2020

A novel framework for standardizing and digitizing clinical pathways in healthcare information systems

Alahmar, Ayman

A Novel Framework for Standardizing and Digitizing Clinical Pathways in Healthcare Information Systems

by

Ayman Alahmar

A thesis
presented to Lakehead University
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Thunder Bay, Ontario, Canada, 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Dr. Carson Kai-Sang Leung Professor, Dept. of Computer Science, University of Manitoba
Supervisor:	Dr. Rachid Benlamri Professor, Dept. of Software Engineering, Lakehead University
Internal Member:	Dr. Abdulsalam Yassine Assistant Professor, Dept. of Software Engineering, Lakehead University
Internal-External Member:	Dr. Salama Ikki Associate Professor, Dept. of Electrical Engineering, Lakehead University

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Most healthcare institutions are reorganizing their healthcare delivery systems based on Clinical Pathways (CPs). CPs are medical management plans designed to standardize medical activities, reduce cost, optimize resource usage, and improve quality of service. However, most CPs are still paper-based and not fully integrated with Health Information Systems (HISs). More CP automation research is therefore required to fully benefit from the practical potentials of CPs. The common theme of current research in this field is to connect CPs with Electronic Medical Record (EMR) systems. Such view positions EMRs at the centre of HISs. A major long-term objective of this research is the placement of CP systems at the centre of HISs, because within CPs lies the very heart of medical planning, treatment and impressions, including healthcare quality and cost factors. An important contribution to the realization of this objective is to develop an international CP-specific digital coding system, and to fully standardize and digitize CPs based on the Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT) medical terminology system. This makes CPs digitally visible and machine-readable. In addition, to achieve semantic interoperability of CPs, we propose a CP knowledge representation using ontology engineering and HL7 standard. Our proposed framework makes CP systems smoothly linkable across various HISs. To show the feasibility and potential of the proposed framework, we developed a prototype Clinical Pathway Management System (CPMS) based on CPs currently in use at hospitals. The results show that CPs can be fully standardized and digitized using SNOMED CT terms and codes, and the CPMS can work as an independent healthcare system, performing novel CP-related functions including useful decision-support tasks. Furthermore, CP data were captured without loss, which contributes to reducing missing patient data and improving the results of data mining algorithms in healthcare. Standardized CPs can also be easily compared for auditing and quality management. The proposed framework is promising, and contributes toward solving major challenges related to CP standardization, digitization, independence, and proper inclusion in today's modern computerized hospitals.

Acknowledgements

I would like to express my gratitude and special thanks to Dr. Rachid Benlamri for his supervision, guidance, and follow-up. I also would like to thank the Graduate Coordinator, Dr. Abdelhamid Tayebi, as well as Ph.D. committee members, Dr. Abdulsalam Yassine and Dr. Salama Ikki, for their valuable comments and suggestions throughout the research. Sincere thanks are also extended to the Dean, to all staff members of the Faculty of Engineering, and to all my friends at Lakehead University (LU) for their constant support. Thanks to Matteo Crupi, who helped in the programming of the prototype system. My sincere appreciation goes to the Dean and all staff members of the Faculty of Graduate Studies, as well as to staff members of the LU library (the Chancellor Paterson Library) for their help throughout my Ph.D. study years. I also want to extend my profound gratitude to the physicians and nurses at the Regional Stroke Unit, who provided expertise and shared data that greatly assisted the research and improved its quality. Last but not least, a heartfelt thank you to my wife for her support, and for being the best mother to our children, always busy with them to let me find the time to succeed in the Ph.D. work.

Dedication

To my parents, my brothers and sisters, my wife, and my children.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Health informatics is the branch of science that deals with computerization and automation in healthcare systems. In health informatics, health data are managed through information technology and Healthcare Information Systems (HISs) to record, organize, and analyze healthcare data in order to facilitate healthcare operations and improve healthcare outcomes. The global health informatics market has an estimated annual growth rate of 13.74%, and is expected to reach US$123 billion by 2025 [1]. The benefits of health informatics include the following [2, 3, 4].

- Cost and time savings: Inefficient processes of healthcare result in unnecessary or wasted spending, which is estimated at 50% of all healthcare dollars spent [4]. Efficient application of health informatics can save costs, increase efficiency, and accelerate healthcare operations.

- Increase in quality of care: Patient treatment can be quicker and improved, raising levels of health and avoiding malpractice claims.

- Increase the safety of healthcare: Health informatics helps reduce human errors and can link information across HISs to keep healthcare on track for providing safer patient healthcare.

- Geographical independence: The internet has made it possible for patients to access digital data across the globe from the comfort of their locations. This creates geographical independence and allows of experts to cooperate remotely.

- Improved patient education: Many healthcare organizations now make the education of patients a key part of their healthcare IT strategy. Health informatics can facilitate the creation and distribution of patient education using a variety of rich media.

- Patient autonomy: Patients can see their own health records and medical information, make appropriate decisions and changes, and better manage their own health and well-being.

The increasing demand for information management and automation in healthcare made the field of health informatics a rapidly-growing domain [5, 6]. Our literature review revealed that the rapid advance in health informatics has been progressing without adequate organization in the field. Few studies were directed towards organizing the way health information systems are interconnected and designed. This has resulted in less efficient health informatics systems, and degradation in quality of healthcare data. This can be observed through various domains in HISs such as the existence of missing data in healthcare.

The ultimate long-term objective of this research is to re-structure the field of health informatics and to modify its norm, which has been followed for years. The norm is that Electronic Medical/Health Record (EMR/EHR) systems are considered as the central component of HISs. We show in our framework that standardizing and digitizing clinical pathways enable CP systems to be positioned at the centre of HISs. This helps in the exploitation of the full potential of computerization in healthcare, and makes it possible to achieve the intended benefits of health informatics. Below, we consider automation and data research challenges in healthcare.

1.1 Automation and Data Research Challenges in Healthcare

The automation of healthcare facilities represents a challenging task of streamlining a highly information-intensive sector. Modern healthcare processes produce large amounts of data that has great potential for health policymakers and data science researchers. However, the current situation in health informatics is that a great portion of such data is not properly captured, missing in electronic format, and hidden inside paperwork and forms. For years, EMR/EHR systems were at the centre of HISs. It was hoped that EMR systems could store that vast amounts of data in digital format. However, that target has never been achieved despite the fact that EMR systems acted as the central component of HISs for decades.

An important study on missing clinical and behavioral health data in a large EMR system revealed that EMRs inadequately capture various healthcare data such as diagnosis, visits, specialty care, hospitalizations, nursing services, and medications [7]. The study concluded that missing data undermine many central functions of EMR and that "missing clinical information raises concerns about medical errors and research integrity"[7]. The authors stressed that "given the fragmentation of health care and poor EHR interoperability, information exchange and usability, priorities for further investment in health IT will need thoughtful reconsideration" [7].

This is not the only study regarding the vast amounts of missing healthcare data in HISs. In [8], the authors presented multiple cases that revealed how missing data in HISs would likely result in medication errors and other issues that could cause harm to patients. Missing data creates obstacles for big data research in healthcare. In [9], the authors indicate that missing patient data are prevalent in HISs, and are "an impedance to utilizing machine learning for predictive and classification tasks in healthcare". The authors presented an imputation method, however, imputation methods only approximate the missing data, whereas our objective in this research is to address the original source (root cause) of missing data in healthcare. This will result in providing improved patient care and in the reduction of missing data in the datasets that are used in data mining applications. For example, hospital Length of Stay (LOS) prediction methods found in machine learning literature operate without considering rehabilitation nursing interventions. This degrades the accuracy of rehabilitation LOS prediction. Although documented in papers, rehabilitation data are rarely captured electronically in patient records [10]. In [11], the authors indicate that "missing data is a frequent occurrence in medical and health datasets. The analysis of datasets with missing data can lead to loss in statistical power or biased results." They also present an imputation method that approximates missing data. As mentioned above, it is the root cause of missing data that must be addressed in order to reduce missing data in healthcare.

Our literature review revealed that the field of health informatics is growing without proper design and organization of HISs. We also discovered that the major sources of missing data in healthcare institutions is paper-based forms and unstructured data. In [7], the authors describe health informatics systems as fragmented systems with poor interoperability. Also, in [12], the authors listed medical data written in an unstructured text format as one of the major sources of missing data in HISs. This is because EMRs are not designed to capture non-standardized data. In support to this analysis, and upon analyzing the literature, we found that an important reason for missing data in HISs is that a primary source of healthcare data is still paper-based and has not yet been fully automated. By this primary data source we mean Clinical Pathways (CPs).

CPs have been defined as optimal sequencing and timing of medical interventions by doctors, nurses, and other caregivers for a particular procedure or diagnosis, developed to minimize delays and resource utilization and to improve the quality of healthcare [13, 14, 15]. Despite the fact that CPs are becoming globally popular in hospitals as main components for patient treatment and follow-up, CPs are still circulated in hospitals as paper-based documents with local ambiguous text that is difficult to computerize. Paper-based CPs have many disadvantages, including difficulty of storing and retrieving CP documents, difficulty of sharing CP data among caregivers and institutions, and the fact that manual CP input is prone to human errors. Human error in hospitals can cause harm (and even death) to patients. In addition, the paper-based nature of CPs forms a great barrier between CPs and their integration with today's automated hospitals. Thus, CP automation is a challenging research topic that remains to be investigated.

Our literature review revealed that the common theme of current research in this field is to consider computerized CP systems as side components in HISs that need only to be connected to EMRs. This view has resulted in partially standardized or digitized CPs which is a major limitation of the research found in the literature. Such view of CP computerization positions EMRs (not CP systems) at the centre of HISs.

CPs are the most important sources of data in hospitals, and must therefore be centralized in HISs. However, centralizing CPs in HISs is a major challenge because of their highly unstructured, non-digitized nature. One of our objectives in this research is to propose a framework that can serve as a base for centralizing CPs in HISs.

A research gap in existing research studies is that the non-standardized nature of CPs has been ignored and was not the focus of research. Research studies also ignored the details of CPs such as local CP terms. This is also a main reason for the fact that only limited CP data is stored in EMR systems, resulting in missing data in healthcare. Even research studies that represented CPs using ontological modeling have done so while keeping the non-standardized nature of CPs. This resulted in non-standardized ontologies that can only be used locally. To achieve the required semantic interoperability with existing HISs, CP ontologies must be internationally standardized so that their vocabulary matches international terminology systems. Standardized ontologies support the ability of computer systems to exchange data with unambiguous, shared meaning. This is the core of semantic interoperability. Thus, standardized ontologies support centralizing CP systems in HISs. Developing a standardized ontological framework is an objective of this research. In addition, CPs lack a digital coding system. Developing a coding system to identify CPs is a challenge because it needs to comply with existing terminology systems which were developed years ago without considering CPs. Without this compliance with established systems, such a coding system might not be accepted in the industry and would eventually fail. The literature review revealed that there was no reported work on establishing a CP-specific coding system. One of the objectives of this research is to develop a CP-specific coding system.

Our research addresses the above-mentioned challenges and as such can be considered a major milestone towards achieving the ultimate objectives of restructuring health information systems by positioning fully digitized CP systems at the centre.

1.2 Motivation and Thesis Contribution

The importance of CP automation in improving healthcare systems and patient treatment in hospitals around the world, as well as the research challenges mentioned above, have motivated us to conduct this research and contribute to this important research area. We hope that our contribution will have a positive impact on the wellness of people and society as a whole, as well as the future of HISs. The summary of our contributions in this thesis includes:

- Proposing a novel CP automation framework that can serve as a base for positioning CPs at the centre of HISs. The modeling of the framework is based on a top-level meta CP ontology that models generic CP knowledge using standardized vocabulary to support semantic interoperability.

- Developing a new and internationally compatible CP identification code by expanding the well-accepted SNOMED CT terminology system with full compliance with its structure.

- A CP standardization method in which CP data are SNOMED CT standardized and computerized disease-specific CPs are independently extended and specialized from the meta CP ontology. CP term standardization helps to eliminate lexical ambiguity and to ensure lexical interoperability (i.e., term-to-term interoperability) between CP systems and existing standardized health information systems.

- A method for merging the CP identification code with the standardized CP data to form a CP-specific digital coding system. This method allows to maintain the link between diseases or medical procedures and their interventions. The digital coding system facilitates CP-based decision support in healthcare, and improves CP data collection, sharing, auditing and quality management.

- Proposing a conceptual design of a model CPMS that integrates CPs with HISs through SNOMED CT and HL7 standards, and ensures the independence of CP management systems by including a data repository and decision-support/data analytics component.

Our research contributes towards a futuristic vision for the field of health informatics because (i) for decades, EMRs occupied the central position of HISs, while in this research, we promote a CP-centric architecture; and (ii) this research addresses an important research challenge related to digitization and full automation of CPs to enable advanced data analytics in healthcare. We consider this research as a starting milestone in CP automation, hoping that it will encourage health informatics researchers around the world to participate in advancing this field through more research efforts.

1.3 Thesis Organization

The thesis is organized as follows. Chapter 2 "Background and Related Work" provides additional details on CPs and their role in healthcare, and presents the literature review along with critical analysis of the literature review. Chapter 3 "Proposed Framework" describes the details of the proposed CP framework through its contributions in CP automation, CPMS integration with HISs, and CPMS independence. Chapter 4 "Prototype Design and Architecture of the Proposed Framework" describes the prototype system that is proposed to realize the CP framework with its structure and components. Chapter 5 "Data Analytics and Decision Support Scenarios" presents various applications and algorithms related to the proposed framework in the field of health informatics, particularly how our framework supports CP-based data analytics and hospital resource management (HRM). Chapter 6 "Conclusions and Future Work" presents concluding remarks, limitations, and directions for future research work.

1.4 List of Publications

1. Ayman Alahmar, Matteo Crupi, and Rachid Benlamri, Ontological Framework for Standardizing and Digitizing Clinical Pathways in Healthcare Information Systems, Computer Methods and Programs in Biomedicine, Elsevier, Vol. 196, 2020, pp. 1-18.

2. Ayman Alahmar and Rachid Benlamri, SNOMED CT-Based Standardized e-Clinical Pathways for Enabling Big Data Analytics in Healthcare, IEEE Access, Vol. 8, 2020, pp. 92765-92775.

3. Ayman Hassan, Rachid Benlamri, Kofi Darko, Ayman Alahmar, Sharon Jaspers, Shelly Brown, Shalyn Littlefield, and Michelle Drombolis, Defining Prevalence, Incidence and Risk Factors of Northwestern Ontario Patients with Ischemic Stroke Secondary to Carotid Artery Disease: A Population-based Study, world stroke congress abstracts, International Journal of Stroke, Vol. 13, No. 2, pp. 23-24, 2018.

4. Ayman Alahmar and Rachid Benlamri, Optimizing Hospital Resources using Big Data Analytics with Standardized e-Clinical Pathways, To appear in Proc. of 6th IEEE Int. Conf. on Cloud and Big Data Computing, Calgary, Canada, August 17-24, 2020.

5. Ayman Alahmar, Emad Mohammed, and Rachid Benlamri, Application of Data Mining Techniques to Predict the Length of Stay of Hospitalized Patients with Diabetes, IEEE, Proc. of 4th International Conference on Big Data Innovations and Applications (Innovate-Data), Barcelona, Spain, August 6-8, 2018, pp. 38–43.

6. Ayman Alahmar, Mohannad AlMousa, and Rachid Benlamri, Clinical Pathway Standardization using SNOMED CT-Based Semantic Similarity and Relatedness, In Progress.

Chapter 2

Background and Related Work

2.1 Clinical Pathways

Since clinical pathways are at the core of this thesis, this section presents additional details on CPs and their application and importance in formalized healthcare systems.

2.1.1 History, Evolution and Definitions of CPs

CPs first emerged in healthcare in the mid-1980s in the USA when Karen Zander, Kathleen Bower, and Mary Etheredge first coined the term at the New England Medical Centre in Boston, USA [16]. The concept itself was not a new one because it has its roots in management theories, and deals with improving the quality of business processes such as Critical Path Method (CPM), Program Evaluation and Review Technique (PERT), and Business Process Reengineering (BPR). These successful management theories were not applied in healthcare; thus, the concept of CP was an initiative to adopt effective management concepts in hospitals [17, 18, 19, 19]. Following the USA, CPs were adopted first in the UK, and then the concept was used internationally with -unfortunately- different views about their development, implementation and evaluation with different CP definitions [20, 21]. Table 2.1 presents some definitions of CPs found in the literature. As noticed from Table

Table 2.1: Definitions of CP in the literature.

Definition	**Reference**
Methodology for the mutual decision making and organization of care for a well-defined group of patients during a well-defined period.	[22, 23, 24]
Optimal sequencing and timing of interventions by physicians, nurses, and other staff for a particular diagnosis or procedure, designed to minimize delays and resource utilization and to maximize the quality of care.	[13, 14, 15]
"Comprehensive methods of planning, delivering and monitoring different healthcare services provided to patients."	[25]
Criteria for an operational definition of CP: (1) A structured multidisciplinary plan of care; (2) Translating guidelines or evidence into local structures; (3) Showing the steps in a course of treatment or care in an inventory of actions (i.e. time-frames or criteria-based progression); and (4) Standardizing care for a specific population.	[21, 26]

2.1, CP is described as "methodology", "comprehensive method", "structured multidisciplinary plan", and the like. These definitions show the importance of CPs in healthcare provision. This is because applying CPs as successful management practices in healthcare has great benefits as addressed below.

2.1.2 Benefits of CP Applications in Healthcare

Clinical pathways are becoming popular in healthcare organizations because their use has been recognized as having several benefits. CP benefits listed in the literature include: reducing patients' Length of Stay (LOS) in hospitals [27], reducing healthcare cost, reducing variations in medical practice [21], optimizing the use of resources [21], improving patient outcomes and reducing treatment complications [28, 29, 30, 31, 32], increasing patient sat-

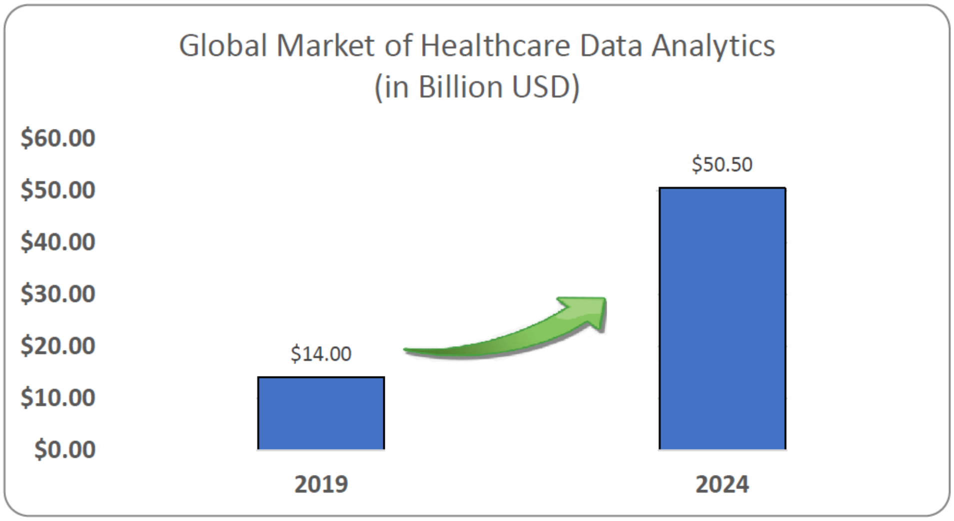

Figure 2.1: Expected increase in expenditures of healthcare data analytics.

isfaction [33, 34], increasing patient participation in health procedures [35], and improving communication between physicians and nurses [36]. CPs are also considered to be an important tool for ensuring that the latest evidence in clinical guidelines is used in the care of patients [37, 38]. It is estimated that healthcare organizations could cut costs by 30% [39] or by 30% to 50% [4] if they adopt the best IT and quality management practices that eliminate waste and discontinue the overuse of resources. To achieve this cost saving, the proper application of CPs is a key factor, since CPs are at the centre of best management practices in healthcare.

CPs are a major source of data in healthcare. Healthcare big data analytics is a growing field [40]. The overall market for healthcare data analytics is expected to reach US$50.5 billion by 2024 from US$14.0 billion in 2019 (see Figure 2.1) [41].

2.1.3 Development of Clinical Pathways

CPs help in optimizing healthcare and reducing costs, however their development is time-consuming and requires commitment and effort from both healthcare staff members and healthcare administrators [42, 43]. The development of CPs undergoes four phases: (1)

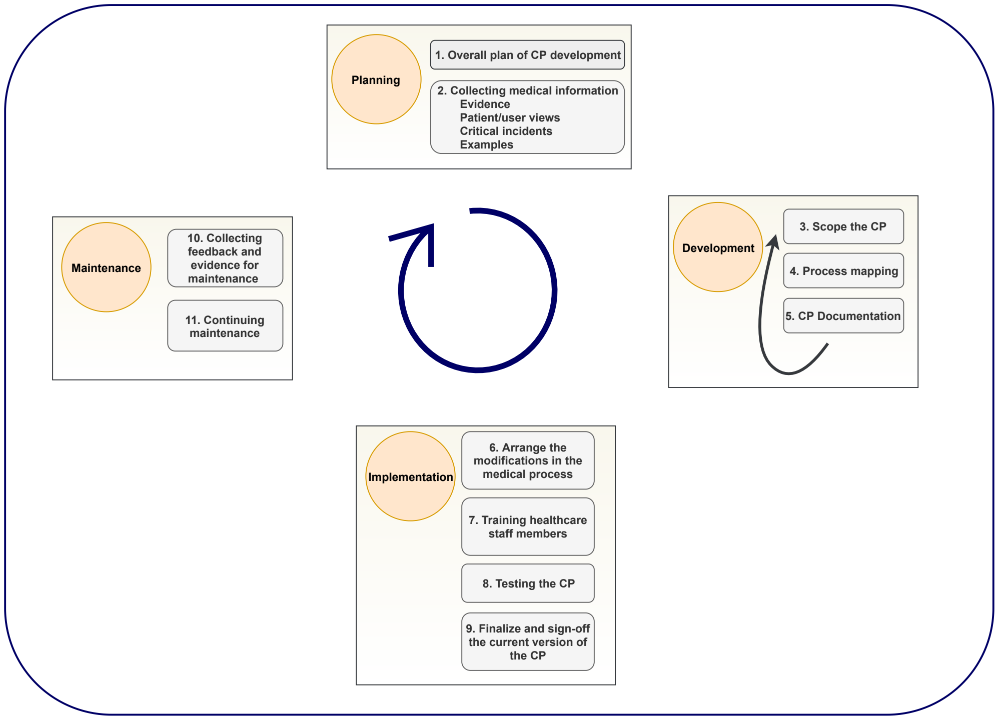

Figure 2.2: CP development phases and steps.

planning phase; (2) development phase; (3) implementation phase; and (4) maintenance phase, with the steps shown in Figure 2.2 [42].

CPs are often developed locally inside hospitals to serve local staff members, which creates challenges in the face of CP computerization, as addressed below.

In Canada, CPs are viewed as a tool to ensure that patients receive the best available care and that they are well-informed about their treatment journey [21, 44, 45]. The 30% saving expected from quality practices and efficient use of CPs [39] means a saving amount of about $72.6 billion Canadian dollars per year, considering that the total health spending in Canada is estimated at $242 billion in 2017 according to the Canadian Institute for Health Information (CIHI) [46]. Provincial health authorities in Canada have independently adopted CPs, and the number of developed CPs is increasing. In order to analyze sample CPs, we contacted Thunder Bay Regional Health Sciences Centre and The Ottawa

Clinical Pathway

Ischemic Stroke

On Admission - Day 1
Is swallowing screen completed? Yes No - Complete swallowing screen
If patient failed swallowing screen, consult SLP
If patient failed initial swallowing screen and SLP unavailable for assessment Repeat swallowing screen within 24 hours of admission
If patient failed repeat swallowing screen: consult with physician regarding insertion of nasoenteric tube/initiation of feeds as per enteral nutrition policies
If on PO diet: Monitor and document tolerance to first meal
Smoking Cessation
Has patient used of any kind of tobacco in the last 6 months? Yes No
If yes, discuss with patient and physician nicotine replacement options
Patient and Family Education
Provide patient and family with: - Your Stroke Journey: TOH Care Companion Booklet - Your Stroke Journey: A guide for people living with stroke

Figure 2.3: Part of an ischemic stroke clinical pathway.

Hospital to request sample CPs. An example CP for stroke is shown in Figure 2.3. We also teamed-up with the Regional Stroke Unit in Thunder Bay. We refer to the cooperating stroke doctors and nurses from the stroke unit as the "domain experts" in this thesis. The Ottawa Hospital states in their Model of Care that CPs "are used to describe and implement clinical standards. They help to provide quality and efficient patient care. The CP documents are part of the patient's permanent record and are integrated into the clinical documentation. Most patients on CPs will receive a patient education booklet explaining their disease or condition and providing them with teaching around their diagnosis" [47].

2.2 The Need for CP Computerization and Automation

The proper automation of paper-based CPs brings great advantages to healthcare because we consider it as the backbone behind the true realization of many of the benefits expected from applying CPs in hospitals. In fact, the advantages of CPs cannot be fully realized without automation. Studies on CPs and CP computerization reveal the following benefits if CPs are correctly practiced and properly computerized [48, 49, 50, 51, 52, 53, 54].

- Benefits of CPs:
  - Reducing Length of Stay (LOS) in hospitals.
  - Optimizing the use of resources.
  - Reducing healthcare costs.
  - Improving patient outcomes.
  - Reducing treatment complications.
  - Encouraging patient participation in health procedures.
  - Improving communication between physicians, nurses, and patients.
  - Increasing patient satisfaction.
- Benefits of CP Computerization:
  - Help medical staff members to share large amounts of medical information which is difficult with paper-based CPs.
  - Easier modification of CPs when required.
  - CP management systems facilitate the automation of CP variance records, which allows for analysis and statistical decision making based on deviations from standardized CPs.
  - Monitoring of CP execution in real-time.
  - Automated error checking for the steps of the treatment.

- Automated time management and temporal data saving.
- Improvement of the efficiency and quality of CP application and patient care in general.
- Integration of Electronic Medical Records (EMR) and CP execution.

The two groups of benefits listed above are interconnected, which implies that CP benefits are better realized when CPs are properly computerized. For example, healthcare errors impact patient safety and are very costly. Automated error checking, provided by computerized CPs, helps prevent medical errors. This improves patient outcomes, decreases cost, and helps reduce the length of stay of patients in hospitals. All these are benefits from the previous lists.

The literature review shows several studies to computerize CPs, however, the existing gap in this research area is that the proper and full computerization has not yet been achieved. In addition, CPs lack a proper coding system. This Ph.D. work fills this research gap.

2.3 Literature Review

Literature review reveals that there are several studies that addressed computerization of CPs. Some research was based on traditional information systems (i.e., non-semantic web approach), while most recent studies adopted a knowledge based approach, relying mainly on semantic web in which CPs were modeled following ontology engineering approaches. Studies found in the literature review have addressed various models and diseases (e.g., chronic kidney disease, breast cancer, diabetes, human papillomavirus, and prostate cancer). Semantic modeling is a relatively recent methodology in software engineering for knowledge representation and is widely used for data management in biomedical informatics [55]. Ontology, Web Ontology Language (OWL), and Semantic Web Rule Language (SWRL) are core components of semantic models [56, 57]. Semantic web components help in using semantic web statements and rules in order to define classes, relationships, and domain constrains in order to model a particular domain (e.g. an ontology to model diabetes) [58, 59]. Literature review revealed that various medical conditions were considered in CP computerization papers. Since, ontology-based articles share similar principles, below is a discussion of selected articles among them, followed by discussing articles among the non-semantic approach. Finally, we discuss literature review related to standardization in healthcare.

2.3.1 Semantic Based Methods

Fudholi et al. [60] proposed a CP ontology model that consists of the following major classes: clinical pathway, person, organizational structure, record, and clinical category. The model is proposed to check whether or not the treatment processes comply with the CP requirements. For example, by querying both the CP ontology and the patient's recorded data, the model checks if the recorded data (e.g., lab tests and assessments) comply with the interventions specified in the CP. To evaluate their model, they developed a dengue fever and typhoid fever CP ontologies and queried them through an ontology query language to perform various compliance checks. Their results show that they have not considered CP standardization in their approach. For example, they refer to the lab tests on the clinical pathway by non-standardized terms like HB, HT, and trombocyt. This shows that they have used a local terminology approach in their model.

Tehrani et al. [61, 62, 63, 64] suggested that the development of CPs in situations where processes are complex, needs to combine ontology-based modeling and organizational semiotics. Organizational semiotics treats organizations (e.g., hospitals and clinics) as information systems in which information is created, processed, distributed, stored and used [65, 66]. In their approach, they interviewed medical staff members and used semantic analysis to develop an ontology that represents the semantics of the CP concepts, their relationships and patterns of behavior of physicians and staff members (see Figure 2.4). They then used norm analysis method to extract and analyze patterns of healthcare activities and informal safety norms that affect CP outcome and patient safety.

Norms in semiotic approaches specify the possible patterns of behaviors. For example, the nurse is "obliged", "permitted", or "prohibited" to do an action (called deontic operators). Norms are described formally using the format:

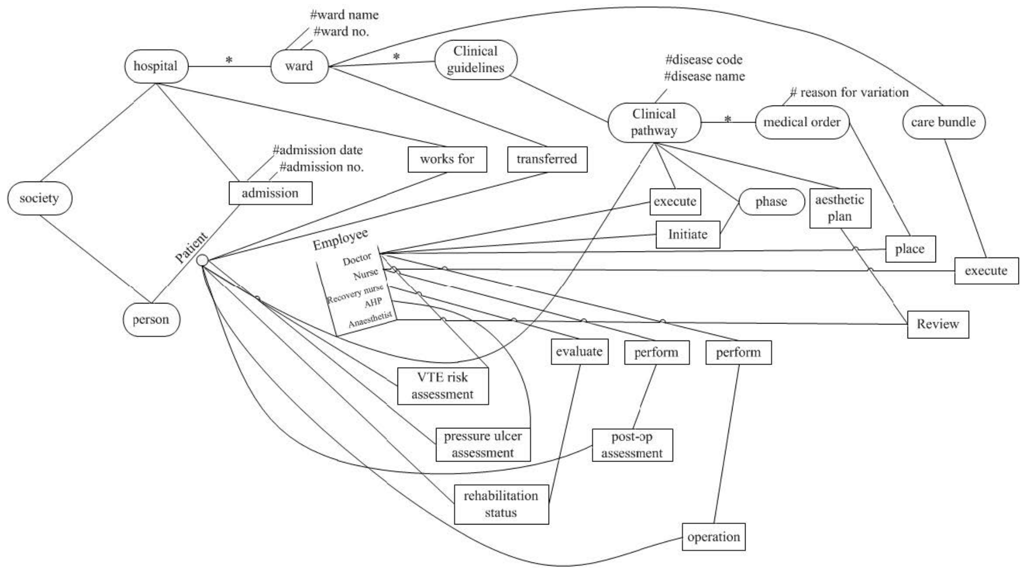

Figure 2.4: The ontology developed by Tehrani, et al. [63]

Whenever ⟨condition⟩ **If** ⟨state⟩ **Then** ⟨agent⟩ **Is** ⟨deontic operator⟩ **To** ⟨action⟩
For example, Norm N1 can be defined as:
Whenever ⟨the patient is assessed for venous thromboembolism⟩
If ⟨there is bleeding risk⟩
Then ⟨doctor⟩ **Is** ⟨permitted⟩ **To** ⟨give prophylaxis⟩
The authors argue that, generating a CP ontology that is enhanced by formal patterns of human behavior and by rules that govern the actions identified in the ontology reduces human errors associated with complex situations that require patient-specific customization and human decisions. The ontology can then be linked to an EMR. Their formal approach is useful to reduce medical errors, however, their ontology development was mainly based on local terminology, through internal staff interviews, without considering terminology standardization.

Wang et al. [67] proposed an EMR-CP integration method that supports EMR systems with a CP system through SNOMED CT linking between equivalent terms. By this way, the CP system can be integrated programmatically with different EMRs. The programmers

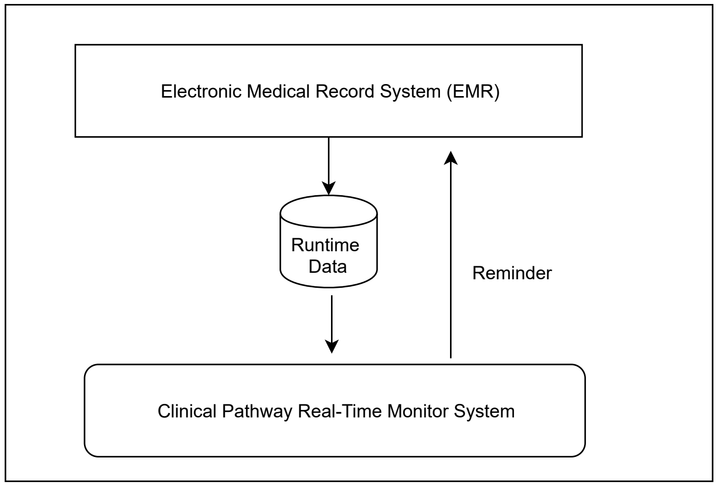

Figure 2.5: Illustration of the system structure of Liu et al. [48]

at a hospital need to generate RDF statements from the EMR database and then use an ontology editing tool to write the statement of relations between EMR terms and basic CP terms. A limitation of their work is the tedious programming work needed to achieve the EMR-CP linking. In addition, future modifications to the EMR dictionary would cause the EMR-CP linking to be lost, which would require the programming tasks to be repeated again. Also, they consider the EMR as the central component, not the CP system.

Liu et al. [48] proposed an ontology-based approach for monitoring of CPs. The main objective was to establish communication between the CP and EMR with the ability to monitor the CP execution and display reminders to clinicians about CP activities. The CP used in their prototype was for unstable angina from the cardiology department of a hospital in China. They could build a CP component that feeds data/reminders to EMR, however, the system was not standardized and did not act as an independent system. It could communicate only with the EMR, and functioned to serve EMR operations, see Figure 2.5.

Abidi et al. [68, 69] proposed an ontology-based prostate cancer CP computerization model and discussed the merging of prostate cancer CPs from different hospitals. In their approach, they represented hospital-specific CPs using an ontological model and then aligned the common activities between multiple CPs from different hospitals to have a com-

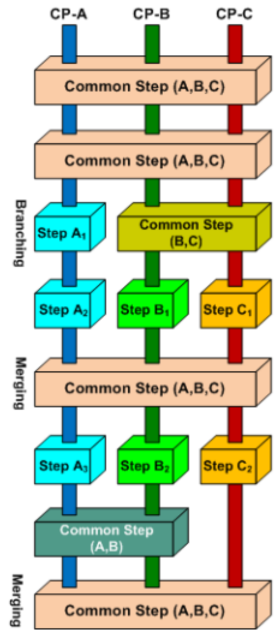

Figure 2.6: A unified CP with merging nodes and branching nodes [68, 69].

mon or unified CP. The resulting care model merges common care activities whilst allowing to have unique hospital-specific activities. For example, Figure 2.6 shows their approach for a merged clinical pathway featuring common-task nodes and institution-specific nodes for prostate cancer CPs in three different hospitals. Their model is useful for performance analysis, however, their method can work only for small-scale unification and is not practical for larger scale multi-hospital approach.

Daniyal et al. [70] followed a similar approach to develop an ontology based prostate cancer CP that integrates multiple localized CPs to have a unified CP for prostate cancer. In addition, they integrated the resulting CP in a prostate cancer computerized system that automated a combined CP flowchart for three different CPs (see Figure 2.7). The system enabled physicians to follow patients as specified, however the computerized CPs were kept local and were not checked for terminology compliance with international reference terminologies.

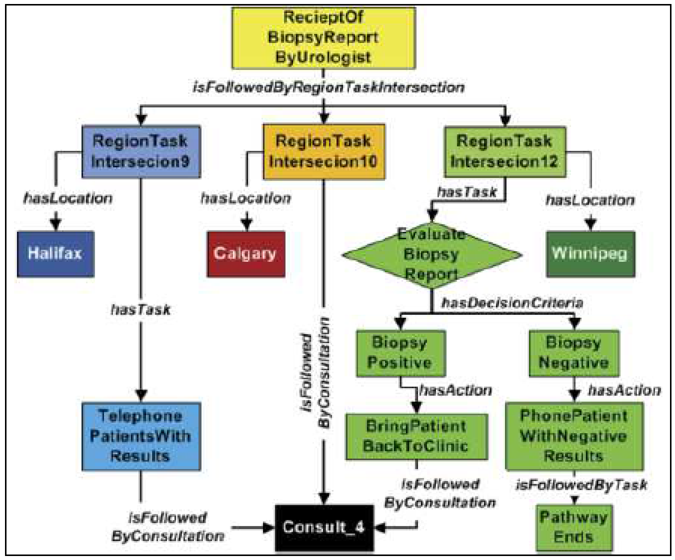

Figure 2.7: Combined CP flowcharts at a branching node [70].

Hu et al. [49] proposed a semantic-based method in which ontology was used to model CPs and SWRL was used to model CP rules. In this way, the application could reason over the rules and information collected. They based their modeling on a CP general ontology that defines common concepts necessary in disease-specific CPs. To evaluate their method, they built a lobectomia pulmonalis CP, and realized it based on an EMR system called IZANAMI such that the CP is noticeable to healthcare providers through the EMR. An illustration of the structure of their system is shown in Figure 2.8. The model was successful in presenting CP steps to healthcare providers; however, the limitation of their approach is that their meta-ontology was a hospital-specific local ontology since its modeling was based on the terms available on local CPs without standardization. In addition, the used CPs were not checked for terminology compliance with international reference terminologies. In addition, their CP system was totally embedded inside the EMR system as shown in the structure of their system (Figure 2.8).

Alexandrou et al. [71] presented a CP ontology model that comprises three parts in a single ontology: the CP part, the quality assurance part, and the business part. They

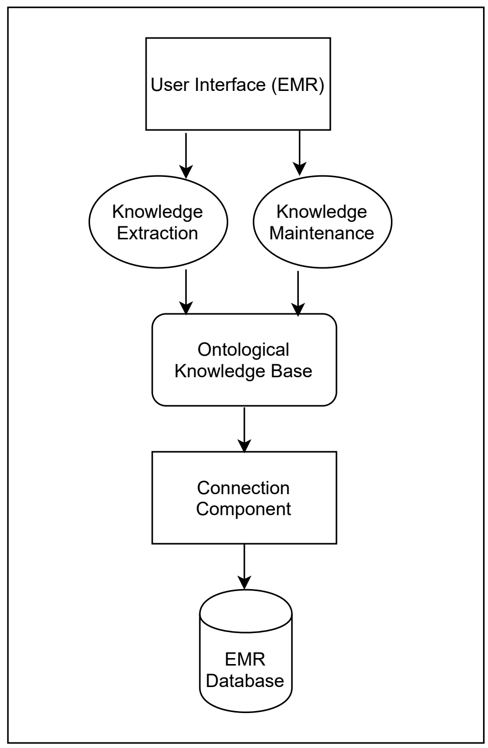

Figure 2.8: Illustration of the system structure of Hu et al. [49].

applied their model on human papillomavirus patients, however, standardization of CPs was not considered in their work. Ye et al. [53] proposed a clinical pathway ontology model in which they included time intervals between tasks using the entry sub-ontology of time. In their case study, they used a CP for cesarean section from a hospital in Shanghai, China, which does not use standard clinical terminologies, and nothing mentioned in the paper related to standardizing and encoding of CP contents.

Hu et al. [50] modeled CP based on an ontology schema with four main units: timeline, category of care, variance record, and outcome criteria. They also showed that the ontology-based approach is suitable to model CPs by conducting a comparison between the concepts of ontology and CP. For example, both ontology and CP are formalization methods (i.e., ontology formalizes concepts in a domain and CP formalizes clinical care processes in healthcare). Their work was limited to ontology-based modeling of CPs and showing that it is a successful modeling technique. No system was built to use the ontology, but their future work was to embed the ontology inside an EMR system.

2.3.2 Non-Semantic Based Methods

Studies that followed the non-semantic modeling approach were mainly adding selected data fields from the CPs programmatically to the EMR/EHR systems to computerize them and to act as reminders of CP steps to medical staff members.

Hoelscher et al. [72] integrated a computerized infectious disease CP within an EMR system. The purpose of the study was to implement an improved rapid-deployment decision support strategy for the detection and treatment of emerging and re-emerging infectious diseases. Using the Plan-Do-Study-Act (PDSA) rapid cycle improvement model, the computerization process was implemented and monitored.

Smulowitz et al. [73] developed an electronic clinical decision support tool within an emergency department system. The goal was limited to flagging patients who were required to follow a chest pain CP called "HEART pathway".

Gibbs et al. [74] presented a framework for developing an online clinical pathway that can be used by patients. They applied their model successfully on a CP for chlamydia infection. However, their approach was limited in scope to only the considered disease.

Blaser et al. [75] developed a prototype system by using an embedded tool within Orbis/OpenMed-system (which is an EMR/EHR system used in Germany). They could add basic CP elements into the EMR, however, the CP functionality worked only within the EMR/EHR and could not be utilized as an independent CP management system.

Bernstein et al. [76] pointed out that clinical pathways are not well integrated with electronic health records. They proposed an integration method that made the patient position in the pathway visible to relevant parties such that each CP would have a SNOMED CT link to the EMR system. Their SNOMED CT linking was limited to a top-level linking between the major steps of the CP and the EMR system. For example, the 'laboratory tests' stage in the CP was considered as a single node linked to the EMR to show that the patient has reached this stage without considering the CP contents or the details of the lab tests. Such top-level linking might help to determine the position of the patient in the CP, however, it cannot help in capturing all CP data to reduce missing data and improve data mining results, as will be addressed in our framework.

Katzan et al. [77] developed, through a collaboration with a company called Epic, an electronic stroke CP program that was integrated within their Epic EMR (which is a commercial EMR developed by Epic Systems Corporation [78]. Epic programming contractors were involved in this project to develop the program and to customize the inpatient Epic EMR screens to include CP-specific options. The modified interface saved time for clinicians by reducing unnecessary data entry based on the CP. Integrating the CP within the EMR also helped to remind healthcare providers (especially trainees) of certain CP guidelines that might have been forgotten, which reduced possible human errors. They reported that the integration of CP with EMR was overall successful, however, not all data fields were captured and not all features worked as planned. For example, an anatomic diagram for stroke location did not function inside the program despite extensive efforts by the programmers to make it work. In addition, some of the discharge checklist items did not auto-populate correctly so care providers were not using them uniformly [77]. The system was programmatically integrated within the EMR, so it was not an independent CP system. In addition, the issues related to standardization of the CP or developing a coding system for CPs were lacking in this work.

2.3.3 SNOMED CT in Healthcare Systems

SNOMED CT is a fast-growing terminology system in healthcare. Many research studies found in the literature have considered SNOMED CT as the adopted terminology for their systems, however, the focus of these studies was on EMR/EHR systems, or other healthcare systems (not on clinical pathways). This is because clinical pathways are mainly paper-based, non-standardized documents that are written in free text formats.

In [79], the authors surveyed the use of SNOMED CT clinical coding in EMR/EHR and Clinical Decision Support Systems (CDSS). Their study focused on preventive care domain, and they found that CDSS built on SNOMED CT support the creation of a high-quality healthcare systems for preventive care. The authors also found that SNOMED CT is a powerful and effective clinical terminology within EHR systems that can be used to reduce medical errors, save lives, advance patient safety, and improve overall quality of healthcare services.

Rai et al. [80] described a large SNOMED-CT project under the supervision of the Ministry of Health and Family Welfare in India to integrate existing EHR with SNOMED-CT. The project started in July 2016 and is still ongoing. Many accomplishments have been achieved including a national drug library database containing 169,000 drugs mapped with SNOMED CT codes. In addition, SNOMED-CT coding for 4000 nationalized lab investigations given from ministry were completed. The project is still ongoing and has been described by the authors as successful.

Hwang et al. [81] described a project for mapping Korean Electronic Data Interchange (EDI) medical procedure to SNOMED CT. EDI is the system used for health insurance claims in Korea. To date, 82.5% of the EDI codes have been mapped to SNOMED CT, and the project is progressing successfully.

Lee et al. [82] considered encoding the terms of a clinical palliative care EMR using SNOMED CT. There were 20 pre-defined diagnoses (e.g., melanoma) and 14 pre-defined problems at referral (e.g., delirium) that could be selected from drop down menus in the EMR interface. Other information was saved as free text entered in an "additional information". Pre-defined values could be encoded in SNOMED CT by mapping them directly to appropriate SNOMED CT terms. Free text values were noisy text that required preprocessing and cleaning in order to extract meaningful clinical terms, and subsequently matching them with standardized terms. For example, processing the free text “Bone mets.” resulted in the clinical term “Bone Metastasis” after extending the abbreviation “mets.” into “metastasis”. After preparing all the clinical terms, some of them had a complete match with SNOMED CT, whereas others had partial match, or no match at all. Partially matched terms could be further analyzed by domain experts to try to map them to SNOMED CT terms, if possible.

Giannangelo et al. [83] used a web-based survey to identify (among other things) the potential and future of SNOMED CT in EMR/EHR systems. They found that survey respondents who were using SNOMED CT indicated an expected future increase in EMR/EHR applications that use SNOMED CT. That expectation is currently being realized, as more EMR systems are adopting SNOMED CT.

2.3.4 Critical Analysis of Literature Review

Most CP studies reported in the literature review, as described in the previous section, consider the CP system as a secondary component in HISs. This is because the final target of the computerization process was the EMR (as the central component) and how to enhance EMR with CPs. This view has resulted in only partially standardized or digitized CPs, which is a major limitation of the research found in the literature.

Unlike research studies available in the literature, our philosophy in this research is that CPs should be fully digitized and positioned at the centre of HISs. This is because CPs are the disease management and treatment plans for patients, thus within CPs lies the very heart of medical planning, including quality and cost factors in healthcare. In addition, CP use is increasing and CPs are becoming popular in health organizations around the world, therefore, we envision that in the future health systems will need to be designed based on the vision that computerized CPs, not EMRs, should be at the centre of HISs, see Figure 2.9. This research is a milestone towards achieving that vision.

Another research gap is that the nature of CPs being ambiguous, non-standardized

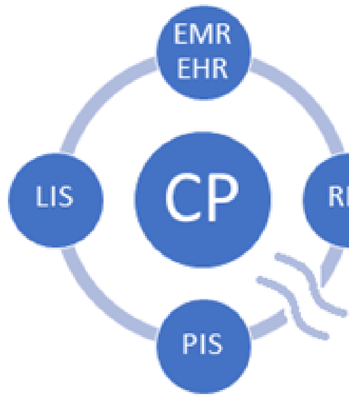

Figure 2.9: Futuristic vision for the future of Health Informatics.

documents was not the focus of research studies. Even for the few studies that have partially used standard terminologies for CP data, the purpose was for programmatically linking basic CP data with EMRs (i.e., programming need). Research studies also ignored the details of CPs such as CP terminology that are country-specific or local to the hospital (i.e., national and local CP terms). The drawback of such approaches is that only limited CP data is stored in EMRs, resulting in missing data in healthcare.

As mentioned in Chapter 1, there are many missing data in HISs and the non-standardized nature of CPs is a main source of missing data. We add here more concerns specific to missing CP data. For example, one of our domain experts estimates that only 60-70% of CP data are captured in EMR systems. This means that 30-40% of CP data are lost and not available in health information systems. Lin et al. [84] mentioned that healthcare activities that are executed on daily basis are not collected from CPs. Therefore, linking such healthcare activities to a particular disease is not possible. Huang et al. [85] addressed the problem related to missing CP traces due to incomplete CP data in EMR, so a complete CP could not be obtained or retrieved from EMR data. Therefore, missing CP data is a major challenge facing the utilization of CPs to their full potential.

An objective of our framework is to make CPs "digitally visible" and to enhance their semantic operability among HISs, thus, the data resulting from CP applications are not lost and can be used whenever needed. This also improves CP data communication between healthcare professionals which reduces human errors in hospitals.

CPs are developed as ambiguous text lines that do not match standard clinical terminologies. In this thesis, we propose a framework that can address this challenge by adopting SNOMED CT for the complete standardization of CPs. Our framework considers various levels of CP data standardization (i.e., global, national and local CP terminology levels).

Aside from standardization of CP content, the present gap between CPs and their integration with various types of HISs can also be linked to the fact that CPs do not have any international digital identifiers to identify and recognize them both digitally and in real life. Therefore, we developed in this research a digital CP identification code (CPID) that is generated by introducing a new SNOMED CT partition identifier. Through the CPID, every CP will have its own digital international identification number. Furthermore, we merged both the standardization of CP data and the CPID to develop a new digital "hyphenated coding system" in healthcare with a novel link between CPs and their data to facilitate data analytics and decision support. The new coding system was developed by extending the SNOMED CT system without violating its logical model which facilitates its acceptance at the international level.

We propose a conceptual design and architecture for a CPMS to realize the framework such that the CPMS has an HL7 engine to communicate with existing HISs (besides the ability to communicate through SNOMED CT standard). The CPMS is an independent system in the sense that it has its own CP-specific database for CP data, and includes data analytics and decision support algorithms. The conceptual design of the system helps in achieving the centralizing of CPs, independence of CPMS, and advancing CPs digital visibility and machine readability.

2.4 Conclusion

This chapter addressed the concept of clinical pathways, their history, development, and benefits in healthcare. Clinical pathways are still used in hospitals as paper-based documents and the need for their proper computerization is addressed. The chapter also discussed the literature review which was divided into three parts: semantic based methods, non-semantic based methods, and SNOMED CT in healthcare systems. The chapter concluded with the critical analysis of the literature review. In the critical analysis, we showed that the common theme in the current clinical pathway research is that computerized CPs are considered as secondary components developed mainly to support electronic medical record systems in their operations. In addition, researchers in this field achieved only partial computerization of CPs because the concepts of standardization, digitization, coding system, and centralizing CPs in health information systems were not addressed. These concepts will be detailed in the next chapter.

Chapter 3

Proposed Framework

The automation revolution in modern healthcare systems has mandated that hospital processes be computerized to streamline healthcare, reduce paperwork, collect digital data, and control costs. CPs are no exception in this regard. Although healthcare has greatly benefited from the introduction of CPs, these benefits cannot be fully realized without properly computerizing CP systems in order to automate their applications within healthcare systems.

3.1 Overview of the Proposed Framework

In order to address the research gaps mentioned in the previous chapters, we propose an ontological framework for standardizing and digitizing clinical pathways in healthcare information systems. In the sections below we address the proposed framework in terms of its contributions to CP automation, CPMS integration with HISs, and the independence of CP management systems. In addition, the framework addresses the issue that CPs are expressed in ambiguous local textual instructions. This not only renders them difficult to understand by medical staff members, but it also creates a digital barrier or "digital divide" between CPs and HISs.

Fig. 3.1 illustrates how unstructured, partially-computerized CPs create a digital divide between CPs and other HISs. This digital divide is a main reason behind the challenge

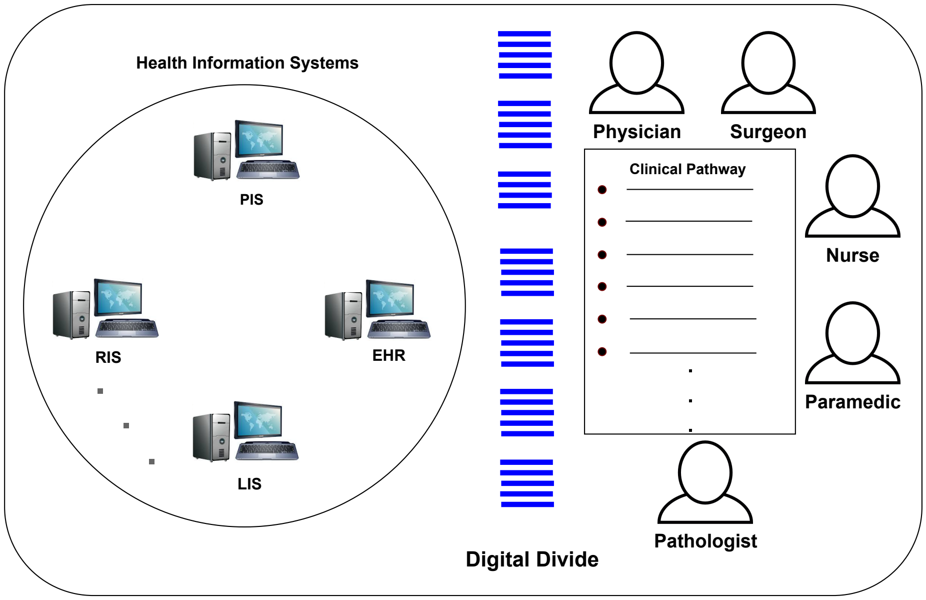

Figure 3.1: Digital Divide caused by unstructured CPs.

of CP automation. To date, CP computerization research has ignored the presence of this digital divide, and most efforts were directed towards "programmatically" linking basic CP data with EMR systems while leaving CPs digitally invisible and distanced from the digital age.

3.2 CP Automation

CPs are populated with data that can be only partially transferred to other HISs. A key factor that impedes the transfer of full CP data is that CPs are prepared in hospitals without attention to standardizing their medical terms. After a thorough review of CP research found in the literature, and from discussions with our domain experts at Thunder Bay Regional Health Sciences Centre, it was clear that most CPs are currently developed using ambiguous local medical terms and abbreviations [86, 87, 88, 89, 90, 91, 92, 93, 94, 95]. This situation makes CPs prone to human errors and creates a challenge of exchanging them across medical institutions. It also causes the loss of valuable CP data because existing HISs use standardized terminology systems in their encoding of medical terms. The solution for this in our framework is that we strongly recommend that CPs be encoded with an international terminology system such as SNOMED CT. The sections below present further details on medical classification and terminology systems and SNOMED CT.

3.2.1 Medical Classification and Terminology Systems

Medical classification is the process of converting descriptions of medical procedures and diagnoses into world-wide standardized codes. During classification, diseases are categorized based on similar properties. A terminology is a group of terms representing the concepts in a domain. Standard reference clinical terminology is crucial for the interoperability between various information systems in healthcare [96, 97, 98].

There are currently several international standards being applied, including Systematized Nomenclature of Medicine-Clinical Terms (SNOMED Clinical Terms terminology, officially abbreviated as SNOMED CT) and the International Statistical Classification of Diseases and Related Health Problems (ICD classification), both widely used systems [99, 100]. In this research, we selected SNOMED CT to be the base terminology for the proposed CP standardization and coding framework because it is considered to be the world's largest and most comprehensive multilingual clinical healthcare terminology [96, 97, 98]. Below is a description of SNOMED CT.

3.2.2 Systematized Nomenclature of Medicine-Clinical Terms

The Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT) is a systematically organized computer-processable collection of medical terms consisting of codes, terms, synonyms and definitions used in clinical documentation and reporting. According to Canada Health Infoway, "SNOMED CT is the largest and most comprehensive medical terminology in the world. The international release of SNOMED CT contains international content and is maintained by SNOMED International" [101]. "The SNOMED CT Canadian Edition contains concepts that are specific to use in Canada and is maintained by Canada Health Infoway" [101].

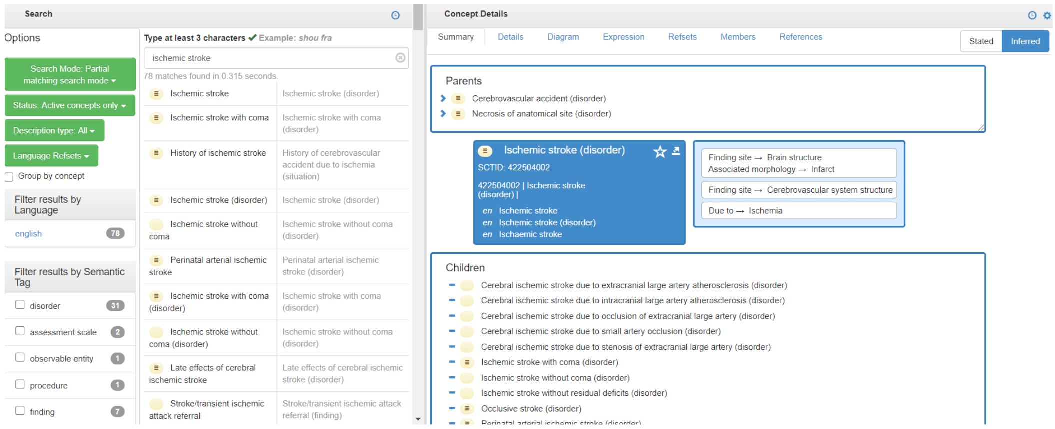

Figure 3.2: The summary of ischemic stroke on SNOMED CT Browser.

The primary purpose of SNOMED CT is to encode the meanings that are used in health informatics, and to support the effective clinical recording of data with the aim of improving patient care. SNOMED CT provides the core general terminology for electronic health records. SNOMED CT comprehensive coverage includes clinical findings, symptoms, diagnoses, procedures, body structures, organisms and other etiologies, substances, pharmaceuticals, devices and specimens. To give SNOMED CT examples from its main repository, Figures 3.2 and 3.3 show the 'summary' and 'diagram' of ischemic stroke as they appear on the SNOMED CT Browser [98]. Table 3.1 lists the names of the nineteen (19) top classes in the structure of SNOMED CT ontology. Top classes have is-a relations with the root class "SNOMED CT Concept" (refer to Figure 3.4). Besides is-a relations, SNOMED CT concepts have many other relations/attributes such as associated-with, contained-in, due-to, finding-site, has-ingredient, and is-about.

The release of the SNOMED CT International Edition on January 31, 2020 includes 352,567 concepts that provide the core general terminology for Electronic Health Records. The SNOMED CT logical model (Figure 3.5) defines the way in which each type of component and derivative is related and represented in SNOMED CT. The core component types are concepts, descriptions and relationships. The logical model therefore specifies a structured representation of the concepts used to represent clinical meanings, the descriptions

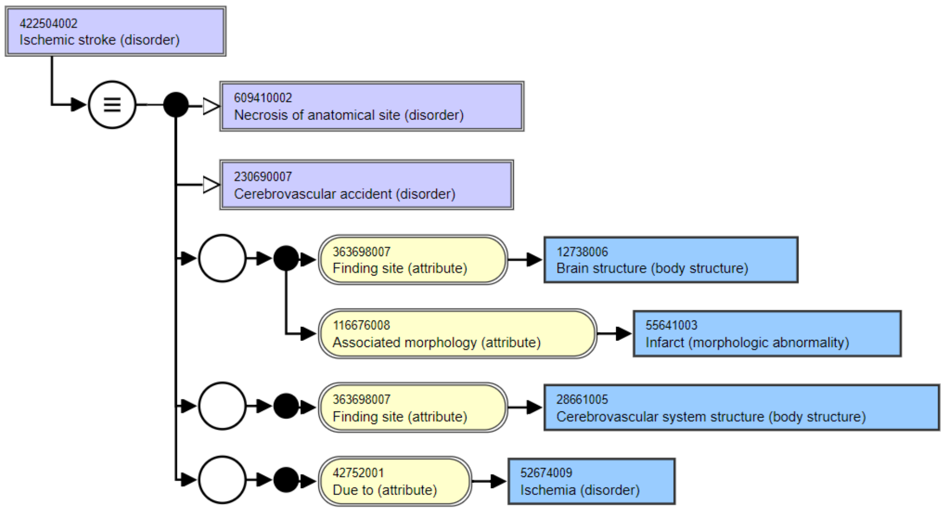

Figure 3.3: The diagram of ischemic stroke on SNOMED CT Browser

Table 3.1: SNOMED CT Structure.

Top classes of SNOMED CT Hierarchy	
Body Structure	Qualifier Value
Clinical Finding	Record Artifact
Environment or Geographical Location	Situation with Explicit Context
Event	SNOMED CT Model Component
Observable Entity	Social Context
Organism	Special Concept
Pharmaceutical/Biologic Product	Specimen
Physical Force	Staging and Scales
Physical Object	Substance
Procedure	

used to refer to them, and the relationships between the SNOMED CT concepts [98].

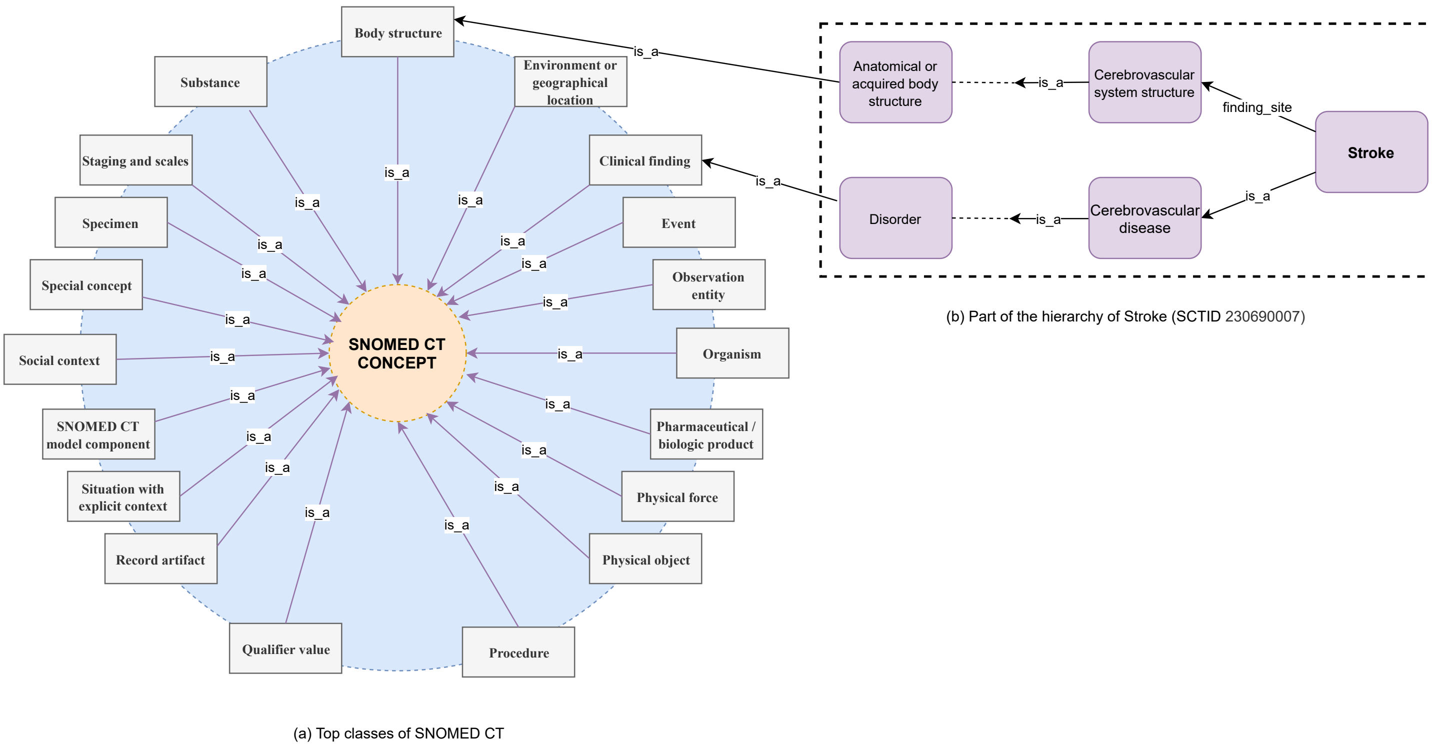

Figure 3.4: Top classes of SNOMED CT taxonomy.

3.2.3 Clinical Pathways Compliance with Terminology Systems

As mentioned earlier, the compliance of CPs with reference terminology systems is a key factor for the integration of computerized CPs with other systems in health informatics. However, our literature review revealed that the compliance is low. It is a well-known fact that “If you cannot measure it, you cannot improve it” [102]. Therefore, there is a need to measure CP compliance with terminology systems. To measure the compliance numerically, we have introduced a metric called Clinical Pathway Compliance Ratio (CPCR), defined

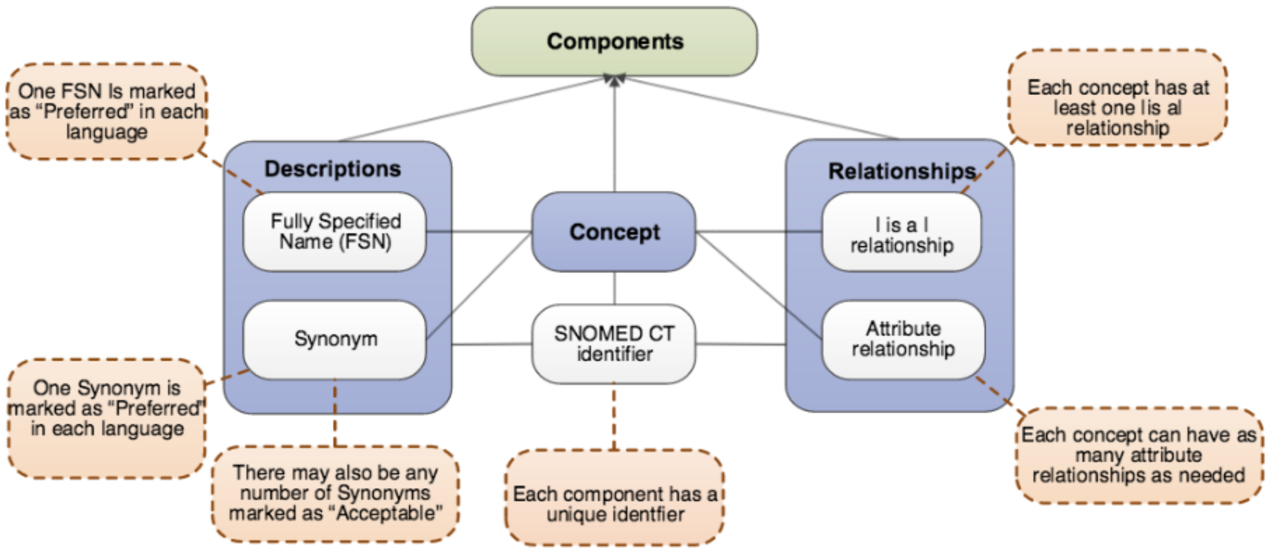

Figure 3.5: SNOMED CT Logical Model.

as follows:

$$CPCR = \frac{\sum_{i=1}^{N} CPTF}{N}, \tag{3.1}$$

where:
N: The number of clinical terms in the CP, $CPTF$: The *CP Term Factor*, defined as:

$$CPTF = \begin{cases} 1, & \text{if CP term complies with terminology} \\ 0, & \text{otherwise.} \end{cases}$$

In the above definition, $CPTF$ is equal to one if the CP term complies with a corresponding term used in the selected reference terminology system (SNOMED CT in our case); otherwise, $CPTF$ is set to zero. This definition agrees with our final objective to standardize CP terms and to make $CPCR$ approach one. Note that partial matching with SNOMED CT is not considered in the above formula because the benefits of terminology standardization are realized only with standardized terms. Our standardization is flexible

Table 3.2: Analyzing terms from Stroke CP

CP Term	SNOMED CT Term	SNOMED CT ID	*CPTF*
Urinary tract infection	Urinary tract infection	68566005	1
Swallowing screen	Screening for dysphagia	431765005	0

regarding the preferred terms in the sense that the exact matching with SNOMED CT preferred terms is not strictly required to achieve the full CP digitization (although preferred). Terminology compliance (not only the exact match) is considered in our framework. For example, the preferred SNOMED CT term 'indwelling urinary catheter' with SNOMED CT Identifier (SCTID) 23973005 has another acceptable term in SNOMED CT system, called 'indwelling bladder catheter'. For both terms, $CPTF = 1$. For an illustrative term compliance example, Table 3.2 considers few terms from the stroke clinical pathway that is used in Ontario hospitals.

We analyzed the main components of the stroke CP and obtained a $CPCR$ close to 61%. This result agrees with our finding (in the critical analysis of the literature review) that non-standardized clinical pathways are a major source of missing data in healthcare information systems. Therefore, there is a need to improve the CP compliance with terminology systems. In our vision, this CP non-compliance with standard terminologies is one of the main reasons to consider CP computerization as still at its infancy. To address the holistic framework for standardization of CPs and making them machine-readable, additional definitions and concepts are introduced below.

3.2.4 Standardization of CP Terminology

In general, CP terms can be classified into the following categories:

- Standardized terms (or initially standardized terms): Clinical terms in the CP that comply with SNOMED CT. These are terms with $CPTF = 1$.
- Non-Standardized terms: Terms in the CP that do not comply with SNOMED CT. These are terms with $CPTF = 0$, and can be further classified into the following:

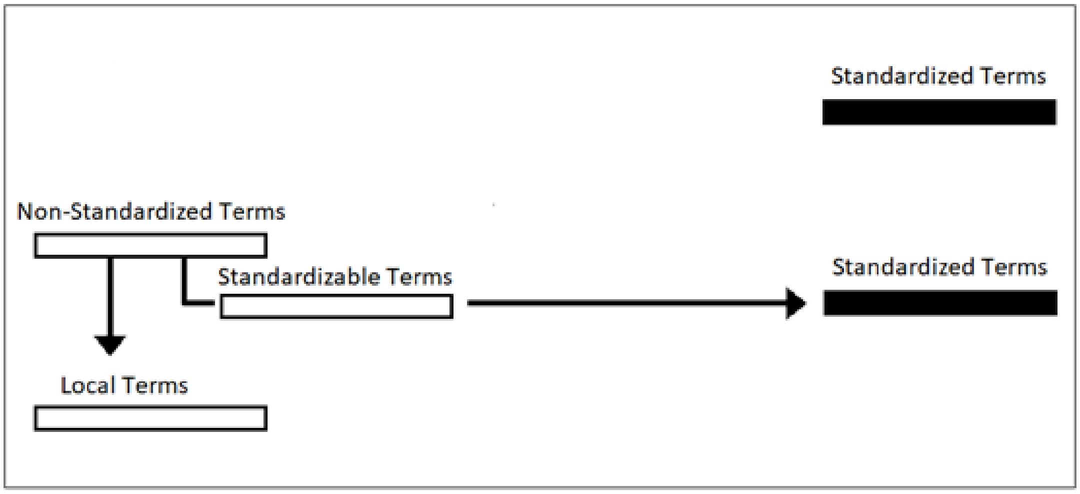

Figure 3.6: Standardization of CP terms as a key step to increase the digital visibility.

- Standardizable terms: Terms that have equivalent terms in SNOMED CT.
- Local terms: Terms that are local to the organization or country/region. Note that whenever possible, it is preferred to use international standardized terms over local terms.

CP term standardization requires the use of standardized terms during new CP developments, or updating the terms for existing CPs. For already developed CPs, term standardization can be achieved by retaining standardized terms and replacing non-standardized terms by their equivalent standardized terms (Figure 3.6). When performing this standardization process, the CP ends up having only two types of terms: Standardized terms and local terms.

This term standardization step can be best performed by physicians and healthcare providers themselves since they are the domain experts. Alternatively, several techniques and resources have been used to achieve term standardization based on methods from Natural Language Processing (NLP) such as studies in [103, 104, 82, 105]. Even when the focus of such studies was not CP standardization, relevant parts can be adopted in the CP domain. For example, in CP domain, we noticed that the use of abbreviations is common, specifically nursing abbreviations in CPs. In such cases, the study by [103] is relevant in the CP domain, whereby the authors proposed an NLP-based abbreviation disambiguation method for nursing notes through an abbreviation normalization module. Using a Python-based web crawler (called Scrapy 1.5), they automatically collected common nursing abbreviations in medical notes from Tabers Medical Dictionary and Nurselabs, and stored them in an abbreviation database together with their complete forms. Consequently, the abbreviation normalization module tokenized the free text in medical notes to single words, and then replaced any occurrences of detected abbreviations with the complete term by consulting with the abbreviation database. For example, if the free text (after tokenization) includes the abbreviation word "CT", then that abbreviation is automatically replaced by "computerized tomography" from the database. In our CP domain, "computerized tomography" is a SNOMED CT term with the identifier 77477000.

It is important to mention that since CPs are treatment plans of interventions and procedures applied on patients, extreme caution should be exercised when utilizing automatic standardization of medical terms in CPs for safety reasons. Therefore, machine-based standardization methods should be considered as decision-support methods rather than decision-making techniques. The final decision makers in medical CP term standardization are terminology-knowledgeable human physicians and domain experts.

Another application of NLP in the field of clinical pathway standardization is using semantic similarity and relatedness to help domain experts in the CP standardization process. Semantic similarity is a metric that can be defined as a quantitative measure of likeness between terms based on their hierarchical distribution within an ontology [106]. Semantic relatedness is a form of measurement that quantitatively identifies the level of connectedness between two concepts based on existing semantic relations [107]. Thus, semantic relatedness is also a metric over the terms; however, semantic relatedness includes any relation between the terms, while semantic similarity includes only "is-a" relations [106, 107]. For example, in SNOMED CT, "ischemic stroke" with SNOMED CT ID (SCTID) 422504002, is similar to "cerebrovascular accident" with SCTID 230690007, but is also related to "ischemia" with SCTID 52674009 (refer to Figure 3.7) [108].

When domain experts have limited time to work on the CP standardization process, junior members of the standardization team can propose an initial SNOMED CT term for

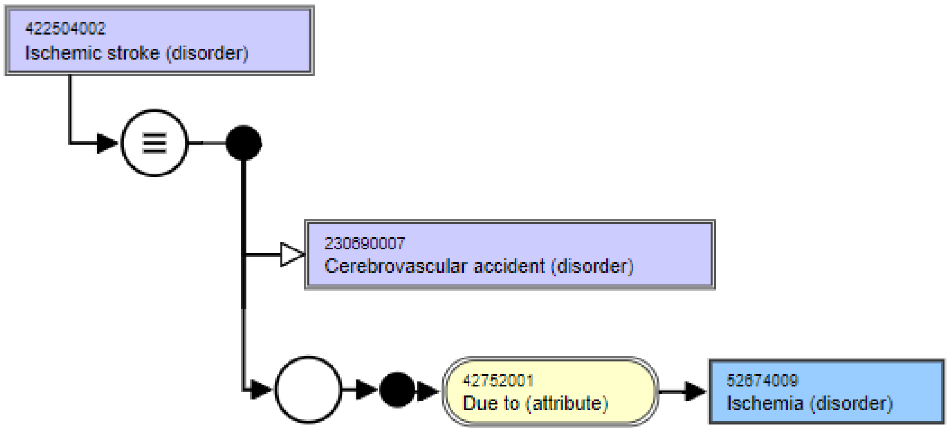

Figure 3.7: Part of the ischemic stroke concept hierarchy in SNOMED CT.

a local CP term. If the domain experts disagree on the accuracy of the proposed term, then in most of the cases the accurate term is a term similar to the chosen one. To automate the process of finding the correct term among similar terms, proper semantic similarity algorithms can be used to find other candidate terms similar to the proposed one.

The goal of algorithm 3.1 is to retrieve the list of candidate SNOMED CT terms. The algorithm limits the list of candidate terms to those that are most semantically similar to the initial term. The input to algorithm is the initial term and the output is the term approved by the domain experts. In steps 5 to 9, the algorithm searches the SNOMED CT ontology starting from the *root* of the initial term, and retrieves the sibling terms of the initial term based on a similarity threshold [1]. All siblings that are similar to the initial term are stored in the array SL. The terms in the array are then presented to domain experts who can either approve a sibling term or expand a sibling term to check its children. This is performed in steps 10 to 16. The "if statement" in step 11 evaluates to "true" only if a sibling term has been approved. The else part in step 13 allows the domain experts to expand a sibling term (c_i). The term c_i becomes the new *root* term, and the *repeat* cycle loops again to explore new possible terms.

It is worth mentioning that the metric ($CPCR$) is a type of metric that we call a

[1] The similarity threshold will be evaluated based on the selected semantic similarity measure. We are currently experimenting with different similarity measures.

Algorithm 3.1: CP SNOMED CT standardization using semantic similarity

Input : $SCTID_{init}$: Initial SCTID

Output: $SCTID_{approved}$: SCTID approved by domain experts

Data Structures:

SL: An array to store the suggested SNOMED CT terms, initially empty

root: A variable to store the $parents\_of(SCTID_{init})$ based on *is_a* relation, initially null

repeat

    foreach *child* $c_i \in child\_of(root)$ **do**

        if $((c_i \neq SCTID_{init}) \wedge (similarity(SCTID_{init}, c_i) < Threshold))$ **then**

            $SL \leftarrow c_i$;   `//` `Append` c_i `to` SL

        end

    end

    `/* Display` SL `to domain experts. Domain experts can either: expand a child node to check its children OR approve child node as a standardized term */`

    foreach *child node* $c_i \in SL$ **do**

        if *(c_i is approved)* **then**

            $SCTID_{approved} \leftarrow c_i$

        else if *(expand c_i)* **then**

            $root \leftarrow c_i$

        end

    end

until $(SCTID_{approved} \neq null) \vee (SL = \emptyset)$;

CP “digital quality” metric because it helps evaluate the “digital visibility” of existing or proposed CPs. A CP with SNOMED-*CPCR = 1* is a totally standardized CP (all CP terms could be made comply with SNOMED CT terms).

3.2.5 Development of a Coding System for Clinical Pathways

The era of computers brought with it the concept of digital coding systems to almost all products and services. A closer look at CPs currently in use reveals that they lack any type of coding system to identify them. In our vision, this is another key reason behind the low machine readability of CPs. The present gap between CPs and their integration with various types of HISs can be linked with this because CPs do not have any digital identifiers to recognize them. Therefore, we propose the development of a coding system specific for CPs to further advance their digital visibility, machine readability, and integration with other information systems.

Furthermore, a proper coding system applicable to CPs facilitates the mathematical modeling of CP data as will be addressed later in this thesis. In order for the suggested standardized coding system to form a foundation for CP digitization, to be accepted by healthcare professionals, and to be operational within existing information systems in health informatics, we propose that it be based on accepted international medical terminology systems. Our proposed universal CP coding system is introduced below.

3.2.6 Coding System of International Data

In the SNOMED CT system, each component has an identification number known as SNOMED Clinical Term Identifier (SCTID). For example, the SCTID for stroke is 230690007, the SCTID for Carotid Artery Disease is 371160000, and so on. A SNOMED CT component can be a concept, a description, or a relationship that conforms to the SNOMED CT logical model (see Figure 3.8).

The SCTID has a structure that includes information about the nature and source of the identified component, as well as the validity of the identifier. This structure supports

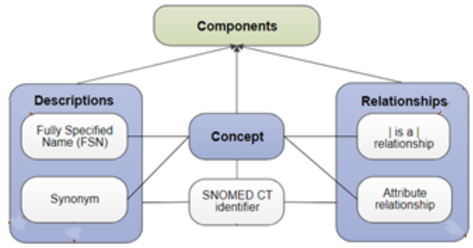

Figure 3.8: SNOMED CT Components.

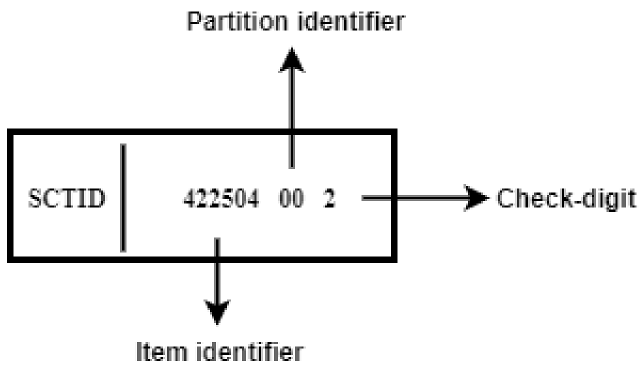

Figure 3.9: SCTID format for international SNOMED CT components.

many features, including the following (see Figure 3.9):

- Item identifier: Used to identify a component so that it is uniquely identified by the complete SCTID.

- Partition identifier: A two-digit identifier that distinguishes the code of different SNOMED CT component types and prevents the identical identifier from being allocated to different components. Thus, when an SCTID is read from a record or other resource, it is possible to determine whether it represents a concept, a description, or a relationship, before searching for the component being identified. Table 3.3 outlines SNOMED CT international partition identifiers (also called short format identifiers).

Table 3.3: SNOMED CT partition identifiers

Partition ID	Component Type
00	Concept
01	Description
02	Relationship

- Check-digit: The last digit of SCTID is generated from the other existing digits, and is used to validate the identifier to minimize errors from human input such as incomplete copy-paste actions.

Based on the structure of SNOMED CT, we propose a CP coding system that adopts the following rules. Every CP should have its identifying code with a CP-specific partition identifier. The new identifier is proposed as follows: There are currently three partition identifiers used (see Table 3.3) for which the values '00', '01', and '02' are allocated to concept, description, and relationship, respectively. Therefore, we propose the use of an agreed-upon, unallocated partition identifier for specifying CP identifiers (CPID). Currently, the next unallocated value is '03'; thus, we present our examples using this value. This agrees with the SNOMED CT logical model for future expansion because it is stated by SNOMED CT international that "all other partition-identifier values are reserved for future use" [98].

As an illustrative example under this proposed coding system, since the SCTID for ischemic stroke is given as 422504002, then the CPID allocated to the ischemic stroke CP is formed by the ischemic stroke item identifier 422504, the new partition identifier 03, and a corresponding check digit 9, to yield a CPID for ischemic stroke CP as 422504039.

For the medical components and terms inside CPs, we propose using the same SCTID codes for each component, and when reporting or communicating a CP component to external computer systems, we propose a code that consists of two SCTIDs separated by a hyphen, as explained in the following example. Considering the ischemic stroke CP, one of the medical steps is to check for the "Screening for dysphagia" of the patient because stroke often causes a swallowing disorder called dysphagia. If not identified and managed properly, it can cause pneumonia, poor nutrition, and increased disability [109].

Table 3.4: Description of medical data coded using the developed framework.

Code	Description
1539035-418426008	X-ray of fingers in the CP of acquired trigger finger
1539035	CPID for acquired trigger finger
73211032	CPID for diabetes mellitus
93870037-418733007	Ultrasound scan of abdominal vessels for CP of liver cancer

We propose the code of screening for dysphagia reported from the CP of stroke to be 422504039-431765005, where 422504039 represents the CPID for ischemic stroke, and 431765005 is the SCTID for screening for dysphagia. Therefore, the general format of the standardized coding system of CP components can be represented by the hyphenated code CPID-SCTID, where the CPID to the left of the hyphen refers to the CP itself (i.e. the international CP identification code), and the SCTID to the right of the hyphen refers to a component inside that CP. The hyphenated coding system for CP contents would be a very novel and useful concept in digital databases because the link between the medical interventions and their associated CPs and diseases is always present and stored in digital format. This facilitates the use of CP data analytics for decision-support. Table 3.4 describes sample medical data that is represented using the developed coding system.

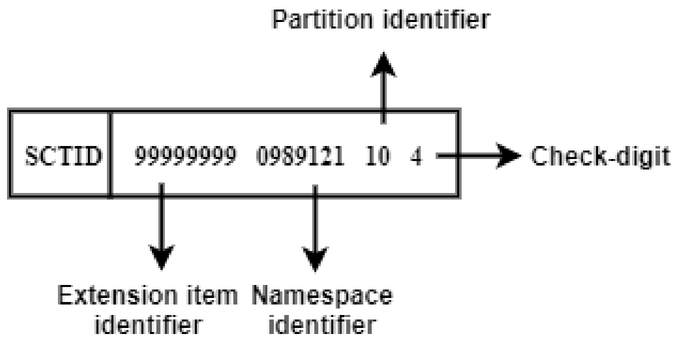

Figure 3.10: SCTID long format - Applicable to components originating from a SNOMED CT extension for local use.

Table 3.5: SNOMED CT local partition identifiers.

Partition ID	**Component Type**
10	Concept
11	Description
12	Relathionship

3.2.7 Standardizing and Coding System of Local Data

The application of CPs results in rich data whose capacities are currently unused because they are not fully captured in digital format. In order to maximize the use of CP data, local components can also be standardized and digitized. SNOMED CT supports an "extension" coding format for local use (called long or local format) in which a namespace identifier is required to indicate the organization responsible for the SCTID long format, such as hospitals, healthcare authorities, provincial governments, etc. [110]. This local SCTID supports the structure shown in Figure 3.10. In this structure, the partition identifier values are shown in Table 3.5.

The namespace identifier is a seven-digit integer number, left padded with zeros as necessary to ensure there are always seven digits in the value. It is allocated to organizations by SNOMED International (the International Health Terminology Standards Development Organization), which is the not-for-profit organization that develops and promotes the use of SNOMED CT to support safe and effective health information exchange codes [111].

Table 3.6: Example SNOMED CT namespace identifiers.

Namespace Identifier	**Organization**
1000087	Canada Health Infoway (English Canadian Extension)
1000112	Alberta Health
1000136	University of Victoria Health Terminology Group
1000038	UK National Health Service
1000026	Cambridge University Hospital
1000032	Australian e-Health Research Centre

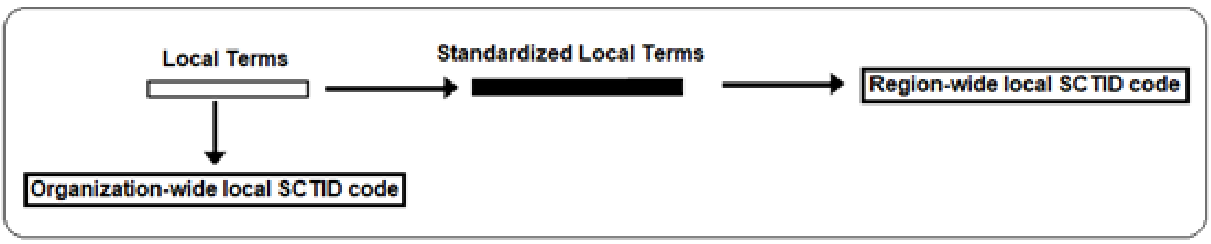

Figure 3.11: Region-wide vs. organization-wide local SCTID codes.

Table 3.6 shows example namespace identifiers [112].

Although the long format SNOMED CT code can be used for local CP terms, we recommend a type of standardization of local terms on a national level for improved semantic interoperability (see Figure 3.11). For an example from Canada, terms that are local to a hospital or a province could be standardized to achieve standardized Canadian local terms (Canada-wide). In cases where this local standardization is not possible, then an organization-wide local SCTID code can be used. Canada Health Infoway has the responsibility of unifying terms to realize a Canada-wide SCTID coding system. We recommend that all hospitals and provinces cooperate with Canada Health Infoway. Note that "region-wide" used in Figure 3.11 could mean city-wide, province-wide, or country-wide, depending on the level required for standardizing local terms.

The CPID addressed above is an international ID derived from the disease (e.g., stroke) of that particular CP. However, in cases where a local CPID is required, then the partition identifier value '13' can be used (or another agreed-upon value) for local CP identification using the local format (see Table 3.5). An example code related to a local component inside the CP of diabetes mellitus would be "73211032-607978811000087105" to denote "Canadian diabetes education program (607978811000087105) inside the CP of diabetes mellitus (73211032)".

Aside from the benefits described above, the standardized terminology and coding system can facilitate (without ambiguity) many computations related to CP analysis and application in hospitals. For example, to verify the conformance of a patient treatment process with the patient's assigned CP (e.g. if all recommended blood tests have been ordered), the following equation can be used: Denoting 'CP blood tests application ratio' as α, then,

$$\alpha = \frac{\text{Number of matching blood test codes in EMR}}{\text{Number of blood tests in CP}}. \tag{3.2}$$

CP conformance analysis is important for both improving the CP itself and for improving the treatment processes [42].

3.2.8 Ontology-Based Modeling and Description Logic

The proposed framework adopts the ontology engineering and semantic web approach in modeling CPs and constructing the knowledge base. Semantic web components like SWRL and Web Ontology Language are based on Description Logic [58, 113, 114]. Description logic is the formal base of the rules to construct useful and valid knowledge representations that are widely used in ontological modeling. Description logic is equipped with a formal semantic that is a precise specification of the meaning of ontologies, which helps in modeling the relationships between ontology entities in a domain of interest.

There are three types of description logic entities used in web ontology language: concepts (or classes), roles (or properties), and individual names. Concepts represent groups or sets of individuals, roles model the binary relationship between the individuals, and individual names represent single individuals in the ontology's domain of interest.

An ontology does not fully describe a particular state; rather, it consists of a set of statements called axioms. Axioms must be a true description of a situation. "These axioms typically capture only partial knowledge about the situation that the ontology is describing, and there may be many different states of the world that are consistent with the ontology." [115]. It is customary to separate axioms into three groups: assertional (ABox), terminological (TBox) and relational (RBox) [115].

3.2.8.1 Assertional (ABox) Axioms

ABox axioms capture knowledge about individuals. They represent how named individuals relate to each other in the domain and the concepts to which they belong. The most common ABox axioms are concept assertions [115]. For example, in a disease hierarchy, there are parent classes, super-classes and individual diseases. The following axiom asserts that the individual named *cerebrovascular_disease* is an instance of the concept or class *Upper_level_disease*.

$$\textit{Upper\_level\_disease(cerebrovascular\_disease)}. \tag{3.3}$$

Property or role assertions describe relations between individuals. For example, the assertion

$$\textit{onSameLevel(cell\_structure, body\_tissue\_structure)}, \tag{3.4}$$

describes that the individual named *cell_structure* is in the relation represented by *onSameLevel* to the individual *body_tissue_structure*. Because description logic does not assume that individual names are unique, different names might refer to the same individual (unless explicitly stated by axioms). The fact that *cell_structure* and *body_tissue_structure* are different individuals does not logically follow from the previous axiom. To ensure that, we must add the *individual inequality* assertion to ensure that *cell_structure* and *body_tissue_structure* are actually different individuals.

$$\textit{cell\_structure} \not\approx \textit{body\_tissue\_structure}. \tag{3.5}$$

In ontology engineering, *individual equality* ABox axiom is often needed for *ontology alignment*. Two ontologies originating from different sources may have two different individual names that refer to the same instance. This causes ontology alignment mismatch.

For example, *cerebrovascular_accident* is another name for *stroke*, thus *individual equality* can be used to indicate that these two different names refer to the same instance, as follows.

$$\textit{cerebrovascular\_accident} \approx \textit{stroke}. \tag{3.6}$$

3.2.8.2 Terminological (TBox) Axioms

Axioms of type Tbox describe relationships between concepts. For example, every disease is a clinical finding. This fact can be expressed by the *concept inclusion* as follows

$$\textit{disease} \sqsubseteq \textit{clinical\_finding}. \tag{3.7}$$

Asserting that two classes have the same instances can also be performed by *concept equivalence* axioms. For example:

$$\textit{disorder} \equiv \textit{disease}. \tag{3.8}$$

This assertion can be used for synonym classes (i.e., equivalent concepts) which is also used in ontology alignment.

3.2.8.3 Relational (RBox) Axioms

In semantic web paradigm, roles (or properties in OWL language) can also be organized in hierarchies, and can have properties. RBox axioms denote properties of roles. For example, the role *parentDiseaseOf* can be a sub-role (or "sub-property") of *ancestorDiseaseOf*. This can be denoted as

$$\textit{parentDiseaseOf} \sqsubseteq \textit{ancestorDiseaseOf}. \tag{3.9}$$

The logical inference of (3.9) is that every pair of individuals related by the property *parentDiseaseOf* is also related by the property *ancestorDiseaseOf*.

3.2.8.4 Description Logic Constructors

Ontologies can more accurately model domains with complex situations using description logic *constructors* by allowing new roles and concepts to be constructed. There are different types of constructors (e.g., Boolean concept constructors and role restrictions).

Boolean concept constructors

Examples from this category are listed below.

$$\top \sqsubseteq \mathit{Male\_patient} \sqcup \mathit{Female\_patient}. \tag{3.10}$$

$$\mathit{Male\_patient} \sqcap \mathit{Female\_patient} \sqsubseteq \bot. \tag{3.11}$$

where the *top concept* $\top$ is a special concept with every individual as an instance, and the *bottom concept* $\bot$ is the special concept with no individuals (ϕ) as instances.

Role restrictions

One of the most interesting features of description logic is the ability to link concepts and roles together. For example:

$$\mathit{Parent} \equiv \exists \mathit{parentOf}.\top. \tag{3.12}$$

This is a *concept equivalence* statement that indicate that a parent is someone who is a parent of at least on individual.

In Protégé, description logic statements are generated using owl statements. For example, the axiom that the classes *disorder* and *disease* are equivalent can be represented by using the following statement:

⟨owl:Class rdf:about="http://www.semanticweb.org/.../ontology#Disease"⟩
⟨owl:equivalentClass rdf:resource="http://www.semanticweb.org/.../ontology#Disorder"⟩
⟨/owl:Class⟩

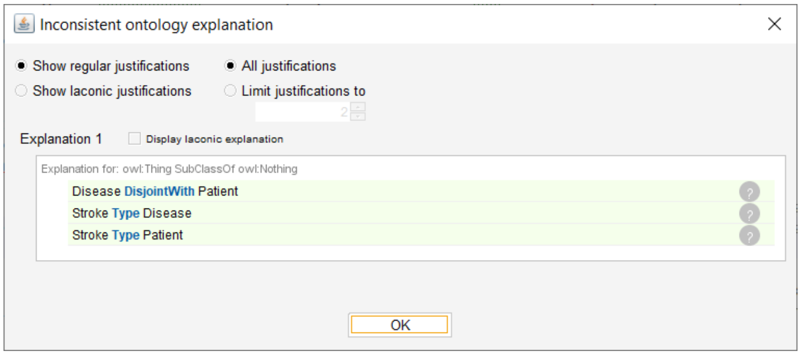

Figure 3.12: Description logic reasoning in Protégé.

The fact that the concepts *Patient* and *Disease* are disjoint classes can be expressed as:

⟨owl:Class rdf:about=“http://www.semanticweb.org/.../ontology#Disease”⟩
⟨owl:disjointWith rdf:resource=“http://www.semanticweb.org/.../ontology#Patient”⟩
⟨/owl:Class⟩

Through description logic reasoning, any attempt to create an individual that belongs to both disjoint classes would be considered as *inconsistency* in Protégé.
For example, Since *Patient* and *Disease* are declared as disjoint classes, declaring *Stroke* (which is a disease) as an individual of the class *Patient* results in the logic reasoning error shown in Figure 3.12.

In summary, ontologies, through their formal logic base, define the terms, their semantics, relations, and constraints describing domain knowledge to provide a shared understanding that can be communicated between people and heterogeneous applications in a machine-understandable way, thus facilitating semantic interoperability among information systems [116, 113].

Another advantage of adopting the ontology-based approach is that ontological modeling facilitates a hierarchical meta-level/disease-level architecture in which the generic and abstract CP domain concepts are modeled at the meta-ontology or upper-ontology level, while disease specific CPs are extended and specialized in the disease-specific ontologies. Furthermore, all CP ontology elements (e.g., interventions, events, observations, outcomes, etc.) are represented textually by SNOMED CT terms, and coded numerically using SNOMED SCTID numbers. This is achieved by integrating a SNOMED CT ontology that enables linking SNOMED CT terms used in CP ontology to their SNOMED CT codes extracted from SNOMED CT ontology.

3.3 CPMS Integration with HISs

In order to address the proper communication level between clinical pathway management systems and HISs, we first briefly consider the major subsystems of HISs [117, 118, 119, 120]. We then consider CPMSs, and analyze the relationship between CP data and HISs data.

3.3.1 Electronic Health/Medical Record Systems (EHR/EMR):

An EMR is a digital version of patients' paper charts. It is typically used by single-practice clinics and small hospitals for their local records of patients. An EMR typically contains the medical history of the patients, diagnoses, and treatments. EMRs provide numerous advantages over paper records including timely reminders for patient appointments and checkups, digital data, and improved patient care. An EHR can be viewed as a 'large-scale' EMR that stores more data and facilitates the sharing of health records across different institutions. It is worth mentioning that modern systems in use today are capable of playing the roles of both EMR and EHR since they are larger systems and offer options of either keeping the patients' data local inside the institution, or sharing it with other systems. Therefore, the terms EMR/EHR or simply EMR (or sometimes EHR) are widely used today to refer to these systems [10, 121, 122, 123].

3.3.2 Laboratory Information Systems (LIS):

LIS are software systems with features that support modern laboratory operations and informatics. The main functions of LIS include recording, managing, and storing clinical laboratory data for patients. LIS have traditionally been most adept at sending laboratory test orders to lab instruments, tracking orders, and recording lab test results. In addition, LIS support the operations of public health institutions and their labs by managing and reporting critical data concerning immunology and infection [124, 125].

3.3.3 Radiology Information Systems (RIS):

RIS are the core systems for the electronic management of imaging departments, and are critical to the efficient workflow of radiology practices. The main functions of RIS can include scheduling of patients, managing resources of radiology departments, image performance tracking, and distribution of results. A central component of RIS is the radiology PACS (Picture Archiving and Communication System), which provides storage and easy access to medical images from various sources (e.g., computed tomography (CT), medical ultrasound, X-ray, magnetic resonance imaging (MRI), computed radiography (CR), etc.) [117].

3.3.4 Pharmacy Information Systems (PIS):

PIS (also called Pharmacy Management System) have various functions to maintain the organization and supply of drugs. A PIS can be a separate system for pharmacy usage, or it can be coordinated with inpatient hospital order entry systems. PIS are used to increase patient safety, report drug usage, reduce medication errors, and track costs. Outpatient PIS have a strong emphasis on medication labeling, drug warnings, and instructions for administration. The effective and safe dispensing of pharmaceutical drugs is the most important function of PIS. During the dispensing process, PIS prompt pharmacists to verify that the medication that they have filled is for the correct patient, that it is of the right quantity and dosage, and that the information on the prescription label was accurately printed [10, 126].

3.3.5 Clinical Pathway Management Systems

The concept of applying CPs in hospitals was a novel initiative to adopt successful management practices in healthcare. Therefore, since their introduction to healthcare institutions, the main objective of CPs was to coordinate and 'manage' healthcare processes as central components. CPs contain all the interventions required to treat patients; thus, within CPs lies the very heart of medical planning and treatment, including cost and quality factors in healthcare. The considerations above suggest that CPs were designed to produce all types of data in healthcare as described above (e.g., EMR data, LIS data, etc.). Fig. 3.13 makes this point clear by illustrating how CPs generate data for all types of HISs discussed above. Two CPs for Diabetes Mellitus and Carotid Artery Disease are illustrated. As shown by the arrows in the figure, both CPs include order instructions that result in data that need to be transferred to all types of HISs.

Thus, computerized CP management systems should be designed and positioned such that they are "centralized" (i.e., positioned at the centre of HISs) and allowed to communicate with all types of HISs (not only EMRs, as was the common theme in CP systems found in the literature). Besides using ontological modeling and SNOMED CT-based communication with other systems, this positioning and communication level of CPMSs can be enhanced by equipping CP management systems with Health Level 7 (HL7) messaging functionality to communicate with existing HISs (Fig. 3.14). HL7 consists of a set of international standards for the transfer of clinical data between software applications [127]. This is achieved through standard, machine-readable HL7 messages. Generation of standard HL7 messages can be automated through application programs in high-level languages such as the Java-based HL7 Application Programming Interface (API) toolkit [128]. Fig. 3.15 shows an illustration of an HL7 observation result message to communicate the result of human immunodeficiency virus status using SNOMED CT encoding.

3.4 Independence of CP Management Systems

Without CPMS independence, the potential for utilizing CP data, and for using computerized CP systems as decision-support systems in healthcare, could not be fully exploited.

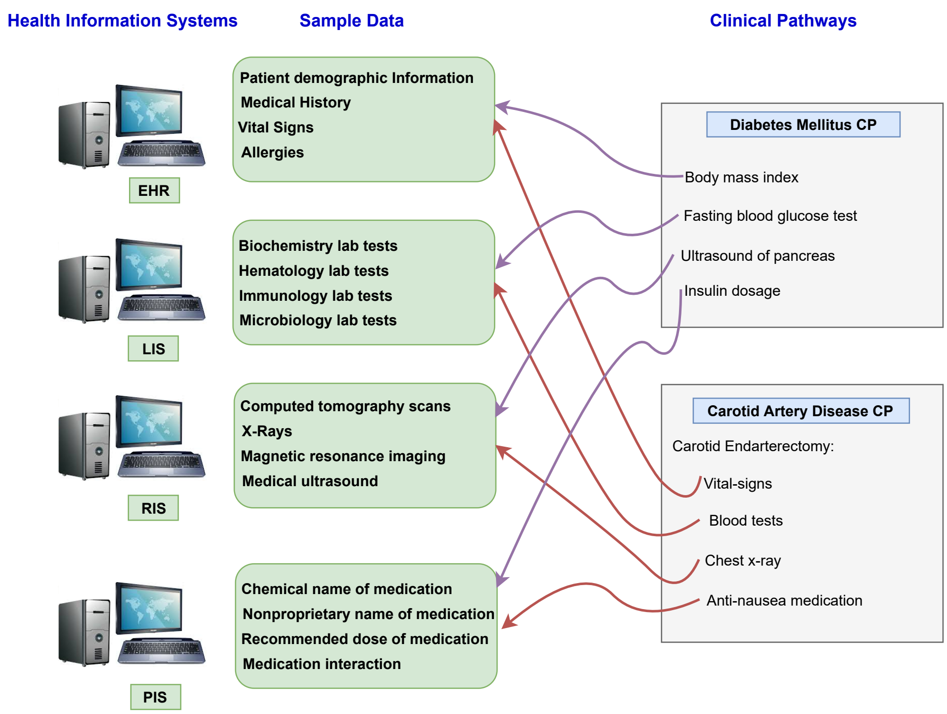

Figure 3.13: CPs produce data for all types of HISs.

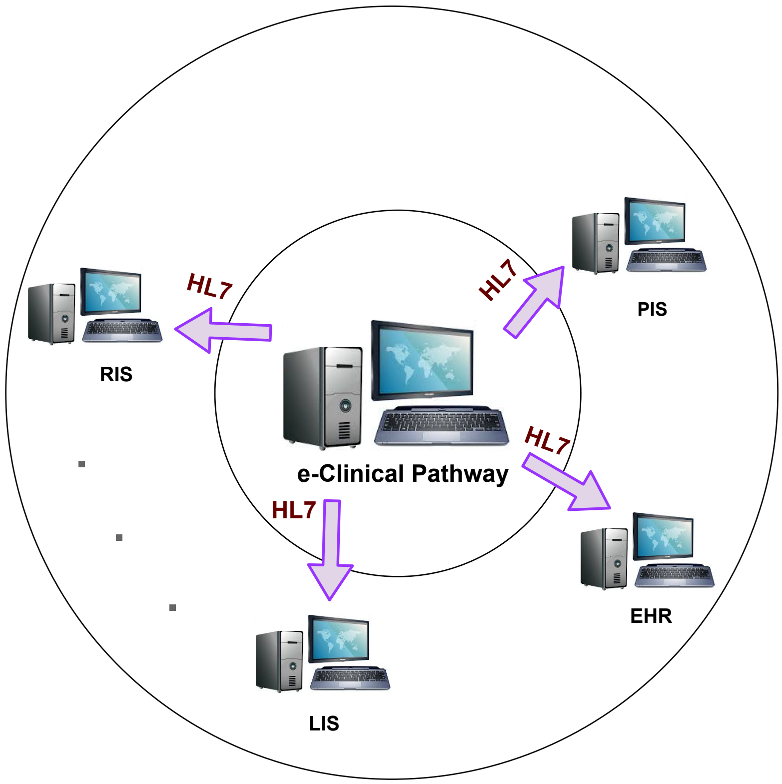

Figure 3.14: HL7 enables CPMSs to communicate with all existing HISs.

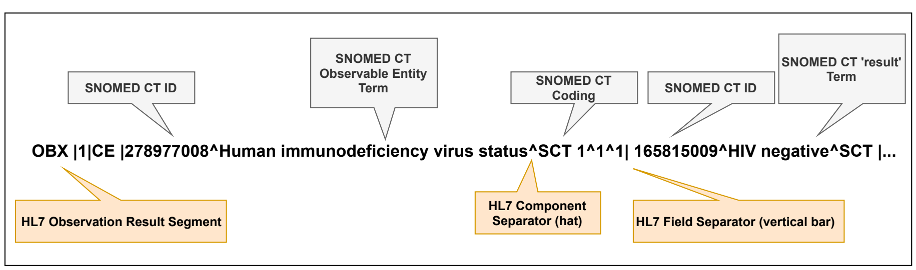

Figure 3.15: Illustration of an HL7 observation result message.

The independence of CP management systems is another contribution of our framework. Independent CP systems have their own CP-specific algorithms and can perform their own CP-related functions, support CP data analytics, and act as decision-support systems in healthcare. One of the core ideas behind CPMS independence is to allocate a specific repository (or database) for CP outcomes and include useful codes based on customized algorithms. Since CPs produce patient treatment paths, allowing CPMSs to have their own customized repository of CP data and paths is a corner-stone to achieve their independence. CP paths (also called CP traces) can be recorded internally by time-stamping healthcare events and storing them in sequence as the patient progresses through the CP treatment. This allows all patient-related CP traces to be recorded in the CP trace repository, and subsequently used for CP data analytics and other useful decision-support functions.

3.5 Conclusion

In this chapter, details of the proposed CP automation framework are addressed. The main components of the framework can be divided into three parts: CP automation, CPMS integration with health information systems, and independence of CP management systems. In CP automation, the concepts of SNOMED CT standardization, CP coding system, and ontology-based modeling were addressed. CPMS integration with health information systems could be realized by using standardized communication methods such as HL7 and standardized ontologies. The independence of CPMS can be achieved by the addressed framework's elements, including a system's component that includes a customized repository for CP data. Independence can also be enhanced by CP-specific algorithms that facilitate data analytics and decision support. Realizing the proposed framework in a working system requires proposing a system architecture that integrates the various components of the framework. This is the topic of the next chapter on the prototype design and architecture.

Chapter 4

Prototype Design and Architecture of the Proposed Framework

The proposed framework is a generic framework in the sense that it can be designed and implemented using different overall structures, ontology designs, and programming languages. The framework can be applied within various types of healthcare systems that are related to CP management. For example, a commercial traditional CP system can be restructured and modified to implement the proposed framework. This creates opportunities for healthcare institutions around the world to adopt and benefit from the framework.

In this chapter, we present our methodology behind developing the proposed framework through addressing the conceptual design of a clinical pathway management system (CPMS) that realizes the CP standardization and digitization framework, and ensures the independence of CPMSs. The proposed prototype CPMS that is based on the framework is conceptually designed in three layers (i) Knowledge Base (KB) layer, (ii) Inference and Data Analytics layer, and (iii) CP Management Tools layer. Figure 4.1 outlines the major components of each layer. In addition to the three layers, our methodology integrates the CPMS with an internal database to store patients' CP data and traces in an internal CP repository. Furthermore, the system can produce sample HL7 messages through an HL7 engine, which enables it to communicate with other HISs. To illustrate the basic functionalities of the framework, the prototype was implemented based on a stroke clinical pathway with the help of domain experts from the Regional Stroke Unit at Thunder Bay Regional Health Sciences Centre (TBRHSC), who thankfully offered to share sample CPs with us for the purpose of this study [129]. A stroke rehabilitation CP was also shared with us by The Ottawa Hospital. Below we describe the structure and main functionality of each layer of the prototype CPMS.

4.1 Knowledge Base

The Knowledge base of the CPMS follows an ontology-based design that includes meta-CP knowledge, disease-specific knowledge, and rules for ontological reasoning expressed in SWRL [57]. The meta-CP knowledge is represented by an upper ontology that is built from three ontologies to represent CPs in a generic meta-level format, namely meta-CP ontology, SNOMED CT ontology, and Time Ontology, as shown in Figure 4.2. This ontological structure is an important component of our methodology as it allows any CP to be expressed as a series of clinical interventions over a specific time period, drawing from the vocabulary and SCTIDs existing in the SNOMED CT standard. The developed knowledge base of the prototype, being integrated with Protégé, provides easy ontology editing tools to facilitate clinical pathway modifications over time, as shown in Figure 4.2.

4.1.1 Meta Clinical Pathway Ontology

An ontology is a formal representation of a set of concepts within a specific domain and the relationships between those concepts. Ontology Engineering is a field of study that focuses on the methodologies and tools for building ontologies. The meta-CP ontology in our system serves as the schema for all possible disease-specific CP ontologies (e.g. ischemic stroke ontology, diabetes mellitus ontology, etc.). The meta-CP ontology is modeled in consultation with domain experts and CP ontologies available in the literature. A generic CP ontology should capture all types of CP data and artifacts that can be encountered in any CP execution regardless of its nature. Therefore, the meta-CP ontology includes the major type of building blocks that one can use to create disease-specific CPs. We

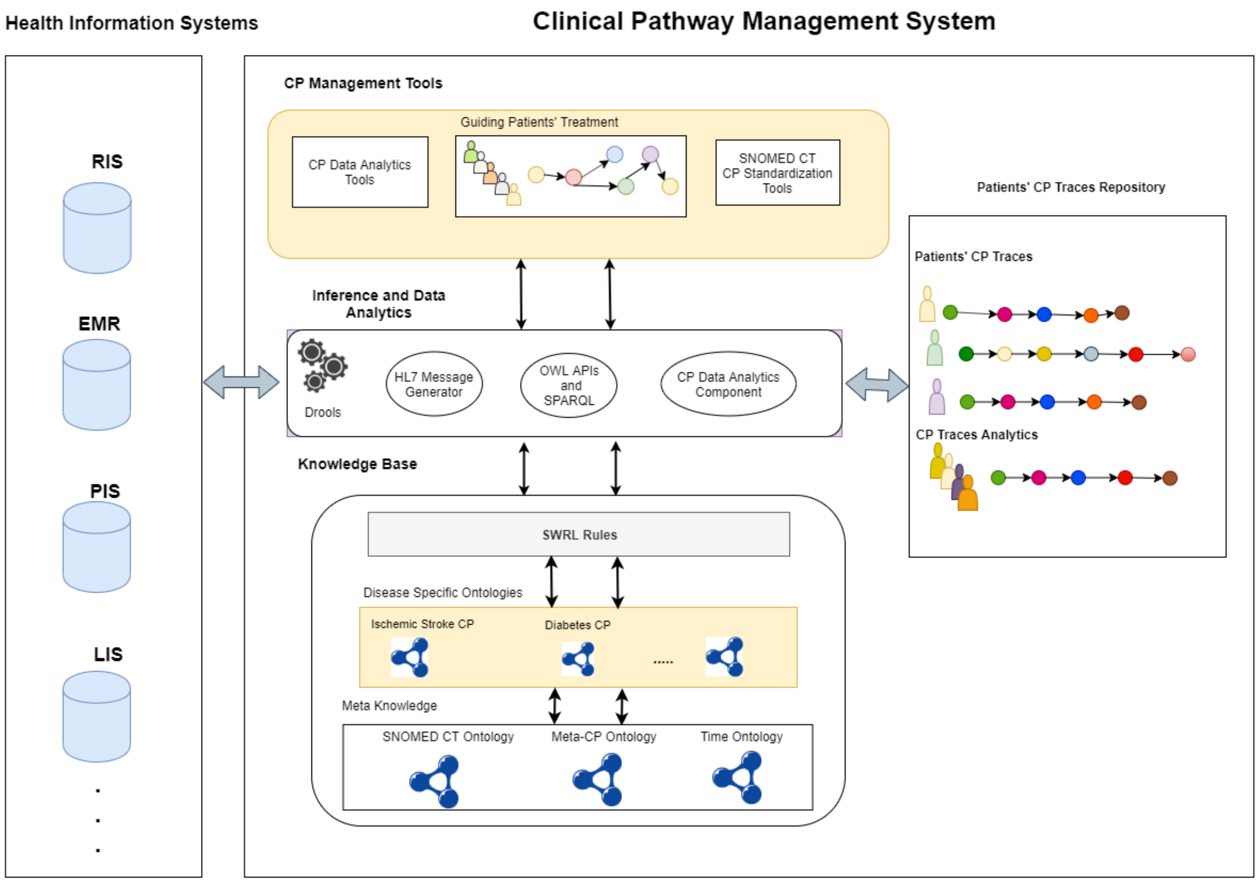

Figure 4.1: Conceptual design of the proposed framework based on our methodology.

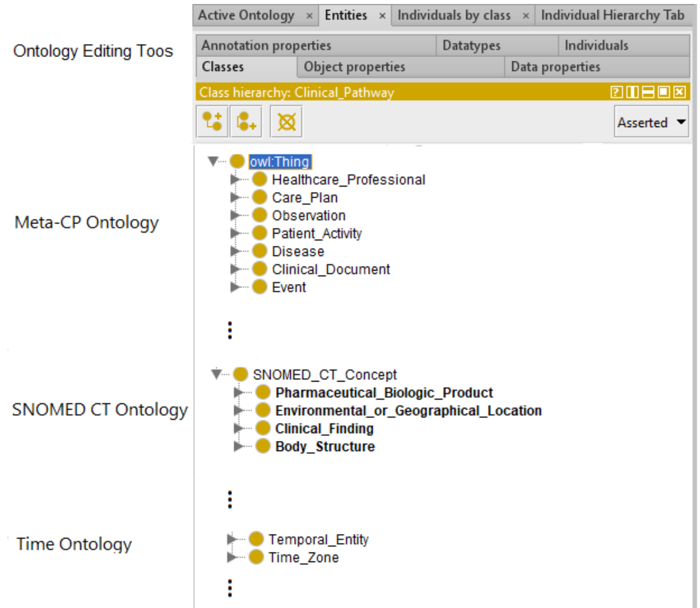

Figure 4.2: Parts of the meta knowledge layer shown in Protégé editor.

followed an ontology engineering process which constitutes the following four main phases, to develop the meta-CP ontology: (A) Domain Understanding, (B) Ontology Design, (C) Ontology Development, and (D) Ontology Evaluation. Note that "evaluation" is used here as a generic term for assessing the ontology, which includes ontology validation by domain domain experts. Figure 4.3 depicts the ontology engineering process, which is described below along with its utilization in our CP automation framework.

4.1.1.1 Domain Understanding

The two major steps of domain understanding include reading about the domain and working/consulting with domain experts. For our research, communicating directly with domain experts, reading literature on stroke clinical pathways and LOS, and conducting relevant research were our major sources for domain understanding. For the domain of stroke, we had the opportunity to work with stroke domain experts in the research related to ischemic stroke incidence and risk factors in Northwestern Ontario [130]. For the domain of hospital LOS, both communicating with stroke experts on stroke LOS and our

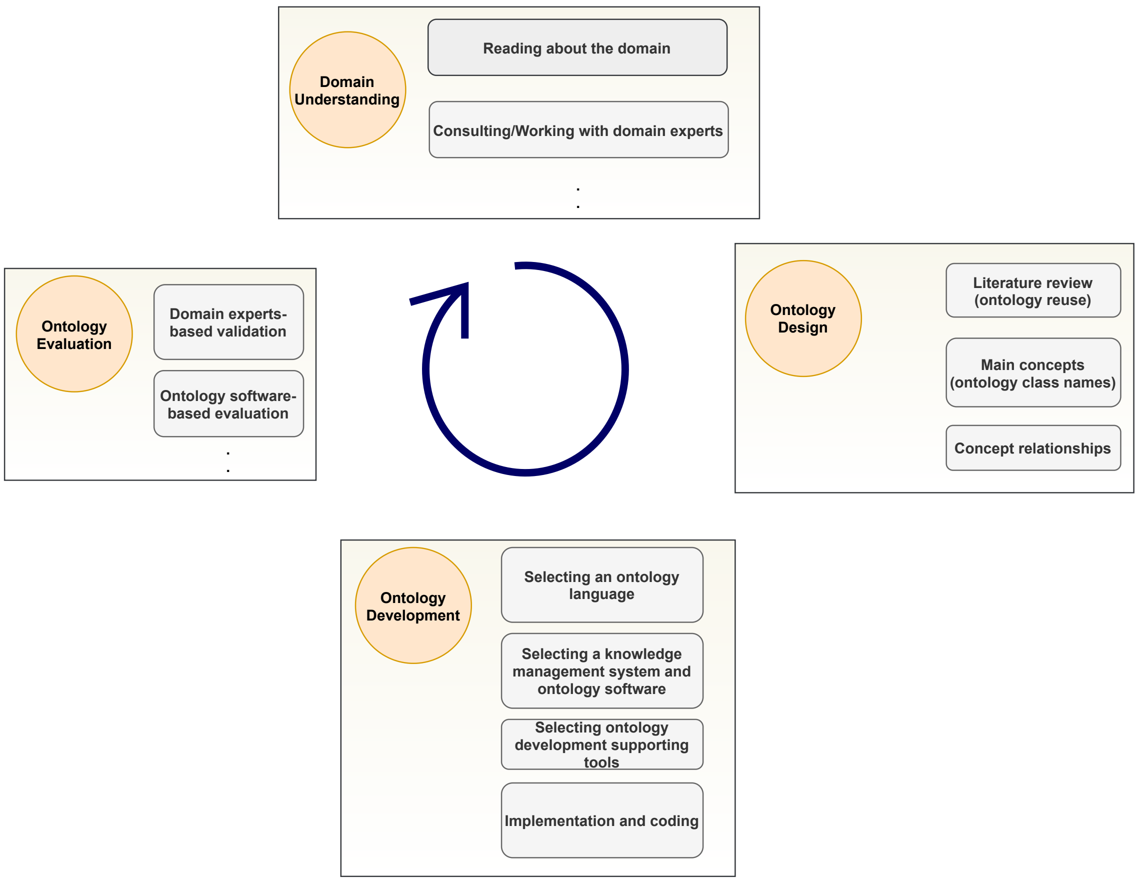

Figure 4.3: Major phases of our Ontology Engineering process.

research in [131] were key stages to understanding the LOS domain. To give some examples, communicating with the domain experts through research meetings allowed us to ask questions about stroke patients in the database to understand the medical data, learn more about the nature of stroke disease and the LOS of stroke patients, and understand the medical vocabulary used in relevant medical domains. This was essential for us to understand research papers of a medical nature (e.g., stroke ontology papers, CP papers, etc.) and to improve the design of the ontology (e.g., adding the CP variance class and variance types). In addition, understanding of vocabulary helped us to understand the structure of SNOMED CT and to differentiate between some SNOMED CT terms (e.g., intracranial hemorrhage (SCTID 1386000), subarachnoid intracranial hemorrhage (SCTID 21454007), and ischemic stroke (SCTID 42250400)).

4.1.1.2 Ontology Design

The first step in designing an ontology is to search for available ontologies, analyze their design, and see how they fit the domain/application being considered. We performed a literature review about CP ontologies to investigate their design and main CP concepts. Our ontology design was inspired by the ontologies already available in the literature. In addition, our stroke knowledge, as well as, the feedback we received from our domain experts, helped us to modify and improve the ontology design.

Figure 4.4 presents a term frequency word cloud for the class names used by researchers on CP ontologies. This text analytics-based visual representation was useful to gain an insight into which classes are used in literature in the context of CP ontology design. We noticed that the main common ontology concepts used include: clinical pathway, observation, patient, intervention, CP trace, event, clinical condition, disease, symptom, outcome, patient education, and procedure.

Another observation from our literature review was the lack of standardization, which prompted researchers to use different class names for the same CP concept in their ontology design (e.g., medical document, health document, or document). This was the main reason that a mismatch was discovered between some ontology classes found in the literature. For example, the standardized SNOMED CT term to represent documents is “clinical document" (SCTID 423876004). Another example is that some authors used the term nursing_procedure as an ontology class name, whereas the same CP concept was named nursing_care by other researchers. In our framework, standardization is the appropriate solution to this inconsistency. Thus, we adopted standardized terms in ontology class design, whenever possible. For instance, the formal SNOMED CT term for the care given by nurses is nursing_service, with the SNOMED CT ID 708170008. Without standardization, the generated ontology cannot be considered as a meta level ontology, but just a "local top ontology", which is the case for CP ontologies found in the literature.

Another observation from the literature review was that although CP variance is an important concept, few authors included it in their ontology design. In addition, the major sub-classes of CP variance were not considered. In our ontology design, the variance was a main ontology class, which we further sub-classed into its main categories: medical evidence, comorbidity, facility resources, declined by patient, and declined by family.

Figure 4.4: Word cloud of CP ontology classes used in literature.

4.1.1.3 Ontology Development

The ontology was developed using Protégé OWL-based ontology software and knowledge management system. As mentioned earlier, OWL stands for Web Ontology Language, which is W3C-approved semantic knowledge representation language for developing ontology-based systems. A statement in OWL consists of multiple elements that form a *triple*. A triple consists of a subject, predicate, and object. The predicate indicates the relationship between the subject and object. For example, in *Laboratory_test isType Intervention*, the subject is *Laboratory_test*, the object is *Intervention*, and the predicate (relationship) is *isType*, see Figure 4.5. An ontology consists of a collection of these statements that define concepts, relationships, and constraints. Protégé also provides various supporting tools that we used in our prototype system development, such as SWRL rule editor for if-then like semantic rules. Java and Java OWL API were also used in Eclipse Java development environment, to connect the system with the ontology.

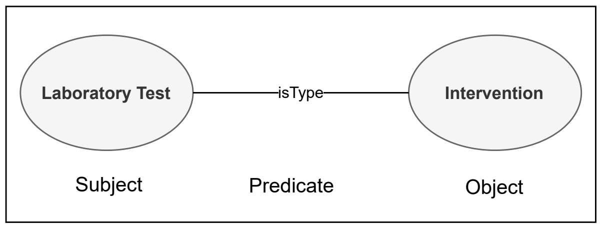

Figure 4.5: An example of an ontology statement (triple-).

4.1.1.4 Ontology Validation

The meta-CP ontology was validated through consultation with domain experts. Figure 4.6 shows the major ontology classes, as well as their child-classes and the relationships between them. For example, the relationship between "Care Plan" and "Clinical Document" is "contains document". The major classes are described below.

- Care Plan: A generic representation of the CP being considered for treatment.
- Trace: All clinical interventions of clinical pathway execution on a specific patient.
- Event: Instructions on CPs to perform an intervention are modeled as medical events. For example, if the same intervention is to be repeated on different days (e.g., perform CBC blood test on day 1 and day 3), there will be a different event of the same intervention for both days.

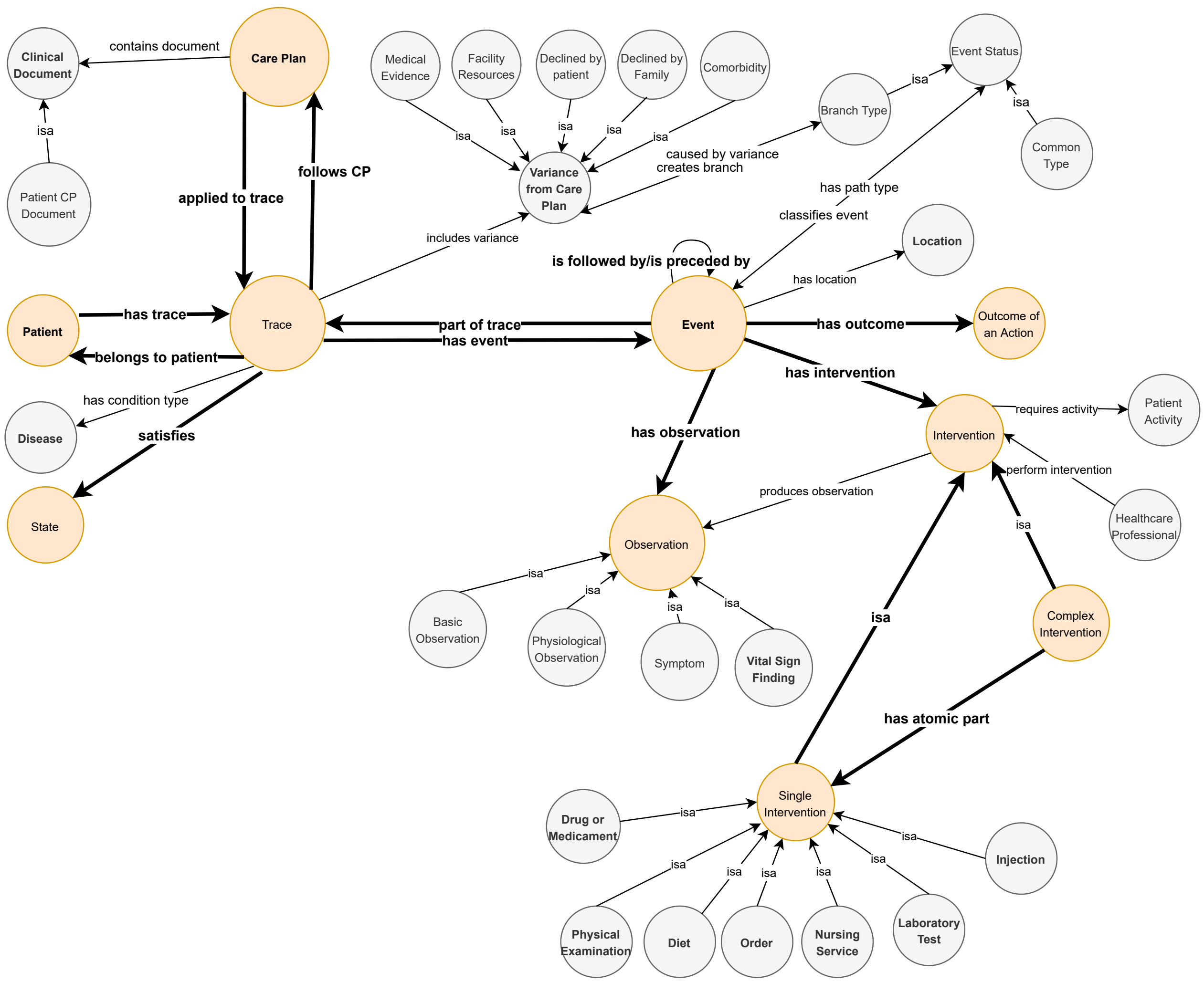

Figure 4.6: Major classes of the meta-CP ontology.

- Intervention: A medical intervention performed during CP execution (e.g. blood glucose level intervention). Several different interventions can be performed. These include single interventions or complex interventions. A single intervention is typically a single clinical activity, whereas a complex intervention is composed of single interventions.

- Observation: An observation made during a CP event.

- Outcome: The result of a certain CP event.

- Patient Disease: Disease for which the patient is admitted and for which the CP is administrated.

- Variance: Deviation from the common CP due to several reasons, such as medical evidence, facility resources, multiple comorbidity, as well as patient or family preferences.

- State: Current state of a trace, which may change based on the patient's progress through the CP.

- Healthcare Professional: Staff members who perform activities related to healthcare, as specified on the CP.

4.1.2 SNOMED CT Ontology

The standard SNOMED CT ontology was included in the system to ensure two main functions. Firstly, to link between standardized CP terms and their equivalent terms in SNOMED CT ontology. This helps in importing SCTID codes to the CP as shown in Figure 4.7. Secondly, to assist members of the medical staff and administrators in finding the correct SNOMED CT terms for non-standardized CP terms. This is essential for adopting international standards in CP documentation. It also allows for clear and accurate documentation of patient assessments, care, and outcomes, thus facilitating communication among caregivers and other healthcare workers.

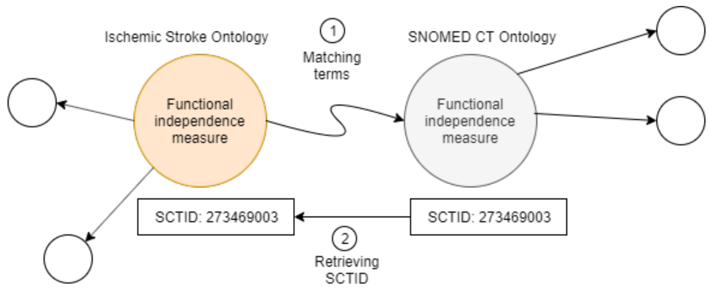

Figure 4.7: Retrieving SCTID from SNOMED CT Ontology.

In the developed prototype system, we used a light-weight SNOMED CT ontology that included only the terms required for stroke CPs. A larger commercial version of the system would have included the entire SNOMED CT ontology, thus allowing mapping to any disease specific CP.

The root class of the SNOMED CT ontology that is modeled in the system is "SNOMED CT Concept". All other details of the SNOMED CT hierarchy are instantiated from the root class. Figure 4.8 shows "Assisting with toileting", which is a descendant of the class "Assisting with activity of daily living". Arranging medical terms using this structure in the system provides additional context to human users to understand the meaning of terms. It can also allow easier inference by machines. For standardization, interventions of the CP ontology reference the corresponding class or individual of the SNOMED CT ontology. This connection is created by using the object property "references SNOMED CT Concept" whose domain is an intervention and range is SNOMED CT Concept. Figure 4.9 shows this object property, as well as other relationships used in the stroke ontology.

4.1.3 Time Ontology

Since CPs contain interventions carried over time, the timing of events of a CP is required for its successful execution. In this research, we adopted the W3C-recommended OWL time ontology to model temporal knowledge [132]. The basic structure of the time ontology is based on an algebra of binary time relations developed by Allen [133]. Allen's interval algebra essentially serves to represent qualitative temporal information and to facilitate reasoning about such knowledge. Figures 4.10 shows the thirteen elementary possible

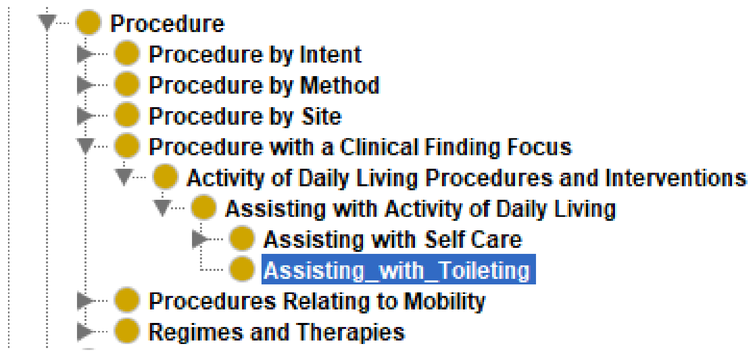

Figure 4.8: A subsection of the SNOMED CT ontology in the prototype.

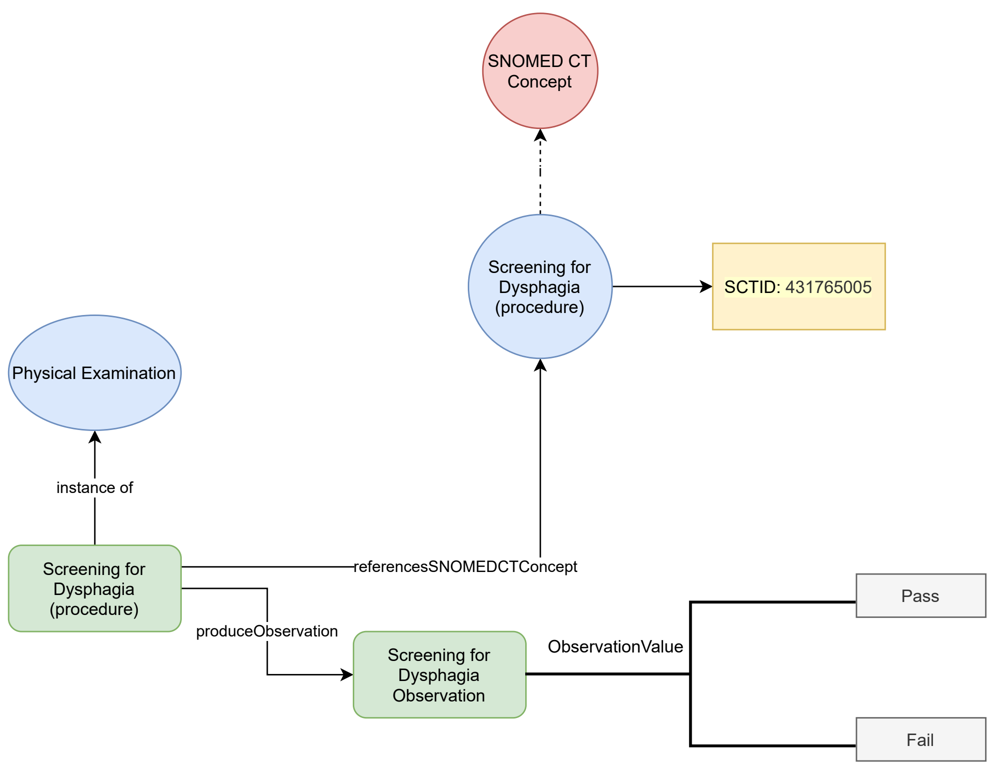

Figure 4.9: Segment of the relationships of screening for dysphagia procedure in the stroke CP.

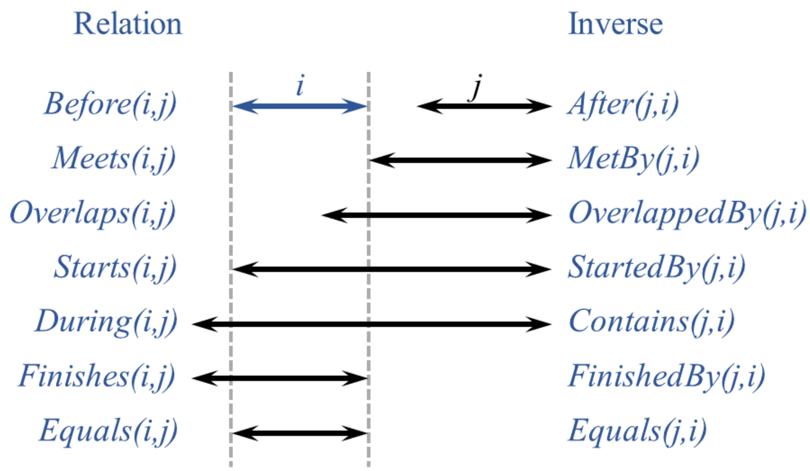

Figure 4.10: Thirteen elementary possible relations between time periods [132].

Interval relation	Equivalent relations on endpoints
t before s (t < s)	t+ < s-
t equals s (t = s)	(t- = s-) & (t+ = s+)
t overlaps s	(t- < s-) & (t+ > s-) & (t+ < s+)
t meets s	t+ = s-
t during s	((t- > s-) & (t+ =< s+)) or ((t- >= s-) & (t+ < s+))

Figure 4.11: Interval relations defined by endpoints [135].

relations between time periods in Allen's algebra. Figure 4.11 depicts the interval relations defined by interval points, where "-" denotes the interval's start time and "+" denotes the interval's end time. The time ontology is an upper level ontology that can be extended and used across several domains. Using Protégé OWL editor [134], the time ontology was integrated with the meta-CP ontology to represent time related data. For example, the CP admission time and discharge time were extended from the "Instant" class of the time ontology so that they represent the patient's hospital admission and discharge times (i.e., year, month, day, and time). The "Instant" class is a subclass of the top class TemporalEntity in the OWL time ontology as shown in Figure 4.12.

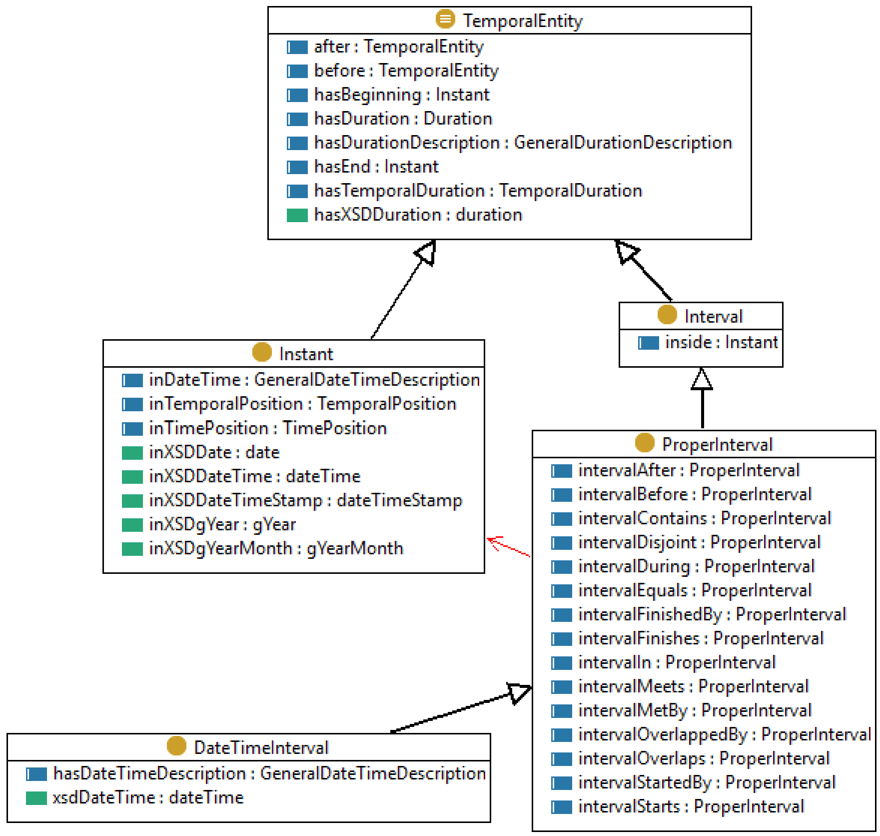

Figure 4.12: Core model of temporal entities in the OWL time ontology [132].

4.1.4 Disease-Specific Ontologies

Disease-specific ontologies inherit the components of the meta-CP ontology and instantiate them with details that are specific to the disease or medical condition under consideration. The details of disease-specific CPs are achieved through either the instantiation of more detailed OWL individuals or by the creation of child classes. In accordance with our framework, every modeled CP is identified by its own CPID. In addition, CP ontology elements are standardized with their SNOMED CT terms. This improves the machine readability of CP contents and facilitates the integration of CPMS with other HISs. Figure 4.13 presents a simplified part of the ischemic stroke ontology. It shows how stroke-specific elements are extended from the meta CP ontology. For example, "Assisting with Toileting", which is a task required for stroke patients, is extended from "Nursing Service". As shown in Figure 4.13, stroke-specific terms are SNOMED CT-compliant. Figure 4.14 shows an OWL code fragment describing a subclass of "Single Intervention" class, denoted by "Nursing Service", and an individual of "Physical Examination", denoted as "Screening for Dysphagia".

4.1.5 SWRL Rules

To describe rules in CP execution, we employed SWRL, which uses the rule syntax "Antecedent $\rightarrow$ Consequent", rendering it more appropriate to model if-then like domain knowledge than using OWL alone. Both antecedent (if-part) and consequent (then-part) are conjunctions of atoms. A variable in SWRL is indicated by a question mark (e.g., "?z"). Moreover, SWRL provides many useful built-in predicates for comparisons (e.g., swrlb:equal, swrlb:greaterThan). SWRL rules are applied on disease-specific CPs using both CP knowledge and patient data. The inference engine used in this work is *Drools* [136], which is a Protégé-embedded rule engine. An example SWRL scenario from the original stroke CP is when a patient fails the intervention Screening for Dysphagia on admission, the Neurology Nurse must then consult with the Speech/Language Therapist if available. This can be represented in SWRL syntax as follows:
Patient(?a) ∧ Screening_for_Dysphagia(?b) ∧ performed_on_patient (?b, ?a) ∧ failed_test(?b,1) ∧ Neurology_Nurse(?d) ∧ Speech_Language_Therapist(?c) ∧ availability (?c, 1) → consult_with (?d, ?c)

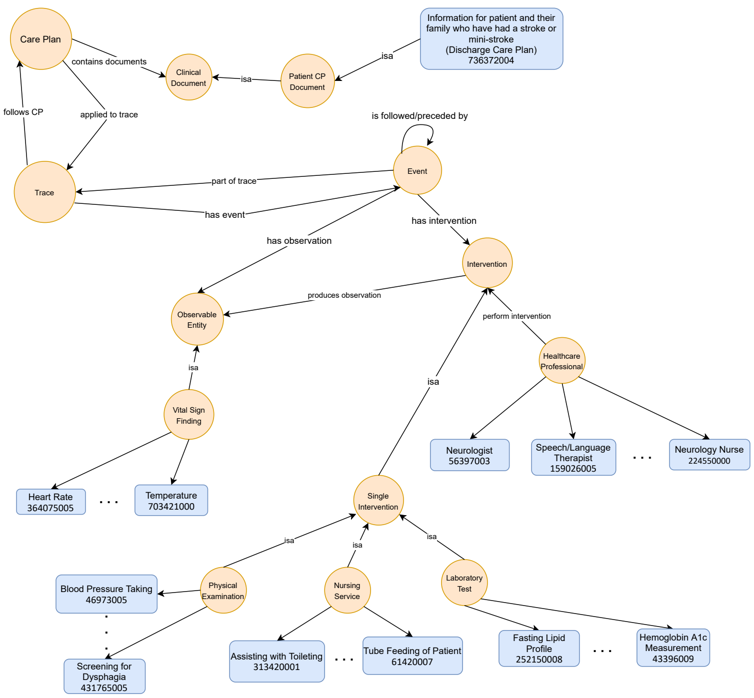

Figure 4.13: An example of disease-specific ontology for stroke.

```
    <owl:Class rdf:about="http://www.semanticweb.org/...#Single_Intervention"/>

      <owl:Class rdf:about="http://www.semanticweb.org/...#Nursing_Service"/>
<rdf:subClassOf rdf:resource="http://www.semanticweb.org/...#Single_Intervention"/>

<owl:NamedIndividual rdf about="http://www.semanticweb.org/...#Screening_for_Dysphagia"/>
    <rdf:type rdf:resource="http://www.semanticweb.org/...#Physical_Examination"/>
```

Figure 4.14: Example OWL code fragments.

A stroke patient who fails the dysphagia test has to repeat the test within 24 hours. The following is the corresponding SWRL rule example:

Trace(?T) ∧ State(Repeat_Screening_for_Dysphagia) ∧ CP_Event(?e) ∧ time:inXSDDateTimeStamp(?e, ?time) ∧ temporal:durationEqualTo(24, ?time, "now", "Hours") → hasState(?T, Repeat_Screening_for_Dysphagia)

4.2 Inference and Data Analytics

The inference and data analytics layer is where system processing takes place. The main functions and the linking between various layers of the system are realized in this layer. For example, this layer performs the task of generating the patients' CP traces and storing them in a SNOMED CT compliant format that can be used for various data analytics and decision support functions. This layer ensures future interoperability between the proposed CPMS and other HISs through SNOMED CT and HL7 messages. Furthermore, the layer includes several algorithms, such as CP cost analysis and CP trace analytics. More details on this layer and the decision support algorithms are addressed in the next chapter.

4.3 Clinical Pathways Management Tools

CP management tools form the user interface of the proposed clinical pathways management system. The Model-View-Controller (MVC) software design pattern was adopted in developing the interface and dividing the program logic into three interconnected elements (see Figure 4.15). Dividing the program logic into elements makes it easier to test and update the code since each modification is organized under its own element. For example, the model contains the stroke ontology which contains the data related to CP execution. The controller calls the Java OWL API functions to update the model or infer new knowledge. The graphical user interface is contained within the view, where, for example, the

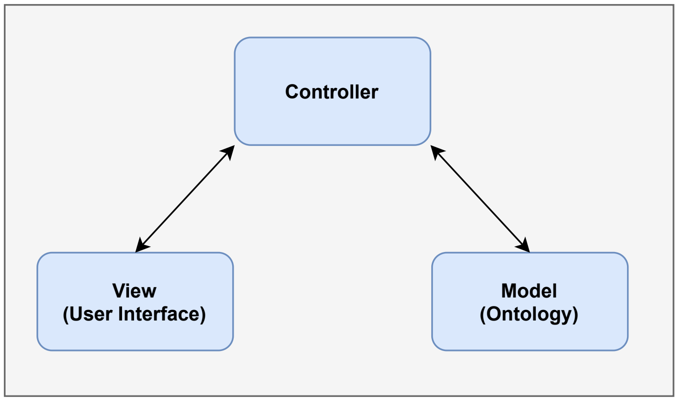

Figure 4.15: Model-View-Controller design pattern of the prototype.

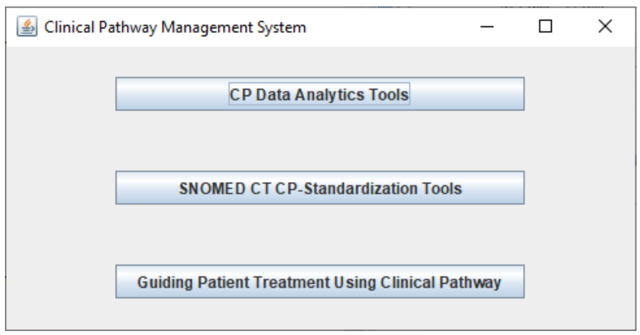

Figure 4.16: Main Screen of the prototype CPMS.

data from the ontology is presented. This allows the user to navigate through the stroke ontology events (through the user interface screens) without the need to deeply understand the semantic structure of the ontology. Figure 4.16 shows the main screen of the system.

The "CP Data Analytics Tools" option provides the user with various CP analytics and management functions. For instance, the user can search, view and compare CPs. Comparing CPs enables healthcare providers to compare standardized CP elements of disease-specific CPs. CP comparison, which is currently a tedious manual task that relies mainly on unstructured text comparison, is made simple by applying the proposed framework. The CP data analytics algorithms can be used from this component.

The "SNOMED CT CP-Standardization Tools" option enables medical staff members

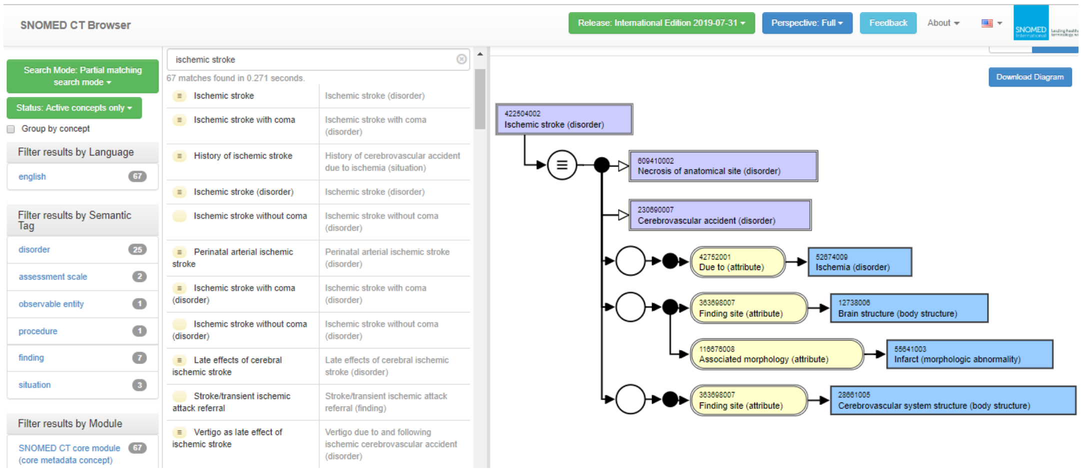

Figure 4.17: Standardizing CPs through SNOMED CT Browser.

to use tools that help them find proper SNOMED CT terms for non-standardized terms appearing in CPs. This option also enables physicians and system administrators to update existing terms that were modified by SNOMED CT. One of these tools enables a live connection with SNOMED CT ontology and invokes the SNOMED CT search engine browser, where up-to-date terminology concepts and codes can be searched by categories (e.g., disorder, assessment scale, observable entity, procedure, etc.). These CP concepts can be displayed with diagrams that show relationships between clinical terms as shown in Fig. 4.17.

Another useful function on the SNOMED CT browser is the check-digit calculator that helps healthcare providers find the correct check-digit to generate a disease-specific CPID. All these tools facilitate standardizing and adding new CPs to the system.

The option "Guiding Patient Treatment Using Clinical Pathway" is where CP execution on patients is realized. Here, the CP-based treatment is recorded through a series of screens that present the SNOMED CT standardized clinical interventions of the CP and allows medical staff to save the performed interventions and their outcomes. It should be noted that CP standardization removes any ambiguity regarding the meaning of the performed interventions.

Figure 4.18 shows a portion of a CP execution screen where clinical activities are

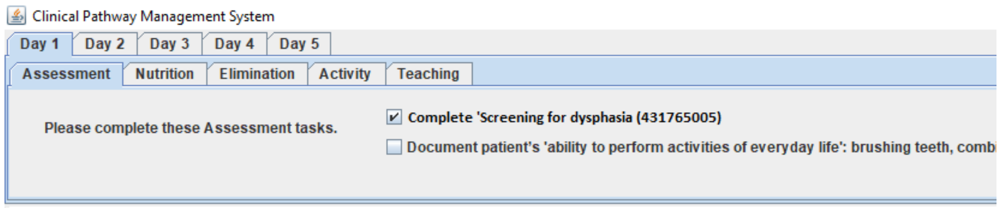

Figure 4.18: A screenshot showing CP guided treatment.

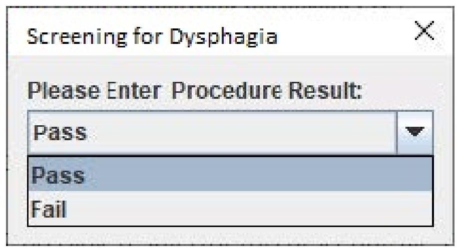

Figure 4.19: A sample CP-progress screenshot from the prototype system.

arranged by category of interventions (e.g., assessment, nutrition, elimination, etc.) for each day in the sample CP. Figure 4.19 illustrates another screen from the system where the result of the screening for dysphagia for a stroke patient is entered by the caregiver as either pass or fail.

Since CP terms are standardized, all medical staff members can realize them, thus facilitating communication and collaboration among the team. As patients progress through the CP treatment, CP interventions are timestamped and recorded in sequence, allowing the system to populate the patients' CP trace repository, which is an output file in our prototype. The output file could then be pre-processed so that it can be used by CP decision support algorithms.

An example fragment of CP data in the output file for ischemic stroke CP patients is shown in Table 4.1. The content of the file is edited and summarized for space limitations. The actual content is much bigger since for each intervention, the system stores all relevant data including its SNOMED CT term, SCTID, outcome, start time, end time, etc.

As mentioned, the output file can be pre-processed and then used in different ways to support decisions and extract useful information. For example, the trace data can be

Table 4.1: Ischemic stroke CP Patients' data (CPID 422504039)

	Real patients' data		Guided Simulation Data	
Patient ID	Speech and Language disorder 422504039-231543005	Smoking 422504039-365981007	Screening for Dysphagia 422504039-431765005	Computed tomography of chest 169069000
378	No	No	Successful	
491	Yes	Yes	Failure	Nodule of lung (variance reason: Comorbidity)
502	No	No	Successful	

extracted by considering only the ordered sequence of medical interventions without their times. This can be useful when comparing categories of patients based on the sequence of medical intervention that they have undergone.

CP variance (if any) is also reported. Table 4.1 shows a case of variance where computed tomography of chest (SCTID 169069000) is ordered for a patient due to comorbidity. Since this procedure is not part of the CP, only the SCTID will be displayed without the CPID hyphenated to it.

4.4 Prototype Validation

The prototype system and its CP algorithms were implemented using the Java programming language in the Eclipse Java development environment, following the prototyping software development methodology. Ontologies were developed in Protégé and integrated with the system using Java OWL APIs. Protégé's Drools inference engine was used for reasoning. The developed system communicates with an EMR that is used for research purposes at the Regional Stroke Unit at TBRHSC.

The Stroke Unit EMR (stroke patients file) contains the data of over 500 stroke patients that were treated and hospitalized at the hospital. Using a medically 'guided' simulation for these patients the output file was generated (see Table 4.1). The left part of the table labeled 'real patients data' represents real data collected inside the hospital, and the right part labeled 'guided simulation data' represent the data resulting from a guided simulation. By medically guided simulation we mean that domain experts have helped us generate simulated CP execution results for some data that were not available for us. The simulation was based on medical insights and knowledge about patient cases (e.g., stroke patients with speech problems fail in the screening for the dysphagia test, a certain patient had a previous stroke incidence, etc.).

Since medical staff members are the potential end users of the system, the prototype was validated with the help of our domain experts. This was performed through several interviews with the domain experts. Some interviews were closed interviews (i.e., there was a pre-defined set of questions) whereas some were open interviews (i.e., open discussion with the domain experts). The questions asked in the interviews were related to various aspects of the system (e.g., CP ontology design, user interface design, standardized terms, clinical pathways, output file, etc.). Appendix A presents sample interview questions.

Domain experts also assisted in evaluating some scenarios using the output file. In the early developmental stage, we experienced issues with the prototype, which our domain experts criticized and disagreed with. For example, the CP traces were lacking the admission and discharge days. Our experts recommended that we include the days in the output file. In addition, finding the proper standardized SNOMED CT terms was challenging in some cases and domain experts helped in the standardization process. We also experienced programming problems related to compatibility issues between Java and Protégé due to the fact that Java updates occur more frequently than Protégé and OWL API updates.

Furthermore, the initial user interface of the prototype was "crowded with buttons" (as commented by the domain experts), and based on their feedback, the user interface was improved throughout the development process by suggesting a less crowded user interface driven by CP daily activities and independent successive small screens for CP interventions and messages, as shown in Figure 4.16 and Figure 4.18.

Our domain experts were helpful throughout the development process by providing advice about various aspects such as the ontology design, SNOMED CT standardization, and retrieving certain required data from the hospital EMR (in cases where the required data were not available in the research-based EMR of the stroke unit).

Due to the privacy of patients and their data, besides real patient data shared with us, we had to perform simulations for data that were not available. Despite this limitation, the simulations were medically guided by our domain experts, as mentioned above.

All in all, although it was a limited proof-of-concept system, the prototype was successful in performing its various intended functions such as simulating the progress of patients through CP sample interventions, timestamping events, storing interventions in the CPMS database (i.e., the output file), running decision-support algorithms on the processed output file, etc. The output file was the main patient data file (or dataset) in this research. The output file contained data for the stroke patients who had sustained a stroke due to Carotid Artery Disease (and other causes). Carotid Artery Disease (CAD) is a chronic vascular disease caused by the formation of plaque in the wall of the carotid artery, causing stenosis and impairing the flow of blood to the brain. In the case of plaque rupture, a blood clot may form and detach, then move with the blood to smaller brain vessels, potentially leading to an ischemic stroke [137, 130].

In the final patients' output file, each patient record contained several characteristics such as demographic data (age, gender, and ethnicity with three categories: Caucasian, Indigenous and Other), disease history (sleep apnea, atrial fibrillation, diabetes, and hypertension), medical history and habits (e.g., previous carotid artery intervention, alcohol consumption, smoking, type of stenosis, admission date, length of stay, speech and language disorder, screening for dysphagia, nicotine withdrawal, assessment of tobacco use, symptoms, doppler results), required nursing services, and stroke cause classification. Stroke cause classification in the dataset was based on the ASCOD phenotyping, whereby A stands for atherosclerosis, S stands for small-vessel disease, C stands for cardiac pathology, O stands for other causes, and D stands for dissection. As mentioned above, not all of the above data were real data since some needed data were simulated to be able to run simulations using the prototype CPMS.

4.5 Conclusion

In this chapter, the conceptual design and architecture of a CP management system were addressed. The proposed system components are structured in such a way that they collectively cooperate to operationalize the proposed framework and ensure the independence of CPMSs. The main components of the prototype system include: The knowledge base, the inference and data analytics, and the CP management tools components. The next chapter presents various CP-based data analytics and decision support scenarios.

Chapter 5

Data Analytics and Decision Support Scenarios

The proposed CP standardization and digitization framework is extremely useful in allowing the recording of observations, laboratory tests, procedures, medication and other CP-related data, as well as linking them to their corresponding CPs. This framework not only enables semantic interoperability among healthcare data, but also provides a rich data source for data analytics and decision support. In this chapter, we address example scenarios that highlight the capabilities of the framework from the data analytics and decision support perspectives. The scenarios cover various health decision support areas (e.g., variance analysis, hospital resource management, etc.); therefore, they are preceded by a background related to the healthcare area under consideration.

5.1 CP Variance Analysis and Action Plan

One of the reasons behind introducing CPs in healthcare is to reduce the variance in medical practice. CPs were successful in this regard because their standardization of patients' treatment (i.e., reduction of variance) has resulted in homogeneous healthcare practices. Nevertheless, variance is inevitable in some patients' cases based on the decision of healthcare specialists. CP variance analysis identifies deviations from the clinical pathway and can be used for clinical auditing and quality improvement. Handling CP variance is an important function of CP management systems. The proposed framework enables healthcare providers to record the variance based on its major reasons. The major reasons are modeled in the meta-ontology and the variance is reported in the output with the intervention's SNOMED CT ID only (i.e., without using the hyphenated coding) to indicate that the intervention is a variance, and not part of the CP.

Recording the variance facilitates the reporting of the statistical analysis required for supporting decisions related to CP "Action Plans" (e.g., an action plan to modify the CP by adding/removing an intervention). Since the action plan related to CP variance affects the safety of patients, action plans are proposed and performed by expert physicians rather than by machines. Action plans may include analyzing the percentage of variance and modifying the CP in cases where the variance reaches an agreed-upon threshold (e.g., 80% of patients deviated from the CP regarding an intervention or procedure). Even in such cases, action plans typically include careful and extensive literature review for medical evidence because CPs are patient treatment plans based on proven medical evidence.

A hypothetical scenario related to this area is for a hospital that uses the proposed framework in their CPMS. Hospital staff can process the output to generate CP variance analysis reports, similar to the one shown in Table 5.1, which demonstrates the variance analysis through intervention application rates for a CP that was not recently updated. Based on analyzing the results and related medical evidence, it was decided to update the CP by adding intervention 699270006 (cerebrovascular accident annual review) as an intervention in the CP under consideration (an intervention that was recommended in 70% of patients).

5.2 Cost Management and Control

The proposed framework allows for improved cost management and control in hospitals. We present a scenario related to managing costs per patient. Healthcare administrators are often interested in comparing the average cost per patient for a particular CP in a single hospital or across several hospitals. This can be determined by tracing all CP-related interventions for all patients enrolled in that CP. A hypothetical scenario is shown

Table 5.1: Variance analysis for population of patients (e.g., for 4000 patients).

Interventions	**Application Rate**
230690030-405035003	100%
230690030-432103005	100%
699270006	70% (Not in CP, but recommended for 70% of patients)
230690030-417986000	100%

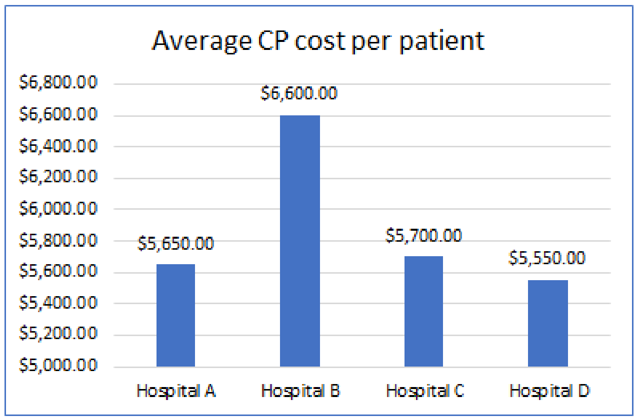

Figure 5.1: An example chart for the average CP cost per patient.

in Figure 5.1, which outlines the average cost per patient for Total Thyroidectomy CP in four different hospitals for a specific period of time, where the average cost per patient in Hospital B was noted to be higher than that of the other hospitals. To facilitate CP cost analytics, we developed a CP cost analytics algorithm based on the new proposed hyphenated coding system. To better illustrate the algorithm, we introduce formal definitions that demonstrate how the framework facilitates the mathematical representation of CPs. This is because the SNOMED CT-based CP identification code (CPID) differentiates between CPs without ambiguity, and it can therefore be used in the mathematical notation of clinical pathways.

Assume that a hospital database has an inventory of n distinct medical interventions and m adopted CPs. Then, let $\mathcal{I}_{Hospital}$ be the set of the medical interventions which can be defined as:

$$\mathcal{I}_{Hospital} = \{\mathcal{I}^1, \mathcal{I}^2, \mathcal{I}^3, ..., \mathcal{I}^n\}, \tag{5.1}$$

where $\mathcal{I}^i$ is the SNOMED CT code (SCTID) of a medical intervention.

Let $\mathcal{I}_{CPID}$ be the set of all medical interventions of a particular CP identified by CPID. This implies that:

$$\mathcal{I}_{CPID} \subset \mathcal{I}_{Hospital}, \tag{5.2}$$

To give some hypothetical examples:

$\mathcal{I}_{422504039} = \{\mathcal{I}^2, \mathcal{I}^7, \mathcal{I}^{54}, ...\}$ denotes the set of medical interventions given in ischemic stroke CP, and

$\mathcal{I}_{73211032} = \{\mathcal{I}^6, \mathcal{I}^{10}, \mathcal{I}^{81}, ...\}$ denotes the set of medical interventions given in diabetes mellitus CP.

Let CP trace be the ordered sequence of medical interventions prescribed for a particular patient enrolled in a CP. Consequently, a CP trace for patient i enrolled in a CP identified by *CPID* can be defined by $T_{CPID,i}$ as:

$$T_{CPID,i} =< \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_{3,}..., \mathcal{I}_k >, \tag{5.3}$$

where $\mathcal{I}_{j=1..k}$ represents the ordered sequence of CP interventions from the set $\mathcal{I}_{CPID}$ taken by patient i. Note that $T_{CPID,i}$ is defined mathematically as a sequence, not a set, since CPs can have the same intervention(s) repeated in a trace at different times.

The cost of trace $T_{CPID,i}$ can be represented as:

$$Cost_{T_{CPID,i}} = \sum_{j=1}^{k} Cost_{I_j}, \tag{5.4}$$

where $Cost_{I_j}$ is the cost of the medical intervention $\mathcal{I}_j$.

Suppose that there are m patients treated through the CP identified by CPID. Since a CP trace is generated for each patient, we can therefore represent the total cost of patients' CP treatments associated to CPID by adding the overall costs of all traces. This can be expressed as:

$$Cost_{CPID} = \sum_{i=1}^{m} Cost_{T_{CPID,i}}. \tag{5.5}$$

The CP cost analytics algorithm (see Algorithm 5.1) considers all traces within a specific time period (e.g., the fiscal year) and applies the equations defined above. Furthermore, the algorithm is generic in the sense that it automatically applies the equations on all patients for each CP, and outputs the total cost per CP, as well as the average CP cost per patient. The inputs to the cost algorithm consist of the file that contains the costs for all interventions, and the file of interventions of patient traces. Patients who have traces in a specified time period, can be determined, for example, by using SPARQL queries.

Figure 5.2 shows an example SPARQL query to extract patients with traces that took place within the fiscal year 2018. The costing algorithm differentiates between different CPs based on their CPID by considering the left part of the hyphenated code; it also differentiates between interventions by considering the right part of the hyphenated code. This allows the algorithm to easily classify and then aggregate CPs, as well as all interventions in every CP (refer to steps 7 to 18 in the algorithm).

5.3 Managing Patient CP Traces

The application of CPs in hospitals results in the generation of CP treatment paths that were followed by patients. In the CP domain, treatment paths are often referred to as "CP traces". Knowledge about CP traces, as well as the ability to perform data analytics on them, provide great support for healthcare decision-makers. In this section, we show how the framework facilitates the development of decision support algorithms related to CP traces for better CP management.

Algorithm 5.1: CP cost analytics

```
input  : Patients' CP traces file (interventions)
input  : Interventions_cost file
output: Cost[CPID]: hash-table with CP cost for each CPID
output: Avg_Cost[CPID]: Hash-table with average cost per patient for each CPID
Data Structures:
HC: A variable representing CPID-SCTID hyphenated code for a CP intervention
Traces: Hash-table with number of patient traces for each CPID, initialized to zero
CP_List: List of all CPIDs in CP traces file, initially empty
Interventions: Hash-table with list of interventions for each CPID, initially empty
begin
    foreach record ∈ CP traces file do
        HC ← read HC of first intervention
        CPID ← extract left part of HC
        if (CPID ∉ CP_List) then
            Add CPID to CP_List
            Traces[CPID] ← Traces[CPID] + 1
        end
        foreach intervention ∈ current record do
            HC ← read HC of current intervention
            SCTID ← extract right part of HC
            Add SCTID to Interventions[CPID]
        end
    end
    foreach CPID ∈ CP_List do
        Cost[CPID] = 0
        foreach SCTID ∈ Interventions[CPID] do
            Cost_SCTID ← Cost of SCTID from interventions_cost file
            Cost[CPID] ← Cost[CPID]+ Cost_SCTID
        end
        Avg_Cost[CPID] ← Cost[CPID] / Traces[CPID]
    end
end
```

```
PREFIX cpms: <http://www.semanticweb.org/...>
SELECT ?patientTrace
WHERE {
{?patientTrace cpms:Admission ?admissionTime}
{?patientTrace cpms:Discharge ?dischargeTime }
FILTER (
?admissionTime >= "2018-01-01T00:00:00"^^xsd:dateTime
&& ?dischargeTime <= "2018-12-31T23:59:59"^^xsd:dateTime)
}
```

Figure 5.2: An example SPARQL query.

A useful algorithm in this context is an algorithm for determining the longest common CP trace shared by all patients from their hospital admissions. A common CP trace from hospital admission starts from hospital admission and extends to the point when the patients start to deviate from each other. The knowledge of the common CP trace from admission enables physicians to optimize CP development by studying the causes of deviations (see Algorithm 5.2). Since all interventions are SNOMED CT standardized, the algorithm loops over the sequences of interventions (starting from the first patient) and keeps monitoring the length of the common sequence among all traces (see steps 6-26 in the algorithm). After determining the length of the common trace from admission, the algorithm retrieves the interventions of the common trace from the trace of the first patient (see steps 27-31 in the algorithm). It is apparent that CP standardization plays a large role in facilitating trace-related algorithms.

Another interesting feature of having standardized and digitized CP traces is the ability to determine the Longest Common Subsequences (LCS) for all patients undergoing a particular CP trace compared to an agreed-upon ideal CP trace, which is a trace of a typical patient that goes smoothly through that part of the CP (e.g., no disease complications). This is performed by Algorithm 5.3, which uses an Apache LCS function (*LongestCommonSubsequence*) in [138] to output a file that contains the LCS for all patients in the CP traces file. Patients' traces and LCS can help classify patients into various categories. Here, we demonstrate patients' CP classification based on their LCS by testing 503 stroke patients' traces and comparing them to an ideal trace from the stroke CP. The sequence

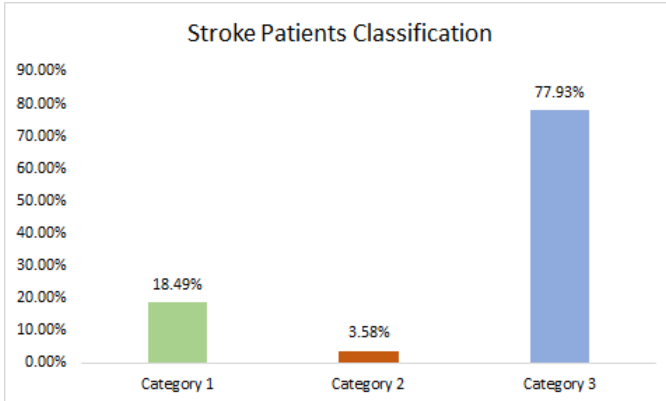

Figure 5.3: Stroke patients categories.

of an ideal trace would be:

screening for dysphagia (pass)→smoking (no)→patient and family education→...,
whereas a sequence of a non-ideal trace would be:
screening for dysphagia (fail)→speech/language therapist→nasoenteric tube→
smoking (yes)→nicotine replacement therapy→ patient and family education→...

Analysis of the results shows three categories of patterns, as can be seen in Figure 5.3. Discussions with domain experts revealed that these categories can be explained by CP artifacts linked to patients' conditions, mainly those related to Speech and Language Disorder (SCTID 231543005), Smoking (SCTID 365981007), and Problems with Balance (SCTID 387603000). As shown in Table 5.2, Category 1 consists of 93 patients who did not have speech and language disorder; did not smoke; and did not have balance problems. Such patients, for example, passed screening for dysphagia and did not need smoking related interventions. Category 2 consists of 18 patients who had all the complications listed above, as shown in Table 5.2, and required additional CP interventions. For instance, they failed screening for dysphagia, and therefore required consultation with a speech language therapist; they were also smokers, which implies that they suffered from nicotine withdrawal. Finally, Category 3, which is the most dominant one with 392 patients, includes patients who experienced some (but not all) complications, as shown in Table 5.2.

Table 5.2: Categories of stroke patients.

Category	Speech and Language Disorder	Smoking	Problem with Balance
Category 1	No	No	No
Category 2	Yes	Yes	Yes
Category 3	Yes/No	Yes/No	Yes/No

Algorithm 5.2: LCS of all patient traces

```
input  : Patients' CP traces file
input  : Ideal_Trace: List of ordered interventions of ideal CP trace
output: LCS: Hash table with longest common subsequence for each patient,
        initially empty
Data Structures:
HC: A variable representing CPID-SCTID hyphenated code for a CP intervention
PatientID: A variable to store ID of the patient under consideration
begin
    foreach record ∈ CP traces file do
        PatientID ← Patient ID from current record
        foreach intervention ∈ record do
            HC← read HC of current intervention
            SCTID ← extract right part of HC
            Add SCTID to Trace[PatientID]
        end
    end
    foreach trace_record ∈ Trace do
        // Calling LongestCommonSubsequence function
        LCS[PatientID] ← longestCommonSubsequence (Ideal_Trace, trace_record)
    end
end
```

Algorithm 5.3: Longest Common Trace (LCT) from hospital admission

```
input  : Patients' CP traces file
output: LCT from hospital admission
Data Structures:
First_Patient_Trace: An array to store the interventions of 1st trace in patient
  trace file, initially empty
Patient_Trace: An array to store current trace
Length: A variable used to store the number of interventions in LCT
begin
    foreach trace_array ∈ CP traces file do
        if (First_Patient_Trace is empty) then
            for i← 0 to length of trace_array do
                First_Patient_Trace[i]←trace_array[i]
            end
            Length ← length of First_Patient_Trace
        else
            for i← 0 to length of trace_array do
                Patient_Trace[i]←trace_array[i]
            end
            for j ← 0 to length of First_Patient_Trace do
                for k ← 0 to length of Patient_Trace do
                    if (j = k AND First_Patient_Trace[j] ≠ Patient_Trace[k]) then
                        if (Length ¿ j) then
                            Length ← j
                        end
                    end
                end
            end
        end
    end
    for x ← 0 to Length do
        LCT [x] ← First_Patient_Trace[x]
    end
    return LCT
end
```

5.4 Hospital Resource Management (HRM)

Healthcare is a booming sector of the economy in most countries around the world. Many challenges are associated with the growth of healthcare, including continuously rising costs and increased pressure on hospitals' limited resources [139]. In Canada, for example, according to the Canadian Institute for Health Information (CIHI) [140], total healthcare costs have been continuously increasing over the years. Figure 5.4 shows the healthcare expenditure trends in billion dollars between 2010 and 2019. Healthcare spending as a share of Canada's Gross Domestic Product (GDP) is also trending upward. In 2019, healthcare spending represented approximately 11.6% of Canada's GDP compared to 7% in 1975. Figure 5.5 compares healthcare expenditures of Canadian provincial governments per capita between 1975 and 2019. For example, in Ontario, the per capita expenditure increased from $378 to $4385 (1060% increase), whereas in Alberta, increased from $384 to $5187 (1251% increase).

The situation is similar worldwide. A recent report by the World Health Organization (Global Spending on Health: A World in Transition, 2019), indicates that global healthcare costs continue to rise rapidly. Financial figures revealed that global healthcare spending increased to US$7.8 trillion in 2017, up from US$7.6 trillion in 2016 [141]. This is a dramatic increase of US$200 billion in just one year. In addition, healthcare spending is growing faster than GDP. Between 2000 and 2017, the global health spending in real terms grew by 3.9% a year, while global GDP grew 3.0% [141]. As Figure 5.6 shows, the increase in healthcare spending was even faster in low-income countries, where it rose 7.8% a year between 2000 and 2017 while the economy grew by 6.4%. In middle-income countries, health spending grew more than 6% a year. In high-income countries, the average annual healthcare spending growth was 3.5%, which is nearly twice as fast as the economic growth represented by GDP.

The recent in-depth statistics above show that endeavors to solve healthcare problems on the top "macro" level are not efficient. The proposed framework helps decision-makers deal in "micro" level by considering the fine details of CP interventions and procedures inside hospitals. Without this fine level of analysis for hospital resource management, healthcare would be a soon-to-be-bankrupt sector in many countries around the world.

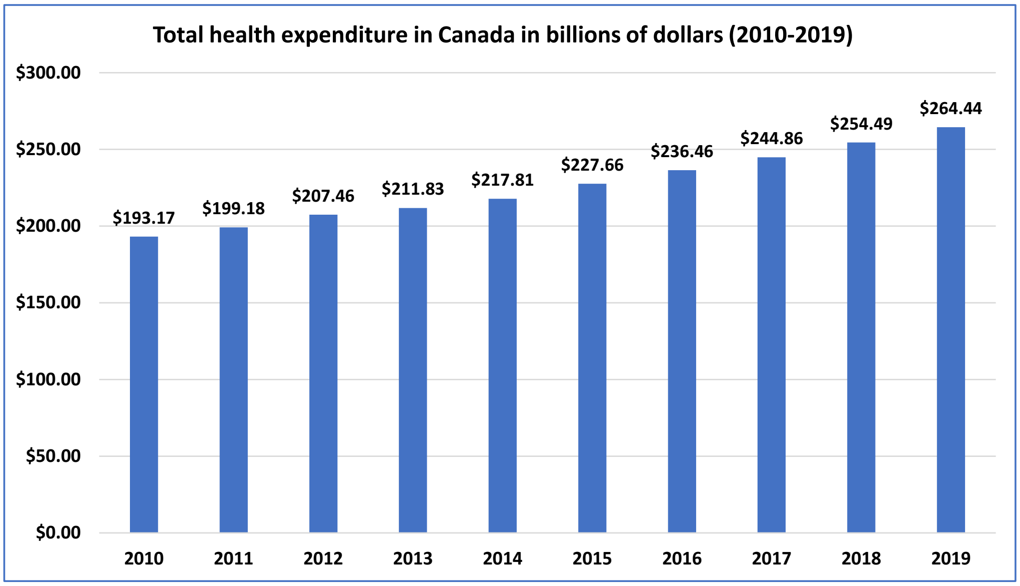

Figure 5.4: Health expenditure in Canada between 2010 and 2019.

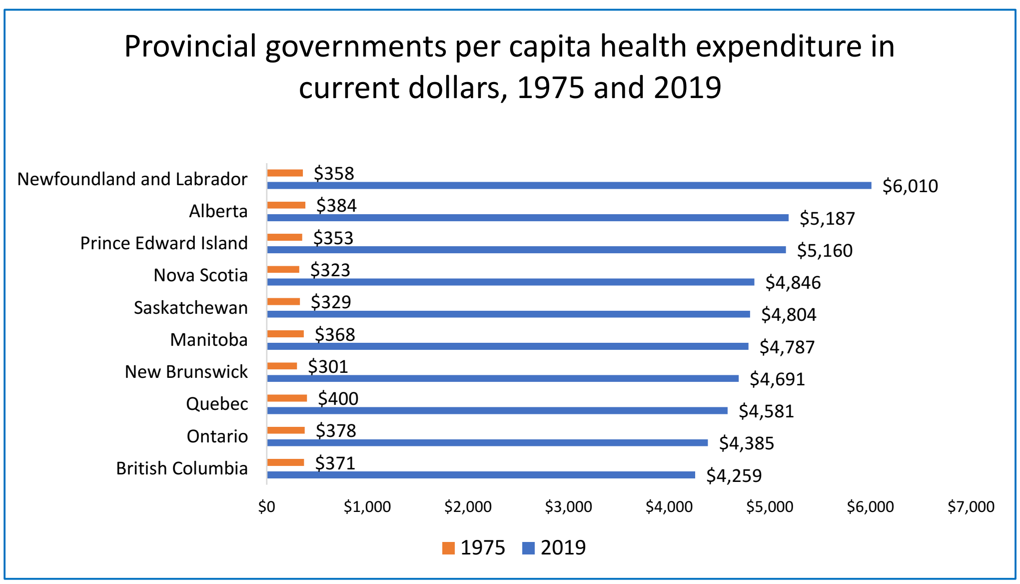

Figure 5.5: Per capita healthcare expenditure increase by Canadian provincial governments.

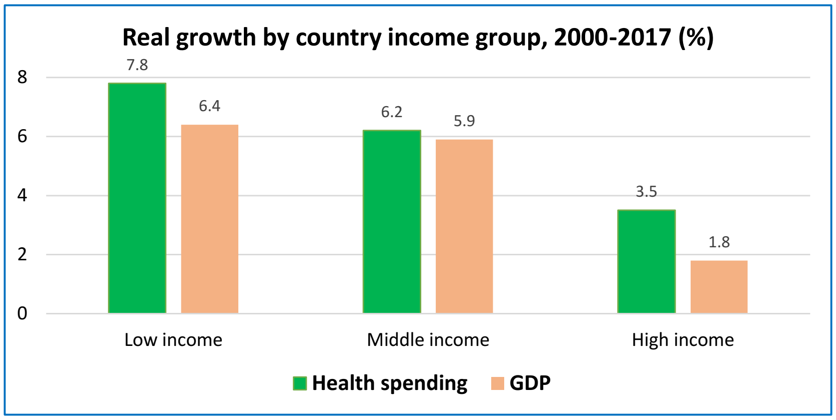

Figure 5.6: International healthcare spending and GDP growth (%) (2000-2017).

Various events and occasions in healthcare uncovered that processes are not optimized, and stressed the importance of hospital resource management. For example, a recent study conducted in Alberta, Canada, revealed that 5% of repeated Complete Blood Count (CBC) tests were repeated inappropriately in Alberta hospitals in 2018 [142]. The same study showed that approximately 36% of repeated electrolyte panel (EP) tests within a 24-hour period were an unnecessary waste of public money [142]. Researchers estimated that the annual cost of unnecessary repeat CBC and EP blood tests was $2.42 million CAD paid by the province of Alberta ($0.52 million CAD in unnecessary CBC tests and $1.90 million CAD in unnecessary EP tests) [142]. The study considered only two types of blood tests. We strongly believe that if all tests were investigated, the study would reveal that a substantial amount of money was wasted in a healthcare system that is already under increasing cost pressure. Such statistics would help convince healthcare practitioners, as well as decision-makers, to adopt our framework of electronic CPs in their hospitals.

Without evidence-based CPs applied in hospitals, the ordering of unnecessary blood tests would be based solely on the judgment of physicians. This is clearly an un-optimized and costly situation. This was the case in the Alberta study mentioned above, where researchers stated that "residents order routine daily CBC and electrolyte panels (EP) more frequently than attending physicians" [142]. Our proposed CP automation framework can contribute positively to improving such situations due to the fact that automated CPs reduce lack of data in healthcare (i.e., data that are not recorded electronically). Without

Table 5.3: Intervention/Cost management through communicating CP data digitally from different hospitals.

Hospital-A	**Hospital-B**	**Hospital-C**	**...**
422504039-405035003	422504039-405035003	422504039-405035003	...
422504039-423103005	422504039-405035003	422504039-405035003	...
422504039-417986000	422504039-417986000	422504039-417986000	...
		422504039-8306009	
422504039-273251005	422504039-273251005	422504039-273251005	...

adequate CP automation, data that are vital for optimizing hospital resources would be lost under piles of paperwork. The following discussion illustrates three hypothetical scenarios in which our proposed framework helps in HRM. The scenarios are presented in a top-down approach, such that the first scenario is related to a country-wide (or province-wide / region-wide) HRM, the second scenario is related to a hospital-wide HRM, and the third scenario addresses an issue at a CP-level HRM.

5.4.1 HRM through communicating best CP practices among hospitals

SNOMED CT-based standardized CPMSs have all their data recorded in an internationally-recognized digital format. This enables the automatic retrieval of CP data from different hospitals from a country or province to perform data analytics on CP data. One important resource management advantage in this context is controlling CP costs and communicating best CP practices. An illustrative scenario would be a government authority (e.g., the Ministry of Health) initiating an HRM study on CPs from different hospitals to have control over treatment costs for a certain medical condition. Table 5.3 shows a sample CP report that can be transmitted and compiled from CP data of different hospitals {Hospital-A, Hospital-B, Hospital-C, ...}. Such a compiled digital report, using the hyphenated coding system, would be greatly beneficial in automatically comparing the contents of specific CPs from different hospitals.

The unified CPID to the left of the hyphen confirms that all interventions originated from CPs related to the same disease or procedure. This compiled report helps in identifying the differences or the sources of the inefficiency or higher costs, such as those generated from unnecessary blood tests or redundant interventions.

For example, Table 5.3 shows an intervention that is administered only in Hospital C. All other hospitals do not apply this intervention. Such unnecessary medical intervention can then be discussed with the relevant hospital administration and eliminated (if not necessary) to improve the efficiency of the CP and achieve cost savings. To roughly estimate possible cost savings, we can refer to the Alberta study and assume that eliminating two unnecessary blood tests would save Alberta Health Services $2.42 million CAD per year. Much greater savings could be realized by initiating a similar CP auditing and quality control process for other diseases.

This scenario illustrates the importance of the proposed framework in facilitating the integration of CPs with strategic healthcare decision-making in the context of managing hospital resources. The objective is to reduce healthcare costs without compromising the quality of service. In addition, this CP comparison capability allows healthcare decision makers to detect inconsistencies, exceptions, and inefficient processes in the CPs. This also helps in detecting health care fraud and abuse (in the case of intentionally adding unnecessary/redundant costly medical interventions).

Another benefit of CP comparison is the desire to standardize and unify patient care among different institutions within a region or across a country [68]. The rationale is that such standardization can lead to (a) transfer of patients between hospitals without disrupting the care process; (b) equity of care across different healthcare providers; (c) understanding of clinical outcomes based on a larger sample of patients who were exposed to similar treatment processes; and (d) uniform implementation of clinical guidelines [68].

This scenario provides a sample HRM advantage of communicating best CP practices between hospitals. The full spectrum of advantages is actually much wider because a well-designed and standardized CPMS allows healthcare decision-makers to detect inconsistencies, poor decisions, and various types of inefficient processes in hospitals' daily operations. Such HRM capability, through comparisons of CPs from different medical institutions, would be a tedious task to realize with today's paper-based CPs or partially standardized electronic CPs.

5.4.2 HRM through blood test/intervention cost and count analytics

The previous scenario was related mainly to HRM across different hospitals within a country. In this scenario, we consider HRM within a single hospital. CPMSs that are modeled according to the proposed framework enable better quality and cost management across all CPs practiced in a single hospital. This facilitates improved cost-savings and budgeting. Since the hyphenated coding system always keeps the link between medical interventions in hospitals (e.g., x-rays, blood tests, etc.) and their originating CPs, the costs incurred across all CPs in a hospital can be traced and analyzed for future budgeting and planning.

As stated by one of our domain experts, many routine blood tests (e.g., CBC) that are recorded in EMRs could not be easily traced to their source medical conditions in some of the hospitals in which they worked. The proposed framework helps solve such situations. Consider, for example, the following simplified dataset of hospital blood tests extracted from a hospital SNOMED CT-standardized CP system during a particular period of time (see Table 5.4). In this dataset, SNOMED CPIDs 422504039, 442338038 and 13619038, refer to three CPs for ischemic stroke, gastric bypass and thyroidectomy, respectively. The blood tests SNOMED CT codes are as follows: complete blood count - CBC (26604007), thyroid stimulating hormone measurement (61167004), triiodothyronine free measurement -T3 (104994008), glucose tolerance test (113076002), and hemoglobin A1c measurement (43396009). Such hyphenated coding system based blood test output facilitates the computation of the costs of blood tests (or other interventions) per CP in a hospital. Because each blood test is linked to its CP, every blood test can now be traced to its medical condition. Considering the actual fees for blood tests from the 2019 Ontario schedule of benefits for laboratory services [143], the blood tests' cost per CP can be determined as follows: thyroidectomy CP (13619038) = CA\$ 25.36, ischemic stroke CP (422504039) =\$61.74, and gastric bypass CP (442338038) = \$72.26 (see Table 5.5). This facilitates data analytics on CP costs for improved HRM.

For example, Figure 5.7 clearly shows the high cost of blood tests of CP-B. This would prompt decision-makers to further investigate the contents of this CP to determine whether it is a proper and accepted practice, or a source of waste. Besides knowing the total cost of blood tests per CP, the frequency of a particular blood test per CP can also be determined. Figure 5.8 illustrates this ability, considering the frequency of CBC blood test per CP. By comparing the CBC test count on the approved CP with the number of performed CBC tests on patient records, healthcare decision-makers could determine if more tests are warranted, and investigate the relevant cases for improved optimization and cost control.

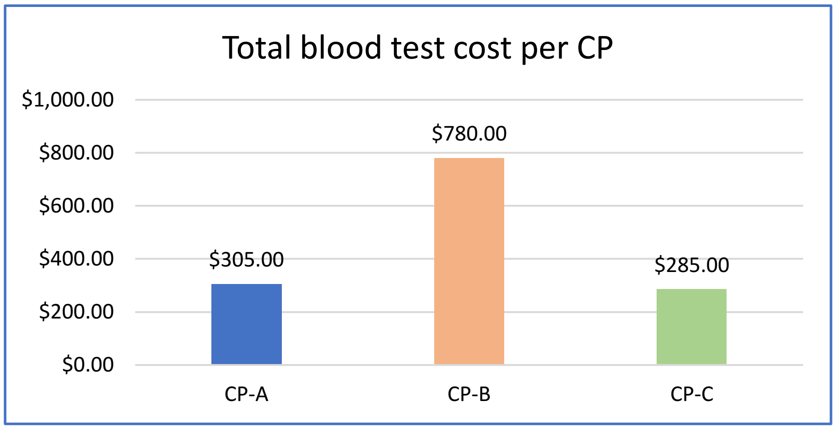

Figure 5.7: Comparison between different CPs based on the cost of blood tests.

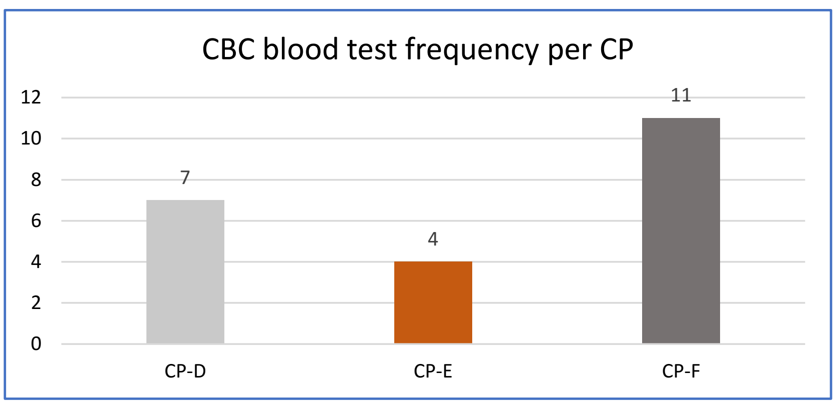

Figure 5.8: Comparison between different CPs based on the frequency of a particular blood test.

Table 5.4: Sample of a hospital's CP-related blood tests encoded using the hyphenated coding system.

August 2019				
13619038-61167004	422504039-113076002	442338038-26604007	422504039-26604007	13619038-61167004
442338038-113076002	13619038-104994008	13619038-104994008	442338038-113076002	422504039-43396009
13619038-61167004	442338038-43396009	442338038-43396009	13619038-26604007	442338038-113076002
422504039-113076002	422504039-113076002	422504039-26604007	442338038-43396009	13619038-61167004

Table 5.5: Blood tests costs scenario.

Ischemic stroke (422504039)	**Gastric bypass** (442338038)	**Thyroidectomy** (13619038)
113076002	26604007	61167004
113076002	113076002	61167004
113076002	43396009	104994008
26604007	113076002	104994008
26604007	43396009	26604007
43396009	113076002	61167004
	43396009	61167004
Cost = \$61.74	Cost = \$72.26	Cost = \$25.36

5.4.3 HRM through CP intervention time analytics

In this scenario, CP-level HRM (i.e., within CP) is considered. The scenario is related to time variation analysis of CP interventions. In today's extremely busy medical work environments, time management in hospitals is important to control costs and save lives. Wasted time may deprive other patients from obtaining the required healthcare service due to lack of nurses and medical staff. This could result in the death of some patients.

Fully standardized and digitized CPs facilitate the recording of the start and end times of any required intervention in the CP. This can be a great help in discovering inefficient practices related to time management. For this objective, we developed the CP time analytics algorithm (Algorithm 5.1). As shown on the algorithm, the Patient_CP_interventions_file is the file that contains the traces of all patients treated by a particular CP with their interventions (intervention_SCTID), intervention_startTime, and intervention_endTime. The output of the algorithm is the maximum duration, minimum duration, and average duration of each intervention of the clinical pathway under consideration. Durations are determined considering all patients who were administered through the clinical pathway within a certain period of time (e.g., one year).

Figure 5.9 illustrates a graph representing a sample output related to the intervention "Screening for Dysphagia" [144] for ischemic stroke patients (SCTID 431765005). The figure shows the time optimization frame that should be considered by healthcare administrators to remove/minimize sources of inefficient time management. The average (and recommended) time for the screening is approximately 30 minutes. A long time for screening (e.g., 57 minutes as shown in the figure), results in the testing room being occupied for longer than necessary, and deprives other patients from getting tested on time. In this particular case, solutions to manage time could be either by preparing the test room/equipment prior to the arrival of patients, by analyzing the cases of longer duration for better control, or by providing more training for nurses on efficient practices in dysphagia screening.

Time variation analysis of CP interventions for all CPs in a hospital helps healthcare managers improve healthcare provision and optimize hospital resources. Such detailed CP analysis is not possible with today's unstructured paper-based CPs or partially com-

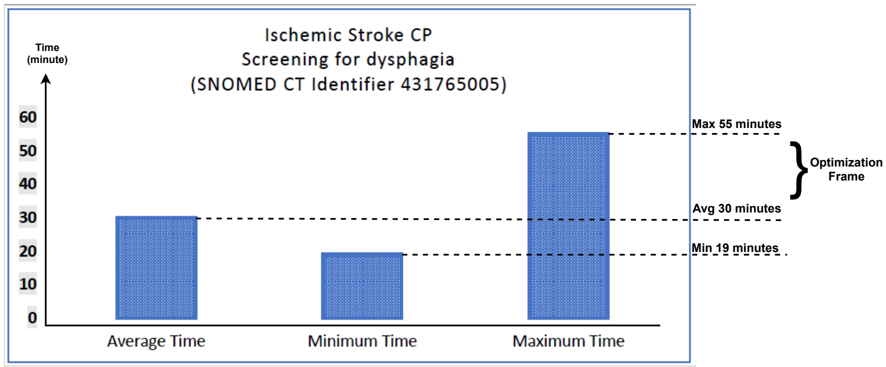

Figure 5.9: CP Intervention time analytics based optimization.

puterized CPs. This shows how fully digitizing CPs could help with hospital resource optimization.

Algorithm 5.4: CP time-based data analytics

```
input  : Patient_CP_interventions_file (contains the sequences of:
         intervention_SCTID, intervention_startTime, and intervention_endTime
         for each patient)
output: max_duration[SCTID]: Maximum duration for each intervention
output: min_duration[SCTID]: Minimum duration for each intervention
output: avg_duration[SCTID]: Average duration for each intervention
Data Structures:
SCTID: A variable to store an intervention's SCTID
intervention_duration[SCTID]: Hash-table with list of all durations for each
  intervention (SCTID)
total_duration[SCTID]: List to store total durations of each intervention (SCTID)
begin
    foreach record ∈ Patient_CP_interventions_file do
        foreach intervention ∈ record do
            SCTID← intervention_SCTID
            if (SCTID ∉ intervention_duration[SCTID]) then
                Add SCTID to intervention_duration[SCTID]
            end
            intervention_duration[SCTID] ← intervention_endTime -
              intervention_startTime
        end
    end
    foreach record ∈ intervention_duration[SCTID] do
        total_duration[SCTID]=0
        max_duration[SCTID] = min_duration[SCTID] = duration [0] (duration of
          1st intervention in the record)
        for i← 0 to length of record do
            total_duration[SCTID]←total_duration[SCTID] + duration[i]
            if (duration[i] > max_duration[SCTID]) then
                max_duration[SCTID] ← duration[i]
            end
            if (duration[i] < min_duration[SCTID]) then
                min_duration[SCTID] ← duration[i]
            end
        end
        avg_duration[SCTID] ← total_duration[SCTID] / length of record
    end
end
```

5.5 Hospital Length of Stay Prediction

Data analytics algorithms prefer rich datasets. They work better when data are as complete as possible [145, 146]. Missing data have always formed a challenge in the face of obtaining good classification and prediction results through machine learning algorithms. This is more critical in the field of healthcare data analytics because patient outcomes are sensitive to data collected in hospitals.

CPs are intended to be the major sources of patient data in hospitals. Since the framework developed in this thesis has the objective of generating computerized CPs that are fully encoded with the hyphenated coding system, this framework contributes to reducing missing healthcare data by providing rich CP-based datasets. This is achieved by making all CP data digitally visible. We illustrate this contribution by an example related to machine learning experiments from the domain of hospital Length of Stay (LOS) prediction.

LOS refers to the number of days that an inpatient remains in a hospital. LOS has long been a crucial metric of hospital efficiency and quality of care. The uncertainty of LOS increases costs and makes it difficult for hospitals to optimize their scheduling process [131]. The clinical and financial consequences of long LOS have made LOS as one of the most observed measures in healthcare systems [147].

LOS predictions that are related to rehabilitation CPs (i.e., CPs applied to patients in their rehabilitation stage) suffer from the fact that many rehabilitation interventions are not stored in EMRs. This makes EMR-based datasets yield less accurate LOS predictions. The challenge with rehabilitation CPs is that they contain many nursing care interventions. We refer to these interventions as “soft” interventions, meaning interventions such as assisting with toileting. Although such interventions are performed on patients and documented on paper, they are rarely recorded in EMRs compared to what we term as “hard” CP interventions, such as X-rays, surgical procedures or injections [10].

In stroke patients, soft interventions have an effect on LOS because patients who require more nursing services show a longer LOS. Thus, we hypothesize that data mining algorithms that work on datasets which do not include soft interventions (i.e., have missing data) yield less accurate LOS prediction results than datasets that include soft interventions. Although terminology systems (like SNOMED CT) have progressed in recent years to include standardized terms and codes for soft interventions like nursing care tasks, current CPs in use at hospitals still present most soft interventions as unstructured text without using their standardized terms and codes. This is a major reason for soft interventions being missed in EMR-based datasets (i.e., datasets obtained from EMRs without capturing all interventions specified on CPs). The framework developed in this research contributes to recording soft interventions by means of SNOMED-CT based standardized CPs. This results in CP-based datasets that are richer in data. To illustrate this, we performed machine learning experiments on the prediction of LOS using two versions (two datasets) of the output dataset (i.e., the output file described in Chapter 4). In the first dataset, we kept the nursing services, whereas they were removed in the second dataset. The objective was to compare LOS prediction results between the two datasets. For a fair comparison between the datasets, we used the same base machine learning algorithm on both datasets. The LOS values in the datasets were between four and nine days. The median LOS was five days.

The objective of the data mining model was to predict short versus long LOS. We used the median LOS as the threshold dividing long vs. short LOS. Thus, patients with LOS less than or equal to five days were labeled as short LOS, while patients with LOS greater than five days were labeled as long LOS. Since more patients had short LOS, the datasets were imbalanced (i.e., contained skewed data). Thus, the median was a representative measure for the central tendency [148]. Since we decided to select the same base machine learning algorithm on both datasets, we selected the decision tree algorithm for the following reasons. Our LOS prediction problem was a non-linear binary classification problem making decision tree-based methods suitable base algorithms for solving this type of problem because they are successful in dealing with non-linear classification; furthermore, many researchers reported that decision tree methods were successful in the domain of LOS prediction [149, 150, 151, 152, 153]. In general, decision tree algorithms use entropy-based methods to form tree nodes [154, 155] by selecting the most informative attributes based on two measures: entropy and information gain, as follows:

- Entropy (H) measures the impurity of a category or class (X), as shown in equation (5.6).

$$H_X = -\sum_{\forall x \in X} P(x) log_2 P(x), \tag{5.6}$$

where $P(x)$ is the probability of label x in X [149].

- Information gain measures the purity of an attribute based on the conditional entropy determined by equation (5.7) below.

$$H_{Y|X} = -\sum_{\forall x \in X} P(x) \sum_{\forall y \in Y} P(y \mid x) log_2 P(y \mid x), \tag{5.7}$$

  where $H_{Y|X}$ is the conditional entropy for each attribute (X) relative to base entropy (Y) which is the entropy of the output variable, LOS in our case. The information gain of an attribute X is defined as the difference between the base entropy and the conditional entropy of the attribute, as shown in equation (5.8).

$$InfoGain_X = H_Y - H_{Y|X}. \tag{5.8}$$

  Information gain compares the degree of purity of the upper node (parent node) before a split with the degree of purity of the lower node (child node) after a split. At every split, an attribute (or predictor) with the highest information gain is considered as the most informative attribute and is chosen for the split [149].

Among the most commonly-used decision tree learning algorithms are ID3 (Iterative Dichotomiser 3), C4.5, and C5.0. ID3 algorithm has the drawback of possibly constructing a complex and deep tree that causes overfitting, leading to poor prediction results. The C4.5 algorithm is an improved ID3 algorithm that addresses the overfitting problem in ID3 by using the technique of pruning to simplify the decision tree. Pruning is achieved by removing the tree nodes and branches that do not provide additional information [149]. C5.0 algorithm offers several improvements over C4.5, including faster processing and more efficient memory usage [156]. Based on the above analysis, we adopted the C5.0 algorithm in our experiments. As will be detailed in the results section, experimental evaluation demonstrates that LOS prediction using CP-based datasets outperformed LOS predictions based on traditional EMR-based datasets.

5.5.1 Results and Discussion

We implemented the experiments using 'RStudio' integrated development environment for R programming language [157, 158]. We performed identical processing and experiments on both datasets. The datasets were split into training/testing sets. In order to avoid generalizing the results from a single split, we conducted experiments with 70:30 and 80:20 training/testing split ratios. Furthermore, we diversified the performance metrics by including multiple major common metrics, including the area under the receiver operating characteristic curve (AUROC), accuracy, sensitivity, specificity, and precision, as shown in equations (5.9), (5.10), (5.11) and (5.12). In the equations, TP, TN, FP, and FN represent True Positive, True Negative, False Positive, and False Negative, respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{5.9}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}. \tag{5.10}$$

$$\text{Specificity} = \frac{TN}{TN + FP}. \tag{5.11}$$

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{5.12}$$

Figs. 5.10 and 5.11 outline the experimental results. As shown in the figures, the performance of the prediction model that included the CP nursing services was better than the model without the nursing services in terms of the considered metrics. The results show better AUROC for CP-based dataset ($\approx$ 88% and 93%) compared to EMR-based dataset ($\approx$ 78% and 84%) for split ratios 70:30 and 80:20, respectively. The most commonly reported measure of a classifier is the accuracy because accuracy evaluates the overall efficiency. The results showed better accuracy for the CP-based dataset ($\approx$ 85% and 92%) compared to EMR-based dataset ($\approx$ 77% and 85%) for split ratios 70:30 and 80:20, respectively.

The results also revealed better performance for the CP-based dataset in terms of sensitivity and equal performance with the EMR-based dataset in terms of specificity. Sensitivity assesses the effectiveness of the classifier on the positive/minority class. In our experiments, this is the class of patients with long LOS. Thus, the CP-based dataset

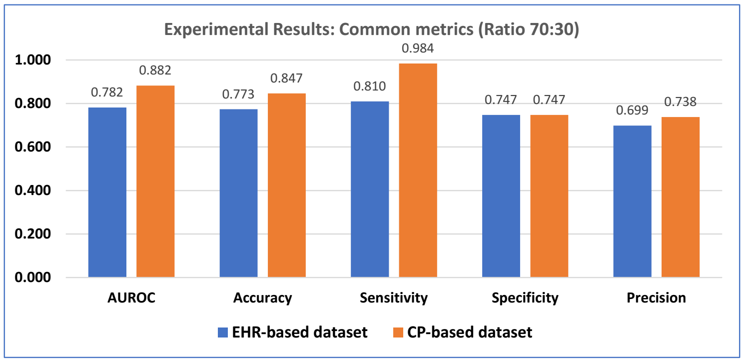

Figure 5.10: Experimental results: Common metrics, 70:30 split ratio.

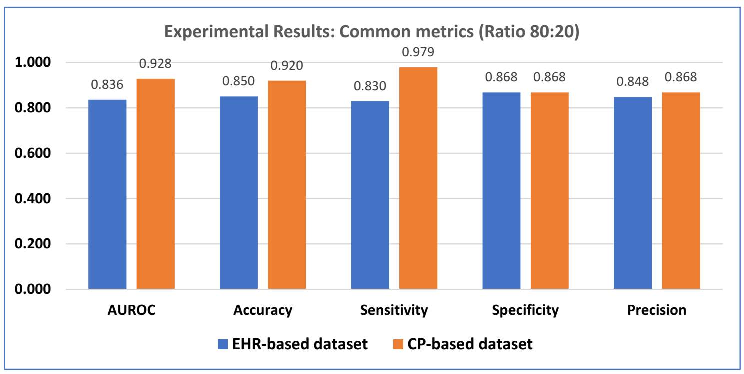

Figure 5.11: Experimental results: Common metrics, 80:20 split ratio.

yielded improved long LOS prediction performance. Specificity, on the other hand, measures the effectiveness of predicting negative cases (short LOS in our experiments). Since fewer nursing services were related to short LOS, it was reasonable that both datasets showed equal specificity, i.e., equal prediction performance on patients with fewer nursing services. Precision is also called positive predictive value. The results revealed that the CP-based dataset provided better precision under both training/testing split ratios.

The above-mentioned metrics are the most used performance measures for such classification problems. However, since our datasets were imbalanced, we decided to investigate additional metrics that consider imbalanced datasets. This helps in generalizing the results by considering performance metrics that combine the previous metrics to account for imbalanced datasets. Therefore, we considered the *balanced accuracy* and *geometric mean (G-mean)*, which are common imbalance-oriented performance metrics [159], as shown in equations (5.13) and (5.14).

$$\begin{aligned}\text{Balanced Accuracy} &= \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right). \\ &= \frac{1}{2}(\textit{sensitivity} + \textit{specificity}).\end{aligned} \tag{5.13}$$

$$\begin{aligned}\text{G-mean} &= \sqrt{\frac{TP}{TP+FN} \times \frac{TN}{TN+FP}}. \\ &= \sqrt{\textit{sensitivity} \times \textit{specificity}}.\end{aligned} \tag{5.14}$$

Figures 5.12 and 5.13 show the experimental results considering the imbalance-oriented metrics. The balanced accuracy is the average between the sensitivity and the specificity, which measures the average accuracy obtained from both majority and minority classes. "This quantity reduces to the traditional accuracy if a classifier performs equally well on either classes. Conversely, if the high value of the traditional accuracy is due to the classifier taking advantage of the distribution of the majority class, then the balanced accuracy will decrease compared to the accuracy" [159].

Our results showed that both the traditional accuracy and the balanced accuracy had similar values for both datasets, with the CP-based dataset showing better performance. This was an indication of good performance of both classifiers on the majority and minority classes with the CP-based dataset yielding improved performance.

G-Mean is a metric suitable for imbalanced datasets since it measures the balance between classification performances on both the majority and minority classes [159]. Our results showed higher G-mean values for the CP-based dataset ($\approx$ 86% and 92%) than the G-mean values of the EMR-based dataset ($\approx$ 78% and 85%), for both 70:30 and 80:20 split ratios, respectively.

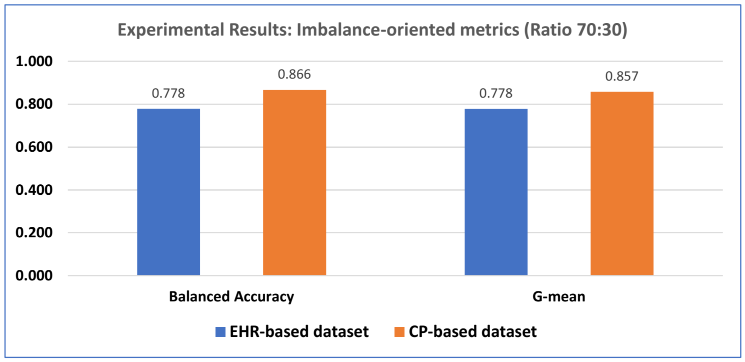

Figure 5.12: Experimental results: Imbalance-oriented metrics, 70:30 split ratio.

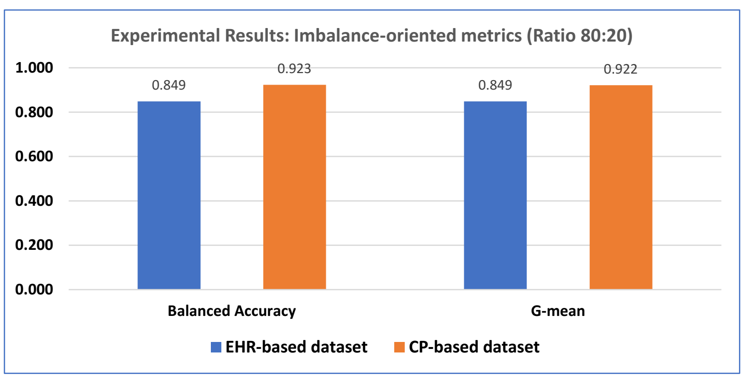

Figure 5.13: Experimental results: Imbalance-oriented metrics, 80:20 split ratio.

As shown in the above analysis, experimental results support the conclusion that CP-based datasets (i.e., datasets without missing interventions) are rich in data and thus, using them improves the performance of data analytics algorithms. It is worth mentioning that there are factors other than nursing services that affect LOS of stroke patients (e.g., comorbidity, diabetes, etc.). However, such data are common to both datasets in our experiments; thus, the only differentiating data are the nursing services available on the clinical pathway.

5.6 Conclusion

Standardization and digitization of clinical pathways enables the capture of all CP related data in hospitals. This provides rich data sources and results in rich datasets that can support healthcare decision making. In this chapter, different data analytics and decision support scenarios were presented and discussed. The scenarios cover various areas in healthcare including: CP variance analysis and action plan, cost management and control, managing patient CP traces, hospital resource management, and hospital length of stay prediction.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

A Clinical Pathway is a multidisciplinary, structured healthcare plan in which therapeutic and diagnostic medical interventions performed by doctors, nurses, and other staff members for a specific disease or procedure are performed in a planned manner [160, 161, 162]. CPs can reduce physicians' mental effort and cognitive load to allow them to focus on thought-provoking, more complex healthcare activities [162]. Therefore, CPs have the potential to improve patient outcomes and satisfaction. CPs also contribute to reducing healthcare costs. Being important components of healthcare systems and important sources of data, CPs deserve more research work towards their automation.

CP automation studies available in the literature have major limitations, such as limiting their role to enhancing EMR functions. In addition, the digital divide between CPs and other healthcare systems was ignored, and no effort was directed towards the standardization, digitization, or independence of clinical pathways in CPMSs. Furthermore, CPs lack any appropriate coding system to identify them digitally, as well as across human networks. In this research, we proposed an ontology-based framework for standardization and digitization of CPs in HISs. An important contribution of the proposed framework is our approach in centralizing CPs in HISs and bridging the digital divide. By centralizing CPs, we mean positioning CPMSs at the centre of HISs. This central position had been occupied by EMRs for decades. The framework would bridge the digital divide by standardizing CPs to convert them from unstructured entities to standardized medical documents.

The SNOMED CT terminology system was adopted for the standardization process. In addition, SNOMED CT was expanded to create an international CP identification code (CPID). The CPID and SNOMED CT coding of the CP data were merged in a hyphenated coding system for CPs that always keeps the link between CPs and their interventions. We proposed the conceptual design of a prototype CPMS that ensures the independence of CPMSs by integrating the components of the framework and allocating CP-specific repository to store CP traces and data analytics algorithms. The proposed framework is a generic framework in the sense that it can be implemented using different designs and programming languages. As a proof of concept, the proposed conceptual design was realized in an sample prototype ontology-based clinical pathway management system, which was developed with the assistance of our domain experts in the Stroke Unit at Thunder Bay Regional Health Sciences Centre.

One major advantage of adopting an ontology-based approach is that ontological modeling facilitates a hierarchical upper-level/disease-level architecture in which abstract CP concepts are modeled at the upper-level ontology, while disease specific CPs are extended and specialized. The hierarchical CP knowledge representation not only renders the framework applicable to any disease-specific CP, but also provides a shared CP standard model that can be communicated among healthcare professionals, as well as through heterogeneous applications in a machine-understandable way, thus facilitating semantic interoperability among healthcare information systems. For example, if a hospital decides to adapt and reuse the CP of another hospital, then what they need to do is to obtain the disease-specific ontology of the other hospital, link it to the common meta-ontology, and modify the CP and its execution to account for its local hospital settings (if required). CP modifications may include changing the CP data (which are SNOMED CT standardized and understandable by all hospitals) or using the appropriate classes from the meta-ontology to add hospital-specific CP sub-classes. Furthermore, the proposed CP digitization framework facilitates the mathematical modeling of CPs and enables a quantitative CP analysis for decision making, CP auditing and quality control.

With respect to big data analytics in healthcare, missing medical data has always impeded the full potential of data mining methods. An objective of the framework proposed in this research is to extract as much data as possible from CPs. This helps in generating rich datasets and consequently improving the performance of data analytics algorithms in healthcare.

6.2 Limitation and Future Work

There are some limitations in this work. The developed CPMS was a small prototype system with basic functionalities implemented to prove that the proposed framework is feasible. A direction for future work would be the development of a complete and extended version of the CPMS by improving its functionality, adding more CPs, and expanding the system's communication messages, which are presently limited to generating sample HL7 messages. Furthermore, additional data analytics/decision support algorithms, such as algorithms that concurrently search for interventions across several CPs, can be developed to benefit from the full digitization of CPs achieved in this research.

SNOMED CT is an enormous terminology system. However, new diseases, procedures, tests, etc. appear continuously, and many terms are still used as local terms. Thus, SNOMED CT might not cover all the required terms, however, SNOMED CT is growing every year and allows the standardization of local terms. Therefore, many new or local terms today that might not have international terminology coverage under SNOMED CT could certainly be covered through its growth and new editions. Moreover, SNOMED CT International (the organization that administers SNOMED CT) always accepts requests to add new terms to the SNOMED CT system. Another limitation is that our ontologies have not yet been connected to other medical ontologies, which limits the semantic behind the used terms. This is a future research direction that will enhance the framework.

At present, only EMR-based datasets are available to researchers. Obtaining CP-based datasets was a challenge in this research, and it is a challenge in general since most CPs in use today are paper-based, and CP-based datasets can only be obtained through simulation or by collaborating directly with cooperating hospitals. Even when a hospital agrees to cooperate, the privacy of patients and their data makes the data available limited, and direct interaction with patients not an option. In order to obtain larger datasets, future research direction would involve collaborating with multiple hospitals.

Another future work is to apply semantic similarity concepts to introduce some automation to the process of CP term standardization. We are currently experimenting with different similarity measures to evaluate their appropriateness for SNOMED CT hierarchy. It is important to understand that AI applications in CP term standardization should not replace human experts as the final decision makers for term standardization. CP instructions are interventions applied on human patients, and therefore extreme caution should be exercised when utilizing automatic standardization methods for safety reasons.

We consider this research to be a starting milestone towards the international standardization and digitization of CPs, as well as towards the complete automation in the CP field, hoping that it will encourage health informatics researchers around the world to participate in advancing this emerging field through more research.

This research movement could definitely entice healthcare policymakers and hospital administrators to appreciate and adopt the proposed approach at both national and international levels. Achieving an international standardization for CPs used in hospitals around the world would have an enormous role in improving the field of health informatics and enhancing the communication of CP practices. It would also help decrease healthcare costs, improve healthcare outcomes, increase patient satisfaction, and achieve healthier people and an overall healthier society.

References

[1] A. C. Machado, M. Martins, B. Cordeiro, and M. Au-Yong-Oliveira, "Where is the health informatics market going?," in *World Conference on Information Systems and Technologies*, pp. 584–595, Springer, 2020.

[2] L. R. Hardy, *Fast Facts in Health Informatics for Nurses*. Springer Publishing Company, 2019.

[3] M. D. Lytras and A. Sarirete, *Innovation in Health Informatics: A Smart Healthcare Primer*. Academic Press, 2019.

[4] "Medicare and Medicaid EHR Incentive Programs." http://fusionppt.com. Accessed: 2020-05-16.

[5] A. K. Manrai and I. S. Kohane, "Bioinformatics and precision medicine," in *Key Advances in Clinical Informatics*, pp. 145–160, Elsevier, 2017.

[6] J. P. Palma and P. Tarczy-Hornoch, "Biomedical informatics in neonatology," in *Avery's Diseases of the Newborn*, pp. 11–19, Elsevier, 2018.

[7] J. M. Madden, M. D. Lakoma, D. Rusinak, C. Y. Lu, and S. B. Soumerai, "Missing clinical and behavioral health data in a large electronic health record (EHR) system," *Journal of the American Medical Informatics Association*, vol. 23, no. 6, pp. 1143–1149, 2016.

[8] J. Leviss and P. Charney, *HIT or miss: lessons learned from health information technology implementations*. AHIMA Press Chicago, 2013.

[9] J. Codella, H. Sarker, P. Chakraborty, M. Ghalwash, Z. Yao, and D. Sow, "eXITs: An Ensemble Approach for Imputing Missing EHR Data," in *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 1–3, IEEE, 2019.

[10] P. A. Potter, A. G. Perry, P. Stockert, A. Hall, B. J. Astle, and W. Duggleby, *Canadian Fundamentals of Nursing*. Elsevier Health Sciences, 2018.

[11] S. Phung, A. Kumar, and J. Kim, "A deep learning technique for imputing missing healthcare data," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6513–6516, IEEE, 2019.

[12] H. Hegde, N. Shimpi, A. Panny, I. Glurich, P. Christie, and A. Acharya, "Mice vs ppca: Missing data imputation in healthcare," *Informatics in Medicine Unlocked*, vol. 17, p. 100275, 2019.

[13] X. Wang, S. Su, H. Jiang, J. Wang, X. Li, and M. Liu, "Short-and long-term effects of clinical pathway on the quality of surgical non-small cell lung cancer care in china: an interrupted time series study," *International Journal for Quality in Health Care*, vol. 30, no. 4, pp. 276–282, 2018.

[14] R. J. Coffey, J. S. Richards, C. S. Remmert, S. S. LeRoy, R. R. Schoville, and P. J. Baldwin, "An introduction to critical paths," *Quality Management in Healthcare*, vol. 14, no. 1, pp. 46–55, 2005.

[15] S. D. Pearson, S. F. Kleefield, J. R. Soukop, E. F. Cook, and T. H. Lee, "Critical pathways intervention to reduce length of hospital stay," *The American journal of medicine*, vol. 110, no. 3, pp. 175–180, 2001.

[16] K. Zander, K. A. Bower, and M. Etheredge, "Nursing case management: blueprints for transformation," *Boston: New England Medical Center Hospitals*, pp. 1–128, 1987.

[17] E. Rooney, "Developing care pathways–lessons from the Steele Review implementation in England," *Gerodontology*, vol. 31, pp. 52–59, 2014.

[18] G. Schrijvers, A. van Hoorn, and N. Huiskes, "The care pathway: concepts and theories: an introduction," *International journal of integrated care*, vol. 12, no. Special Edition Integrated Care Pathways, 2012.

[19] R. S. Russell and B. W. Taylor, *Operations and Supply Chain Management, 9th Edition*. John Wiley & Sons, 2017.

[20] K. Vanhaecht, M. Panella, R. Van Zelm, and W. Sermeus, "An overview on the history and concept of care pathways as complex interventions," *International Journal of Care Pathways*, vol. 14, no. 3, pp. 117–123, 2010.

[21] A. K. Lawal, T. Rotter, L. Kinsman, A. Machotta, U. Ronellenfitsch, S. D. Scott, D. Goodridge, C. Plishka, and G. Groot, "What is a clinical pathway? refinement of an operational definition to identify clinical pathway studies for a cochrane systematic review," *BMC medicine*, vol. 14, no. 1, p. 35, 2016.

[22] "European Pathway Association." http://e-p-a.org. Accessed: 2020-01-30.

[23] M. Donald, K. McBrien, W. Jackson, B. J. Manns, M. Tonelli, K. King-Shier, K. Jindal, R. Z. Lewanczuk, N. Scott-Douglas, T. Braun, *et al.*, "Development and implementation of an online clinical pathway for adult chronic kidney disease in primary care: a mixed methods study," *BMC medical informatics and decision making*, vol. 16, no. 1, p. 109, 2016.

[24] L. De Bleser, R. Depreitere, K. D. WAELE, K. Vanhaecht, J. Vlayen, and W. Sermeus, "Defining pathways," *Journal of nursing management*, vol. 14, no. 7, pp. 553–563, 2006.

[25] M. Khalifa and O. Alswailem, "Clinical pathways: Identifying development, implementation and evaluation challenges.," in *ICIMTH*, pp. 131–134, 2015.

[26] L. Kinsman, T. Rotter, E. James, P. Snow, and J. Willis, "What is a clinical pathway? development of a definition to inform the debate," *BMC medicine*, vol. 8, no. 1, p. 31, 2010.

[27] M. Renholm, H. Leino-Kilpi, and T. Suominen, "Critical pathways: a systematic review," *JONA: The Journal of Nursing Administration*, vol. 32, no. 4, pp. 196–202, 2002.

[28] B. J. Gebhardt, J. Thomas, Z. D. Horne, C. E. Champ, G. M. Ahrendt, E. Diego, D. E. Heron, and S. Beriwal, "Standardization of nodal radiation therapy through changes to a breast cancer clinical pathway throughout a large, integrated cancer center network," *Practical radiation oncology*, vol. 8, no. 1, pp. 4–12, 2018.

[29] D. Ogilvie-Harris, D. Botsford, and R. W. Hawker, "Elderly patients with hip fractures: improved outcome with the use of care maps with high-quality medical and nursing protocols.," *Journal of orthopaedic trauma*, vol. 7, no. 5, pp. 428–437, 1993.

[30] M. Panella, S. Marchisio, R. Brambilla, K. Vanhaecht, and F. Di Stanislao, "A cluster randomized trial to assess the effect of clinical pathways for patients with stroke: results of the clinical pathways for effective and appropriate care study," *BMC medicine*, vol. 10, no. 1, p. 71, 2012.

[31] M. Panella, S. Marchisio, A. Barbieri, and F. Di Stanislao, "A cluster randomized trial to assess the impact of clinical pathways for patients with stroke: rationale and design of the clinical pathways for effective and appropriate care study [nct00673491]," *BMC health services research*, vol. 8, no. 1, p. 223, 2008.

[32] S. Preston, S. Markar, C. Baker, Y. Soon, S. Singh, and D. Low, "Impact of a multidisciplinary standardized clinical pathway on perioperative outcomes in patients with oesophageal cancer," *British journal of surgery*, vol. 100, no. 1, pp. 105–112, 2013.

[33] L. Stead, C. Arthur, and A. Cleary, "Do multidisciplinary pathways of care affect patient satisfaction," *Health Care Risk Report*, vol. 11, pp. 13–5, 1995.

[34] P. A. Van Dam, G. Verheyden, A. Sugihara, X. B. Trinh, H. Van Der Mussele, H. Wuyts, L. Verkinderen, J. Hauspy, P. Vermeulen, and L. Dirix, "A dynamic clinical pathway for the treatment of patients with early breast cancer is a tool for better cancer care: implementation and prospective analysis between 2002–2010," *World journal of surgical oncology*, vol. 11, no. 1, p. 70, 2013.

[35] J. Williams, R. Roberts, and M. Rigby, "Integrated patient records: another move towards quality for patients?," *Quality in Health Care*, vol. 2, no. 2, p. 73, 1993.

[36] C. Mosher, P. Cronk, A. Kidd, P. McCormick, S. Stockton, and C. Sulla, "Upgrading practice with critical pathways," *The American journal of nursing*, vol. 92, no. 1, pp. 41–44, 1992.

[37] J. MacDermid, "Practice guidelines, algorithms, and clinical pathways," *Evidence-based rehabilitation: a guide to practice*, pp. 227–261, 2008.

[38] P. Bjurling-Sjöberg, *Clinical Pathway Implementation and Teamwork in Swedish Intensive Care: Challenges in Evidence-Based Practice and Interprofessional Collaboration*. PhD thesis, Acta Universitatis Upsaliensis, 2018.

[39] W. G. Carnett, "Clinical practice guidelines: a tool to improve care," *Journal of nursing care quality*, vol. 16, no. 3, pp. 60–70, 2002.

[40] Y. Wang, L. Kung, and T. A. Byrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations," *Technological Forecasting and Social Change*, vol. 126, pp. 3–13, 2018.

[41] Markets and Markets, "Healthcare analytics market." https://www.marketsandmarkets.com/Market-Reports/. Accessed: 2020-04-16.

[42] K. De Luc and J. Todd, *e-pathways: Computers and the patient's journey through care*. Radcliffe Publishing, 2003.

[43] J. G. Yetzer, P. Pirgousis, Z. Li, and R. Fernandes, "Clinical pathway implementation improves efficiency of care in a maxillofacial head and neck surgery unit," *Journal of Oral and Maxillofacial Surgery*, vol. 75, no. 1, pp. 190–196, 2017.

[44] J. L. Baumbusch, S. R. Kirkham, K. B. Khan, H. McDonald, P. Semeniuk, E. Tan, and J. M. Anderson, "Pursuing common agendas: a collaborative model for knowledge translation between research and practice in clinical settings," *Research in nursing & health*, vol. 31, no. 2, pp. 130–140, 2008.

[45] "eHealth Ontario." https://www.ehealthontario.on.ca. Accessed: 2020-01-30.

[46] "Canadian Broadcasting Corporation." http://www.cbc.ca/news/health/cihi-health-costs-canada-report-prescriptions-pharmacare-1.4390945. Accessed: 2020-01-30.

[47] "Clinical Pathways, Ottawa Hospital." http://www.ottawahospital.on.ca/en/our-model-of-care/clinical-pathways. Accessed: 2018-01-30.

[48] J. Liu, Z. Huang, X. Lu, and H. Duan, "An ontology-based real-time monitoring approach to clinical pathway," in *2014 7th International Conference on Biomedical Engineering and Informatics*, pp. 756–761, IEEE, 2014.

[49] Z. Hu, J.-S. Li, T.-S. Zhou, H.-Y. Yu, M. Suzuki, and K. Araki, "Ontology-based clinical pathways with semantic rules," *Journal of medical systems*, vol. 36, no. 4, pp. 2203–2212, 2012.

[50] Z. Hu, J.-S. Li, H.-y. Yu, X.-g. Zhang, M. Suzuki, and K. Araki, "Modeling of clinical pathways based on ontology," in *2009 IEEE International Symposium on IT in Medicine & Education*, vol. 1, pp. 1170–1174, IEEE, 2009.

[51] W. Fan, X. Lu, Z. Huang, W. Yu, and H. Duan, "Constructing clinical pathway ontology to incorporate patient state, intervention and time," in *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, vol. 3, pp. 1697–1701, IEEE, 2011.

[52] Y. Ye, X. Diao, Z. Jiang, and G. Du, "A knowledge-based variance management system for supporting the implementation of clinical pathways," in *2009 International Conference on Management and Service Science*, pp. 1–4, IEEE, 2009.

[53] Y. Ye, Z. Jiang, X. Diao, D. Yang, and G. Du, "An ontology-based hierarchical semantic modeling approach to clinical pathway workflows," *Computers in biology and medicine*, vol. 39, no. 8, pp. 722–732, 2009.

[54] A. Alahmar, E. Mohammed, and R. Benlamri, "Application of data mining techniques to predict the length of stay of hospitalized patients with diabetes," in *2018 4th International Conference on Big Data Innovations and Applications (Innovate-Data)*, pp. 38–43, IEEE, 2018.

[55] C. J. Baker and e. Kei-Hoi Cheung, *Semantic web: Revolutionizing knowledge discovery in the life sciences*. Reading, Massachusetts: Springer Science and Business Media, 2007.

[56] "W3C Web Ontology Language (OWL)." https://www.w3.org. Accessed: 2020-01-30.

[57] "SWRL: A Semantic Web Rule Language." https://www.w3.org/Submission/SWRL. Accessed: 2020-01-30.

[58] J. Hebeler, M. Fisher, R. Blace, and A. Perez-Lopez, *Semantic web programming*. John Wiley & Sons, 2011.

[59] J. Davies, R. Studer, and P. Warren, *Semantic Web technologies: trends and research in ontology-based systems*. John Wiley & Sons, 2006.

[60] D. H. Fudholi and L. Mutawalli, "An ontology model for clinical pathway audit," in *2018 4th International Conference on Science and Technology (ICST)*, pp. 1–6, IEEE, 2018.

[61] J. Tehrani, "Computerisation of clinical pathways," in *Healthcare Ethics and Training: Concepts, Methodologies, Tools, and Applications*, pp. 1050–1074, IGI Global, 2017.

[62] J. Tehrani, "Computerisation of clinical pathways: Based on a semiotically inspired methodology," in *E-Health and Telemedicine: Concepts, Methodologies, Tools, and Applications*, pp. 25–48, IGI Global, 2016.

[63] J. Tehrani, K. Liu, and V. Michell, "Semiotics-oriented method for generation of clinical pathways," in *LISS 2012*, pp. 477–482, Springer, 2013.

[64] J. Tehrani, K. Liu, and V. Michell, "Ontology modeling for generation of clinical pathways," *Journal of Industrial Engineering and Management (JIEM)*, vol. 5, no. 2, pp. 442–456, 2012.

[65] J. Effah, P. K. Senyo, and S. Opoku-Anokye, "Business intelligence architecture informed by organisational semiotics," in *International Conference on Informatics and Semiotics in Organisations*, pp. 268–277, Springer, 2018.

[66] M. S. Santos, B. S. M. Bertãozini, and V. Neris, "Studies in organisational semiotics: A systematic literature review," 2016.

[67] H.-Q. Wang, T.-S. Zhou, Y.-F. Zhang, L. Chen, and J.-S. Li, "Research and development of semantics-based sharable clinical pathway systems," *Journal of medical systems*, vol. 39, no. 7, p. 73, 2015.

[68] S. R. Abidi and S. S. R. Abidi, "An ontological modeling approach to align institution-specific clinical pathways: Towards inter-institution care standardization," in *2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 1–4, IEEE, 2012.

[69] S. R. Abidi, S. S. R. Abidi, L. Butler, and S. Hussain, "Operationalizing prostate cancer clinical pathways: An ontological model to computerize, merge and execute institution-specific clinical pathways," in *Workshop on Knowledge Management for Health Care Procedures*, pp. 1–12, Springer, 2008.

[70] A. Daniyal, S. R. Abidi, and S. S. R. Abidi, "Computerizing clinical pathways: ontology-based modeling and execution.," in *MIE*, pp. 643–647, 2009.

[71] D. A. Alexandrou, K. V. Pardalis, T. D. Bouras, P. Karakitsos, and G. N. Mentzas, "Sempath ontology: Modeling multidisciplinary treatment schemes utilizing semantics," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 2, pp. 235–240, 2011.

[72] S. H. Hoelscher and S. McBride, "Digitizing infectious disease clinical guidelines for improved clinician satisfaction," *CIN: Computers, Informatics, Nursing*, 2020.

[73] P. B. Smulowitz, Y. Dizitzer, S. Tadiri, L. Thibodeau, L. Jagminas, and V. Novack, "Impact of implementation of the heart pathway using an electronic clinical decision support tool in a community hospital setting," *The American journal of emergency medicine*, vol. 36, no. 3, pp. 408–413, 2018.

[74] J. Gibbs, L. J. Sutcliffe, V. Gkatzidou, K. Hone, R. E. Ashcroft, E. M. Harding-Esch, C. M. Lowndes, S. T. Sadiq, P. Sonnenberg, and C. S. Estcourt, "The eclinical care pathway framework: a novel structure for creation of online complex clinical care pathways and its application in the management of sexually transmitted infections," *BMC medical informatics and decision making*, vol. 16, no. 1, p. 98, 2016.

[75] R. Blaser, M. Schnabel, C. Biber, M. Bäumlein, O. Heger, M. Beyer, E. Opitz, R. Lenz, and K. A. Kuhn, "Improving pathway compliance and clinician performance by using information technology," *International journal of medical informatics*, vol. 76, no. 2-3, pp. 151–156, 2007.

[76] K. Bernstein and U. Andersen, "Managing care pathways combining snomed ct, archetypes and an electronic guideline system.," *Studies in health technology and informatics*, vol. 136, p. 353, 2008.

[77] I. L. Katzan, Y. Fan, M. Speck, J. Morton, L. Fromwiller, J. Urchek, K. Uchino, S. D. Griffith, and M. Modic, "Electronic stroke carepath: integrated approach to stroke care," *Circulation: Cardiovascular Quality and Outcomes*, vol. 8, no. 6_suppl_3, pp. S179–S189, 2015.

[78] "Epic Systems Corporation." `https://www.epic.com`. Accessed: 2020-01-30.

[79] B. Al-Hablani, "The use of automated snomed ct clinical coding in clinical decision support systems for preventive care," *Perspectives in health information management*, vol. 14, no. Winter, 2017.

[80] S. Rai, O. M. Siddiqui, S. K. Meher, A. Shariff, and S. Banga, "Role of nursing informatics in implementation of snomed-ct in india.," *Studies in health technology and informatics*, vol. 264, pp. 1718–1719, 2019.

[81] E. J. Hwang, H.-A. Park, S. K. Sohn, H. B. Lee, H. K. Choi, S. Ha, H. J. Kim, T. W. Kim, and W. Youm, "Mapping korean edi medical procedure code to snomed ct.," in *MedInfo*, pp. 178–182, 2019.

[82] D. H. Lee, F. Y. Lau, and H. Quan, "A method for encoding clinical datasets with snomed ct," *BMC medical informatics and decision making*, vol. 10, no. 1, p. 53, 2010.

[83] K. Giannangelo and S. H. Fenton, "Snomed ct survey: an assessment of implementation in emr/ehr applications," *Perspectives in Health Information Management/AHIMA, American Health Information Management Association*, vol. 5, 2008.

[84] F.-r. Lin, S.-c. Chou, S.-m. Pan, and Y.-m. Chen, "Mining time dependency patterns in clinical pathways," *International journal of medical informatics*, vol. 62, no. 1, pp. 11–25, 2001.

[85] Z. Huang, X. Lu, and H. Duan, "On mining clinical pathway patterns from medical behaviors," *Artificial intelligence in medicine*, vol. 56, no. 1, pp. 35–50, 2012.

[86] A. L. Hilario, J. D. H. Oruga, M. P. B. Turqueza, and D. V. Hilario, "Utilization of clinical pathway on open appendectomy: A quality improvement initiative in a private hospital in the philippines," *International journal of health sciences*, vol. 12, no. 2, p. 43, 2018.

[87] P. Ibeziako, K. Brahmbhatt, A. Chapman, C. De Souza, L. Giles, S. Gooden, F. Latif, N. Malas, L. Namerow, R. Russell, *et al.*, "Developing a clinical pathway for somatic symptom and related disorders in pediatric hospital settings," *Hospital pediatrics*, vol. 9, no. 3, pp. 147–155, 2019.

[88] N. A. Patel, R. A. Bly, S. Adams, K. Carlin, S. R. Parikh, J. P. Dahl, and S. Manning, "A clinical pathway for the postoperative management of hypocalcemia after pediatric thyroidectomy reduces blood draws," *International journal of pediatric otorhinolaryngology*, vol. 105, pp. 132–137, 2018.

[89] S. Ovaere, I. Boscart, I. Parmentier, P. J. Steelant, T. Gabriel, J. Allewaert, H. Pottel, F. Vansteenkiste, and M. D'Hondt, "The effectiveness of a clinical pathway in liver surgery: A case-control study," *Journal of Gastrointestinal Surgery*, vol. 22, no. 4, pp. 684–694, 2018.

[90] I. S. Lanig, P. W. New, A. S. Burns, G. Bilsky, J. Benito-Penalva, D. Bensmail, and M. Yochelson, "Optimizing the management of spasticity in people with spinal cord damage: a clinical care pathway for assessment and treatment decision making from the ability network, an international initiative," *Archives of physical medicine and rehabilitation*, vol. 99, no. 8, pp. 1681–1687, 2018.

[91] C. R. Shubert, M. L. Kendrick, E. B. Habermann, A. E. Glasgow, B. J. Borah, J. P. Moriarty, S. P. Cleary, R. L. Smoot, M. B. Farnell, D. M. Nagorney, *et al.*, "Implementation of prospective, surgeon-driven, risk-based pathway for pancreatoduodenectomy results in improved clinical outcomes and first year cost savings of $1 million," *Surgery*, vol. 163, no. 3, pp. 495–502, 2018.

[92] V. Wylde, W. Bertram, A. D. Beswick, A. W. Blom, J. Bruce, A. Burston, J. Dennis, K. Garfield, N. Howells, A. Lane, *et al.*, "Clinical-and cost-effectiveness of the star care pathway compared to usual care for patients with chronic pain after total knee replacement: study protocol for a uk randomised controlled trial," *Trials*, vol. 19, no. 1, p. 132, 2018.

[93] M. Li, J. Zhang, T. J. Gan, G. Qin, L. Wang, M. Zhu, Z. Zhang, Y. Pan, Z. Ye, F. Zhang, *et al.*, "Enhanced recovery after surgery pathway for patients undergoing cardiac surgery: a randomized clinical trial," *European Journal of Cardio-Thoracic Surgery*, vol. 54, no. 3, pp. 491–497, 2018.

[94] W. Kamil, O. Y. Hian, S. Mohd-Said, S. L. A. Zainuddin, H. Ramli, E. Noor, R. Ayob, A. F. A. Aziz, A. Ismail, S. Sulong, *et al.*, "Development of clinical pathway for non-surgical management of chronic periodontitis," *Malaysian Journal of Public Health Medicine*, vol. 2018, no. Specialissue1, pp. 26–32, 2018.

[95] "Clinical pathways, thunder bay regional health sciences centre." https://tbrhsc.net/, 2019. Accessed: 2019-01-23.

[96] "SNOMED CT at Canada Health Infoway." http://infocentral.infoway-inforoute.ca/en/standards/international/snomed-ct. Accessed: 2018-04-8.

[97] S. Madani, J. Henderson, and K. W. Fung, "Development of an oncology subset of snomed ct based on patient notes.," in *AMIA*, 2016.

[98] "SNOMED CT Browser." http://snomed.org. Accessed: 2020-01-30.

[99] J. Jamie, *Starting Snomed: A Beginner's Guide to the Snomed CT Healthcare Terminology*. Amazon Publishing, 2018.

[100] M. J. Bowie and R. M. Schaffer, *Understanding ICD-10-CM and ICD-10-PCS: A Worktext*. CENGAGE Learning Custom Publishing, 2020.

[101] "Canada Health Infoway." https://www.infoway-inforoute.ca/en, 2020. Accessed: 2020-01-30.

[102] J. Kaufman and K. Reichert, "What is measured gets improved (or if you cannot measure it, you cannot improve it)," *Pediatric Critical Care Medicine*, vol. 19, no. 3, pp. 267–268, 2018.

[103] J. Y. Nakayama, V. Hertzberg, and J. C. Ho, "Making sense of abbreviations in nursing notes: A case study on mortality prediction," *AMIA Summits on Translational Science Proceedings*, vol. 2019, p. 275, 2019.

[104] Y. Fan, A. Wen, F. Shen, S. Sohn, H. Liu, and L. Wang, "Evaluating the impact of dictionary updates on automatic annotations based on clinical nlp systems," *AMIA Summits on Translational Science Proceedings*, vol. 2019, p. 714, 2019.

[105] J. Patrick, Y. Wang, and P. Budd, "An automated system for conversion of clinical notes into SNOMED clinical terminology," in *Proceedings of the fifth Australasian symposium on ACSW frontiers-Volume 68*, pp. 219–226, Australian Computer Society, Inc., 2007.

[106] S. Harispe, S. Ranwez, S. Janaqi, and J. Montmain, "Semantic similarity from natural language and ontology analysis," *Synthesis Lectures on Human Language Technologies*, vol. 8, no. 1, pp. 1–254, 2015.

[107] Y. Feng, E. Bagheri, F. Ensan, and J. Jovanovic, "The state of the art in semantic relatedness: a framework for comparison," *The Knowledge Engineering Review*, vol. 32, 2017.

[108] "SNOMED International." https://confluence.ihtsdotools.org/, 2019. Accessed: 2020-02-18.

[109] "Stroke Association." http://www.strokeassociation.org. Accessed: 2020-01-30.

[110] "SNOMED International's Confluence." https://confluence.ihtsdotools.or. Accessed: 2020-01-30.

[111] "SNOMED International." https://www.snomed.org. Accessed: 2020-01-30.

[112] "SNOMED CT International Release." http://confluence.ihtsdotools.org/display/DOCGLOSS/International+release. Accessed: 2020-01-5.

[113] D. Nardi, R. J. Brachman, *et al.*, "An introduction to description logics.," *Description logic handbook*, vol. 1, p. 40, 2003.

[114] D. Allemang and J. Hendler, *Semantic web for the working ontologist: effective modeling in RDFS and OWL*. Elsevier, 2011.

[115] M. Krötzsch, F. Simancik, and I. Horrocks, "A description logic primer," *arXiv preprint arXiv:1201.4089*, 2013.

[116] H. S. Pinto and J. P. Martins, "Ontologies: How can they be built?," *Knowledge and information systems*, vol. 6, no. 4, pp. 441–464, 2004.

[117] K. A. Wager, F. W. Lee, and J. P. Glaser, *Health care information systems: a practical approach for health care management*. John Wiley & Sons, 2017.

[118] "The Office of the National Coordinator for Health Information Technology." http://www.healthit.gov. Accessed: 2020-01-30.

[119] "Athena Health." http://www.athenahealth.com. Accessed: 2020-01-30.

[120] R. Evans, "Electronic health records: then, now, and in the future," *Yearbook of medical informatics*, vol. 25, no. S 01, pp. S48–S61, 2016.

[121] J. Pepper, *The Electronic Health Record for the Physician's Office for SimChart for the Medical Office-E-Book*. Elsevier Canada, 2019.

[122] "University of South Florida, Health Informatics Program." http://www.usfhealthonline.com. Accessed: 2020-01-30.

[123] "OpenEMR." https://www.open-emr.org/. Accessed: 2020-03-22.

[124] J. W. Park, A. Shah, R. I. Arriaga, and S. Vempala, "Redesigning a basic laboratory information system for the global south," in *2019 ITU Kaleidoscope: ICT for Health: Networks, Standards and Innovation (ITU K)*, pp. 1–8, IEEE, 2019.

[125] The Laboratory, Health, and Science Informatics Encyclopedia, "Laboratory Information System," 2020.

[126] S. Goundrey-Smith, *Information Technology in Pharmacy: An Integrated Approach*. Springer Science & Business Media, 2012.

[127] T. Benson and G. Grieve, *Principles of health interoperability: SNOMED CT, HL7 and FHIR*. Springer, 2016.

[128] "University health network, hapi fhir." http://hapifhir.io, 2019. Accessed: 2019-07-08.

[129] "Regional stroke unit, thunder bay regional health research centre." https://tbrhsc.net/programs-services/cardiovascular-and-stroke-program/regional-stroke-unit, 2019. Accessed: 2019-04-30.

[130] A. Hassan, R. Benlamri, K. Darko, A. Alahmar, S. Jaspers, S. Brown, S. Littlefield, and M. Drombolis, "Defining prevalence, incidence and risk factors of northwestern ontario patients with ischemic stroke secondary to carotid artery disease: A population-based study, world stroke congress abstracts," *International journal of Stroke*, vol. 13, no. 2, pp. 23–24, 2018.

[131] A. Alahmar, E. Mohammed, and R. Benlamri, “Application of data mining techniques to predict the length of stay of hospitalized patients with diabetes,” in *2018 4th International Conference on Big Data Innovations and Applications (Innovate-Data)*, pp. 38–43, IEEE, 2018.

[132] “World wide web consortium, time ontology in owl.” https://www.w3.org/TR/owl-time/, 2019. Accessed: 2019-01-30.

[133] J. F. Allen, “Towards a general theory of action and time,” *Artificial intelligence*, vol. 23, no. 2, pp. 123–154, 1984.

[134] M. A. Musen, “The protégé project: a look back and a look forward,” *AI Matters*, vol. 1, no. 4, pp. 4–12, 2015.

[135] J. F. Allen, “Maintaining knowledge about temporal intervals,” *Communications of the ACM*, vol. 26, no. 11, pp. 832–843, 1983.

[136] M. Proctor, “Drools: a rule engine for complex event processing,” in *Proceedings of the 4th international conference on Applications of Graph Transformations with Industrial Relevance*, pp. 2–2, Springer-Verlag, 2011.

[137] G. M. Karageorgos, I. Z. Apostolakis, P. Nauleau, V. Gatti, R. Weber, E. S. Connolly, E. C. Miller, and E. E. Konofagou, “Arterial wall mechanical inhomogeneity detection and atherosclerotic plaque characterization using high frame rate pulse wave imaging in carotid artery disease patients in vivo,” *Physics in Medicine & Biology*, vol. 65, no. 2, p. 025010, 2020.

[138] “Apache.org.” https://www.apache.org, 2019. Accessed: 2019-08-20.

[139] A. Wagstaff, G. Flores, J. Hsu, M.-F. Smitz, K. Chepynoga, L. R. Buisman, K. van Wilgenburg, and P. Eozenou, “Progress on catastrophic health spending in 133 countries: a retrospective observational study,” *The Lancet Global Health*, vol. 6, no. 2, pp. e169–e179, 2018.

[140] “Canadian Institute for Health Information (CIHI).” https://www.cihi.ca/en, 2020. Accessed: 2020-03-06.

[141] W. H. Organization *et al.*, “Global spending on health: a world in transition,” tech. rep., World Health Organization, 2019.

[142] V. Kandalam, C. K. Lau, M. Guo, I. Ma, and C. Naugler, “Inappropriate repeat testing of complete blood count (cbc) and electrolyte panels in inpatients from alberta, canada,” *Clinical Biochemistry*, vol. 77, pp. 32–35, 2020.

[143] “Ontario Ministry of Health.” http://health.gov.on.ca/en, 2019. Accessed: 2019-11-20.

[144] “Stroke association.” http://www.strokeassociation.org, 2019. Accessed: 2019-02-20.

[145] R. J. Roiger, *Data mining: a tutorial-based primer*. CRC press, 2017.

[146] S. P. Kudyba, *Healthcare informatics: improving efficiency through technology, analytics, and management*. CRC Press, 2018.

[147] M. Shulan and K. Gao, “Revisiting hospital length of stay: what matters?,” *The American journal of managed care*, vol. 21, no. 1, pp. e71–7, 2015.

[148] F. J. Gravetter and L. B. Wallnau, *Statistics for the behavioral sciences*. Cengage Learning, 2016.

[149] E. E. Services, *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. Wiley, 2015.

[150] P. Liu, L. Lei, J. Yin, W. Zhang, W. Naijun, and E. El-Darzi, “Healthcare data mining: Prediction inpatient length of stay,” in *2006 3rd International IEEE Conference Intelligent Systems*, pp. 832–837, IEEE, 2006.

[151] L. Breiman, *Classification and regression trees*. Routledge, 2017.

[152] K. B. Highland, M. R. Cole, S. P. Bilir, J. Pruett, and W. Drake, “Decreasing length of stay and pulmonary arterial hypertension-related hospitalizations with macitentan using a decision tree model in a medicare population,” in *A68. WOW: PHARMACOLOGICAL TREATMENT OF PULMONARY HYPERTENSION*, pp. A2286–A2286, American Thoracic Society, 2017.

[153] M. A. Rahman, B. Honan, T. Glanville, P. Hough, and K. Walker, "Using data mining to predict emergency department length of stay greater than 4 hours: Derivation and single-site validation of a decision tree algorithm," *Emergency Medicine Australasia*, 2019.

[154] B. Lantz, *Machine learning with R: expert techniques for predictive modeling*. Packt Publishing Ltd, 2019.

[155] A. Burkov, *The hundred-page machine learning book*, vol. 1. Andriy Burkov Quebec City, Can., 2019.

[156] M. Kuhn and K. Johnson, *Applied predictive modeling*, vol. 26. Springer, 2013.

[157] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019.

[158] RStudio Team, *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2019.

[159] J. Akosa, "Predictive accuracy: a misleading performance measure for highly imbalanced data," in *Proceedings of the SAS Global Forum*, pp. 2–5, 2017.

[160] R. H. Al-Ashwal and E. Supriyanto, "Clinical pathway in cardiovascular disease management," in *Cardiovascular Engineering*, pp. 143–153, Springer, 2020.

[161] C. Zhang and J. Xiao, "Application of fast-track surgery combined with a clinical nursing pathway in the rehabilitation of patients undergoing total hip arthroplasty," *Journal of International Medical Research*, vol. 48, no. 1, p. 0300060519889718, 2020.

[162] M. Jabbour, A. S. Newton, D. Johnson, and J. A. Curran, "Defining barriers and enablers for clinical pathway implementation in complex clinical settings," *Implementation Science*, vol. 13, no. 1, p. 139, 2018.

APPENDICES

Appendix A

Sample Interview Questions

Q1. Does the meta-ontology represent the generic knowledge about clinical pathways?

Q2. Is there any intervention type that is missing from the ontology design? If yes, what is the missing intervention type?

Q3. How do you evaluate the system screens in general (e.g., user messages, error messages)?

Q4. Is the first screen of the user interface well designed?

Q5. Are the instructions on the user screens easy to understand?

Q6. Does the stroke ontology represent the stroke clinical pathway?

Q7. Is there any CP variance type that is missing from the ontology design? If yes, what is the missing variance type?

Q8. How do you evaluate the user interface screens related to the progress of the patients through the clinical pathway? [Excellent, Good, Fair, Poor].

Q9. Do the SNOMED CT terms and codes used on the system screens match the intended terms on the clinical pathway?

Q10. Questions related to SNOMED CT standardization (e.g., what is the correct SNOMED CT terms for swallowing screen, hemorrhagic stroke, etc?).

Q11. How do you evaluate the use of the terminology throughout the user interface screens? [All are correct, Some errors, All require corrections].