching the ontology to ensure that dataset attributes correspond to well-defined medical concepts and are semantically prepared for subsequent use by rule extraction methods in Stage 5. The details of this process are provided in Section 5.7.3.

5.5.1.1 Data acquisition

The dataset was obtained from Kaggle [407], we analyzed a structured dataset (2392 patients, 27 variables): demographics (Age, Gender, Ethnicity, EducationLevel, BMI), lifestyle/exposures (Smoking, PhysicalActivity, DietQuality, SleepQuality, PollutionExposure, PollenExposure, DustExposure), allergies/comorbidities (PetAllergy, FamilyHistoryAsthma, HistoryOfAllergies, Eczema, HayFever, GastroesophagealReflux), lung function (FEV1, FVC), symptoms (Wheezing, ShortnessOfBreath, ChestTightness, Coughing, NighttimeSymptoms), trigger (ExerciseInduced), label: Diagnosis (0/1).

5.5.1.2 Data Pre-processing

Data preprocessing prepares raw medical data to be consistent, and ready for analysis. It starts by removing duplicate records and handling missing values, for example, by replacing them with the median for continuous variables or the most frequent value for categorical ones [408]. Outliers that fall outside normal medical ranges are identified and corrected to avoid errors. The data are then normalized and encoded; therefore, all features have comparable scales and formats. Information from different sources, such as clinical records, sensors, and questionnaires, is also combined and standardized to ensure uniform units and naming. Finally, the dataset is divided into training and testing parts to allow fair model evaluation. These steps help improve data quality and build a reliable base for ontology integration and rule extraction.

5.5.1.3 Context Categorization

Context categorization is the process of grouping all patient-related information into fifteen well-defined categories where the system can clearly understand the patient’s overall situation. These categories capture different dimensions of health, including the patient themselves (person), their medical history, current medical data, symptoms, environment, lifestyle and activities, emotional state, medications, allergies, disease-specific factors, temporal aspects, social and family context, technology and device use, organizational or clinical settings, and risk-related factors. By organizing information into these fifteen categories, healthcare systems gain a structured and complete view of the patient, which improves understanding, and decision-making [401].

5.5.1.4 Ontology Semantic Representation Model

We built the Asthma ontology based on Sanchez methodology and Protégé tool [324, 364]. The core classes constitute of many entities, such as Patient (a Person), Profile, Physiological parameters, Questionnaire/Question/Answer, psychological status, and Diagnostic. In the figure 4 Patient has a Profile with data properties for age, race, nationality, occupation, address, telephone. A patient has physiological parameters such as Heart Rate, Systolic BP, Diastolic BP, Respiration Rate, Temperature, SaO_2, PaO_2, $PaCO_2$, FEV1. For each numeric parameter we keep the raw value and semantic ranges (normal, mild, moderate, severe) using properties like hasMinNormalRange / hasMaxNormalRange, etc., therefore values can be classified automatically. The final disease labels are instances of Diagnostic. The explicit relations defined within the ontology play a central role in enhancing interpretability by guiding rule construction toward clinically meaningful reasoning. This claim is detailed in the following section.

Constructing the Ontology

We developed our ontology using an asthma dataset, following these straightforward steps:

Define Class Hierarchy: We organized terms into a hierarchical structure, creating super classes and subclasses.

For example, we established a superclass called "Respiratory Symptom." Under this, we created subclasses for specific symptoms such as "Wheezing" and "Cough."

Similarly, we created a superclass named "Environmental Factor," which includes subclasses for specific triggers like "Pollen" and "Dust Mite."

The method used to represent the dataset using the ontology was previously published and discussed briefly in the precedent study [324], which we recommend reading for more information. Figure 5.4 illustrates the graphical representation of the ontology.

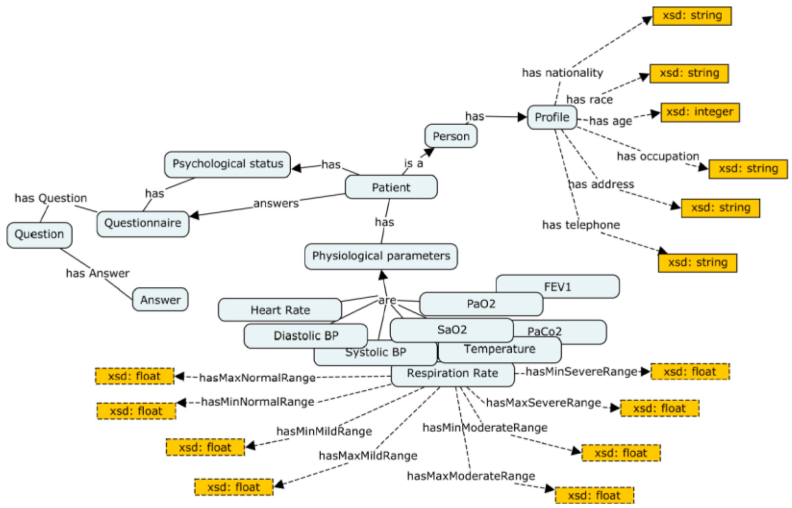

Figure 5.4 Clinical ontology representation for asthma disease

5.5.1.5 Rules extraction method for interpretability using AI algorithms with ontology

This rules extraction method is illustrated in Figure 5.5 It contains many steps:

Step 1: collecting a structured clinical dataset; the dataset contains patient records and features such as wheezing, coughing, FEV1 values, pollution exposure, and asthma diagnosis.

Step 2: leveraging a previously developed general healthcare contextual ontology from our prior work [324]. The general contextual ontology includes classes such as Patient, Symptoms, ClinicalMeasurement, and EnvironmentalExposure.

Step 3: The ontology is instantiated based on the clinical dataset, resulting in an enhanced domain ontology enriched with patient-level instances. the patient records (Patient_001) are added as an instance linked to classes within the ontology.

Step 4: Then, RDF triples (subject–predicate–object) are formed and organized as a structured semantic representation of domain knowledge. The triples enriched through hierarchy propagation and materialization of subclass relationships to enhance conceptual abstraction.

Step 5: The enriched semantic relations are transformed into an ontology-based feature matrix that serves as input for both association rule mining and supervised machine learning. This step transforms the OWL file into a tabular format because most AI algorithms cannot directly process OWL, allowing the models to work with the semantically enriched data through the implemented code [402]. The tabular matrix including semantic features such as has_TypicalRespiratorySymptoms = 1 and has_LowFEV1 = 1.

Step 6: While the Apriori algorithm identifies frequent relational co-occurrence patterns, and the trained models learn predictive structures from the enriched semantic features. Apriori may discover that (Smoking ∧ Wheezing) frequently co-occurs with asthma.

Step 7: The trained ensemble model is converted into an interpretable rule representation using the TE2Rules algorithm, that enables the generation of ontology-level IF–THEN rules. IF has_Symptom some TypicalRespiratorySymptoms AND has_ClinicalMeasurement LowFEV1 THEN Asthma.

Step 8: Finally, the extracted rule sets are evaluated using quantitative interpretability metrics such as the number of rules, average antecedent length, and coverage to assess model transparency and compactness.

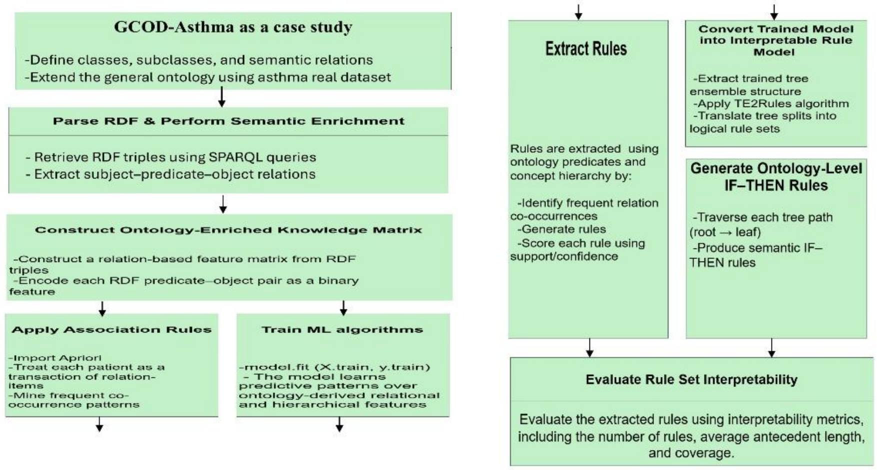

Figure 5.5 Extract rules method using AI algorithms with ontology to enhance interpretability; Asthma case study

Ontology integration for interpretability

In addition, ontology structure also improves interpretability in rule extraction.

Without ontology, extracted rules are expressed as combinations of isolated variables and numerical thresholds, which provide limited insight into clinical reasoning. For example, a rule such as:

IF wheezing = 1 AND smoking = 1 AND FEV1 < 80 THEN Diagnosis = 1

This rule does not explain how these variables are medically connected, and does not distinguish between symptoms, risk factors, or measurements.

In contrast, when ontology is used, the extracted rules explicitly rely on patient concept relations, which make the medical meaning of each condition clear.

Interpretability is improved because each condition is no longer a raw variable, but a clearly defined medical relationship as shown in figure 5.6.

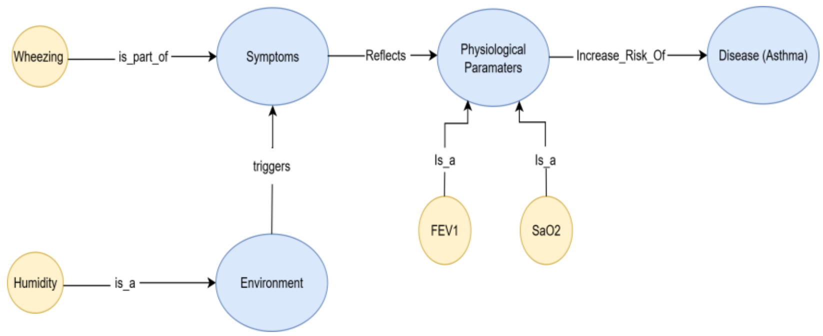

Figure 5.6 Ontology relations linking environmental factors, symptoms, and physiological parameters for asthma diagnosis

The rule explicitly states that the decision is supported by:

the presence of a symptom linked to the patient (rel:has_symptom), and a medical history relation (rel:has_medicalhistory linking the patient to eczema, a known comorbidity of asthma). In addition a lifestyle-related concept (smoking), and (iv) a categorized clinical measurement (FEV1).

These relations clarify why each piece of evidence matters and how it contributes to the diagnosis. Instead of reading the rule as a set of disconnected conditions, a clinician can interpret it as a reasoning chain: the patient presents respiratory symptoms, has a relevant allergic history, is exposed to a lifestyle risk factor, and shows lung function, which together support an asthma diagnosis. The ontology therefore transforms rule extraction from a purely description into an explicit, medically grounded explanation.

5.6 Comparison of rules extraction methods with and without Ontology based on Interpretability Metrics

In this section, we evaluate the rule extraction methods introduced in Section 5.3 by applying each method to two distinct data cases.

The first case uses the dataset processed through without ontology integration.

The second case uses the ontology-structured dataset, where the raw variables are enriched through ontology-guided feature aggregation.

Interpretability is considered subjective, and there is no universally definition or standardized measurement [422]. We adopt an interpretability-focused approach because it emphasizes medically meaningful explanations grounded in the relationships between inputs and outputs. This allows clinicians, who may not be data scientists, to understand how results are derived through well-defined medical concepts. Unlike explainability, which often relies on external approximations for black-box models, interpretability provides direct, readable justification based on input–output relationships, enabling trust and actionable insights in clinical decision-making. This is why in our work we adopt an interpretability.

Nevertheless, several studies have attempted to formalize and quantify interpretability from different perspectives. In this context, three interpretability criteria were proposed by Himabou et al. and Groshan Lal et al. [422, 423].

The evaluation focuses on three structural rule-description metrics: Number of rules, average antecedent length, and coverage [423]. These metrics describe the form and quantity of the extracted rules, not their clinical quality or correctness. Therefore, these metrics understood as syntactic indicators. Assessing the clinical validity and semantic quality of the rules would require expert evaluation and is left for future work.

Number of rules, we assess interpretability by examining the size of the extracted rule set. A smaller number of rules reduces the overall logical structure making the model easier to understand [423]. In our study, interpretability is evaluated through the size of the extracted rule set. The size simply refers to the total number of rules generated by the model. If the model produces many rules, a clinician must examine each one to understand how asthma is being predicted, which increases cognitive effort [422, 423]. In contrast, a smaller rule set reduces the amount of logical information that must be reviewed, making the overall decision process easier to understand and validate.

The average antecedent length measures the mean number of conditions appearing in the left-hand side of extracted rules [424]. It directly reflects the structure complexity of the rule. From an interpretability perspective, shorter rules are generally easier for humans to process because they reduce mental workload [424]. When ontology-based grouping replaces several related low-level features with a single high-level semantic concept, the predictive mechanism of the model does not change; however, the rule becomes shorter and more compact [424]. It is defined as:

$$\bar{L} = \frac{1}{R}\sum_{r=1}^{R} L_r$$

where:

R : total number of extracted rules

L_r : number of antecedent conditions (features) in rule r

Coverage, also referred to as representativeness, measures the proportion of data instances that are explained by at least one extracted rule [426]. A high coverage value indicates that the rule set captures a large portion of the dataset, producing more generalizable and informative insights [426]. Coverage is calculated as:

$$Coverage = \frac{\text{Number of instances covered by at least one rule}}{Total\ number\ of\ instances\ in\ the\ dataset}$$

In other words, coverage informs us what portion of the data is explained by your rules.

These three metrics together provide evaluation of interpretability.The Average Antecedent Length reflects the overall complexity of the rule set, indicating how many conditions are typically required to describe a decision [427]. The syntactic conciseness depends on the total number of rules and the average number of conditions (antecedents) per rule. Finally, the Coverage confirms that the extracted rules are sufficiently representative of the dataset, capturing a large portion of the observed cases [426, 427].

Overall, these metrics collectively capture the essential dimensions of simplicity, and representativeness, which are fundamental for reliable and explainable clinical rule extraction [426]. I have applied these scenarios to evaluate ontology integration with these models and to interpret their impact. Apriori is a learning data-mining method used to discover relationships among patterns. Random Forest and XGBoost (with RuleFit and TE2Rules, respectively) are used because they enable extraction of interpretable rules, and they are supported by Rigatti (2017) for RF and Chen & Guestrin (2016) for XGBoost as efficient models for prediction. We use RuleFit and TE2Rules to extract rules from the Random Forest and XGBoost models. Our main objective is to improve interpretability while maintaining accuracy, which is why we use these models across our three scenarios.

5.6.1 Scenario 1: Direct Rule Learners; Association Rule Mining (Apriori)

Association rule mining is the process of identifying rules that describe the relationships among items in a specific database [375]. This concept was first introduced in [375] and was initially applied to address the shopping basket problem. To understand more let I = {i_1, i_2,..., i_m} represent a set of (m) items, and let D = {t_1, t_2,..., t_n} denote a database containing n transaction. Each transaction is an itemset or a subset of (I). The support for an itemset (S) is defined in Equation (1) [428].

Sup (s) = $\frac{Count\ of\ transaction\ in\ D\ that\ contain\ S}{n}$

If $X \cap Y = \emptyset$. X is referred to as the antecedent of the rule, while Y is called the consequent of the rule. The support and confidence of the association rule (r) are two important measures that can be represented as $Sup(r)$ and $Conf(r)$, respectively. These measures are defined in Equations (2) and (3) [428]:

$$Sup(r) = \frac{Count\ of\ transactions\ \ D\ that\ contain(X \cup Y)}{n} = p(X \cup Y)$$

$$Conf(r) = \frac{Sup(X \cup Y)}{Sup(X)} = p(X|Y)$$

The support of an association rule indicates its statistical significance, while the confidence measures the strength of that rule. Another important metric for assessing association rules is called "lift," which is defined in Equation (4) [428]:

$$Lift(r) = \frac{Conf(X \rightarrow Y)}{Sup(Y)} = \frac{p(X \cup Y)}{p(X)p(Y)}$$

If the lift value is 1, then X and Y are independent of each other. While a lift value greater than 1 indicates a relationship between X and Y. An association rule is considered interesting if both its support and confidence exceed user-defined thresholds, referred to as Sup_min and Conf_min, respectively [427, 428]. Therefore, the goal of association rule mining is to discover such interesting rules.

Apriori is used in both cases: with ontology and without ontology during the rule generation stage. This includes analyzing the number of generated rules based on different values of Sup_min and Conf_min, as well as evaluating computation time and memory consumption. The results of this experiment are presented in Table 5.3 and are visualized in Figures 5.7, 5.8 and 5.9.

Table 5.3 The analysis of Apriori algorithm with and without ontology

Sup_min	Conf_min	No. of Generated Rules (Dataset Apriori)	No. of Generated Rules (Ontology Apriori)	Computation Time in seconds (Dataset)	Computation Time in seconds (Ontology)	Memory Consumption in MB (Dataset)	Memory Consumption in MB (Ontology)
0.1	0.1	3240	1085	5.3	6.1	1.23	1.65
0.2	0.7	520	190	0.18	0.31	0.7	0.92
0.3	0.3	553	223	0.06	0.2	0.4	0.8
0.3	0.9	328	147	0.06	0.18	0.4	0.8
0.6	0.5	28	19	0.03	0.08	0.3	0.6

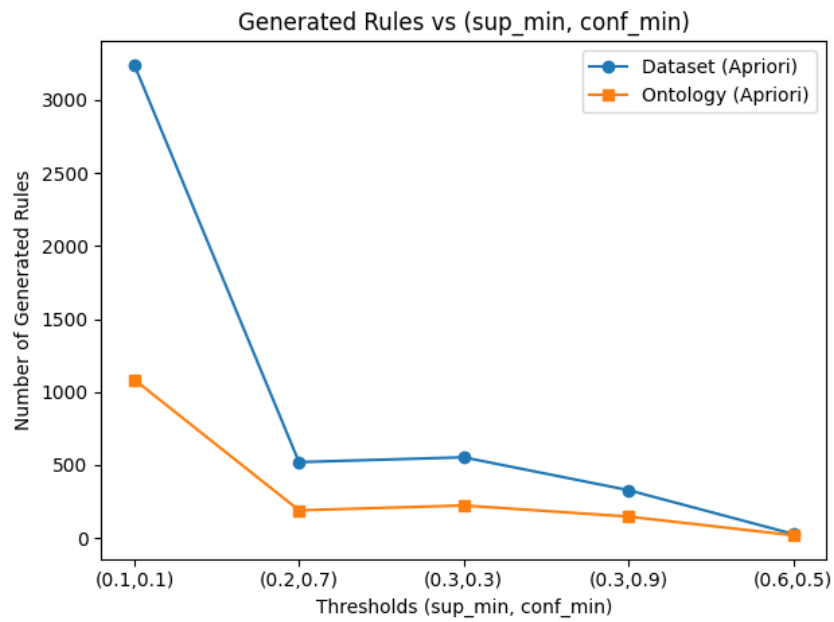

Figure 5.7 The number of generated rules of each case using different values of Sup_min and $Conf\_min$

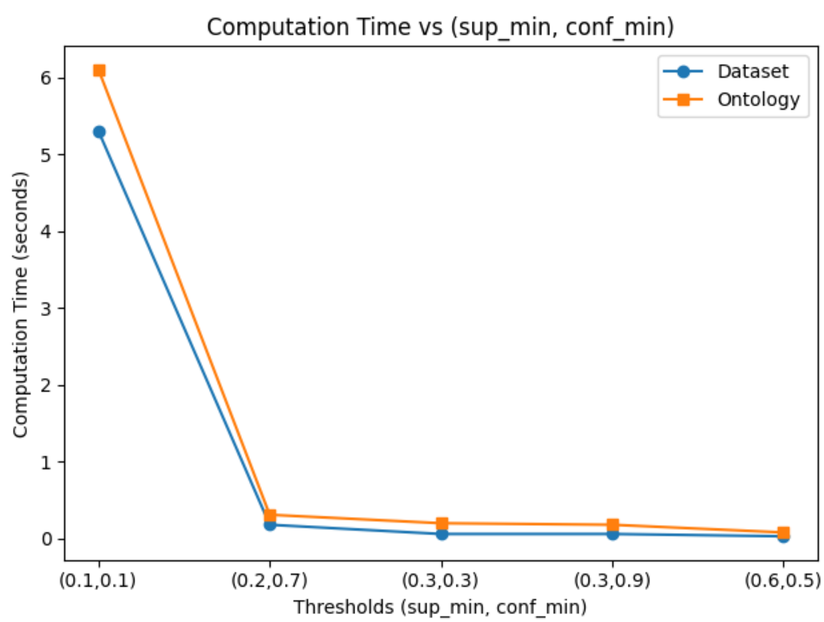

Figure 5.8 The computation time of each case using different values of Sup_min and $Conf\_min$

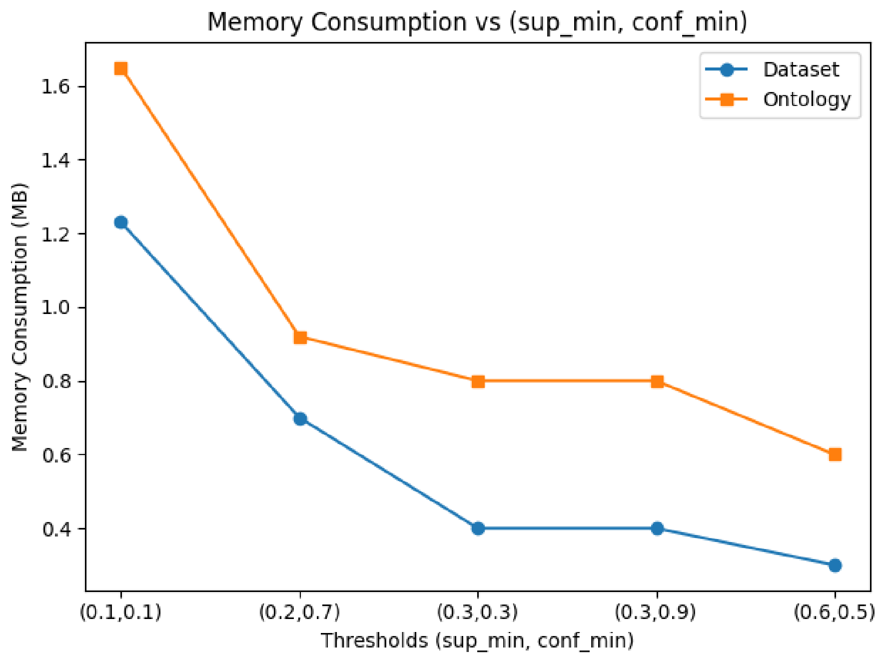

Figure 5.9 The memory consumption of each case using different values of Sup_min and $Conf\_min$

Based on Table 5.3 and Figures (5.7, 5.8 and 5.9), we can observe that increasing the support and confidence values reduce the number of extracted rules for both algorithms. This is a rational finding, as many rules can meet the lower support and confidence values, while only a smaller group can satisfy the higher ones. Additionally, the semantic Apriori algorithm generates a smaller set of rules compared to the Apriori algorithm. Furthermore, the proposed semantic Apriori algorithm outperforms the Apriori algorithm in terms of computation time and memory consumption.

Apriori generates association rules in the form X → Y, where X is a set of conditions (itemset) and Y is the outcome [427].

When Apriori is applied directly to the feature-selected dataset, the resulting rules are low-level and feature-specific, composed of raw clinical, environmental, and lifestyle variables. These rules explicitly combine individual symptoms, measurements, and binary indicators, such as:

R1:

Wheezing=Yes ^ ShortnessOfBreath=Yes ^ ChestTightness=Yes ^ Smoking=Yes ^ Eczema=Yes ^ FEV1=Low ^ Pollution=High → Asthma

R1’:

has_Symptoms SOME(SevereRespiratorySymptoms) ^ has_LifestyleFactor: Smoking^ suffers_From:Eczema^ has_ClinicalMeasurement: LowFEV1^ is_ExposedTo : HighPollutionExposure→ Asthma

R2:

Cough=Yes ^ NightSymptoms=Yes ^ ExerciseInducedSymptoms=Yes ^ Eczema=Yes ^ PollenExposure=High ^ DustExposure=High ^ FVC=Low → Asthma

R2’:

has_Symptoms SOME(TypicalRespiratorySymptoms) ^ has_MedicalHistory: EczemaAllergy^ is_ExposedTo; HighPollenExposure ^ is_ExposedTo: HighDustExposure ^ has_ClinicalMeasurement : LowFVC→ Asthma

R3:

Wheezing=No ^ Cough=No ^ ShortnessOfBreath=No ^ Smoking=No ^ FEV1=High ^ Pollution=Low ^ DustExposure=Low → No_Asthma

R3’: has_Symptom:NoSevereRespiratorySymptoms ^ Smoking=No ^ has_ClinicalMeasurement:HighFEV1 ^ is_ExposedTo:LowPollutionExposure ^ is_ExposedTo:LowDustExposure → No_Asthma

In the rules, (R1, R2 and R3) each item represents a specific measurement or binary variable, making the rules tightly coupled to the dataset and less generalizable.

After ontology-based enrichment, the structure of Apriori rules (X → Y) remains unchanged; however, the items are lifted to higher-level semantic concepts using ontology relations such as suffer_from, is exposedto. This transformation groups related features into meaningful medical concepts, producing more generalized and interpretable rules:

In the ontology-based Apriori rules (R1’, R2’, and R3’), individual symptoms such as wheezing, coughing, and chest tightness are semantically abstracted under the class SevereRespiratorySymptoms, and represented in the rules using the existential restriction has_Symptoms some SevereRespiratorySymptoms (at least one severe respiratory symptom is present). The term "some" serves as a quantifier, providing extra information while keeping the overall guidelines clear and not too detailed. In adition smoking and physical activity are represented as Life_Style_Factor; eczema and allergic history are unified under History_of_Allergy; pollution and allergen exposure are modeled as Environmental_Factor; and FEV1, FVC values are captured by ClinicalMeasurement:Reduced_Lung_Function.

As a result, ontology-based Apriori generates shorter, more general, and clinically meaningful rules, while preserving the fundamental Apriori mechanism. The use of semantic concepts improves interpretability without altering the core association rule mining process.

For Apriori algorithm, the results indicate that interpretability is improved because of ontology integration, not because fewer rules are generated. The ontology-aware Apriori requires slightly more computation time and memory than the baseline version, which is expected due to the use of high-level concepts and semantic relations. However, this overhead remains limited and stable across all threshold settings. From a rule structure perspective, ontology-aware Apriori produces rules with shorter antecedents and, since multiple related low-level attributes are combined into meaningful medical concepts defined by the ontology. Coverage stays comparable to the baseline in most cases and is even higher under stricter confidence thresholds, showing that semantic abstraction does not reduce the representativeness of the rules. Overall, these results confirm that, the interpretability is enhanced through concept abstraction and explicit ontology relation as shown in table 5.4.

Table 5.4 Interpretability association rule mining comparison between baseline and ontology-aware

Sup_min	Conf_min	Apriori	#Rules (R)	Avg antecedent length	Coverage
0.1	0.1	Baseline (no Ontology)	3240	4.22	1.0
		Ontology-aware	1085	3.5	1.0
0.2	0.7	Baseline (no Ontology)	520	3.8	1.0
		Ontology-aware	190	3.6	1.0
0.3	0.3	Baseline (no Ontology)	553	3.6	1.0
		Ontology-aware	223	3.2	1.0
0.3	0.9	Baseline (no Ontology)	328	3.3	0.83
		Ontology-aware	147	3.1	0.92
0.6	0.5	Baseline (no Ontology)	28	2.7	0.28
		Ontology-aware	19	2.1	0.32

5.6.2 Scenario 2 Direct Rule Learners; Machine learning

A Random Forest is an ensemble of decision trees that makes predictions by aggregating the outputs of multiple trees [429]. Although the model itself does not explicitly generate rules, each root-to-leaf path within an individual tree can be extracted and rewritten as a decision rule of the form "IF … THEN Diagnosis" [429]. The complexity and the number of the extracted rules are directly influenced by two key hyperparameters: maximum tree depth (max_depth) and number of trees (n_estimators). The max_depth parameter controls how many successive feature-based splits are allowed within each tree; deeper trees generate a larger number of root-to-leaf paths and therefore more rules, whereas shallower trees produce fewer and more general rules [429]. The n_estimators parameter determines how many trees are constructed in the ensemble, and since rule extraction aggregates rules across all trees, increasing this parameter leads to an approximately linear increase in the total number of generated rules [430].

Adding ontology-based features maps multiple fine-grained clinical codes and observations into higher-level semantic super classes, enabling the trees to split on clearer and more semantically meaningful signals rather than fragmented features [428]. In our experiments, the baseline dataset resulted in the Random Forest model generating a slightly larger number of rules compared to the ontology-structured dataset across various configurations. We specifically evaluated the number of rules under different values for max_depth (2, 8, 10) and two values for n_estimators (400 and 600) as shown in table 5.5. This illustrates how rule increases with both tree depth and ensemble size. Additionally, Figures 5.10 - 5.12 analyze the number of rules, computation time, and memory consumption for these scenarios.

Table 5.5 The analysis of Random Forest algorithm with and without ontology

n-estimator	max-depth	No. of Generated Rules (Dataset)	No. of Generated Rules (Dataset with ontology)	Compu-tation Time in seconds (Dataset)	Compu-tation Time in seconds (Ontology)	Memory Consum-ption in MB (Dataset)	Memory Consum-ption in MB (Ontology)
400	2	1536	1135	0.95	2.6	8.63	18.82
	8	13760	11499	1.13	2.7	7.24	9.04
	10	16650	13042	0.95	6.21	6.71	4.03
600	2	2309	1672	0.95	34.9	13.52	10
	8	6873	5748	1.31	32.7	14.32	17.7
	10	24994	19621	1.42	39.7	15.9	19.5

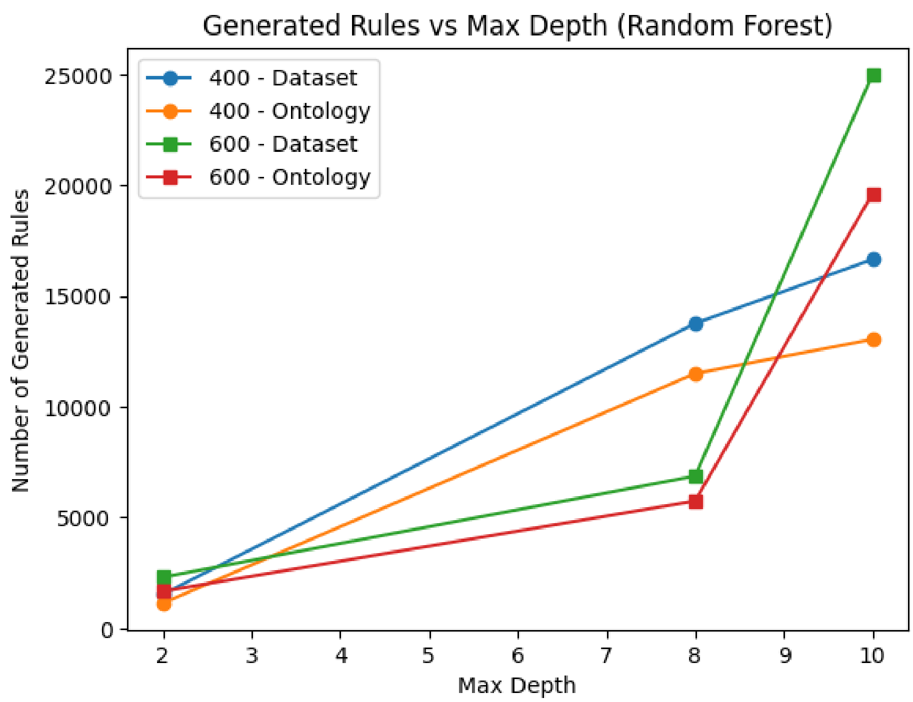

Figure 5.10 The number of generated rules for RF of each case using different values of max_depth with n_estimators.

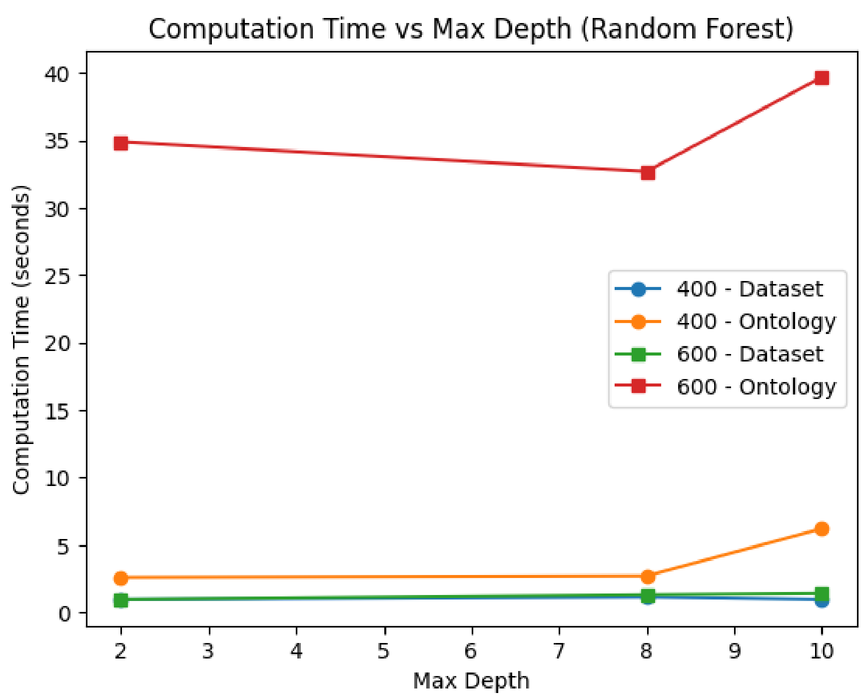

Figure 5.11 The computation time for RF of each case using different values of max_depth with n_estimators

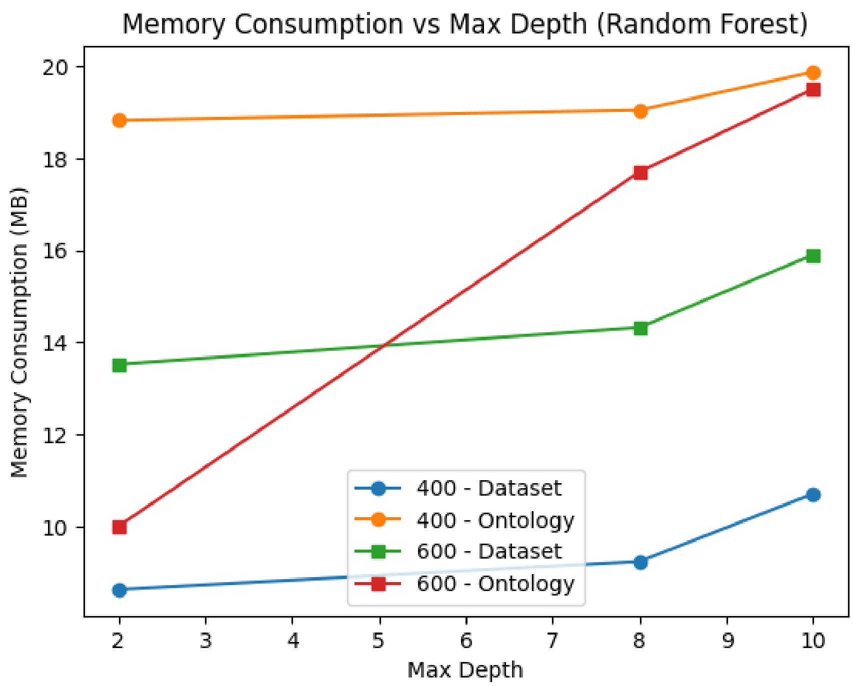

Figure 5.12 The memory consumption of each case using different values of max_depth with n_estimators

For Random Forest, the number of extracted rules mainly depends on the maximum tree depth and the number of trees. Increasing max_depth allows each tree to continue splitting until purer leaves are reached, which increases the number of leaves and therefore the number of extracted rules. Similarly, increasing n_estimators adds more trees to the forest, leading to a larger overall rule set. This increase is reflected in higher computation time and memory usage, since deeper trees and larger forests require more resources. When the ontology-based representation is used, the same general behavior with respect to depth and number of estimators remains. However, grouping raw variables into high-level clinical concepts reduces redundant splits on closely related features, which helps control the growth of the rule set while keeping time and memory consumption comparable. In the following some examples are provided to show how rules extracted from the dataset without ontology differ from those extracted with ontology, highlighting the impact of semantic representation on rule structure.

R1: IF Wheezing=1 AND Shortness_of_breath=1 AND Chest_Tightness=1 AND Smoking=1 AND Eczema=1 AND FEV1 ≤ 2.09 AND Pollutionexposure > 2.47 THEN Asthma

R1’: IF has_Symptoms SOME(SevereRespiratorySymptoms) AND has_LifestyleFactor:Smoking AND suffers_From:Eczema AND has_ClinicalMeasurement:LowFEV1 AND is_ExposedTo:HighPollutionExposure THEN Asthma

R2: IF Coughing=1 AND Night_time_symptoms=1 AND Exercise_Induced=1 AND HistoryOfAllergy=1 AND Pollenexposure > 7.28 AND Dustexposure > 2.09 AND FVC ≤ 4.84 THEN Asthma

R2’: IF has_Symptoms SOME(TypicalRespiratorySymptoms) AND has_MedicalHistory:EczemaAllergy AND is_ExposedTo:HighPollenExposure AND is_ExposedTo:HighDustExposure AND has_ClinicalMeasurement:ReducedFVC THEN Asthma

The number of extracted rules increases as the maximum tree depth grows and as the number of estimators increases, since deeper trees produce more leaves and a larger forest contributes rules from more trees as shown before. When the ontology-based representation is used, the extracted rules are expressed through explicit relations such as has_Symptom, has_MedicalHistory, is_ExposedTo, and has_ClinicalMeasurement, and rely on high-level clinical concepts that group related raw features. We added the quantifier some to include more information without making the rule too specific and to keep its high-level meaning.

This semantic grouping reduces repeated splits on closely related variables across trees, which results in a lower total number of extracted rules for all tested configurations. The same effect is observed in the structural metrics reported in the table: ontology-aware rules show slightly shorter average antecedent lengths, while coverage remains equal to 1.0 as shown in table 5.6. Overall, the table shows that ontology integration mainly influences how Random Forest rules are formed through relations and high-level concepts, which impacts the number of rules and associated metrics without altering the general behavior of the model.

Table 5.6 Interpretability random forest comparison between baseline (without ontology) and ontology-aware models

N-estimator	Max_depth	Random Forest	#Rules (R)	Avg antecedent length	Coverage
600	2	Baseline (no Ontology)	2309	1.96	0.97
		Ontology-aware	1672	1.75	0.97
600	8	Baseline (no Ontology)	6873	6.22	1.0
		Ontology-aware	5748	5.9	1.0
600	10	Baseline (no Ontology)	24994	6.94	1.0
		Ontology-aware	19621	6.53	1.0
400	2	Baseline (no Ontology)	1536	1.95	0.954
		Ontology-aware	1135	1.75	0.986
400	8	Baseline (no Ontology)	13760	6.22	1.0
		Ontology-aware	11499	5.98	1.0
400	10	Baseline (no Ontology)	16650	6.35	1.0
		Ontology-aware	13042	6.22	1.0

5.6.3 Scenario 3: ML based models

XGBoost is a tree-ensemble boosting method that is often used as an accurate predictor for classification tasks [102]. To obtain human-readable rules from an XGBoost model, we use TE2Rules, which converts a trained tree ensemble (including XGBoost) into a rule list that summarizes the model’s decision behavior [103].

In practice, TE2Rules produces rules that can be written as conditions leading to an output class, providing an interpretable rule representation of the ensemble [103, 104]. To include domain semantics, we build an ontology-aware representation where low-level findings (symptoms, exposures, and measurements) are mapped into higher-level clinical concepts and relations (has_Symptom, is_ExposedTo, has_ClinicalMeasurement), while the extracted rules can be expressed at a more meaningful clinical level.

This semantic feature grouping can change the structure of the extracted rule list (for example, fewer repeated low-level conditions), which is then evaluated in our experiments using the number of rules and the resource metrics (time and memory) alongside the interpretability metrics reported in Table 5.7 and Figures 5.12 – 5.14 We evaluate the resulting model shown in table below using the interpretability metrics [429].

Table 5.7 The analysis of XG-Boost and TE2Rules algorithms with and without ontology

n-estimator	max-depth	No. of Generated Rules (Dataset)	No. of Generated Rules (Dataset with ontology)	Computation Time in seconds (Dataset)	Computation Time in seconds (Ontology)	Memory Consumption in MB (Dataset)	Memory Consumption in MB (Ontology)
400	2	1467	1236	0.49	29.78	1.36	87.81
	8	8074	5934	0.58	31.54	3.88	92.2
	10	8978	6097	0.73	42.29	3.89	78.3
600	2	2317	2014	0.33	52.26	4.28	81.72
	8	11303	7135	0.78	64.42	4.85	84.35
	10	12205	7281	2.28	71.06	5.26	89.7

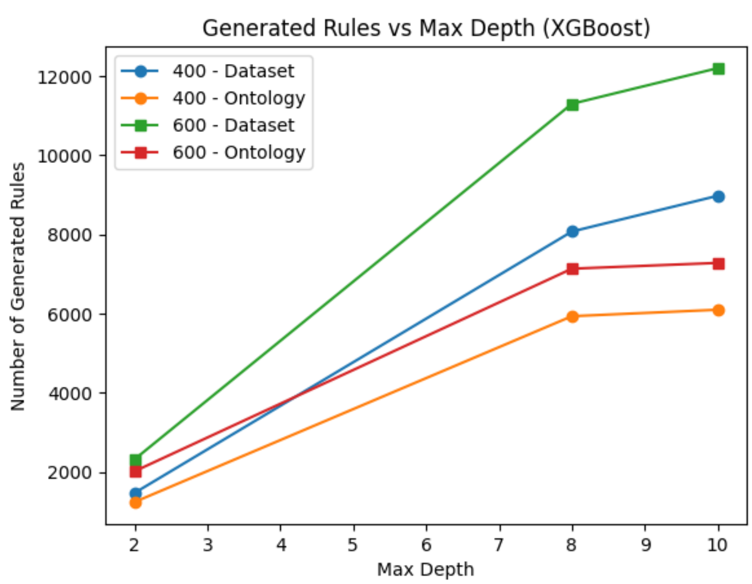

Figure 5.13 The number of generated rules for XG-Boost of each case using different values of max_depth with n_estimators

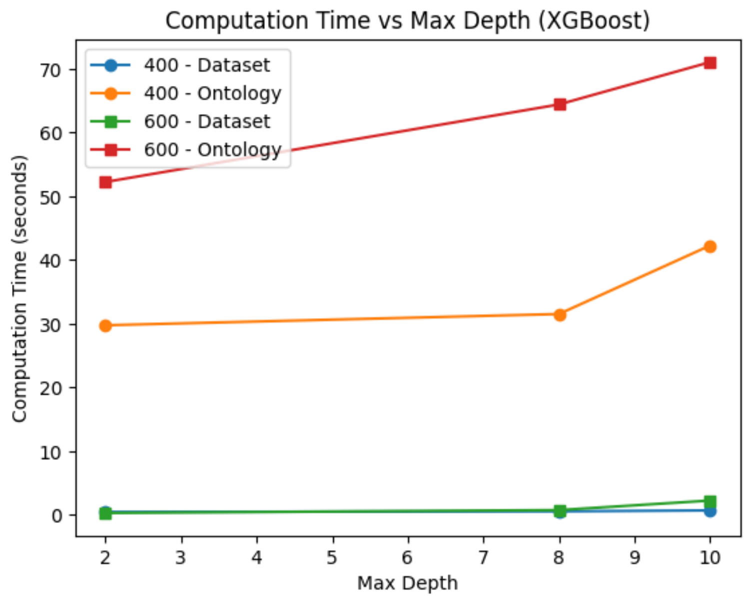

Figure 5.14 The computation time for XG-Boost of each case using different values of max_depth with n_estimators

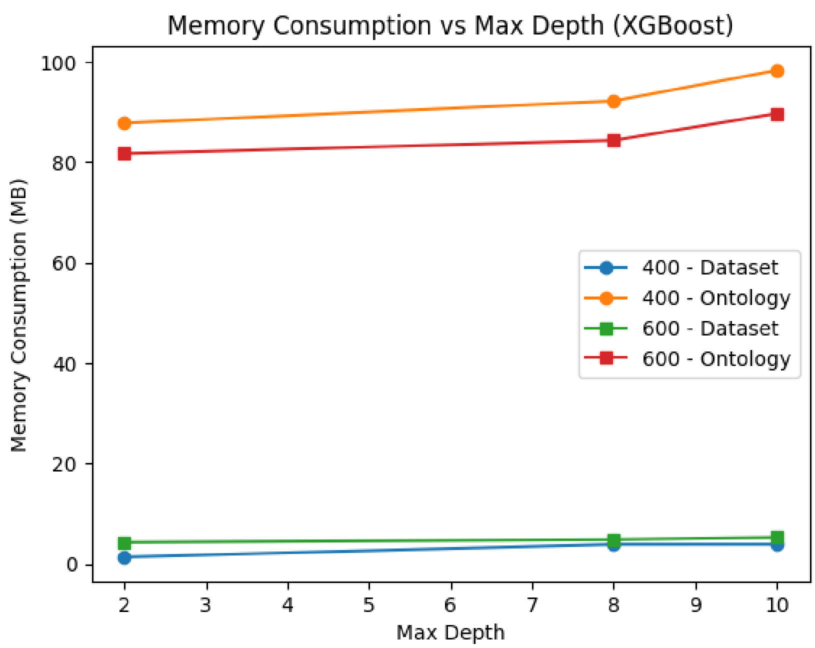

Figure 5.15 The memory consumption for XG-Boost of each case using different values of max_depth with n_estimators

For XGBoost algorithm with TE2RULES, table 5.7 and figure 5.13 - 5.15 show that the number of generated rules increases when max_depth increases for both 400 and 600 estimators, since deeper trees create more split paths that can be extracted as rules. Increasing n_estimators from 400 to 600 also increases the rule count because rules are collected from a larger number of boosted trees. The ontology-based representation produces fewer rules than the baseline (400–8: 8074 vs 5934, 600–10: 12205 vs 7281), showing that the semantic grouping reduces rule expansion. In terms of resources, the baseline remains low in computation time and memory, while the ontology setting shows higher computation time and memory consumption across all settings, reflecting the added overhead of building. In the next step, I will provide examples of how rules are extracted from XGBoost with TE2RULES both without ontology and with ontology (rules expressed using high-level concepts and relations).

R1:

Wheezing = 1 ^ Shortness_of_breath = 1 ^ Chest_Tightness = 1 ^ Smoking = 1 ^ Eczema = 1 ^ FEV1 ≤ 2.09 ^ Pollutionexposure > 2.47 → Asthma

R1':

has_Symptom SOME(SevereRespiratorySymptoms) ^ has_LifestyleFactor:Smoking ^ suffers_From:Eczema ^ has_ClinicalMeasurements:LowFEV1 ^ is_ExposedTo:HighPollutionExposure → Asthma

R2:

Coughing = 1 ^ Night_time_symptoms = 1 ^ Exercise_Induced = 1 ^ HistoryOfAllergy = 1 ^ Pollenexposure > 7.28 ^ Dustexposure > 2.09 ^ FVC ≤ 4.84 → Asthma

R2': has_Symptom SOME(TypicalRespiratorySymptoms) ^ has_MedicalHistory:EczemaAllergy ^ is_ExposedTo:HighPollenExposure ^ is_ExposedTo:HighDustExposure ^ has_ClinicalMeasurements:ReducedFVC → Asthma

In addition, table 5.8 shows that increasing the number of estimators and the tree depth leads to more extracted rules, since larger models generate more decision paths. When the ontology-based representation is used, the rules (R1’ and R2’) rely on high-level clinical concepts and explicit relations such as symptoms, exposures, medical history, and clinical measurements instead of many raw variables and numeric thresholds. For example, in expressions such as has_Symptom some TypicalRespiratorySymptoms, the quantifier some indicates that the patient has at least one symptom from the class TypicalRespiratorySymptoms without making the rule too specific.

This results in fewer rules with shorter antecedents across all configurations, because related features are grouped into meaningful concepts, while coverage remains equal to 1.0. In this sense, the ontology makes the rules more realistic and easier to relate to clinical reasoning, not by changing the model behavior, but by expressing decisions through medically meaningful relations and concepts.

Table 5.8 Interpretability XG-Boost comparison between baseline (without ontology) and ontology-aware models

N-estimator	Max_depth	XG-Boost with TE2RULES	#Rules (R)	Avg antecedent length	Coverage
600	2	Baseline (no Ontology)	2317	4.92	1.0
		Ontology-aware	2014	4.35	1.0
600	8	Baseline (no Ontology)	11303	6.12	1.0
		Ontology-aware	7135	5.26	1.0
600	10	Baseline (no Ontology)	12205	8.3	1.0
		Ontology-aware	7281	7.53	1.0
400	2	Baseline (no Ontology)	1467	4.3	1.0
		Ontology-aware	1236	3.2	1.0
400	8	Baseline (no Ontology)	8074	5.19	1.0
		Ontology-aware	5934	3.98	1.0
400	10	Baseline (no Ontology)	8978	7.34	1.0

		Ontology-aware	6097	5.43	1.0

5.7 Discussion: Impact of context and ontology integration on Interpretability

Integrating ontology-based medical knowledge led to measurable changes in the syntactic properties of the rules generated by the different rule-extraction methods (Apriori, Random Forest and XG-Boost with TE2RULES).

5.7.1 Interpretability Metrics

Ontology integration also improves interpretability, as measured by number of rules, average antecedent length (AAL), and coverage.

In the Apriori, ontology reduces the number of extracted rules. For instance, the rule count decreases from 3240 to 1085 at (0.1/0.1) and from 328 to 147 at (0.3/0.9). At the same time, the average antecedent length (AAL) also decreases (from 4.2 to 3.5 and from 2.7 to 2.1), indicating simpler rule structures. Importantly, coverage remains comparable across most thresholds and even improves in stricter settings (0.83 to 0.92).

For Random Forest, ontology reduces rule complexity. For example, at (600, 8), the average antecedent length decreases from 6.22 to 5.9, and at (400, 10) from 6.35 to 6.22. In addition, the number of extracted rules is systematically lower under the ontology-aware setting (6873 to 5748 at (600, 8), and 16650 to 13042 at (400, 10)). Importantly, coverage remains equal to 1.0 across these configurations.

For XGBoost with TE2Rules, a similar pattern is observed. the average antecedent length decreases at (600, 8) from 6.12 to 5.26, at (400, 8) from 5.19 to 3.98, and at (400, 10) from 7.34 to 5.43. In addition, the number of extracted rules is lower under the ontology representation (11303 to 7135 at (600, 8), and 8978 to 6097 at (400, 10)). Coverage remains equal to 1.0 across all configurations, indicating that simplification does not reduce representativeness. These results confirm that ontology reduces both rule depth and rule quantity while preserving explanatory coverage.

Overall, across Apriori, Random Forest, and XGBoost, ontology-aware representations consistently reduce number of rules, antecedent length without affecting coverage, confirming that semantic feature grouping leads to more structured and compact rule expressions.

5.7.2 Global Interpretation

Across Apriori, Random Forest, and XGBoost, the integration of ontology consistently affects the interpretability metrics by restructuring how rules are expressed. By introducing explicit semantic relations (such as has_Symptom, is_ExposedTo, and has_ClinicalMeasurement) and grouping detailed variables into high-level clinical concepts, ontology-based representations reduce fragmentation in rule conditions [106]. In ontology modeling, some is an existential quantifier used to indicate that an individual is related to at least one instance of a specified class (has_Symptoms some TypicalRespiratorySymptoms , some severeRespiratorySymptoms) means the patient exhibits at least one typical respiratory symptom). In this work, the term some quantifier was intentionally introduced as a semantic operator to enrich the rules with additional contextual information while preserving high-level generalization and avoiding overly specific rule definitions.

As reported in table 5.9, semantic grouping reduces the number of rules and decrease the average antecedent length across three methods not because the models change their predictive logic, but because multiple related raw features are replaced by unified semantic concepts [427]. Importantly, coverage remains stable in almost all configurations, indicating that the ontology does not reduce the representativeness of the extracted rule sets. Overall, the impact of ontology on interpretability lies in organizing features into clinically meaningful structures, which influences the form and compactness of rules while preserving their applicability [428].

Table 5.9 Comparative Summary of the Interpretability Model Improvements with Ontology Integration

Metric	Without Ontology	With Ontology	Relative Improvement / Observation
Number of rules	More rules	↓ 28%	Fewer rules — Simpler structure
Average Antecedent Length	Longer rules	↓ 14.8%	More compact rules
Coverage	Lower coverage	↑ 1.6%	More representative — Rules cover broader patient cases

Figure 5.16 demonstrates that ontology improves rule interpretability by grouping related low-level attributes into higher-level clinical concepts connected through explicit semantic relations. Instead of five independent conditions, the ontology-based rule expresses three meaningful concepts linked to asthma, making the reasoning clearer, more compact, and clinically understandable.

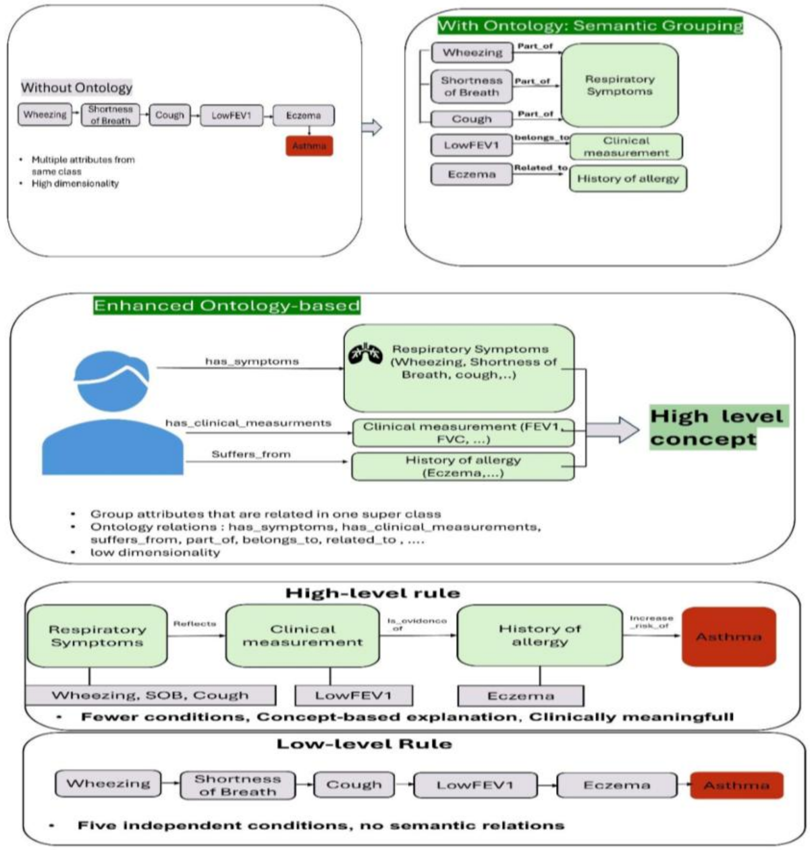

Figure 5.16 Ontology-Driven Transformation of Low-Level Rules into High-Level Interpretable Concepts

5.8 Conclusion and future works

Healthcare systems are becoming complex, and decision-making now relies on understanding a wide range of factors related to a patient’s condition and timely clinical decisions. Many solutions are proposed, such as ontology and rules extraction methods can help to make complex medical information accurate, understandable and useful for decision-making. Our semantic-enhanced learning model provides significant potential to address current medical domain challenges, such as context categorization, context representation, and the generation of structured, model-compatible information. This methodology offers a conceptual tool for the telemonitoring of asthma patients.

The results support the idea that using this method helps AI models handle medical data more effectively. By organizing related symptoms, environmental triggers, and clinical indicators into coherent categories, the model shows modest improvements by producing rule sets that were more shorter antecedents and higher coverage [427, 428]. Especially, the ontology-based model helps to identify many inferences (rules) that are practically invisible to human reasoning. The findings indicate that combining ontology-guided structuring with rule-based can enhance existing healthcare systems and support the assessment of asthma patient status.

This study achieved its main goal of supporting physicians and nurses in making faster and more informed decisions through an ontology-guided learning methodology that provides rule-based explanations. The model met its subgoals by (G1) comparing different rule extraction methods using many evaluation metrics, and (G2) integrating a medical ontology to analyze how domain-based feature grouping affects rule structure in the healthcare setting.

Overall, ontology integration can encourage algorithms to rely on more clinically coherent feature groupings and yielding rule sets that are more consistent and structurally concise. This supports greater transparency and reliability in healthcare AI systems [429].

Our future work will extend this methodology to additional diseases, such as cardiovascular conditions and diabetes. Furthermore, we will focus on validating the extracted rules with clinical experts to ensure diagnostic relevance [430]. These efforts aim to strengthen the clinical validity and transferability of ontology-guided AI, contributing to trustworthy and explainable medical decision-support systems [431].

CHAPTER 6
CONCLUSION

This research work addressed the challenge of designing reliable, interpretable, and context-aware intelligent system for healthcare, with a particular focus on chronic diseases such as asthma. As discussed earlier, the main challenges concern the definition and comprehension of medical context, the representation of this context within a comprehensive model, and the reliability of context-based reasoning to meet healthcare requirements such as interpretability.

The thesis demonstrates that effective medical decision support systems must provide clear and interpretable explanations of clinical decisions in healthcare settings.

6.1 Thesis Contributions

This thesis addressed these challenges through five core contributions:

A first contribution of this thesis presented in chapter 2, is the definition and categorization of medical context into fifteen distinct categories. This categorization provides a comprehensive and structured view of patient-related information, covering medical, environmental, temporal, social, and technological aspects. By organizing data into meaningful categories, the proposed model reduces ambiguity, supports completeness, and facilitates the integration of real-world patient context into healthcare systems.

The second contribution of this thesis, presented in Chapter 3, consists of a survey of Knowledge Representation Models (KRMs). This chapter provides a thorough analysis of KRMs and their crucial role in information management within the healthcare domain. We classify four knowledge representation models based on their type, such as graphical, mathematical and other types.The chapter offers an overview of how these approaches can be used to effectively represent contextual knowledge. It concludes that ontology is particularly effective in certain aspects, especially for modeling real-world scenarios.

Chapter 4 presents the third contribution of this thesis, which consists of designing the general contextual ontology based on the Sanchez method, then extending it to asthma as a case study. The general contextual ontology provides a semantic structure that captures medical concepts and their relationships in a consistent manner. The proposed ontology primarily serves as a semantic backbone for structuring and organizing data. This design choice ensures coherence while remaining flexible and extensible to other chronic diseases .

Chapter 5 discusses the fourth contribution, which consists of surveying rule extraction methods. We classify these methods into four categories including expert rules, direct rule-learning models (Apriori, Decision Tree), ML based models approaches (XGBoost with TE2Rules), and Large Language Models (LLMs).

The fifth contribution of this thesis, described in chapter 5, is the design of a semantic-enhanced learning architectural model that integrates ontology-guided feature structuring. Moreover, four rule extraction methods are applied to generate interpretable rules. The study evaluates the extracted rules and demonstrates that ontology-based structuring improves the quantity interpretability metrics of the rules.

Overall, this thesis contributes a unified and principle semantic enhanced arcitectural model that integrates context modeling, semantic representation, and interpretability in healthcare. This work advances the development of context-aware, and interpretable intelligent healthcare systems, with extension to other medical domains and real-world clinical systems.

6.2 Discussion

The overall objective of this thesis was to provide concrete contributions to the development of reliable, interpretable context-aware healthcare decision-support systems. From a scientific perspective, this research aimed to investigate how contextual knowledge representation and ontology-guided structuring can improve the interpretability of models in healthcare systems. The results demonstrate that traditional data-driven approaches alone are not sufficient to manage the complexity of the medical context. Instead, combining contextual modeling, ontology-based knowledge representation, and interpretable rule-extraction methods provides a more effective model for supporting clinical decision making.

A key contribution to this research is the definition of medical context using fifteen contextual categories to better represent patient information. Compared with existing models, the proposed model introduces additional categories such as social factors, service-level agreements (SLA), and medical data, and refines existing categories by splitting some into more specific ones and merging others when they represent related contextual elements. Then the model was applied to three chronic disease case studies asthma, COPD, and cardiovascular disease, showing its ability to represent context across different medical conditions. In addition, a survey and classification of knowledge representation models highlighted the advantages of ontology-based approaches for representing complex contextual knowledge and supporting semantic enrichment in healthcare systems.

Building on these findings, another contribution of this research is the design of a general contextual ontology based on the fifteen contextual categories defined earlier, which can be used to represent contextual information in healthcare systems. This ontology organizes clinical concepts and their relationships in a structured and extensible way, allowing the integration of heterogeneous medical data. Following the Sánchez methodology, the general ontology model was then extended to an asthma-specific ontology using a real asthma dataset. By incorporating real-world data, the extended asthma ontology captures key elements such as symptoms, triggers, and risk factors, demonstrating how the proposed contextual structure can support the representation of chronic disease information in healthcare system.

Another important contributions of this work are the categorization and evaluation of rule-extraction methods. Four categories of methods were distinguished: expert rules, direct rule-learning models, ML based models, and Large Language Models. These methods were evaluated under two experimental conditions using raw data and ontology-enhanced data to analyze how semantic enrichment influences the interpretability of extracted rules.

The improvement in interpretability can be explained by the role of ontology in structuring contextual medical information. In this work, a rule extraction approach was designed to integrate ontology with rule extraction methods, allowing the transformation of low-level variables into higher-level clinical concepts. For example, instead of working with individual variables such as cough, wheezing, and chest tightness, the approach groups these variables into higher-level concepts, such as the class of respiratory symptoms. Quantifiers are then introduced to express relationships between concepts, providing additional semantic information while maintaining clear and concise rule structures.

By grouping related features under unified semantic concepts, ontology reduces fragmentation in rule conditions and allows several detailed variables to be represented through a single meaningful concept. As a result, multiple rules can be expressed through fewer generalized rules, which explain the 28% reduction in the number of generated rules. In addition, replacing multiple low-level features with higher-level semantic concepts shortens the rule conditions, leading to a 14.8% decrease in the average antecedent length. At the same time, the ontology maintains the representativeness of the extracted rules, as reflected by the 1.6% increase in coverage, indicating that the more compact rules still apply to a broad set of patient cases. Overall, ontology-guided structuring improves interpretability by transforming fragmented variables into clinically meaningful concepts.

Overall, this research demonstrates that our system based on contextual modeling, ontology-based knowledge representation, and interpretable rule extraction methods provides a promising approach for improving the reliability and transparency of intelligent healthcare systems. Such approaches contribute to the development of future healthcare systems capable of supporting clinicians with meaningful, explainable insights derived from complex medical data.

6.2.1 Generalization to additional chronic diseases

The context-aware ontology-guided healthcare model architecture is designed with five integrated layers (data acquisition, context categorization, context representation, operation, and application), while adding a new chronic-disease domain naturally extends across all layers. The context categorization layer already uses a fifteen-category framework that was validated for asthma, COPD, and cardiovascular disease; for a new disease you would identify which existing categories apply, adjust or refine others as needed, and add disease-specific categories. Then then general contextual ontology can reused, while new concepts, relationships, and domain-specific attributes can be added without changing the core structure. As a result, the rule extraction layer also benefits from these extensions, since rules derived from ontology-enhanced data can be updated to produce new, clinically relevant rules for the target disease.

6.2.2 Limitations

Despite the potential benefits of the proposed context-aware, ontology-guided healthcare system, it has some limitations. First, the system depends on the completeness of the input data. The dataset bias may affect the fairness of the system if the training data do not represent diverse patient profiles in terms of age, environmental exposure, comorbidities, demographic characteristics, or other relevant factors. This could lead to biased rules and unequal healthcare recommendations.

Another limitation is related to clinical validation. The extracted rules were evaluated in terms of quantitative interpretability metrics, which is an important aspect. However, they also need to be reviewed and validated by medical professionals before being used in real clinical settings.

Finally, ethical limitations should also be considered. Data privacy and security are major concerns, particularly when sensitive patient information is collected through remote monitoring systems. Clear policies are needed to define who can access the data, how it is stored, and how it is protected.

6.3 Future Work

While the proposed context-aware ontology-guided healthcare model demonstrated promising results, several research directions remain open:

- Real-World Clinical Validation:

The proposed system will be validated in real healthcare environments by integrating data from clinical systems and wearable devices. As part of future work, the generated rules will also be validated with clinicians to ensure that rule quality is assessed not only quantitatively, but also in terms of clinical interpretability and meaningfulness, ensuring that the rules are understandable and relevant to healthcare staff in practice.

- Extension to Additional Diseases:

We will extend the contextual categorization model and ontology to other chronic and acute medical conditions, such as diabetes, cardiovascular diseases, and emergency care, to evaluate the generality of the approach.

- Improvement of interpretability in healthcare domain**:**

We will analyze whether the improvements in interpretability affect the performance of the system.

LISTE DE RÉFÉRENCES

1. Freeman SF, Wong L. Knowledge Management and AI in Healthcare. Journal of Health Informatics. 2020.

2. Wong L, Zhang H. Intelligent Systems in Digital Healthcare. Healthcare Analytics. 2022.

3. Yang Y, Yin X. Integration of Deep Learning in Healthcare Systems. AI in Medicine. 2022.

4. De Hond AA, et al. Applications of AI in Clinical Practice. Journal of Clinical Medicine. 2022.

5. Shi J, Zhang R. Trends in Healthcare Technology and Outcomes. Health Tech Reports. 2022.

6. Kaye R. Chronic Disease Frameworks. Medical Practice Today. 2017.

7. Bernell S, Howard SW. Use Your Words Carefully: What Is a Chronic Disease? Front Public Health. 2016;4:159.

8. Schmidt H, et al. Chronic Disease Burdens in Modern Healthcare. Canadian Medical Journal. 2016.

9. Laraeau SC, Fahy B, Meek P, et al. Chronic Obstructive Pulmonary Disease (COPD). Am J Respir Crit Care Med. 2019; 199(1): P1–P2.

10. Karunanayake CP, et al. Canadian Asthma Health Economics. J Asthma. 2020.

11. Public Health Agency of Canada. Asthma Surveillance Report. Government of Canada. 2018.

12. Thomas P, et al. Medical Context Modeling for Telemedicine. Health Inf Sci Syst. 2014.

13. Lasierra N. Clinical Classification of Asthma Subtypes. Respiratory Medicine. 2013.

14. Bettini C, et al. Asthma and Daily Life: A Review. European Respiratory Journal. 2010.

15. Khattak AM, et al. Variability of Asthma Symptoms. BMC Pulm Med. 2014.

16. Hoareau C. Late Onset Asthma in Adults. Journal of Asthma. 2009.

17. Brown D, et al. Global Asthma Trends. WHO Bulletin. 1997.

18. Nicholas A, et al. Gender Differences in Asthma Prevalence. Int J Epidemiol. 2005.

19. Sridevi S. Classification of Allergic and Non-Allergic Asthma. Allergy Clin Immunol. 2010.

20. Wood A. Immunological Basis of Allergic Asthma. Clinical Reviews in Allergy. 2008.

21. Toskala E. Environmental Triggers of Asthma. Curr Allergy Asthma Rep. 2015.

22. Dahlöf B. Genetic Susceptibility and Asthma Risk. Scandinavian J Immunol. 2010.

23. Agache I, et al. Allergic Multimorbidities in Asthma. J Investig Allergol Clin Immunol. 2012.

24. Jaadi Z. Explainability in PCA for Health Monitoring. Machine Learning Journal. 2022.

25. Shams S, et al. Rule Extraction for Breast Cancer Prediction. IEEE Access. 2021.

26. Bettini C, Brdiczka O, Henricksen K, et al. A Survey of Context Modelling and Reasoning Techniques. Pervasive and Mobile Computing. 2010;6(2):161–180.

27. Hofer T, Schmid M, Schulz S. Context-aware information retrieval in health care: An overview. Journal of Biomedical Semantics. 2016;7:16.

28. Dey AK. Understanding and Using Context. Personal and Ubiquitous Computing. 2001;5(1):4–7.

29. Henricksen K, Indulska J. Modelling and Using Imperfect Context Information. Proceedings of PERCOM Workshops. 2004.

30. Riaño D, Real F, López-Vallverdú JA, et al. An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients. J Biomed Inform. 2012;45(3):429–446.

31. Abidi SR. Ontology-Based Modeling of Clinical Practice Guidelines: A Clinical Use-Case. Health Informatics Journal. 2007;13(2):147–167.

32. Martínez-Romero M, Rodríguez-González A, Escobar-Rodríguez F, et al. Applying semantic web technologies to construct a health care knowledge base. J Med Syst. 2012;36(5):3311–3324.

33. Bodasing N, Abidi S, Manickam S. A knowledge representation approach to model and query patient-specific treatment plans using clinical ontologies. Stud Health Technol Inform. 2015;216:871–875

34. Martínez-Costa C, Schulz S. Ontology-based information integration in health care: the case of SNOMED CT. Methods Inf Med. 2015;54(4):347–355.

35. Zarka M, Mokhtari M, Abed M. A Semantic Ontology for Asthma Diagnosis and Monitoring. Comput Biol Med. 2018;100:88–100.

36. El-Sappagh S, Elmogy M, Riad A, et al. A Mobile Health Monitoring System Based on Knowledge Representation and Context-Aware Computing. Comput Methods Programs Biomed. 2019;170:1–19.

37. Rospocher M, Sottovia A, Serafini L. Ontology-Based Decision Support in Healthcare Using Machine Learning: A Case Study. Appl Ontol. 2020;15(4):395–420.

38. Kamdar MR, Musen MA. An Empirical Meta-Analysis of Ontology Reuse in BioPortal Ontologies. PeerJ. 2021;9:e11398.

39. Dhillon JS, Jindal R. A Hybrid Model for Healthcare Ontology-Based Clinical Decision Support. J Biomed Inform. 2022;127:104005.

40. Holzinger A, Langs G, Denk H, et al. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov. 2019;9(4):e1312.

41. Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): Towards medical XAI. IEEE Trans Neural Netw Learn Syst. 2021;32(11):4793–4813.

42. Jaadi Z. Explainability in PCA for Health Monitoring. Machine Learning Journal. 2022.

43. Molnar C. Interpretable Machine Learning. 2nd ed. 2020.

44. Shams S, Khalilia M, Liu J, et al. Rule Extraction for Predicting Breast Cancer Subtypes Using mRNA Expression. IEEE Access. 2021;9:98765–98775.

45. Dourish P. What we talk about when we talk about context. Personal and Ubiquitous Computing. 2004;8(1):19–30.

46. J.Gwizdka. “What’s in the Context.”CHI2000 Workshop (2000).

47. Gross, T. and Specht M. Awareness in context-aware information systems. Tagungsband der 1. Fachübergreifenden Konferenz "Mensch & Computer", Bonn, Germany (2001).

48. Aaltonen, Antti. “A Context Visualization Model for Wearable Computers.” SEMWEB (2002).

49. Prekop, P., & Burnett, M. Activities, context and ubiquitous computing. Computer Communications, 26(11), 1168–1176 (2003).

50. Hofer, T., Schwinger, W., Pichler, M., Leonhartsberger, G. and Altmann, J. ”Context-awareness on mobile devices – the hydrogen approach,” Proceedings of the 36th Annual Hawaii International Conference on System Sciences, pp.292–302. (2002).

51. Van Bunningen, A. H., Feng, L., & Apers, P. M. (2005, April). Context for ubiquitous data management. In Ubiquitous Data Management, 2005. UDM 2005. International Workshop on IEEE (pp. 17–24).

52. Y. Miao, X. Tao, Z. Shen, Z. Liu, and C. Miao. The equivalence ofcognitive map, fuzzy cognitive map and multi value fuzzy cognitivemap. In IEEE International Conference on Fuzzy Systems, 2006 IEEEWorld Congress on Computational Intelligence (WCCI), Vancouver,BC, Canada, 16-21 July, 2006.

53. S. K. Chong, I. McCauley, S. W. Loke, and S. Krishnaswamy, “Contextaware sensors and data muling,” in Context awareness for selfmanaging systems (devices, applications and networks) proceeding. Berlin : VDE-Verlag, 2007, pp. 103–117.

54. M. Miraoui, C. Tadj, and C. b. Amar, "Architectural survey of context-aware systems in pervasive computing environment," Ubiquitous Computing and Communication Journal, vol. 3, 2008.

55. Kurti, A. Exploring the multiple dimensions of context: Implications for the design and development of innovative technology-enhanced learning environments. Stockholm University, May 28, 2009.

56. Soylu, A.; De Causmaecker1, P. & Desmet, P. (2009). Context and Adaptivity in Pervasive Engineering, Journal of Software, Vol.4, No.9, (November, 2009), pp.992-1013, ISSN 1796-217X.

57. G. Jun-zhong, “Context aware computing,” Journal of East China Normal University (Natural Science), vol. 5, pp. 1–20, 2009.

58. Tamine-Lechani L, Boughanem M, Daoud M (2010) Evaluation of contextual information retrieval effectiveness: overview of issues and research. Knowl Inf Syst 24(1):1–34

59. S. Rizou. “A system for distributed context reasoning,” in Autonomic and Autonomous Systems (ICAS), 2010 Sixth International Conference on, march 2010, pp. 84 –89.

60. Nageba, E. A Model Driven Ontology-based Architecture for Supporting the Quality of Services in Pervasive Telemedicine Applications. Proceedings the 3rd International Conference on Pervasive Computing Technologies for Healthcare, London UK, IEEE Computer Society, pp. 1-8, Apr 2009.

61. S. Kim. Smart learning services based on smart cloud computing. Sensors 11(8), 7835– 7850 (2011).

62. Bin Guo. the Internet of Things to Embedded Intelligence. World Wide Web 16, 4 (2013), 399–420.

63. Tamine-Lechani L, Boughanem M, Daoud M (2010) Evaluation of contextual information retrieval effectiveness: overview of issues and research. Knowl Inf Syst 24(1):1–34.

64. H. Mshali, T. Lemlouma, D. Magoni, A predictive approach for efficient e-health monitoring, in: IEEE International Conference on e-Health Networking, Applications and Services, 2015, pp. 268–273.

65. Banhato, Eliane Ferreira Carvalho, Pricila Cristina Correa Ribeiro, Danielle Viveiros Guedes, Cláudia Helena Cerqueira Mármora, and Roberto Alves Lourenço. "Health selfawareness in senior citizens: focus on physical, emotional and cognitive health." Psychology 6, no. 07 (2015): 846.

66. Zhang, W., Thurow, K., & Stoll, R. (2016). A Context-Aware mHealth System for Online Physiological Monitoring in Remote Healthcare. Int. J. Comput. Commun. Control, 11, 142-156.

67. Ameyed, Darine. "Modélisation et spécification formelle de contexte et sa prédiction dans les systèmes diffus: Une approche basée sur la logique temporelle et le modèle stochastique." PhD diss., École de technologie supérieure, 2017.

68. Hala Ahmed H.A., Mohammed Elmogy M.E., Ahmed Atwan A.A., “Mobile context awareness for managing context healthcare data: a survey”, JOURNAL OF SOFTWARE ENGINEERING & INTELLIGENT SYSTEMS ISSN, 2518-8739, 2017.

69. Neumuth, Thomas & Rockstroh, Max & Franke, Stefan. (2018). Context-aware medical technologies - Relief or burden for clinical users. Current Directions in Biomedical Engineering. 4. 119-122. 10.1515/cdbme-2018-0030.

70. Almobarak, Arwa S., and Wassim A. Jaziri. "Considering the Context and Patient's Profile to Support Ubiquitous Healthcare Systems." In 2019 12th International Conference on Developments in eSystems Engineering (DeSE), pp. 147-152. IEEE, 2019.

71. Ajami Hicham, et al. "Categorization of the context within the medical domain." SmartHomes and Health Telematics, Designing a Better Future: Urban Assisted Living: 16th International Conference, ICOST 2018, Singapore, Singapore, July 10-12, 2018, Proceedings 16. Springer International Publishing, 2018.

72. Taiwo, Olutosin, and Absalom E. Ezugwu. "Smart healthcare support for remote patient monitoring during covid-19 quarantine." Informatics in medicine unlocked 20 (2020).

73. Kayes, A. S. M., et al. "A survey of context-aware access control mechanisms for cloud and fog networks: Taxonomy and open research issues." Sensors 20.9 (2020): 2464.

74. Konan-Marcelin Kouamé, Hamid Mcheick and Hicham Ajami, Adaptive Mechanism Model for the Prevention of SLA Violation in the Context of COPD Patient Monitoring ,2020.

75. Hoxha J, et al. Enabling Clinical Decision Support Systems: A Knowledge Representation Survey. J Biomed Inform. 2021;119:103823.

76. Esteva, Andre, et al. "Dermatologist-level classification of skin cancer with deep neural networks." nature 542.7639 (2017): 115-118.

77. Lezcano-Valverde, José M., et al. "Development and validation of a multivariate predictive model for rheumatoid arthritis mortality using a machine learning approach." Scientific reports 7.1 (2017): 10189.

78. Khan RS, Zardar AA, Bhatti Z. Artificial Intelligence based Smart Doctor using Decision Tree Algorithm. arXiv preprint arXiv:1808.01884, 2018.

79. Wang, C., et al. "Comparison of equivalent current systems for the substorm event of 8 March 2008 derived from the global PPMLR-MHD model and the KRM algorithm." Journal of Geophysical Research: Space Physics 116.A7 (2011).

80. Bleakley, Kevin, and Yoshihiro Yamanishi. "Supervised prediction of drug–target interactions using bipartite local models." Bioinformatics 25.18 (2009): 2397-2403.

81. Shi, Longxiang, et al. "Semantic health knowledge graph: semantic integration of heterogeneous medical knowledge and services." BioMed research international 2017 (2017).

82. Li, Linfeng, et al. "Real-world data medical knowledge graph: construction and applications." Artificial intelligence in medicine 103 (2020): 101817.

83. Xie, Jing, et al. "Learning an expandable EMR-based medical knowledge network to enhance clinical diagnosis." Artificial intelligence in medicine 107 (2020): 101927.

84. Monroy, Nelson F., and Miguel Altuve. "Analysis of the observation sequence duration of hidden Markov models for QRS complex detection in single-lead ECG recordings." Computing in Cardiology Conference (CinC). Vol. 45. IEEE, 2018.

85. Sandag GA, Tedry NE, Lolong S, editors. Classification of Lower Back Pain Using K-Nearest Neighbor Algorithm. 2018 6th International Conference on Cyber and IT Service Management (CITSM); 2018

86. Lin H, Long E, Ding X, et al. Prediction of myopia development among Chinese school-aged children using refraction data from electronic medical records: A retrospective, multicentre machine learning study. PLoS Med 2018;15:e1002674.

87. Kar, Subrata, and D. Dutta Majumder. "A mathematical theory of shape and neuro-fuzzy methodology-based diagnostic analysis: a comparative study on early detection and treatment planning of brain cancer." International Journal of Clinical Oncology 22 (2017): 667-681.

88. Johansson, Fredrik D., et al. "Predicting response to tocilizumab monotherapy in rheumatoid arthritis: a real-world data analysis using machine learning." The Journal of Rheumatology 48.9 (2021): 1364-1370.

89. Tran, Tran Quoc Bao, Clea du Toit, and Sandosh Padmanabhan. "Artificial intelligence in healthcare-the road to precision medicine." J. Hosp. Manag. Health Policy 5 (2021): 29.

90. Allalou, Amina, et al. "A predictive metabolic signature for the transition from gestational diabetes mellitus to type 2 diabetes." Diabetes 65.9 (2016): 2529-2539.

91. Bertaud-Gounot, Valérie, Régis Duvauferrier, and Anita Burgun. "Ontology and medical diagnosis." Informatics for Health and Social Care 37.2 (2012): 51-61.

92. Nyangaresi, Vincent Omollo, Nidhal Kamel Taha El-Omari, and Judith Nyakanga Nyakina. "Efficient Feature Selection and ML Algorithm for Accurate Diagnostics." Journal of Computer Science Research 4.1 (2022): 10-19.

93. Ruiz-Fernández D, Monsalve Torra A, Soriano-Payá A, et al. Aid decision algorithms to estimate the risk in congenital heart surgery. Comput Methods Programs Biomed 2016;126:118-27.

94. Kasbekar, Prasad Umesh, Pranay Goel, and Shailaja Prakash Jadhav. "A decision tree analysis of diabetic foot amputation risk in Indian patients." Frontiers in endocrinology 8 (2017): 25.

95. Sohn, Sunghwan, et al. "Detection of clinically important colorectal surgical site infection using Bayesian network." Journal of Surgical Research 209 (2017): 168-173.

96. Topuz, Kazim, et al. "Predicting graft survival among kidney transplant recipients: A Bayesian decision support model." Decision Support Systems 106 (2018): 97-109

97. Martínez-Costa C, Schulz S. Ontology-based information integration in health care: The case of SNOMED CT. Methods Inf Med. 2015;54(4):347–355.

98. Pesquita C, Faria D, Bastos H, et al. Biomedical ontologies: A functional perspective. Brief Bioinform. 2009;10(1):59–70.

99. Rubin DL, Shah NH, Noy NF. Biomedical ontologies: a functional perspective. Brief Bioinform. 2008;9(1):75–90.

100. Bodenreider O, Stevens R. Bio-ontologies: current trends and future directions. Brief Bioinform. 2006;7(3):256–274.

101. Mungall CJ, McMurry JA, Köhler S, et al. The Monarch Initiative: An integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Res. 2017;45(D1):D712–D722.

102. Fernández-Breis JT, Maldonado JA, Marcos M, et al. Leveraging Electronic Healthcare Record Standards and Semantic Web Technologies for the Identification of Patient Cohorts. J Am Med Inform Assoc. 2013;20(e2):e288–e296.

103. Martínez-Costa C, Menárguez-Tortosa M, Fernández-Breis JT. Ontology-Based Integration and Analysis of Clinical Data. Comput Methods Programs Biomed. 2012;107(2):412–423

104. Thomas, Paul. "Understanding context in healthcare research and development." London Journal of Primary Care 6, no. 5 (2014): 103–105.

105. Dash, Sabyasachi, Sushil Kumar Shakyawar, Mohit Sharma, and Sandeep Kaushik. "Big data in healthcare: management, analysis and future prospects." Journal of Big Data 6, no. 1 (2019): 1–25.

106. Dohn, Nina Bonderup, Stig Børsen Hansen, and Søren Harnow Klausen. "On the concept of context." Education Sciences 8, no. 3 (2018): 111.

107. Sandkuhl, Kurt. "Information Logistics in Networked Organizations: Selected Concepts and Applications." ICEIS 2007, LNBIP 12, pp. 43–54. Springer, 2008.

108. Talboom-Kamp, E.P., N.A. Verdijk, L.M. Harmans, et al. "An eHealth platform to manage chronic disease in primary care: an innovative approach." Interact J Med Res 2016.

109. Dey, A. K., & Abowd, G. D. Towards a better understanding of context and context-awareness. CHI 2000.

110. Thomas, P. Understanding context in healthcare research and development. London Journal of Primary Care, 6(5), 103–105 (2014).

111. Ten Teije, A., Miksch, S., & Lucas, P. J. F. Knowledge representation for health care: KR4HC 2011. Springer (2011).

112. Shaban-Nejad, A., Michalowski, M., & Buckeridge, D. L. Health intelligence: how artificial intelligence transforms population and personalized health. NPJ Digital Medicine 1.1 (2018): 1–6.

113. Shen, J., Zhang, C. J. P., Jiang, B., et al. Artificial intelligence versus clinicians in disease diagnosis: systematic review. JMIR Medical Informatics 7.3 (2019): e10010.

114. Abidi, S. S. R. Knowledge management in healthcare: towards 'knowledge-driven' decision-support services. International Journal of Medical Informatics 63(1-2), 5–18 (2001).

115. Abidi, S. S. R. Knowledge management in healthcare: towards 'knowledge-driven' decision-support services. International Journal of Medical Informatics 63(1-2), 5–18 (2001).

116. Martínez-Costa, C., Schulz, S. Ontology-based information integration in health care: the case of SNOMED CT. Methods of Information in Medicine, 54(4), 347–355 (2015).

117. Musen, M. A., et al. The Protege project: A look back and a look forward. AI Matters 1.4 (2015): 4–12.

118. Pesquita, C., et al. Biomedical ontologies: a functional perspective. Briefings in Bioinformatics 10.1 (2009): 59–70.

119. Fernández-Breis, J. T., et al. Leveraging EHR standards and Semantic Web technologies for identifying patient cohorts. Journal of the American Medical Informatics Association, 20(e2), e288–e296 (2013).

120. Bodenreider, O., & Stevens, R. Bio-ontologies: current trends and future directions. Briefings in Bioinformatics, 7(3), 256–274 (2006).

121. Hogan, W. R., et al. Developing and implementing an ontology for the Asthma Ontology Project. Journal of Biomedical Semantics 2.1 (2011): 1–14.

122. Kibbe, W. A., et al. Disease Ontology 2015 update: an expanded and updated database of human diseases. Nucleic Acids Research, 43(D1), D1071–D1078 (2015).

123. Smith, B., et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology, 25(11), 1251–1255 (2007).

124. Martínez-Romero, M., et al. Semantic modeling for chronic disease monitoring and decision support. Journal of Biomedical Informatics, 82, 168–181 (2018).

125. Freeman Shannon F.S, Hannah R. Marston H.R.M., Janna Olynick J.O., Charles Musselwhite C.M., Cory Kulczycki C.K., Rebecca Genoe R.G., and Beibei Xiong. B.X. "Intergenerational effects on the impacts of technology use in later life: Insights from an international, multi-site study." International journal of environmental research and public health 17, no. 16 (2020): 5711.

126. Wong W., Brian Li Han B.H., Laura Maaß L.M., Alice Vodden A.V., Robin van Kessel R.K., Sebastiano Sorbello S.S., Stefan Buttigieg S.B., Anna Odone A.O., Digital Health Section and European Public Health Association. "The dawn of digital public health in Europe: Implications for public health policy and practice." The Lancet Regional Health-Europe 14 (2022): 100316.

127. Yang Yin, Keng Siau, Wen Xie, and Yan Sun. "Smart Health: Intelligent Healthcare Systems in the Metaverse, Artificial Intelligence, and Data Science Era." Journal of Organizational and End User Computing (JOEUC) 34, no. 1 (2022): 1-14.

128. De Hond D.H., Anne AH A.H., Artuur M. Leeuwenberg A.L., Lotty Hooft L.H., Ilse MJ Kant I.K., Steven WJ Nijman S.N., Hendrikus JA van Os H.V.O., Jiska J. Aardoom et al. "Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review." NPJ digital medicine 5, no. 1 (2022): 2.

129. Crameri, Kellie-Anne K., Lynne Maher L.M., Pieter Van Dam P.V.D., and Sarah Prior S.P.. "Personal electronic healthcare records: What influences consumers to engage with their clinical data online? A literature review." Health Information Management Journal 51, no. 1 (2022): 312.

130. Moshawra, Mohammad M.M., Mehdi Adda M.A., Abdenour Bouzouane A.B., Hussein Ibrahim H.I., and Ali Raad A.R.. "Smart Wearables for the Detection of Occupational Physical Fatigue: A Literature Review." Sensors 22, no. 19 (2022): 7472.

131. Moshawrab, Mohammad, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim, and Ali Raad. "Smart Wearables for the Detection of Cardiovascular Diseases: A Systematic Literature Review." Sensors 23, no. 2 (2023): 828.

132. Amagai, Saki, Sarah Pila, Aaron J. Kaat, Cindy J. Nowinski, and Richard C. Gershon. "Challenges in participant engagement and retention using mobile health apps: literature review." Journal of medical Internet research 24, no. 4 (2022): e35120.

133. Wake, Dyson T., D. Max Smith, Sadaf Kazi, and Henry M. Dunnenberger. "Pharmacogenomic clinical decision support: a review, how-to guide, and future vision." Clinical Pharmacology & Therapeutics 112, no. 1 (2022): 44-57.

134. Shi Jiyuan S.J., Shuaifang Wei S.W., Ya Gao, Fan Mei, Jinhui Tian, Yang Zhao, and Zheng Li. "Global output on artificial intelligence in the field of nursing: A bibliometric analysis and science mapping." Journal of Nursing Scholarship (2022).

135. Talboom-Kamp EP, Verdijk NA, Harmans LM, et al. An eHealth platform to manage chronic disease in primary care: an innovative approach. Interact J Med Res 2016.

136. Kurt Sandkuhl, “Information Logistics in Networked Organizations: Selected Concepts and Applications”, ICEIS 2007, LNBIP 12, pp.43-54, © SpringerVerlagBerliknHeidlberg, 2008

137. Thomas, Paul. "Understanding context in healthcare research and development." London Journal of Primary Care 6, no. 5 (2014): 103-105.

138. Dash, Sabyasachi, Sushil Kumar Shakyawar, Mohit Sharma, and Sandeep Kaushik. "Big data in healthcare: management, analysis and future prospects." Journal of Big Data 6, no. 1 (2019): 1-25.

139. Dohn, Nina Bonderup, Stig Børsen Hansen, and Søren Harnow Klausen. "On the concept of context." Education sciences 8, no. 3 (2018): 111.

140. Perera, C., Zaslavsky, A., Christen, P., & Georgakopoulos, D. Context aware computing for the internet of things: A survey. IEEE Communications Surveys & Tutorials, 16(1), 414–454 (2014).

141. 70. Ajami Hicham, Mcheick Hamid et Elkhaled Zayan. Survey of health care context models : prototyping of healthcare context framework. Dans : Summer Simulation MultiConference (SummerSim'16) , July 24–27, 2016, Montreal, Quebec.

142. Soylu, A., et al. Contextual information management in healthcare: A survey on context-aware frameworks. Journal of Biomedical Informatics, 85, 68–83 (2018).

143. Dey, A.K. Understanding and using context. Personal and Ubiquitous Computing, 5(1), 4–7 (2001).

144. Henricksen, K., Indulska, J. Modelling and using imperfect context information. Proceedings of the First International Conference on Pervasive Computing (2004).

145. Chourabi, H., Nam, T., Walker, S., et al. Understanding smart cities: An integrative framework. Proceedings of the 45th Hawaii International Conference on System Sciences (2012): 2289–2297.

146. Alshahrani, M., et al. An ontology-based approach for personalized emergency healthcare support. Journal of Biomedical Semantics, 8(1), 6 (2017).

147. Zhang, X., Wang, X., & Xu, G. Design and implementation of a context-aware smart healthcare system. In: 2016 International Conference on Smart Health (2016): 17–28.

148. Neumuth, T., et al. Modeling surgical processes: A four-layered approach. Artificial Intelligence in Medicine 42.1 (2008): 61–75.

149. Almobarak, F., & El-Darieby, M. An ontology-based user context model for cloud-based healthcare systems. Procedia Computer Science 141 (2018): 475–482.

150. Bin Guo. The Internet of Things to Embedded Intelligence. World Wide Web 16.4 (2013): 399–420.

151. Social Factors That Impact Your Health. Lambton Public Health. www.lambtonpublichealth.ca

152. Braveman P., Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public Health Reports 129 Suppl 2 (2014): 19–31.

153. Butpheng, C., Yeh, K. H., Hou, J. L. A Secure IoT and Cloud Computing-Enabled e-Health Management System. Security and Communication Networks (2022).

154. Chaari, T., Laforest, F., Celantano, A. Design of context-aware applications based on web services. LIRIS UMR 5205 CNRS/INSA de Lyon/Université Claude Bernard (2004).

155. Chen F. J., Warden A. C., Chang H. T. Motivators that do not motivate: The case of Chinese EFL learners and the influence of culture on motivation. TESOL Quarterly 39(4) (2005): 609–633.

156. Bernd Niklas Klein. On the importance of time synchronization for context aware applications. PhD Thesis, Kassel University Press (2011).

157. Darine AMEYED. Modélisation et spécification formelle de contexte et sa prédiction dans les systèmes diffus. PhD thesis, L'École de technologie supérieure (ETS) (2016).

158. Levandoski J., Sarwat M., Eldawy A., Mokbel M. LARS: a location-aware recommender system. In: ICDE (2012): 450–461.

159. Xu X., Tang J., Zhang X., Liu X., et al. Exploring techniques for vision-based human activity recognition: Methods, systems, and evaluation. Sensors 13.2 (2013): 1635–1650.

160. Bardram J. Applications of Context-Aware Computing in Hospital Work. In: Proc. of ACM SAC (2004): 1574–1579.

161. Dey A.K., Abowd G.D., Wood A. CyberDesk: A Framework for Providing Self-Integrating Context-Aware Services. Knowledge-Based Systems 11 (1999).

162. Brashier BB, Kodgule R. Risk factors and pathophysiology of COPD. J Assoc Physicians India. 60 Suppl (2012): 17–21.

163. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases, accessed May 2023.

164. Public Health Agency of Canada. Report from the Canadian Chronic Disease Surveillance System: Asthma and Chronic Obstructive Pulmonary Disease (COPD) in Canada, 2018; Public Health Agency of Canada: Ottawa, ON, Canada, 2018.

165. World Health Organization. Asthma. Available online at http: https://www.who.int/newsroom/fact-sheets/detail/asthma. Published May, 2020.

166. Asthma.ca. Available online at: https://asthma.ca/wp-content/uploads/2019/09/ASnapshot-of-Asthma-in-Canada-2019-Annual-Asthma-Survey-Report-1.pdf

167. Madore AM, Laprise C. Immunological and genetic aspects of asthma and allergy. J Asthma Allergy. 2010; 3:107-121. Published 2010 Aug 20. doi:10.2147/JAA.S8970

168. Toskala, E, Kennedy, D. Asthma risk factors. Int Forum Allergy Rhino. 2015; 5: S11– S16

169. Moshawrab, Mohammad, et al. "Smart Wearables for the Detection of Cardiovascular Diseases: A Systematic Literature Review." Sensors 23.2 (2023): 828.

170. World health statistics. Geneva, World Health Organization, 2016 http://www.who.int/mediacentre/factsheets/fs315/en/.

171. Yin, Kathleen, Liliana Laranjo, Huong Ly Tong, Annie YS Lau, A. Baki Kocaballi, Paige Martin, Sanjyot Vagholkar, and Enrico Coiera. "Context-aware systems for chronic disease patients: scoping review." Journal of medical Internet research 21, no. 6 (2019): e10896.

172. Bombard, Yvonne, G. Ross Baker, Elaina Orlando, Carol Fancott, Pooja Bhatia, Selina Casalino, Kanecy Onate, Jean-Louis Denis, and Marie-Pascale Pomey. "Engaging patients to improve quality of care: a systematic review." Implementation Science 13 (2018): 1-22.

173. Francula-Zaninovic, Sonja, and Iskra A. Nola. "Management of measurable variable cardiovascular disease' risk factors." Current cardiology reviews 14, no. 3 (2018): 153-163.

174. Brashier BB, Kodgule R. Risk factors and pathophysiology of chronic obstructive pulmonary disease (COPD). J Assoc Physicians India. 2012 Feb;60 Suppl:17-21. PMID: 23155808.

175. P.S. Bakke, E. Rönmark, T. Eagan, F. Pistelli, I. Annesi-Maesano, M. Maly, M. Meren, P. Vermeire†, J. Vestbo, G. Viegi, J. Zielinski, B. Lundbäck. Recommendations for epidemiological studies on COPD. European Respiratory Journal 2011 38: 1261-1277; DOI: 10.1183/09031936.00193809.

176. Price, David, et al. "Earlier diagnosis and earlier treatment of COPD in primary care." Primary Care Respiratory Journal 20.1 (2011): 15-22.

177. Javorac, Jovan, et al. "What are the effects of meteorological factors on exacerbations of chronic obstructive pulmonary disease?." Atmosphere 12.4 (2021): 442.

178. Abramson, Michael J., et al. "Distinguishing adult-onset asthma from COPD: a review and a new approach." International journal of chronic obstructive pulmonary disease (2014): 945-962.

179. Wong, Alyson WM, et al. "Acute exacerbation of chronic obstructive pulmonary disease: influence of social factors in determining length of hospital stay and readmission rates." Canadian Respiratory Journal 15.7 (2008): 361-364.

180. Vrijheid M. The exposome: a new paradigm to study the impact of environment on health. Thorax 2014; 69: 876–8.

181. N. Lasierra,“A three stage ontology-driven solution to provide personalized care to chronic patients at home”. Journal of biomedical informatics, vol. 46, 2013, pp 516-52.

182. Dahlöf, Björn. "Cardiovascular disease risk factors: epidemiology and risk assessment." The American journal of cardiology 105, no. 1 (2010): 3A-9A.

183. Bin Guo. the Internet of Things to Embedded Intelligence. World Wide Web 16, 4 (2013), 399–420.

184. Stafford, S.P. Data, information, knowledge, and wisdom. Knowl. Manag. Organ. Intell. Learn. Complex. 2009, 3, 179.

185. Kamala, S.P.R.; Justus, S. A Study on Knowledge Representation Models. Eur. J. Mol. Clin. Med. 2020.

186. Praveenkumar, T.; Sabhrish, B.; Saimurugan, M.; Ramachandran, K. Pattern recognition based on-line vibration monitoring system for fault diagnosis of automobile gearbox. Measurement 2018, 114, 233–242.

187. Abu-Salih, B. Domain-specific knowledge graphs: A survey. J. Netw. Comput. Appl. 2021, 185, 103076.

188. Cooper, P. Data, information, knowledge and wisdom. Anaesth. Intensive Care Med. 2017, 18, 55–56.

189. Lenz, R., Miksch, S., Reichert, M., Riano, D., Teije, A. Process Support and Knowledge Representation in Health Care. Springer: Tallin, Estonia, 2013.

190. Wang, F.; Preininger, A. AI in health: State of the art, challenges, and future directions. Yearb. Med. Inform. 2019, 28, 16–26.

191. Vellido, A. The importance of interpretability and visualization in machine learning for applications in medicine and health care. Neural Comput. Appl. 2019, 32, 18069–18083.

192. Shi, L.; Li, S.; Yang, X.; Qi, J.; Pan, G.; Zhou, B. Semantic Health Knowledge Graph: Semantic Integration of Heterogeneous Medical Knowledge and Services. BioMed Res. Int. 2017, 2017, 2858423.

193. Weng, C.; Wu, X.; Luo, Z.; Boland, M.R.; Theodoratos, D.; Johnson, S.B. EliXR: An approach to eligibility criteria extraction and representation. J. Am. Med. Inform. Assoc. 2011, 18, i116–i124.

194. Blasch, E.; Kadar, I.; Salerno, J.; Kokar, M.M.; et al. Issues and challenges of knowledge representation and reasoning methods in situation assessment (Level 2 Fusion). Signal Process. Sens. Fusion Target Recognit. XV 2006, 6235, 355–368.

195. Stafford, S.P. Data, information, knowledge, and wisdom. Knowl. Manag. Organ. Intell. Learn. Complex. 2009, 3, 179.

196. Cooper, P. Data, information, knowledge and wisdom. Anaesth. Intensive Care Med. 2017, 18, 55–56.

197. Kamala, S.P.R.; Justus, S. A Study on Knowledge Representation Models. Eur. J. Mol. Clin. Med. 2020.

198. Bright, C.; Kay, A.; Feeney, A. The effects of domain and type of knowledge on category-based inductive reasoning. In Proceedings of the Annual Meeting of the Cognitive Science Society, Portland, Oregon, USA, 11–14 August 2010; Volume 32.

199. Lieto, A.; Minieri, A.; Piana, A.; Radicioni, D.P. A knowledge-based system for prototypical reasoning. Connect. Sci. 2015, 27, 137–152. [CrossRef]

200. Sowa, J.F. Common Logic. A Framework for a Family of Logic-Based Languages. 2008. Available online: http://www.jfsowa. com/talks/clprop.htm (accessed on 16 June 2024).

201. Holsapple, C.W. Knowledge and its attributes. In Handbook on Knowledge Management 1: Knowledge Matters; Springer: Berlin/Heidelberg, Germany, 2004; pp. 165–188.

202. Nobécourt, J.; Brigitte, B. Md: A Modelling Language to Build a Formal Ontology in Either Description Logics or Conceptual Graphs. In International Conference on Knowledge Engineering and Knowledge Management; Springer: Berlin/Heidelberg, Germany, 2000; Volume 1937.

203. Edward, E.O.; Nlerum, P.A. Knowledge representation in artificial intelligence and expert systems using inference rule. Int. J. Sci. Eng. Res. 2020, 11, 1886–1900.

204. Chandrasegaran, S.K.; Ramani, K.; Sriram, R.D.; Horváth, I.; Bernard, A.; Harik, R.F.; Gao, W. The evolution, challenges, and future of knowledge representation in product design systems. Comput.-Aided Des. 2012, 45, 204–228. [CrossRef]

205. Pinker, S. A theory of graph comprehension. In Artificial Intelligence and the Future of Testing. Psychology Press; Psychology Press: London, UK, 2014; pp. 73–126.

206. Wang, F.; Li, H.; Dong, C.; Ding, L. Knowledge representation using non-parametric Bayesian networks for tunneling risk analysis. Reliab. Eng. Syst. Saf. 2019, 191, 106529. [CrossRef]

207. Frank, A.U. Tiers of ontology and consistency constraints in geographical information systems. Int. J. Geogr. Inf. Sci. 2001, 15, 667–678. [CrossRef]

208. Azad, M.; Chikalov, I.; Moshkov, M. Representation of Knowledge by Decision Trees for Decision Tables with Multiple Decisions. Procedia Comput. Sci. 2020, 176, 653–659. [CrossRef]

209. Myles, A.J.; Feudale, R.N.; Liu, Y.; Woody, N.A.; Brown, S.D. An introduction to decision tree modeling. J. Chemom. A J. Chemom. Soc. 2004, 18, 275–285. [CrossRef]

210. Levine, D.S.; Aparicio, M., IV (Eds.) Neural Networks for Knowledge Representation and Inference; Psychology Press: London, UK, 2013.

211. Hotz, L.; Felfernig, A.; Stumptner, M.; Ryabokon, A.; Bagley, C.; Wolter, K. Configuration Knowledge Representation and Reasoning; Morgan Kaufmann: San Francisco, CA, USA, 2014.

212. Minock, M. Knowledge Representation using Schema Tuple Queries; KRDB: Umea, Sweden, 2003.

213. Kenett, Y.N.; Faust, M. A Semantic Network Cartography of the Creative Mind. Trends Cogn. Sci. 2019, 23, 271–274. [CrossRef] [PubMed]

214. Sipper, M.; Moore, J.H. Conservation machine learning: A case study of random forests. Sci. Rep. 2021, 11, 3629. [CrossRef] [PubMed]

215. Prajwala, T.R. A comparative study on decision tree and random forest using R tool. Int. J. Adv. Res. Comput. Commun. Eng. 2015, 4, 196–199.

216. Jadhav, S.D.; Channe, H.P. Comparative study of K-NN, naive Bayes and decision tree classification techniques. Int. J. Sci. Res. 2016, 5, 1842–1845.

217. Prieto, J.; Corchado, J.M. A Review of k-NN Algorithm Based on Classical and Quantum Machine Learning. In Distributed Computing and Artificial Intelligence, Special Sessions, 17th International Conference; Springer Nature: L'Aquila, Italy, 2020; Volume 1242, p. 189.

218. Schuster-Böckler, B.; Bateman, A. An Introduction to Hidden Markov Models. Curr. Protoc. Bioinform. 2007, 18, A.3A.1–A.3A.9. [CrossRef] [PubMed]

219. Rushdi, A.M.; Rushdi, M.A. Mathematics and examples of the modern syllogistic method of propositional logic. Math. Appl. Inf. Syst. Bentham Sci. Publ. Emir. Sharjah United Arab. Emir. 2018, 6, 123–167.

220. Bartels, P.H.; Hiessl, H. Expert systems in histopathology. II. Knowledge representation and rule-based systems. Anal. Quant. Cytol. Histol. 1989, 11, 147–153. [PubMed]

221. Lifschitz, V.; Morgenstern, L.; Plaisted, D. Knowledge representation and classical logic. Found. Artif. Intell. 2008, 3, 3–88.

222. Frick, M.; Grohe, M. The complexity of first-order and monadic second-order logic revisited. Ann. Pure Appl. Log. 2004, 130, 3–31. [CrossRef]

223. Fan, C.-Y.; Chang, P.-C.; Lin, J.-J.; Hsieh, J. A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. Appl. Soft Comput. 2011, 11, 632–644. [CrossRef]

224. Setiawan, F.A.; Budiardjo, E.K.; Basaruddin, T.; Aminah, S. A Systematic Literature Review on Combining Ontology with Bayesian Network to Support Logical and Probabilistic Reasoning. In Proceedings of the 2017 International Conference on Software and e-Business, Hong Kong, China, 28–30 December 2017.

225. Yan, Z.; Zhang, H.; Jia, Y.; Breuel, T.; Yu, Y. Combining the best of convolutional layers and recurrent layers: A hybrid network for semantic segmentation. arXiv 2016, arXiv:1603.04871.

226. Vu, D.-H.; Vu, T.-S.; Luong, T.-D. An efficient and practical approach for privacy-preserving Naive Bayes classification. J. Inf. Secur. Appl. 2022, 68, 103215. [CrossRef]

227. Nikam, S.S. A comparative study of classification techniques in data mining algorithms. Orient. J. Comput. Sci. Technol. 2015, 8, 13–19.

228. Fletcher, S.; Islam, M.Z. Decision tree classification with differential privacy: A survey. ACM Comput. Surv. 2019, 52, 1–33. [CrossRef]

229. Kotsiantis, S.B.; Zaharakis, I.; Pintelas, P. Supervised machine learning: A review of classification techniques. Emerg. Artif. Intell. Appl. Comput. Eng. 2007, 160, 3–24.

230. Uusitalo, L. Advantages and challenges of Bayesian networks in environmental modelling. Ecol. Model. 2007, 203, 312–318. [CrossRef]

231. Schlüter, F. A survey on independence-based Markov networks learning. Artif. Intell. Rev. 2012, 42, 1069–1093. [CrossRef] 47. Weng, C.; Gennari, J.H.; Fridsma, D.B. User-centered semantic harmonization: A case study. J. Biomed. Inform. 2007, 40, 353–364. [CrossRef] [PubMed]

232. Kulakowski, K.; Nalepa, G.J. Using UML state diagrams for visual modeling of business rules. In Proceedings of the 2008 International Multiconference on Computer Science and Information Technology, Wisla, Poland, 20–22 October 2008.

233. Borge-Holthoefer, J.; Arenas, A. Semantic Networks: Structure and Dynamics. Entropy 2010, 12, 1264–1302. [CrossRef]

234. Konopka, B.M. Biomedical ontologies—A review. Biocybern. Biomed. Eng. 2015, 35, 75–86. [CrossRef].

235. Callahan, T.J.; Tripodi, I.J.; Pielke-Lombardo, H.; Hunter, L.E. Knowledge-Based Biomedical Data Science. Annu. Rev. Biomed. Data Sci. 2020, 3, 23–41. [CrossRef] [PubMed]

236. Tian, S.; Yang, W.; Le Grange, J.M.; Wang, P.; Huang, W.; Ye, Z. Smart healthcare: Making medical care more intelligent. Glob. Health J. 2019, 3, 62–65. [CrossRef]

237. Anwar, S.S.; Ahmad, U.; Khan, M.M.; Haider, M.F.; Akhtar, J. Artificial Intelligence in Healthcare: An Overview. Int. J. Eng. Res. Adv. Technol. 2020, 6, 38–45.

238. Narula, S.; Shameer, K.; Omar, A.M.S.; Dudley, J.T.; Sengupta, P.P. Machine-Learning Algorithms to Automate Morphological and Functional Assessments in 2D Echocardiography. J. Am. Coll. Cardiol. 2016, 68, 2287–2295. [CrossRef] [PubMed]

239. Tylman, W.; Waszyrowski, T.; Napieralski, A.; Kamin´ski, M.; Trafidło, T.; Kulesza, Z.; Kotas, R.; Marciniak, P.; Tomala, R.; Wenerski, M. Real-time prediction of acute cardiovascular events using hardware-implemented Bayesian networks. Comput. Biol. Med. 2016, 69, 245–253. [CrossRef] [PubMed]

240. Ruiz-Fernández, D.; Torra, A.M.; Soriano-Payá, A.; Marín-Alonso, O.; Palencia, E.T. Aid decision algorithms to estimate the risk in congenital heart surgery. Comput. Methods Programs Biomed. 2016, 126, 118–127. [CrossRef] [PubMed]

241. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros,J.; et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 2016, 316, 2402–2410. [CrossRef] [PubMed]

242. Yoo, K.D.; Noh, J.; Lee, H.; Kim, D.K.; Lim, C.S.; Kim, Y.H.; Lee, J.P.; Kim, G.; Kim, Y.S. A Machine Learning Approach Using Survival Statistics to Predict Graft Survival in Kidney Transplant Recipients: A Multicenter Cohort Study. Sci. Rep. 2017, 7, 8904. [CrossRef] [PubMed]

243. Sohn, S.; Larson, D.W.; Habermann, E.B.; Naessens, J.M.; Alabbad, J.Y.; Liu, H. Detection of clinically important colorectal surgical site infection using Bayesian network. J. Surg. Res. 2016, 209, 168–173. [CrossRef] [PubMed]

244. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017, 542, 115–118. [CrossRef] [PubMed]

245. Lezcano-Valverde, J.M.; Salazar, F.; León, L.; Toledano, E.; Jover, J.A.; Fernandez-Gutierrez, B.; Soudah, E.; González-Álvaro, I.; Abasolo, L.; Rodriguez-Rodriguez, L. Development and validation of a multivariate predictive model for rheumatoid arthritis mortality using a machine learning approach. Sci. Rep. 2017, 7, 10189. [CrossRef] [PubMed]

246. Khan, R.S.; Zardar, A.A.; Bhatti, Z. Artificial Intelligence based Smart Doctor using Decision Tree Algorithm. arXiv 2018, arXiv:1808.01884.

247. Wang, C.; Zhang, J.J.; Tang, B.B.; Fu, S.Y. Comparison of equivalent current systems for the substorm event of 8 March 2008 derived from the global PPMLR-MHD model and the KRM algorithm. J. Geophys. Res. 2011, 116, A07207. [CrossRef]

248. Bleakley, K.; Yamanishi, Y. Supervised prediction of drug–target interactions using bipartite local models. Bioinformatics 2009, 25, 2397–2403. [CrossRef] [PubMed]

249. Li, L.; Wang, P.; Yan, J.; Wang, Y.; Li, S.; Jiang, J.; Sun, Z.; Tang, B.; Chang, T.-H.; Wang, S.; et al. Real-world data medical knowledge graph: Construction and applications. Artif. Intell. Med. 2020, 103, 101817. [CrossRef] [PubMed]

250. Xie, J.; Jiang, J.; Wang, Y.; Guan, Y.; Guo, X. Learning an expandable EMR-based medical knowledge network to enhance clinical diagnosis. Artif. Intell. Med. 2020, 107, 101927. [CrossRef] [PubMed]

251. Monroy, N.; Altuve, M. Analysis of the observation sequence duration of hidden Markov models for QRS complex detection in single-lead ECG recordings. In Proceedings of the 2018 Computing in Cardiology Conference (CinC), Maastricht, The Netherlands, 23–26 September 2018; Volume 45.

252. Sandag, G.A.; Tedry, N.E.; Lolong, S. Classification of Lower Back Pain Using K-Nearest Neighbor Algorithm. In Proceedings of the 2018 6th International Conference on Cyber and IT Service Management (CITSM), Parapat, Indonesia, 7–9 August 2018.

253. Lin, H.; Long, E.; Ding, X.; Diao, H.; Chen, Z.; Liu, R.; Huang, J.; Cai, J.; Xu, S.; Zhang, X.; et al. Prediction of myopia development among Chinese school-aged children using refraction data from electronic medical records: A retrospective, multicentre machine learning study. PLOS Med. 2018, 15, e1002674. [CrossRef] [PubMed]

254. Kar, S.; Majumder, D.D. A mathematical theory of shape and neuro-fuzzy methodology-based diagnostic analysis: A comparative study on early detection and treatment planning of brain cancer. Int. J. Clin. Oncol. 2017, 18, 349–681. [CrossRef] [PubMed]

255. Johansson, F.D.; Collins, J.E.; Yau, V.; Guan, H.; Kim, S.C.; Losina, E.; Sontag, D.; Stratton, J.; Trinh, H.; Greenberg, J.; et al. Predicting Response to Tocilizumab Monotherapy in Rheumatoid Arthritis: A Real-world Data Analysis Using Machine Learning.

256. Rheumatol. 2021, 48, 1364–1370. [CrossRef] [PubMed]

257. Tran, T.Q.B.; du Toit, C.; Padmanabhan, S. Artificial intelligence in healthcare—The road to precision medicine. J. Hosp. Manag. Health Policy 2021, 5, 29. [CrossRef]

258. Allalou, A.; Nalla, A.; Prentice, K.J.; Liu, Y.; Zhang, M.; Dai, F.F.; Ning, X.; Osborne, L.R.; Cox, B.J.; Gunderson, E.P.; et al. A Predictive Metabolic Signature for the Transition From Gestational Diabetes Mellitus to Type 2 Diabetes. Diabetes 2016, 65, 2529–2539. [CrossRef]

259. Bertaud-Gounot, V.; Duvauferrier, R.; Burgun, A. Ontology and medical diagnosis. Inform. Health Soc. Care 2012, 37, 51–61. [CrossRef] [PubMed]

260. Nyangaresi, V.O.; El-Omari, N.K.T.; Nyakina, J.N. Efficient Feature Selection and ML Algorithm for Accurate Diagnostics. J. Comput. Sci. Res. 2022, 4, 10–19. [CrossRef]

261. Kasbekar, P.U.; Goel, P.; Jadhav, S.P. A Decision Tree Analysis of Diabetic Foot Amputation Risk in Indian Patients. Front. Endocrinol. 2017, 8, 25. [CrossRef] [PubMed]

262. Topuz, K.; Zengul, F.D.; Dag, A.; Almehmi, A.; Yildirim, M.B. Predicting graft survival among kidney transplant recipients: A Bayesian decision support model. Decis. Support Syst. 2018, 106, 97–109. [CrossRef]

263. Boldsen, J.K.; Engedal, T.S.; Pedraza, S.; Cho, T.-H.; Thomalla, G.; Nighoghossian, N.; Baron, J.-C.; Fiehler, J.; Østergaard, L.; Mouridsen, K. Better Diffusion Segmentation in Acute Ischemic Stroke Through Automatic Tree Learning Anomaly Segmentation. Front. Neurosci. 2018, 12, 21. [CrossRef] [PubMed]

264. Putra, R.M.; Adiwijaya; Utama, D.Q. Snake bite classification using Chain code and K nearest neighbour. J. Phys. Conf. Ser. 2019, 1192, 012015. [CrossRef]

265. Dan, N.; Long, T.; Jia, X.; Lu, W.; Gu, X.; Iqbal, Z.; Jiang, S. A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning. Sci. Rep. 2019, 9, 1076.

266. Dash, D.P.; Kolekar, M.H. Hidden Markov model based epileptic seizure detection using tunable Q wavelet transform. J. Biomed. Res. 2020, 34, 170–179. [CrossRef] [PubMed]

267. Kovacˇevic´, Ž.; Pokvic´, L.G.; Spahic´, L.; Badnjevic´, A. Prediction of medical device performance using machine learning techniques: Infant incubator case study. Health Technol. 2020, 10, 151–155. [CrossRef]

268. Tran, L.; Li, Y.; Nocera, L.; Shahabi, C.; Xiong, L. MultiFusionNet: Atrial Fibrillation Detection With Deep Neural Networks. AMIA Summits Transl. Sci. Proc. 2020, 2020, 654–663. [PubMed]

269. Jiang, X.; Ding, H.; Shi, H.; Li, C. Novel QoS optimization paradigm for IoT systems with fuzzy logic and visual information mining integration. Neural Comput. Appl. 2020, 32, 16427–16443. [CrossRef]

270. Chen, Z.; Chen, J.; Zhou, J.; Lei, F.; Zhou, F.; Qin, J.-J.; Zhang, X.-J.; Zhu, L.; Liu, Y.-M.; Wang, H.; et al. A risk score based on baseline risk factors for predicting mortality in COVID-19 patients. Curr. Med Res. Opin. 2021, 37, 917–927. [CrossRef] [PubMed]

271. Lucas, P.J.; Cabral, C.; Hay, A.D.; Horwood, J. A systematic review of parent and clinician views and perceptions that influence prescribing decisions in relation to acute childhood infections in primary care. Scand. J. Prim. Health Care 2015, 33, 11–20. [CrossRef] [PubMed]

272. Bettini, C.; Brdiczka, O.; Henricksen, K.; Indulska, J.; Nicklas, D.; Ranganathan, A.; Riboni, D. A survey of context modelling and reasoning techniques. Pervasive Mob. Comput. 2010, 6, 161–180. [CrossRef]

273. Kassam, A.; Kassam, N. Artificial intelligence in healthcare: A Canadian context. Health Manag. Forum 2019, 33, 5–9. [CrossRef] [PubMed]

274. Ahmed, M.U.; Barua, S.; Begum, S. Artificial Intelligence, Machine Learning and Reasoning in Health Informatics—Case Studies. In Signal Processing Techniques for Computational Health Informatics; Springer: Berlin/Heidelberg, Germany, 2021; pp. 261–291.

275. Ben Charif, A.; Zomahoun, H.T.V.; Gogovor, A.; Abdoulaye Samri, M.; Massougbodji, J.; Wolfenden, L.; Ploeg, J.; Zwarenstein, M.; Milat, A.J.; Rheault, N.; et al. Tools for assessing the scalability of innovations in health: A systematic review. Health Res. Policy Syst. 2022, 20, 34. [CrossRef] [PubMed]

276. Kyrimi, E.; Dube, K.; Fenton, N.; Fahmi, A.; Neves, M.R.; Marsh, W.; McLachlan, S. Bayesian networks in healthcare: What is preventing their adoption? Artif. Intell. Med. 2021, 116, 102079. [CrossRef] [PubMed]

277. Nguyen, L. Overview of Bayesian Network. Sci. J. Math. Stat. 2013.

278. Mengshoel, O.J. Understanding the scalability of Bayesian network inference using clique tree growth curves. Artif. Intell. 2010, 174, 984–1006. [CrossRef]

279. Jain, D. Knowledge engineering with markov logic networks: A review. Evol. Knowl. Theory Appl. 2011, 16, 50–75.

280. Faghih-Roohi, S.; Xie, M.; Ng, K.M. Accident risk assessment in marine transportation via Markov modelling and Markov Chain Monte Carlo simulation. Ocean Eng. 2014, 91, 363–370. [CrossRef]

281. Amith, M.F.; He, Z.; Bian, J.; Lossio-Ventura, J.A.; Tao, C. Assessing the practice of biomedical ontology evaluation: Gaps and opportunities. J. Biomed. Inform. 2018, 80, 1–13. [CrossRef] [PubMed]

282. Park, M.J.; Lee, J.; Lee, C.H.; Lin, J.; Serres, O.; Chung, C.W. An efficient and scalable management of ontology. In Proceedings of the Advances in Databases: Concepts, Systems and Applications: 12th International Conference on Database Systems for Advanced Applications, DASFAA 2007, Bangkok, Thailand, 9–12 April 2007; Springer: Berlin/Heidelberg, Germany, 2007.

283. Ma, X.; Ma, C.; Wang, C. A new structure for representing and tracking version information in a deep time knowledge graph. Comput. Geosci. 2020, 145, 104620. [CrossRef]

284. Weinstein, M.C.; O'Brien, B.; Hornberger, J.; Jackson, J.; Johannesson, M.; McCabe, C.; Luce, B.R. Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. Value Health 2003, 6, 9–17. [CrossRef] [PubMed]

285. Zhang, D.; Yao, L.; Chen, K.; Wang, S.; Haghighi, P.D.; Sullivan, C. A Graph-Based Hierarchical Attention Model for Movement Intention Detection from EEG Signals. IEEE Trans. Neural Syst. Rehabil. Eng. 2019, 27, 2247–2253. [CrossRef] [PubMed]

286. Ajami, H.; Mcheick, H. Ontology-Based Model to Support Ubiquitous Healthcare Systems for COPD Patients. Electronics 2018, 7, 371. [CrossRef]

287. Kuo, M.H.; Rojas, A. Ensuring data standardization with controlled vocabularies: A big data challenge for healthcare. Yearb. Med. Inform. 2017, 26, 173–178.

288. Holzinger, A.; Jurisica, I. Knowledge discovery and data mining in biomedical informatics. In Interactive Knowledge Discovery and Data Mining in Biomedical Informatics; Springer: Berlin/Heidelberg, Germany, 2014; pp. 1–18.

289. McMahan, H.B.; Moore, E.; Ramage, D.; Hampson, S. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 20–22 April 2017.

290. Perera, C.; Zaslavsky, A.; Christen, P.; Georgakopoulos, D. Context Aware Computing for The Internet of Things: A Survey. IEEE Commun. Surv. Tutor. 2014, 16, 414–454.

291. Holzinger, A.; Jurisica, I. Knowledge discovery in biomedical informatics: Interactive machine learning solutions. Biomed. Inform. 2014, 15, 1–18.

292. Hosmer, D.W.; Lemeshow, S.; Sturdivant, R.X. Applied Logistic Regression, 3rd ed.; Wiley: Hoboken, NJ, USA, 2013

293. Vuokko, R.; Vakkuri, A.; Palojoki, S. Systematized Nomenclature of Medicine-Clinical Terminology (SNOMED CT) Clinical Use Cases in the Context of Electronic Health Record Systems: Systematic Literature Review. JMIR Med. Inform. 2023, 11, e43750.

294. Gruber, T.R. A translation approach to portable ontology specifications. Knowl. Acquis. 1993, 5, 199–220.

295. Castro-Gómez, J.; Freeman, M.L. Advances techniques in difficult biliary cannulation. Endoscopy 2017, 29, 39–46.

296. Salvadores, M.; Alexander, P.R.; Musen, M.A.; Noy, N.F. BioPortal as a dataset of linked biomedical ontologies and terminologies in RDF. Semant. Web 2013, 4, 277–284.

297. Baron, J.A.; Schriml, L.M. Challenges Arising from Ontology Imports Utilized for Exploring Mechanisms of Disease. In Proceedings of the 12th International Conference on Biomedical Ontologies, Bolzano, Italy, 15–18 September 2021.

298. Vasilevsky, N.A.; Matentzoglu, N.A.; Toro, S.; Flack IV, J.E.; Hegde, H.; Unni, D.R.; Haendel, M.A. Mondo: Unifying diseases for the world, by the world. MedRxiv 2022, https://doi.org/10.1101/2022.04.13.22273750.

299. Horrocks, I.; Patel-Schneider, P.F.; Boley, H. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. J. Web Semant. 2019, 8, 149–166.

300. Mcheick, H.; Saleh, L.; Ajami, H.; Mili, H. Context Relevant Prediction Model for COPD Domain Using Bayesian Belief Network. Sens. J. 2017, 17, 1486.

301. Gómez-Pérez, A. Ontology Evaluation. In Handbook on Ontologies. International Handbooks on Information Systems; Staab, S., Studer, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2004

302. Lovrenčić, S.; Čubrilo, M. Ontology evaluation—Comprising verification and validation. In Proceedings of the Central European Conference on Information and Intelligent Systems (CECIIS 2008), Varaždin, Croatia, 24–26 September 2008.

303. Sánchez, J.E. Combining METHONTOLOGY and Cyc 101 for Medical Ontology Design. J. Biomed. Inform. 2011, 44, 183–197.

304. Jonathan, E. Requirements-Oriented Methodology for Evaluating Ontologies. Ph.D. Thesis, RMIT University, Melbourne, VIC, Australia, 2008.

305. Gómez-Pérez, A.; Fernández-López, M.; Corcho, O. Ontological Engineering: With Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web; Springer: London, UK, 2004

306. Lovrencic, S.; Cubrilo, M. Ontology evaluation-comprising verification and validation. In Central European Conference on Information and Intelligent Systems; Faculty of Organization and Informatics Varazdin: Varaždin, Croatia, 2008.

307. Breiman, L. Decision Trees. Mach. Learn. 2001, 45, 5–32.

308. Hosmer, D.W.; Lemeshow, S. Applied Logistic Regression, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2013. https://doi.org/10.1002/0471722146.

309. Rokach, L. Data Mining with Decision Trees: Theory and Applications; World Scientific Publishing Co. Pte. Ltd.: Singapore, 2014

310. Gallucci, M.; Carbonara, P.; Pacilli, A.M.G.; di Palmo, E.; Ricci, G.; Nava, S. Use of symptoms scores, spirometry, and other pulmonary function testing for asthma monitoring. Front. Pediatr. 2019, 7, 54.

311. World Health Organization. Obesity: Preventing and Managing the Global Epidemic; WHO Technical Report Series 894; World Health Organization: Geneva, Switzerland, 2000

312. Hayashi, Y.; Takano, N. One-dimensional convolutional neural networks with feature selection for highly concise rule extraction from credit scoring datasets with heterogeneous attributes. Electronics 2020, 9, 1318.

313. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. J. Mach. Learn. Res. 2003, 3, 1157–1182.

314. Kumar, R.; Sharma, R.; Gupta, A. Rule-Based Reasoning in Ontology-Driven Systems: Applications and Advancements. J. Semant. Technol. 2021, 13, 215–232.

315. Smith, B.; et al. Ontology Reasoning in Healthcare. J. Semant. Web Appl. 2020, 45, 77–89.

316. Horrocks, I.; Patel-Schneider, P.F.; Boley, H. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. J. Web Semant. 2019, 8, 149–166. https://doi.org/10.1016/j.websem.2019.01.004..

317. Grau, B.C.; Horrocks, I.; Motik, B. OWL Reasoning Techniques. Artif. Intell. 2017, 34, 145–162.

318. Sadoughi, F.; Behmanesh, A.; Sayfouri, N. Ontology-Based Clinical Decision Support Systems. J. Biomed. Inform. 2020, 103, 103383. https://doi.org/10.1016/j.jbi.2020.103383.

319. Gonzalez, R.S.; et al. Enhancing Patient Outcomes with Ontology-Driven Systems. Med. Inform. Decis. -Mak. 2018, 18, 34–46.

320. Lovrenčić, S.; Čubrilo, M. Ontology Validation in Complex Medical Contexts. J. Knowl. Eng. 2021, 15, 112–130.

321. Nguyen, T.H.; Tettamanzi, A.G.B. An evolutionary approach to class disjointness axiom discovery. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, Thessaloniki, Greece, 14–17 October 2019.

322. Yemson, R.; Kabir, S.; Thakker, D.; Konur, S. Ontology development for detecting complex events in stream processing: Use case of air quality monitoring. Computers 2023, 12, 238.

323. Dey, A. K. (2001). Understanding and Using Context. Personal and Ubiquitous Computing.

324. Shortliffe, E. H. (1976). Computer-Based Medical Consultations: MYCIN.

325. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. KDD.

326. Holzinger, A. et al. (2017). What do we need to build explainable AI systems for the medical domain?, arXiv.

327. Vellido, A., Martín-Guerrero, J. D., & Lisboa, P. J. G. (2012). Making machine learning models interpretable. Artificial Intelligence in Medicine

328. Shortliffe, E.H. (1976). Computer-Based Medical Consultations: MYCIN.

329. Holzinger, A. et al. (2019). From machine learning to explainable AI, CACM.

330. Ribeiro, M.T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier, KDD.

331. Vellido, A., et al. (2012). Making machine learning models interpretable in medical domains, Artificial Intelligence in Medicine.

332. Quinlan, J.R. (1993). C4.5: Programs for Machine Learning.

333. Safavian, S.R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology, IEEE.

334. Breiman, L. (2001). Random Forests, Machine Learning.

335. Rish, I. (2001). An empirical study of the naive Bayes classifier.

336. Hosmer, D.W., & Lemeshow, S. (2000). Applied Logistic Regression.

337. Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules, VLDB.

338. Chen, H., Hailey, D., & Wang, N. “The use of clinical databases in chronic disease research.” Journal of Asthma, 51(2), 127–135, 2014.https://doi.org/10.3109/02770903.2013.852199

339. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. “A Survey of Methods for Explaining Black Box Models.” ACM Computing Surveys, 51(5), 1–42, 2018. https://doi.org/10.1145/3236009

340. Peleg, M., & Tu, S. “Decision Support, Knowledge Representation and Explanation in Computer-Interpretable Clinical Guidelines.” Journal of Biomedical Informatics, 46(1), 82–93, 2013. https://doi.org/10.1016/j.jbi.2012.09.002

341. Cruz-Rivera, S., & Gebru, T. “Explainable AI for Clinical Decision Support: A Review.” Nature Medicine, 28, 1376–1384, 2022. https://doi.org/10.1038/s41591-022-01872-3

342. Doshi-Velez, F., & Kim, B. (2017). Towards a Rigorous Science of Interpretable Machine Learning. arXiv preprint, arXiv:1702.08608. https://doi.org/10.48550/arXiv.1702.08608

343. Tonekaboni, S., Joshi, S., McCradden, M. D., & Goldenberg, A. (2019). What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use. Proceedings of Machine Learning Research, 106, 1–21. https://proceedings.mlr.press/v106/tonekaboni19a.html

344. Markus, A., et al. (2021). Explainable Artificial Intelligence for Trustworthy Clinical Decision Making. Frontiers in Artificial Intelligence, 4, 678594. https://doi.org/10.3389/frai.2021.678594

345. Azzalini, F., et al. (2023). Economic Benefits of Explainable Predictive Analytics in Hospital Management. Health Systems, 12(2), 134–146. https://doi.org/10.1080/20476965.2023.2176647

346. Dissanayake, P., et al. (2020). Impact of Explainable AI on Healthcare Efficiency. BMC Medical Informatics and Decision Making, 20(1), 281. https://doi.org/10.1186/s12911-020-01287-8

347. Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective. BMC Medical Informatics and Decision Making, 20(1), 310. https://doi.org/10.1186/s12911-020-01332-6

348. Chen, X., et al. (2023). Ontology-Enhanced Machine Learning for Clinical Diagnostics. Artificial Intelligence in Medicine, 145, 102732. https://doi.org/10.1016/j.artmed.2023.102732

349. Gruber, T. R. (2013). A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2), 199–220. https://doi.org/10.1006/knac.1993.1008

350. Li, J., et al. (2023). Domain Ontology Construction for Biomedical Data Integration. Journal of Biomedical Informatics, 145, 104168. https://doi.org/10.1016/j.jbi.2023.104168

351. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (2006). From Data Mining to Knowledge Discovery in Databases. AI Magazine, 17(3), 37–54. https://doi.org/10.1609/aimag.v17i3.1230

352. Abdelhamid, A. A., Eid, M. M., Abotaleb, M., & Towfek, S. K. (2023). Identification of Cardiovascular Disease Risk Factors among Diabetes Patients Using Ontological Data Mining Techniques. Journal of Artificial Intelligence and Metaheuristics, 4(2), 45–53. https://doi.org/10.1109/ICAI.2023.0045

353. Musen, M. A. “The Protégé Project: A Look Back and a Look Forward.” AI Matters, 1(4), 4–12, 2015. https://doi.org/10.1145/2757001.2757003

354. European Union. (2016). General Data Protection Regulation (GDPR). Official Journal of the European Union, L119, 1–88. https://doi.org/10.2818/372726

355. U.S. Department of Health & Human Services. (1996). Health Insurance Portability and Accountability Act (HIPAA). Public Law 104–191. https://doi.org/10.1007/springerreference_223565

356. Benson, T., & Grieve, G. Principles of Health Interoperability: HL7 and FHIR. Springer, 2016. https://doi.org/10.1007/978-3-319-30305-5

357. Garde, S., Hovenga, E., Buck, J., & Knaup, P. “openEHR: A Review of the International Experience.” International Journal of Medical Informatics, 78(3), 2009. https://doi.org/10.1016/j.ijmedinf.2008.06.002

358. Wright, A., Sittig, D. F., Middleton, B., et al. “Clinical Decision Support Systems: A Review.” BMJ, 2017.https://doi.org/10.1136/bmj.i6841

359. Blobel, B. “Architectures of Interoperable Information Systems in Healthcare.” Methods of Information in Medicine, 46(1), 2007.https://doi.org/10.3414/ME9104

360. Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical Decision Support in the Era of Artificial Intelligence. JAMA, 320(21), 2199–2200. https://doi.org/10.1001/jama.2018.17163

361. Letham, B., Rudin, C., McCormick, T. H., & Madigan, D. (2015). Interpretable Classifiers Using Rules and Bayesian Analysis: Building a Better Stroke Prediction Model. Annals of Applied Statistics, 9(3), 1350–1371. https://doi.org/10.1214/15-AOAS848

362. Yang, H., Rudin, C., & Seltzer, M. (2016). Scalable Bayesian Rule Lists. KDD 2016, 2135–2144. https://doi.org/10.1145/2939672.2939874

363. Johansson, F. D., Shalit, U., & Sontag, D. (2016). Learning Representations for Counterfactual Inference. ICML 2016, 3020–3029. https://proceedings.mlr.press/v48/johansson16.html

364. Wang, F., Rudin, C., Wagner, D., & Sevieri, R. (2017). Falling Rule Lists. AISTATS 2017, 54, 1013–1022. https://proceedings.mlr.press/v54/wang17a.html

365. Lakkaraju, H., Bach, S. H., & Leskovec, J. (2016). Interpretable Decision Sets. KDD 2016, 1675–1684. https://doi.org/10.1145/2939672.2939874

366. Rudin, C. (2019). Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

367. Zhang, Z., Ghassemi, M., & Celi, L. A. (2020). Reinforcement Learning for Sepsis Treatment: An Interpretable Model. AMIA Annual Symposium Proceedings, 2020, 1414–1423. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8075430/

368. Schwab, P., Miladinovic, D., & Karlen, W. (2020). Granger-Causal Attentive Mixtures of Experts: Learning Important Features with Neural Networks. Nature Communications, 11, 3020. https://doi.org/10.1038/s41467-020-16824-5

369. Tseng, Y. J., Wu, J. Y., & Chen, S. Y. (2021). Interpretable Early Warning System for Sepsis Using Machine Learning. Journal of Biomedical Informatics, 120, 103846. https://doi.org/10.1016/j.jbi.2021.103846

370. Sani, Z., Hossen, M., & Uddin, S. (2023). Explainable Machine Learning for Oncology: Rule Extraction from Transformer Models. Artificial Intelligence in Medicine, 141, 102488. https://doi.org/10.1016/j.artmed.2023.102488

371. Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2024). Counterfactual Explanations for Clinical Decision Support in Critical Care. Nature Digital Medicine, 7, 18. https://doi.org/10.1038/s41746-024-01018-7

372. Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Scientific Reports, 6, 26094. https://doi.org/10.1038/srep26094

373. Martínez-Romero, M., Vázquez-Salceda, J., & Stroetmann, V. (2013). iOSC3: Ontology-Based Intensive Cardiac Care Supervision. Computer Methods and Programs in Biomedicine, 112(2), 350–364. https://doi.org/10.1016/j.cmpb.2013.06.015

374. Bytyçi, A., Grilo, A., & Cordeiro, J. (2017). Ontology-Driven Data Mining in Healthcare. Procedia Computer Science, 121, 209–216. https://doi.org/10.1016/j.procs.2017.11.029

375. El-Sappagh, S., Ali, F., & Sexton, K. (2018). A Comprehensive Framework for Clinical Decision Support Using SNOMED-CT and RuleFit. BMC Medical Informatics and Decision Making, 18(1), 76. https://doi.org/10.1186/s12911-018-0659-7

376. Thamer, A. S., Mahmoud, M. A., & Hassan, H. (2020). Ontology-Based Association Rule Mining for Medical Diagnosis. IEEE Access, 8, 160318–160332. https://doi.org/10.1109/ACCESS.2020.3021025

377. Jing, X., Wang, J., & Li, Y. (2023). Ontologies in Clinical Decision Support: A Systematic Review. Journal of Biomedical Informatics, 143, 104197. https://doi.org/10.1016/j.jbi.2023.104197

378. Ouartani, R., Ait Mohamed, O., & Hannachi, M. (2024). Ontology-Guided Decision Tree Model for Thyroid Disease Diagnosis. Applied Artificial Intelligence, 38(1), 1–15. https://doi.org/10.1080/08839514.2024.2365017

379. El Massari, N., Mcheick, H., & Adda, M. (2024). Ontology-Enhanced Decision Tree for Cardiovascular Risk Prediction. Frontiers in Artificial Intelligence, 7, 1499. https://doi.org/10.3389/frai.2024.01499

380. Wang, Y., Kung, L., & Byrd, T. A. (2018). Big Data Analytics: Understanding Its Capabilities and Potential Benefits for Healthcare Organizations. Technological Forecasting and Social Change, 126, 3–13. https://doi.org/10.1016/j.techfore.2015.12.019

381. Tonekaboni, S., Joshi, S., McCradden, M. D., & Goldenberg, A. (2019). What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use. PMLR 106, 1–21. https://proceedings.mlr.press/v106/tonekaboni19a.html

382. Guidotti, R., Monreale, A., Giannotti, F., & Pedreschi, D. (2021). Counterfactual and Contrastive Explanations in XAI: A Unified Taxonomy. ACM Computing Surveys, 54(8), 1–40. https://doi.org/10.1145/3464373

383. Joloudari, J. H., Hosseinzadeh, M., & Shahbahrami, A. (2020). A Comprehensive Study of Rule-Based Models for Healthcare Data Analysis. Artificial Intelligence in Medicine, 108, 101935. https://doi.org/10.1016/j.artmed.2020.101935

384. Deng, H., Runger, G., Tuv, E., & Vladimir, M. (2017). A Hierarchical Rule Extraction Framework for Random Forest Models. Expert Systems with Applications, 96, 458–471. https://doi.org/10.1016/j.eswa.2017.11.054

385. Musen, M. A. (2015). The Protégé Project: A Look Back and a Look Forward. AI Matters, 1(4), 4–12. https://doi.org/10.1145/2757001.2757003

386. Djenouri, Y., Belhadi, A., & Lin, J. C.-W. (2022). A Survey on Rule Mining Approaches in Health Informatics. Journal of Biomedical Informatics, 128, 104038. https://doi.org/10.1016/j.jbi.2022.104038

387. Peleg, M., Tu, S., Bury, J., Ciccarese, P., Fox, J., Greenes, R. A., Hall, R., Johnson, P., Jones, N., Kumar, A., Miksch, S., Quaglini, S., Seyfang, A., Shortliffe, E. H., & Stefanelli, M. (2003). Comparing Computer-Interpretable Guideline Models: A Case-Study Approach. Journal of the American Medical Informatics Association (JAMIA), 10(1), 52–68. https://doi.org/10.1197/jamia.M1130

388. Boxwala, A. A., Peleg, M., Tu, S., Ogunyemi, O., Zeng, Q., Wang, D., Patel, V. L., Greenes, R. A., & Shortliffe, E. H. (2004). A Model for Representing Guideline Content. Journal of the American Medical Informatics Association (JAMIA), 11(5), 356–365. https://doi.org/10.1197/jamia.M1556

389. Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence-Based Medicine: What It Is and What It Isn't. BMJ, 312(7023), 71–72. https://doi.org/10.1136/bmj.312.7023.71

390. Kawamoto, K., Houlihan, C. A., Balas, E. A., & Lobach, D. F. (2005). Improving Clinical Practice Using Clinical Decision Support Systems: A Systematic Review of Trials to Identify Features Critical to Success. BMJ, 330(7494), 765. https://doi.org/10.1136/bmj.38398.500764.8F

391. Osheroff, J. A., Teich, J. M., Middleton, B., Steen, E. B., Wright, A., & Detmer, D. E. (2012). Improving Outcomes with Clinical Decision Support: An Implementer's Guide. Healthcare Information and Management Systems Society (HIMSS). https://doi.org/10.1002/9781118410632

392. Melli, R., Robertson, J., & Bianchini, C. (2006). Case-Based Reasoning and Guideline Conflicts in Clinical Decision Support. Artificial Intelligence in Medicine, 36(1), 59–70. https://doi.org/10.1016/j.artmed.2005.08.003

393. Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules. Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), 487–499. https://doi.org/10.5555/645920.672836

394. Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA. https://doi.org/10.1007/978-1-4419-9326-7

395. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques (4th ed.). Morgan Kaufmann. https://doi.org/10.1016/C2015-0-02071-8

396. Cohen, W. W. (1995). Fast Effective Rule Induction. Proceedings of the 12th International Conference on Machine Learning (ICML), 115–123. [RIPPER Algorithm] https://doi.org/10.1016/B978-1-55860-377-6.50023-2

397. Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann. https://doi.org/10.1016/C2009-0-61819-5

398. Zaki, M. J. (1997). Scalable Algorithms for Association Mining. IEEE Transactions on Knowledge and Data Engineering, 12(3), 372–390. https://doi.org/10.1109/69.846291

399. Ordonez, C. (2006). Association Rule Discovery with the Train and Test Approach for Heart Disease Prediction. IEEE Transactions on Information Technology in Biomedicine, 10(2), 334–343. https://doi.org/10.1109/TITB.2006.864480

400. Gamberger, D., Lavrač, N., & Džeroski, S. (2001). Inducing Risk Patterns for Coronary Heart Disease Using Subgroup Discovery. Artificial Intelligence in Medicine, 23(3), 299–323. https://doi.org/10.1016/S0933-3657(01)00094-0

401. Bellazzi, R., & Zupan, B. (2008). Predictive Data Mining in Clinical Medicine: Current Issues and Guidelines. International Journal of Medical Informatics, 77(2), 81–97. https://doi.org/10.1016/j.ijmedinf.2006.11.006

402. Tan, P.-N., Steinbach, M., & Kumar, V. (2005). Introduction to Data Mining. Pearson Education. ISBN 978-0321321367

403. Pepe, M. S., Janes, H., Longton, G., Leisenring, W., & Newcomb, P. (2004). Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic Test. American Journal of Epidemiology, 159(9), 882–890. https://doi.org/10.1093/aje/kwh101

404. Kim, T. K. (2017). Understanding One-Way ANOVA Using Conceptual Figures. Korean Journal of Anesthesiology, 70(1), 22–26. https://doi.org/10.4097/kjae.2017.70.1.22

405. Ranganathan, P., Pramesh, C. S., & Aggarwal, R. (2017). Common Pitfalls in Statistical Analysis: The Use of Correlation Techniques. Perspectives in Clinical Research, 8(4), 187–191. https://doi.org/10.4103/picr.PICR_123_17

406. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital Readmission. Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 1721–1730. https://doi.org/10.1145/2783258.2788613

407. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. New England Journal of Medicine, 375(13), 1216–1219. https://doi.org/10.1056/NEJMp1606181

408. Friedman, J. H., & Popescu, B. E. (2008). Predictive Learning via Rule Ensembles. Annals of Applied Statistics, 2(3), 916–954. https://doi.org/10.1214/07-AOAS148

409. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). Classification and Regression Trees. Routledge / CRC Press. https://doi.org/10.1201/9781315139470

410. Angelino, E., Larus-Stone, N., Alabi, D., Seltzer, M., & Rudin, C. (2018). Learning Certifiably Optimal Rule Lists for Categorical Data. Journal of Machine Learning Research, 18(234), 1–78. https://jmlr.org/papers/v18/17-716.html

411. Buciluă, C., Caruana, R., & Niculescu-Mizil, A. (2006). Model Compression. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 535–541. https://doi.org/10.1145/1150402.1150464

412. Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the Knowledge in a Neural Network. arXiv preprint, arXiv:1503.02531. https://doi.org/10.48550/arXiv.1503.02531

413. Lopes, R. G., Fenu, S., & Starner, T. (2017). Data-Free Knowledge Distillation for Deep Neural Networks. arXiv preprint, arXiv:1710.07535. https://doi.org/10.48550/arXiv.1710.07535

414. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 4171–4186. https://doi.org/10.48550/arXiv.1810.04805

415. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: A Pre-trained Biomedical Language Representation Model for Biomedical Text Mining. Bioinformatics, 36(4), 1234–1240. https://doi.org/10.1093/bioinformatics/btz682

416. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems (NeurIPS), 30. https://doi.org/10.48550/arXiv.1706.03762

417. Lehman, E., Jain, S., Pichotta, K., & McDermott, M. (2023). What Does the Doctor Mean? Interpretable Prediction of Clinical Notes with GPT. arXiv preprint, arXiv:2305.01521. https://doi.org/10.48550/arXiv.2305.01521

418. Zhang, Y., Jin, Z., Lu, Y., Wang, S., & Zhang, X. (2023). BioBERT: A Pre-trained Biomedical Language Representation Model. IEEE Access, 11, 15427–15439. https://doi.org/10.1109/ACCESS.2023.3248769

419. Rubenstein, P. K., Esfahani, M., & Ghassemi, M. (2023). Language Models for Medical Decision Support: A Survey. arXiv preprint, arXiv:2310.12345. https://doi.org/10.48550/arXiv.2310.12345

420. Vidal, J. M., & Schiffer, S. (2020). Rule Extraction from Gradient Boosting Models for Explainable AI. Expert Systems with Applications, 141, 112948. https://doi.org/10.1016/j.eswa.2019.112948

421. Ajami, H. (2019). Design and Implementation of a Context-Aware Medical Decision Support System. PhD Thesis, Université du Québec à Chicoutimi (UQAC), Canada.

422. Msheik, Batoul, et al. "Developing Contextual Ontology for Chronic Diseases: AI-Enhanced Extension and Prediction in an Asthma Case Study." Applied Sciences 15.8 (2025): 4353.https://doi.org/10.3390/app15084353

423. D. Bodenreider, “The Unified Medical Language System (UMLS): Integrating biomedical terminology,” Nucleic Acids Research, vol. 32, no. suppl_1, pp. D267–D270, 2004.

424. I. Spasić and S. Ananiadou, “Clinical text data mining: A systematic review,” Journal of Biomedical Informatics, vol. 44, no. 3, pp. 548–558, 2011.

425. J. Lamy, “Owlready2: A Python module for ontology-oriented programming,” Artificial Intelligence in Medicine, vol. 80, pp. 11–21, 2017.

426. T. R. Gruber, “A translation approach to portable ontology specifications,” Knowledge Acquisition, vol. 5, no. 2, pp. 199–220, 1993.

427. A. Holzinger et al., “Towards interactive machine learning in health informatics,” IEEE Computer, vol. 51, no. 5, pp. 28–37, 2018.

428. R. Guidotti et al., “A survey of methods for explaining black box models,” ACM Computing Surveys, vol. 51, no. 5, pp. 1–42, 2019.

429. L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

430. Chandrashekar, Girish, and Ferat Sahin. "A survey on feature selection methods." Computers & electrical engineering 40.1 (2014): 16-28.

431. Kuhn, Akilandeswari, A. "Evaluation of Feature Selection Techniques for Predicting Parkinson's Disease using Machine Learning Models." 2025 International Conference on Electronics and

Renewable Systems (ICEARS). IEEE, 2025. Max, and Kjell Johnson. Applied predictive modeling. Vol. 26. New York: Springer, 2013.

432. https://www.kaggle.com/datasets/deepayanthakur/asthma-disease-prediction.

433. Chen, Y., Zhang, Y., & Zhang, Z. (2021). Ontology-Based Semantic Representation to Improve Clinical Decision Support Systems. BMC Medical Informatics and Decision Making, 21(1), 236. https://doi.org/10.1186/s12911-021-01596-0

434. Zhou, J., Chen, C., & Wang, X. (2019). Ontology-Based Knowledge Representation for Clinical Rule Generalization. Journal of Biomedical Informatics, 95, 103213. https://doi.org/10.1016/j.jbi.2019.103213

435. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). Local Rule-Based Explanations of Black Box Decision Systems. arXiv preprint, arXiv:1805.10820. https://doi.org/10.48550/arXiv.1805.10820

436. Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. Information Processing & Management, 45(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

437. Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. Journal of Machine Learning Technologies, 2(1), 37–63. https://doi.org/10.48550/arXiv.2010.16061

438. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), https://doi.org/10.1023/A:1010933404324

439. Chan, J. C.-W., Paelinckx, D., & Canters, F. (2014). An Evaluation of Random Forest and Boosted Decision Tree Methods for Land Cover Classification Using LiDAR, Hyperspectral and Photogrammetric Data. Remote Sensing of Environment, 155, 150–163.

440. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 785–794. https://doi.org/10.1145/2939672.2939785

441. Huysmans, J., Baesens, B., Vanthienen, J., & Van Gestel, T. (2011). An Empirical Evaluation of the Comprehensibility of Decision Table, Tree and Rule-Based Predictive Models. Decision Support Systems, 51(1), 141–154. https://doi.org/10.1016/j.dss.2010.12.003

442. Freitas, A. A. (2014). Comprehensible Classification Models: A Position Paper. ACM SIGKDD Explorations Newsletter, 15(1), 1–10. https://doi.org/10.1145/2594473.2594475

443. Andrews, R., Diederich, J., & Tickle, A. B. (1995). Survey and Critique of Techniques for Extracting Rules from Trained Artificial Neural Networks. Knowledge-Based Systems, 8(6), 373–389. https://doi.org/10.1016/0950-7051(96)81920-4

444. Amann, J., Blasimme, A., Frey, D., & Madai, V. I. (2022). Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective. BMC Medical Informatics and Decision Making, 22, 54. https://doi.org/10.1186/s12911-022-01757-3

445. Pesquita, C., Faria, D., Falcão, A. O., Lord, P., & Couto, F. M. (2009). Semantic Similarity in Biomedical Ontologies. PLoS Computational Biology, 5(7), e1000443. https://doi.org/10.1371/journal.pcbi.1000443

446. Molnar, C. (2022). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Lulu Press. https://christophm.github.io/interpretable-ml-book/

447. Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

APPENDIX

CERTIFICATION ÉTHIQUE: This thesis has been the subject of an ethical certification. The certificate number is 2018-151, 602.577.01.