Thesis for the Degree of Doctor of Philosophy

HYBRID KNOWLEDGE MODELING FOR CASE SELECTION AND ADAPTATION IN CDSS

Syed Imran Ali

Department of Computer Science and Engineering
Graduate School
Kyung Hee University
Yongin Korea

August 2022

HYBRID KNOWLEDGE MODELING FOR CASE SELECTION AND ADAPTATION IN CDSS

Syed Imran Ali

Department of Computer Science and Engineering
Graduate School
Kyung Hee University
Yongin Korea

August 2022

HYBRID KNOWLEDGE MODELING FOR CASE SELECTION AND ADAPTATION IN CDSS

by

Syed Imran Ali

Supervised by

Prof. Sungyoung Lee

Prof. TaeChoong Chung

Submitted to the Department of Computer Science and Engineering and the Faculty of Graduate School of Kyung Hee University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Dissertation Committee:

Prof. Sung-Ho Bae

Prof. Seok-Won Lee

Prof. Sang-Ho Lee

Prof. Seong Tae Kim

Prof. TaeChoong Chung

Prof. Sungyoung Lee

I dedicate my thesis to the dreams and aspirations of my beloved parents, to my sister for her much-needed wise counsel, and my wife for her persistent support through trying times, they all played a pivotal role by encouraging and supporting me in all hard times to keep me in the position to complete my Ph.D. degree.

Acknowledgement

Allahumma salli 'ala Muhammad wa Ali Muhammad.

(O Allah! I invoke You to bless Muhammad and the Progeny of Muhammad).

I am eternally indebted to Allah Almighty for blessing me and my family with his grace. He gave me the strength, courage, and patience to honorably complete my doctoral dissertation.

I would especially like to extend my gratitude to my respected advisors, Professor Sungyoung Lee and Professor TaeChoong Chung for giving me the opportunity to complete the doctoral studies under their able guidance, support, and encouragement throughout the ebbs and flows of this journey.

I appreciate my dissertation evaluation committee for their valuable feedback and insightful recommendations during the dissertation defense. Their input enhanced the overall quality of the presentation and contents of the dissertation.

I would like to thank Mrs. Kim, Dr. Tae Ho Hur, and Dr. Jaehun Bang for helping me out with all sorts of administrative responsibilities despite their busy schedule, and always extending their support whenever I needed it.

I would like to acknowledge my colleagues and seniors, especially Mr. Zain Abbas, Dr. Wajahat Ali Khan, Dr. Hafiz Syed Muhammad Bilal, Dr. Jamil Hussain, Dr. Taqdir Ali, Dr. Muhammad Afzal, Dr. Maqbool Hussain, Dr. Bilal Amin, Dr. Shujaat Husasain, Mr. Muhammad Sadiq, Dr. Musarrat Hussain, Mr. Fahad Ahmed Satti, Mr. Ubaid Ur Rehman, Dr. Muhammad Asif Razzaq, Dr. Usman Akhtar, Dr. Rehman Ali, Dr. Maqbool Ali, Dr.Hua-Cam Hao, Dr. Tri, Mr. Asim Hussain, Dr. Ahsan Raza Kazmi, Mr. Muhammad Zaki, Mr. Anees Ul Hassan, Dr. Kifayat Ullah, Dr. Saeed Ullah, Dr. Aftab Alam, Dr. Jawad Khan, Dr. Omar Farooq, Mr. Abdul Sattar, and Dr. Thien Huynh The.

I am thankful to my MS advisor Professor Waseem Shahzad in NUCES, Islamabad for his guidance that led me on the path to machine learning.

In the end, I would like to express my sincere gratitude to my family and relatives. My dear uncles Mr. Muhamamd Husnain and Mr. Muhamamd Saqlain, and my elder cousins Mr. Nayyar Kazim, Mr. Irfan Kazim, and Appi, and a series of younger cousins including Mehr Zahra and others, for their unwavering moral support and heartfelt prayers throughout this journey.

Last but not the least, my sister Syeda Malika Zahra and her husband Mr. Imran Haider for taking care of our parents in my absence. I am also grateful to my in-laws for their unwavering support, patience and persistent prayers without which surmounting this journey would have been extremely unpleasant.

Syed Imran Ali

August, 2022

Abstract

Clinical decision support system (CDSS) represents the latest technological transformation in healthcare for assisting clinicians in complex decision-making. Several CDSS are proposed to deal with a range of clinical tasks such as disease diagnosis, prescription management, and medication ordering. Although a few CDSSs have focused on treatment selection, medication selection along with dosing selection remained an under-researched area. In this regard, this study presents one of the first studies in which a CDSS is proposed for clinicians who manage patients with end-stage renal disease undergoing maintenance hemodialysis, almost all of whom have some manifestation of chronic kidney disease-mineral and bone disorder (CKD-MBD). In this thesis, we have endeavored to complement both consensus-based domain knowledge (reflecting general framework provided in clinical practice guidelines) and experiential knowledge of domain experts (that can be acquired from clinical cases). Clinical Practice Guidelines provide a general framework to guide clinicians but lack operational details. In this regard, this thesis addresses the problem of how to complement case-base and domain model in order to generate domain compliant complex recommendation generation that reflect both clinicians' consensus-based training as well as clinician's experience?

The primary objective of the system is to aid clinicians in dosage prescription by levering medical domain knowledge as well existing practices. The proposed CDSS is evaluated with a real-world hemodialysis patient dataset acquired from Kyung Hee University Hospital, South Korea. Our evaluation demonstrates overall high compliance based on the concordance metric between the proposed CKD-MBD CDSS recommendations and the routine clinical practice. The concordance rate of overall medication dosing selection is 78.27%. Furthermore, the usability aspects of the system are also evaluated through the User Experience Questionnaire method to

highlight the appealing aspects of the system for clinicians. The overall user experience dimension scores for pragmatic, hedonic, and attractiveness are 1.53, 1.48, and 1.41, respectively. The service reliability for the Cronbach's alpha coefficient greater than 0.7 is achieved using the proposed system whereas the dependability coefficient of the value 0.84 revealed a significant effect. The proposed CDSS serves as a valuable tool in selecting appropriate treatment regimens based on domain knowledge and past experiences. It also helps in reducing the cognitive load of clinicians at the point of care. Furthermore, it can also be used as an educational resource for training purposes.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The main focus on this dissertation is to propose a hybrid knowledge modeling framework through which both explicit knowledge (domain model) and implicit knowledge (cases) can be used in a complementary manner for complex recommendation generation tasks. In this regard, chronic kidney disease - mineral and bone disorder (CKD-MBD) treatment case study is selected for demonstrating the efficacy of the proposed approach. Moreover, the proposed approach is used for developing a clinical decision support system (CDSS) that is thoroughly evaluated from both system-centric and user-centric perspectives.

1.1 Background

Clinical decision support systems (CDSSs) play an important role in enhancing the overall capabilities of healthcare providers [1, 2]. In a rapidly changing healthcare landscape, CDSSs are emerging as inevitable applications for informed decision-making. CDSSs are software application systems that provide time-critical, valuable, and relevant information to doctors, paramedical staff, and patients in order to help them deal with complex medical cases. CDSSs are sophisticated systems that encompass a variety of tools, such as alerts/reminders to the caregivers, knowledge extraction from clinical guidelines, medication and test ordering, diagnosis automation, and prescription management [3].

1.1.1 Benefits of Clinical Decision Support System (CDSS)

CDSSs provides assistance to medical practitioners in critical decision-making pertaining to patients' medical situations. The adoption of computerized systems in healthcare has a positive impact on the quality of patient care. In this regard, CDSSs provide medical practitioners with the

necessary support in terms of pointers to relevant domain knowledge, highlighting relevant patient information, and decision support to deal with complex situations requiring expert intervention [4]. Although most CDSSs are geared towards medical professionals, some systems also provide support to the patients in terms of education and awareness regarding their medical condition.

It is widely indicated in the literature that CDSSs can positively impact the overall quality of healthcare by leveraging state-of-the-art technologies which result in effective and efficient decision management without hindering the established clinical/healthcare workflows. In this regard, it is of the utmost importance that CDSSs provide services without becoming overtly bothersome to the clinicians [5,6]. Therefore, the usability aspects of CDSSs are also an important consideration. The application of these systems can be justified on the basis of their impact on the following: increased quality of service, reliable and transparent decision support, real-time situational awareness, enhanced health outcomes, user satisfaction e.g., of healthcare personnel and/or patients, and time-saving [7]. In this study, we designed a CDSS to assist the management of chronic kidney disease–mineral and bone disorder (CKD–MBD) in patients undergoing maintenance hemodialysis. The kidneys keep blood levels of electrolytes including calcium and phosphate within the normal range by finely handling their urinary excretion. Dysregulation of serum calcium, phosphate, and parathyroid hormone (PTH) begins far before reaching end-stage renal disease, even when kidney function is declined by half [8]. Importantly, biochemical abnormalities are closely interrelated to altered bone turnover and mineralization, and vascular calcification, leading to fractures [9, 10] and cardiovascular disease [11], both of which are serious and highly prevalent morbidities in dialysis patients. CKD–MBD is an insidious pathological process complicated in CKD that encompasses biochemical abnormalities, bone abnormalities, and vascular calcification.

Kidney Disease: Improving Global Outcomes (KDIGO) guidelines recommend serial assessments of serum phosphorus, calcium, and PTH and present their target ranges [12]. However, these three key laboratory values are hard to control within the target range at the same time despite technical advancements in dialysis-related apparatus and the introduction of new medications. Indeed, more than half of patients on dialysis do not achieve recommended target ranges of serum phosphorus, calcium, and PTH levels [13]. Since maintaining serum phosphate, calcium, and PTH within target ranges may reduce cardiovascular events and mortality [14, 15] in ESRD

patients, their optimal control is of paramount importance. One of the major barriers to correct laboratory abnormalities is that the medication prescribed to control one parameter may cause other parameters to fall out of the target range. To help clinicians prescribe the best set of medications, we developed a computerized decision support system which provides recommendations mostly regarding medication adjustment based on domain knowledge and past patient cases. A generic process flow for the CKD–MBD evaluation and treatment is presented in Figure 1.1.

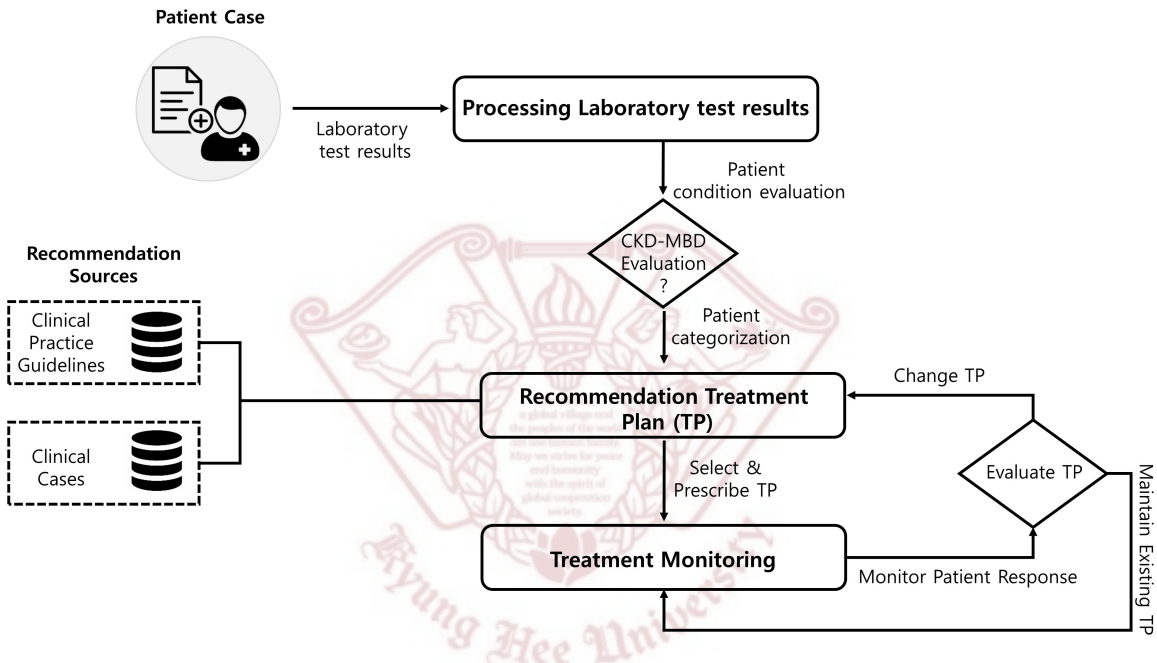


Figure 1.1: A generic process flow for the chronic kidney disease–mineral and bone disorder (CKD–MBD) treatment regimen selection.

1.1.2 CDSS for CKD-MBD Treatment

As CKD–MBD is not a single disease entity but encompasses a variety of altered mineral and bone metabolisms, a patient is initially evaluated for CKD–MBD by measuring serum calcium, phosphate, and PTH levels, and examining ectopic calcification with lateral abdominal radiography and echocardiography. Then, considering CKD–MBD status and associated clinical situations, an appropriate treatment plan consisting of dietary modification and medications is established. The objective of the proposed CKD–MBD CDSS is to assist clinicians in the selection of an appropriate

treatment regimen, i.e., medication selection and dosage recommendations for the management of CKD–MBD in ESRD patients. In this regard, we have focused on the expert-in-the-loop approach, i.e., clinicians provide essential domain knowledge for decision modeling and recommendation generation.

1.2 Case-based Reasoning (CBR)

1.2.1 An overview of CBR

Case-based reasoning (CBR) provides a methodology for solving computational problems through analogy with past solved cases [16, 17]. CRB can be applied to an array of different computational problems hence it is versatile and its recommendations are inherently interpretable by the domain expert [17, 18]. The main approach adopted in CBR is to model past successful experiences through cases. A case captures a problem component and a solution component with an optional outcome component as well [17, 19]. The problem component specifies operating conditions under which a solution is required. The solution component deals with providing solutions that resolve the given problem [20]. The outcome component is useful when success of a solution can be quantified and supplied along with the proposed solution for subsequent reuse for a similar problem. In CRB, the reasoning is applied on the basis of cases and similar cases are selected with an operating assumption that the cases which are similar to each other would have similar solutions [20]. Unlike record retrieval operational in database where given a formal query an exact solution is provided that is already available in the database. In CBR, it is not assumed that exact solution should be already available in the case-base, therefore, an approximate solution may be selected that can be further fine-tuned on the basis of peculiar details of the test case [21]. The similar selected case should be relevant, therefore, the notion of similarity is an important consideration in CBR. In this regard, CBR provides flexible solutions due to employed approximate reasoning. Furthermore, the CBR methodology can be applied to different kinds of data such as tabular data, images, textual data, etc [16]. In this thesis, we have used CRB for tabular data that is comprised of patients' measurements of laboratory test results along with prescribed medications. The outcome component is added to each case based on each treatments efficacy as recorded by the system.

1.2.2 Past events as Cases

The CRB methodology encodes past events as cases. Past events result in both positive experiences and negative experiences in terms of the efficacy of their solutions for given problems [17]. Generally, those cases are deemed relevant that have had resulted in positive experiences. But it is also argued that similar cases with negative experience enrich the case-base and point towards potential choices that needs to be avoided [17, 22]. The problem of encoding cases with most important aspects of past experiences is a non-trivial task and should be performed carefully. In this regard, the task of feature selection also plays an important role as it identifies a set of suitable features that have statistically meaningful impact on the decision [17]. Moreover, the domain expert may also be consulted to identify the core aspects of their past experiences so as to model CBR as closely to the operational conditions as possible. In Figure 1.2 a generic CBR framework is depicted that is comprised of a number of steps. A new case essentially specifies a computational problem to be solved, based on the problem represented in the form of a case, a set of similar cases are retrieved from a case-base. This case-base serves as a memory where different cases are stored. The case-base has an important role in the CBR as it specifies the problem and solution space for the application.

1.2.3 Components of a Case

A case representation can take many forms, in this thesis, we have selected a feature-based representation as the required source data is already stored in an structured form within hospital's information management system. As earlier mentioned, a case is composed of problem and solution components whereas the problem component may be characterized by a set of features that encompass the necessary aspect of the problem to be solved [23–26]. In the treatment for CKD-MBD patients, the scope of the problem deals with patients' laboratory test results. Therefore, with consultation with domain experts a set of medical tests are identified that are followed under operational protocols for treatment prescription by physicians. Each medical test is encoded as a single feature. Moreover, age and gender of the patients are also made part of the problem component. In this regard, the biographical information and medical test results form the problem component for our CKD-MBD application. The solution parts is encompass a multi-attribute rec-

ommendation in which a set of different medications may be prescribed to a patient. Hence, each feature in the solution component represent a pre-specified medication and its dosage. For such cases, where a certain medication is not prescribed, the dosage of all unprescribed medication is set to 0. As aforementioned, a case within CRB methodology can be divided as follows:

- A problem component that provides necessary information e.g. encoded as a set of features
- A solution component that may be comprised of a single feature e.g. disease diagnosis, or a set of features e.g. medication prescription, that charts a recommended solution for a given problem.
- An optional outcome component that quantifies the efficacy of the solution through a subsequent feedback

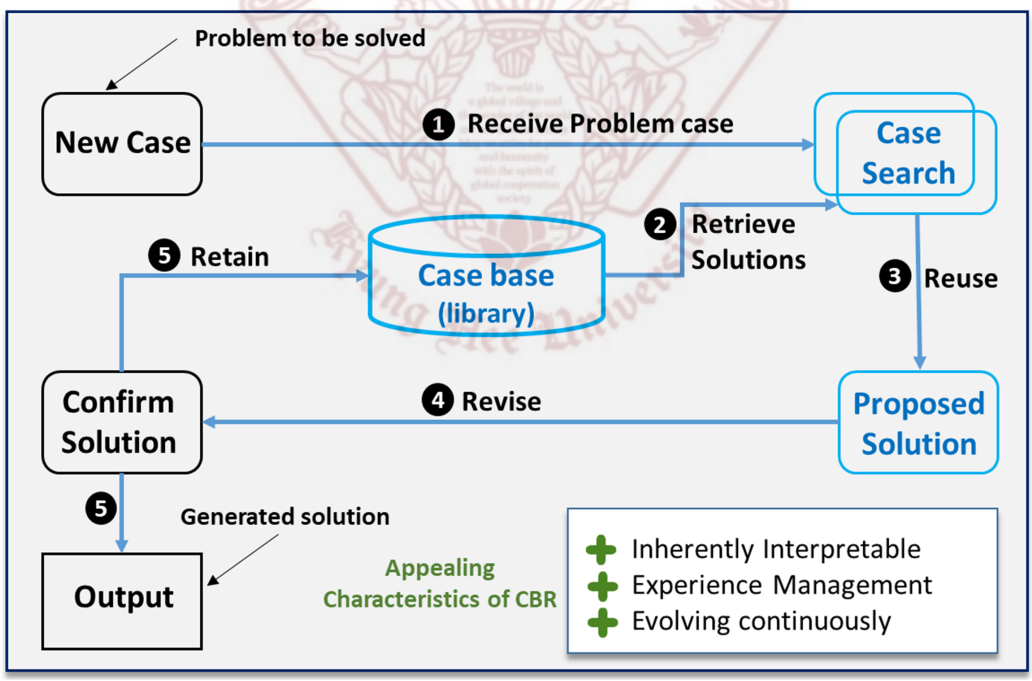


Figure 1.2: Case-based Reasoning (CBR) schematic representation.

Case-base can consists of multiple cases, carefully selected a reference for future reuse [17]. The initial set of cases can be curated with the help of domain experts [21]. Moreover, as mentioned earlier, both positive experience and negative experience may be encoded into cases for

further reuse. In this regard, a recommendation may be comprised for suggestion pertaining to things advised to do and things not to do. So long a case is used for forming a recommendation, it is deemed as a relevant case and therefore a methodology maybe devised in which both positive experiences and negative experiences can be used in tandem for recommendation generation [17].

When positive and negative cases occur one can introduce two sets of cases: C^+ (positive) and C^- (negative) cases. Negative cases occur often in the context of decision making where one has to choose from different alternatives or when advice has to be given.

A CRB framework can be adopated for genreating a recommendation for a wide variety of scanarios, such as:

- Diagnosis: To perform screening of patients and divide patients who are susceptible of disease and those that are deemed otherwise.
- Prediction: Given the current health condition of a patient and treatment, the system may provide prognosis for the wellbeing of patient over a period of time e.g. 5 years outlook of survivability for a cancer patient undergoing chemotherapy.
- Planning: Device treatment regimens for a patient based on successful results of patients having similar conditions.
- Physician selection: Given a set of patient's signs and symptoms, recommend a suitable physician for further consultation.

1.2.4 Case Representation

The case representation is one of the most important consideration in designing a CBR solution [27]. Case representation deals with explicitly stating different aspects of problem and solution components so that similar cases can be searched and reasoning operation can be performed in order to generate a solution [27]. In this regard, the necessary context of the encoded experience also need to be recorded. Generally, a case is represented as feature-value pairs [28]. A feature represents an entity or a quantifiable concept that can take on a number of distinct values as applicable in the domain such as color of skin. In this example, the skin color is a concept that is associated with a set of several possible distinct values, restricted by the application domain [29].

Therefore, case represent should closely model concepts and their corresponding values for a given application domain. Moreover, explicitly characterizing different features also specify what kind of values these features can take on e.g. a numeric value may not be applicable for a feature expecting a nominal value [29–31]. Please note that in literature, the word feature and attribute are used interchangeably. Therefore, we may also switch between these words. A complete case, may represent one encounter of patient or multiple encounters can be accounted into a single case. Therefore, it is also important to specify the scope of a case. In this thesis, one encounter of a patient, i.e. a single visit, constitutes a case. Table 1.1 depicts one case with feature-value pairs for problem and solution components. As it can be seen that feature value representations are, in fact, just an attribute-value vector. More precisely, it can be stated as follows:

Definition 1 For a given set U of objects, an attribute A assigns to each object O $\in$ U some value taken from a set dom(A), the domain of A.

1.2.5 Case Base for case storage

The case-base serves as a collective memory of the recommendation system. Generally, cases are placed in the case-base in an unordered manner [16]. It is effectively a repository of cases that define the scope of problems that can be effectively resolved with a set of solutions already available in the case base. Adapting a solution from existing solutions retrieved from the case-base is not handled by the case base [32].

Definition 2 A *case base* is a collection of cases.

The major consideration in designing a case-base is to effectively retrieve similar cases, therefore, indexing plays an important role in case base designing. Moreover, based on the nature of applicable, case-base may contain cases in a flat file structure, unstructured textual documents or images [16, 33]. In this thesis we are dealing with heterogeneous data such as laboratory test results, biological information, and medication dosages, therefore, a flat file structure is selected for case representation.

In this thesis we are dealing with heterogeneous data such as laboratory test results, biological information, and medication dosages, therefore, a flat file structure is selected for case representation.

Table 1.1: A sample case specifying problem and solution components

Case Attributes	PTH	Albumin	Calcuim	Phosphate	Calcification	Cinacalcet	ClacitriolPro	CalcitriolIV	Paricalcitol	Calcium Carbonate	Calcium Acetate	Sevelamer	Lanthanum
Sample Case	641	4	8.4	8.6	Yes	24	0	0	10	0	1000	3200	0
Case Components	Problem Component (measurements for laboratory test results)					Solution Component (medication dosages that can be prescribed by the physician)							

1.2.6 Case Retrieval

The CBR is a passive modeling approach i.e. once an input case is received then the case-base is searched for relevant candidate solutions. Therefore, the case retrieval strategy is of profound importance in the CBR methodology [16, 34–36]. The notion of case similarity may differ with respect to the nature of cases stored i.e. in case of tabular, the feature similarity may play an important role in defining the notion of similarity whereas for the problem of content-based image retrieval the notion of similarity would be based on different characteristics of stored images that are not pre-coded into features. Moreover, the selection of similarity measure is also informed by the nature of features i.e. nominal features, numeric features, or a mixed set of features [16]. The objective of case retrieval task is to find a solution that is most suitable candidate for a given problem. Since, no explicit model is available that be directly be used for generating a suitable solution, therefore, all the cases stored in the case-base are traversed in order to specify the context through which a single relevant solution or a set of relevant solutions may be selected [37–39]. In order to define an ad-hoc context of relevance, the similarity of features present in the problem component of the new case are matched with that of features stored in the case-base. As a convention, if two cases are similar then the similarity score is 1 else it is 0 for the given feature and then the accumulated score of all the features is taken as a proximity of an stored case with that of the new case [16, 37, 40]. Once a set of stored cases are identified, they define a solution scope in which a set of candidate solutions can be further refined or provided directly to the decision maker.

Definition 3 Let *CB* be a set of objects and *p* be an object; then some *s* of CB is a nearest neighbour to *p* if there is no object in the *CB* that has a higher similarity to *p* than *s*.

The nearest neighbor cases are deemed more suitable for providing solution to the given problem with an assumption that the solutions to similar problems can be reused. The notion of similarity can be extended from the domain knowledge perspective, as it is proposed in this dissertation.

1.2.7 Reuse and Adaptation

Once a set of candidate solutions are identified and retrieved then these can be used for generating solution to the new problem. The reuse of solution of previous cases mean that certain aspects of their solutions can be reused e.g. for a medication prescription problem, a set of common medicines selected from previous cases can be suggested to a physician for a treatment selection task [23, 25, 41]. For complex problems comprised of multiple-attributes of a recommendation, i.e. as medication prescription, the previous experiences provide a partial solution that is required to be incorporated along with external knowledge that may not be part of the case base [41]. It is rare that a retrieve solution is directly applicable to a new case, except from cases where the retrieved case is very similar to the new case. In general, the retrieved solutions are adapted within the context of the new case e.g. a patient maybe allergic to a certain medication, therefore, an alternative needs to be prescribed for the patient although the medication had positive effect on similar patients. The recommendation adaptation can be done either manually or automatically. In this thesis we propose a hybrid case adaptation technique that automatically incorporates two different models (i.e. on different levels of granularity) to generate a complex recommendation that is composed of multiple attributes.

In an abstract way the CBR problem-solving procedure (as shown in Figure 1.2) can be described by the following steps:

1. First explicitly characterize past experiences through encoding them into cases
2. Select a similarity assessment technique that can efficiently compare new case with stored cases in the case-base
3. Select relevant cases (single case or a set of cases) that can provide a solution for the given new case
4. Adapt the retrieved solutions to the specific needs of the new case

Unlike database retrieval where an exact solution is sought for a given query, in CBR the reasoning provides an approximate solution to a queried problem, therefore it doesn't guarantee the exact solution (as it may not be available in the case-base) [**?**, 42]. With the added flexibility

of selecting relevant solutions that can be adopted for the queried problem, the CBR is more closer to experience-based human thinking and it serves as a competitive approach to address an array of computational problems. Once an adapted solution is validated by the domain expert, then the new solution becomes part of the case-base in this manner the knowledge evolves and new knowledge is accumulated over time. It may give rise to problem of extended search time, once the case-base becomes large, and redundant or obsolete knowledge, these aspects are not within the scope of this thesis. Hence, the feedback mechanism provides flexibility to the CBR framework to include solutions that were not part of the initial case-base. It is one of the key advantages of CRB methodology in which through the feedback loop an incremental learning takes place. CBR is applied in a wide variety of applications such as energy optimization [43], injection molding manufacturing [44], recommendation of baking products [45], diagnosis of mobile phone faults [27], internet of things [46], and real state valuation [47], among others.

1.2.8 Medical Case-based Reasoning

This section mentions describes some of the typical applications of CBR in medical and healthcare domain. It is to indicate where CBR methodology can be of use and what aspects in each application are suitable for the methodology. In medical domain, disease diagnosis and prescription are very common, and the way they seem to be approached by physicians and nurses, diagnoses tend to be based on previous experiences and on evidences. These reasoning approaches are consistent with CBR [36,38,48]. CRB methodology is successfully applied to a number of medication applications such as supporting explainable artificial intelligence for breast cancer [49], risk prediction in surgery [50], and medication management [51], among others.

In medical applications there is an intrinsic connection between the diagnosis and prescription tasks [48]. CRB can be used to address both of these tasks or it can be used in a diagnosis and prescription management pipeline along with other modeling approaches [52–54]. Such a configuration can be considered as a hybrid approach. Another way to characterize a hybrid approach is when different modeling techniques are used in parallel to solve a common problem i.e. a single model fails to provide a complete high quality solution. In this thesis, we have opted for the latter case of hyrbdization, where two different models at different levels of granularity are combined

in such a manner that a domain-compliant complex solution (multi-attribute recommendation) can be generated as shown in Table 1.1.

1.3 Role of hybrid approach in decision modeling

Machine learning model's for classification and regression tasks are based on inductive learning in which specific data instances are used to determine general outcomes e.g. benign or malignant tumor classification or dosage estimation for radiotherapy, etc. In this regard, the induced model is a generalization of the specific examples in the training dataset.

Different machine learning algorithms applied on the same training dataset generally result in different models due to the consideration of different inductive biases while learning the resultant model. The prior knowledge incorporated into the learning strategy of the model is called inductive bias that enables the model to select a specific set of functions [55].

For example in CBR the operating assumption is this that cases that are closer to each other in the problem space would also be close in the solution space. Therefore, a new case with an unknown solution (e.g. a label for a diagnosis problem) is placed in a close proximity to a set of cases that taken on a certain label e.g. benign tumor, then the test case may be labeled as a benign case as well. This is an example of inductive bias used in algorithms that source solutions from nearest neighbors.

In order to improve the robustness of the final output a set of multiple independent models are employed. The prediction that is common across all the models in the set is regarded as the final outcome. Therefore, the hybrid approach generally used in designing machine learning algorithms deals with configurations in which different models interact with each other i.e. bagging, boosting, stacking, etc. The output of each model is compatible with that of the other model in the solution space, therefore, through various aggregation operators e.g. weighted majority voting or mean can be directly applied to obtain the final outcome. One important distinction between the aforementioned model hybridization and the proposed hybridization approach is in the incompatibility of the model's outcome. The domain model provides a consensus-based generic recommendation that lacks operational details while the case-based reasoning provides detailed recommendation but may lack the compliance with the consensus recommendation. In this regard, the important

overall contribution of the proposed approach is hybrid both the models in such a manner that the resultant recommendation with compliant with the domain knowledge and also contains operational details to make the recommendation actionable.

Hybrid techniques along with case-based reasoning are proposed to deal with complex cases, e.g. [56] used case-base reasoning along with Bayesian reasoning for prescription recommendation, [57] used hybrid case-base maintenance strategies that take into account both addition and deletion of cases in the case repository. These approaches work under an assumption that the independent models provide comparable recommendations that can be directly compared and a final recommendation can be generated based on an aggregation operation. Our proposed approach deals with a novel situation where two models provide recommendation at different level of granularity and therefore can not be directly aggregated. In this case, recommendation from model 2 is adapted with respect to recommendation from model 1, essentially generating a sequential pipeline of case hybridization.

1.4 Motivation

DSSs have become ubiquitous in clinical settings. CDSS support in a wide variety of cognitive-intensive decision making tasks such as disease diagnosis and medication prescription. There are either expert-driven approaches or data-driven approaches to model knowledge required for medical decision making. In this thesis, we endeavor to synthesize both in a complimentary manner for complex decision making tasks.

Medical domain is characterized by complex interaction of multiple contributing factors and scarcity of data, where domain experts prefer automated decision support approaches that are domain compliant and have higher degree of transparency in the decision making process.

The overall motivation of the proposed approach is to provide both higher-level of domain compliance along with higher-degree of transparency in the decision making process. Following are some of the key benefits of the proposed methodology and system:

1. Align consensus-based domain knowledge and experience-based routine clinical practice for multi-factor recommendations.

2. Provide recommendation for complex cases with inherent data scarcity issues.
3. Provide transparency and knowledge-based interpretability for the domain expert in recommendation generation.
4. Reducing cognitive-load on the physician with intuitive and reliable decision support system.
5. Patient Improvement Indicator provides treatment efficacy assessment based on patient's physiological status after the treatment.
6. Valuable educational resource for medical trainees for handling complex CKD-MBD cases with evidence from clinical practice.
7. Preventing accidental dosage errors by generating appropriate alerts to the physician.

1.5 Problem Statement

Clinical cases provide us with operational details of the medication prescription. These operational details are in terms of what medicines are prescribed along with their dosages. These clinical cases reflect the subjective experience of particular clinicians. In order to generate recommendations for complex cases (i.e. multi-attribute recommendation), we need to know operational details of the recommendation as well as general consensus of the decision makers (i.e. a domain model)to deal with such cases as provided in clinical practice guidelines and other such clinical references.

The challenge faced in this regard is due to the incomparable level of granularity of information provided by the domain model and information provided in clinical cases. Domain knowledge acquired from clinical practice guidelines is in abstract form and for complex problems it results in partial models that are not expressive enough to provide a detailed recommendation such as medical dosing support. Clinical Practice Guidelines provide a general framework to guide clinicians but lack operational details. How to incorporate expert-model in case-based reasoning for domain compliant complex recommendation generation e.g. medication prescription?

- **Active case base partitioning:** A clinical case base is a homogeneous entity in while all cases are indistinguishable until a test case is provided. Therefore, at the stage of case

acquisition there is little guidance to specify what type of cases are required and whether the case-base at hand is capable of providing solutions to certain types of problems (e.g. rare events)?

- **Multi-level case selection:** Once the suitable partitions in the case base are identified a priori. All the cases within a partition have same degree of membership i.e. for a given test case, a partition is selected to provide a set of candidate cases, then all cases within that partition are retrieved to provide candidate solutions. It is important to note that cases within the partition denote both positive experience and negative experience. Without explicitly stated information about the outcome of the cases, how to utilize the longitudinal case data to infer the outcome associated with each case so that a final set of reference cases are selected?

- **Hybrid case-adaptation:** Domain model provides a general framework that is agreed upon by decision makers but for complex cases it lacks operational details. The required operational details are available in clinical cases but they reflect the subjective experience of a particular decision maker. In this regard, the major challenge is how to align knowledge provided by the domain model and supply operational details sourced from relevant clinical cases?

Specifically, the main aim of this thesis is to address the following challenges:

1. How to a priori identify distinct neighborhoods within a single clinical case base?
2. How to identify reference cases within a set of candidate case?
3. How to combine incomparable models providing knowledge at different levels of granularity for a consolidated recommendation generation?

1.6 Key Contributions

The key contributions of the thesis are as follows:

1. Leverage domain knowledge to actively partition the case base in an a priori manner for distinct neighborhood identification for case insufficiency detection.
2. Leverage case outcome information from the feedback for candidate solutions selection
3. Leverage domain knowledge with selected cases for recommendation generation e.g. medication dosage.
4. The proposed system is validated on real-world clinical cases of CKD-MBD patients.
5. Evaluation of usability aspects to demonstrate higher user experience utility for clinicians.

In order to address the aforementioned points, the study is designed in a collaborative manner in which both domain experts (i.e., clinicians) and knowledge engineers work in tandem to realize the proposed CDSS. The domain experts provide the relevant domain knowledge, which is in turn modeled and enhanced by knowledge engineers. Most of the CDSSs in the domain of CKD are based on black-box machine learning models [58–60]. Systems built on such models are generally applied for diagnostic applications where the system provides a prediction along with a confidence score [61, 62]. Although black-box models generally exhibit higher accuracy, such data-driven models have limited utility due to sparsity of data, such as in the case of medication intake where a small subset of medications are prescribed more frequently than the others. Therefore, in this thesis, we have focused on hybrid knowledge modeling to combine expert knowledge with that of clinical cases of patients for decision support in medication recommendations. The abstract idea of the proposed case-based hybridization approach is depicted in Figure 1.4. The proposed approach is comprised of three major operations, i.e., case-base partitioning, case selection, and case adaptation. The major emphasis of the proposed approach is to synthesize abstract domain knowledge with specific domain cases in order to generate a comprehensive recommendation for a complex scenario.

The main contribution of the study is a proposed hybrid methodology that combines both explicit knowledge (i.e., acquired from domain experts in the form of a partial domain model) and implicit knowledge (i.e., in the form of clinical cases) for complex multi-factor recommendations. Therefore, the medication and dosing selection for CKD–MBD patients is adopted as a case study. The proposed approach is based on the CBR framework, which imitates a clinician's thinking and

attempts to solve new problems by reusing solutions that have been used to address similar problems in the past. CBR works with specific cases from past scenarios and adapts the outcomes and experiences to an unseen problem. The greater interpretability of the recommendation is a key benefit of the CBR framework. Therefore, clinicians can easily evaluate the CDSS recommendation and follow the line of reasoning followed by the system. Major differences in this case from a conventional CBR are the development of a domain model and the leveraging of it for case-base partitioning. The case selection is further refined using reference case selection using a Patient Improvement Indicator (PII). A hybrid approach is used for case adaptation using domain based rules and statistical techniques, such as interquartile range. The domain model only partially captures the solution component (i.e., only medication selection is covered by the domain model), we therefore demonstrate how to utilize the partial domain model in conjunction with clinical cases for medication dosage selection. Figure 1.3 summarizes key challenges and their proposed solutions. In this regard, both the domain model and clinical cases are employed in a complementary manner through the hybridization pipeline proposed in this thesis. The proposed pipeline can be applied to any other medical treatment domains that include medication prescription and dosing adjustment.

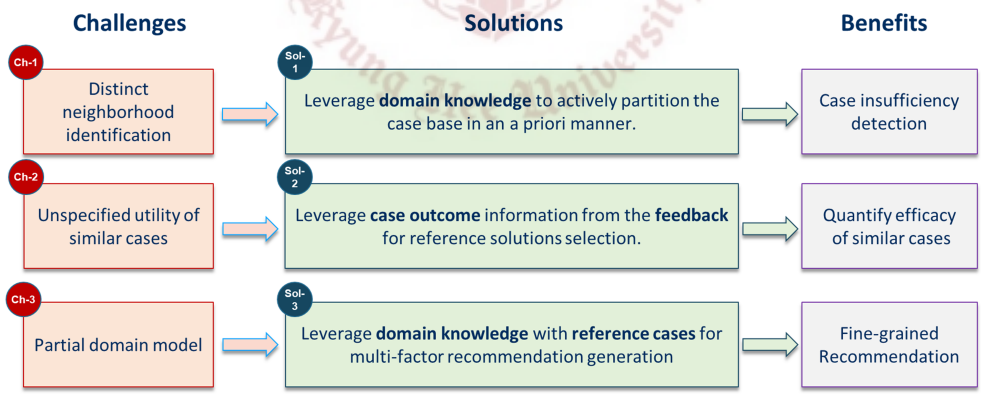


Figure 1.3: Summarized challenges along with proposed solutions and their benefits

1.7 Thesis Organization

This dissertation investigates the syntheize between domain knowledge and data-driven approaches i.e. case-based reasoning for complex recommendation tasks. More specifically, the aim of this research is to explore approaches that can leverage the expert-based partial domain knowledge model into case-base partitioning, case selection and case adaptation tasks.

This dissertation is organized into following chapters:

- **Chapter 1: Introduction.** In this chapter we provide the overview of the role of decision support systems and their importance in clinical decision making tasks. Moreover, we provide the overall motivation of the thesis along with the key contributions of the proposed approach.
- **Chapter 2: Related Work.** In this chapter, we survey the related work in the domain of expert-based knowledge acquisition approaches for medical decision making tasks along with related work on medical prescription methodologies employing both expert-driven and data-driven approaches.
- **Chapter 3: Proposed Methodology.** In this chapter the focus on the acquiring domain knowledge from experts and utilizing the acquired decision model for active-based partitioning, case, case selection and case adaptation. Moreover, this chapter also covers automated knowledge generation through feature selection.
- **Chapter 4: Experiments and Discussion.** This chapter provides details of two types of experimentation i.e. system-centric and user-centric. Moreover, the relevance of the results for clinical decision making are also underlined in this chapter.
- **Chapter 5: Conclusion and Future Directions.** This chapter concludes the dissertation along with identifying future research directions for extending the hybrid knowledge modeling research.

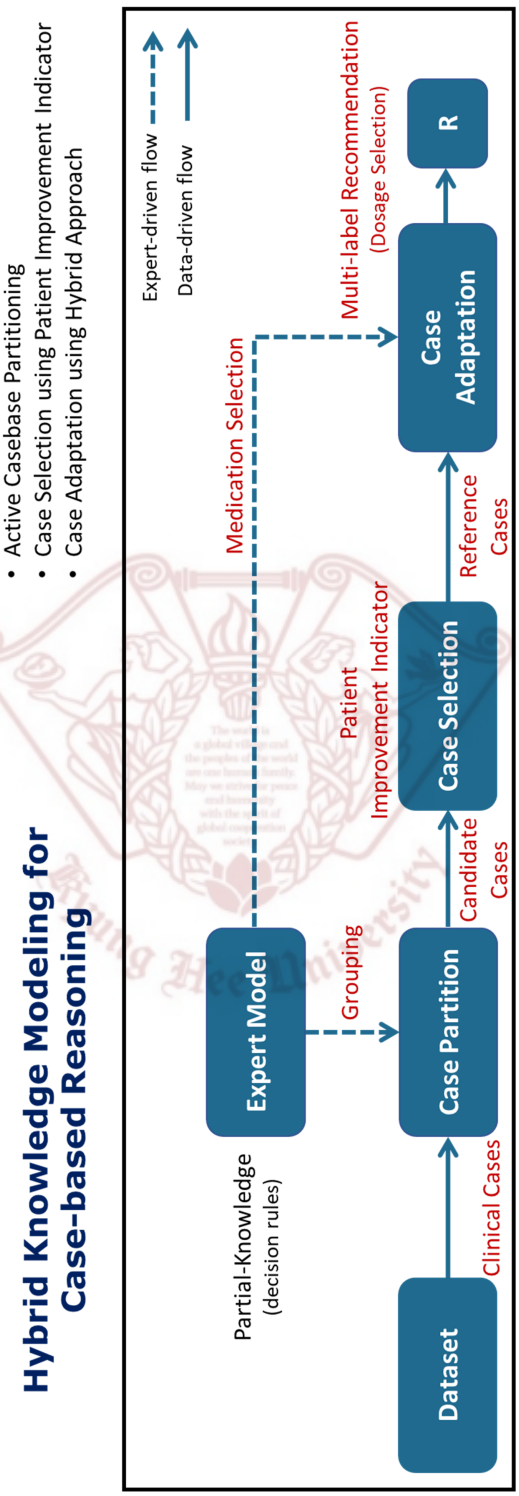


Figure 1.4: Abstract diagram depicting the role of domain knowledge and clinical cases in the proposed approach.

Chapter 2

Related Work

2.1 Chapter Overview

Medication management is a tedious and error-prone task for both clinicians and patients. Deep learning-based approaches are generally employed for processing unstructured data, such as medication images and clinical texts, for the purpose of correctly identifying medication information. To reduce medication identification errors by the patients, deep learning-based techniques are leveraged that aid in prescription pill identification from mobile images [63–67]. Similarly, deep learning techniques are also successfully applied to the task of medication and dosage extraction from clinical texts, such as clinical notes [68–70] and social media texts [71–73]. Moreover, some studies have explored deep learning applications for medication selection focusing on drug–drug interaction [74], dosage selection from free clinical text processing published literature [33] [75], and electronic health records [76], selecting discharge medications based on patient information documented in admission notes [77], among other sources. The aforementioned approaches benefited from training models on a huge amount of data and/or leveraging pre-trained models already available for similar tasks. However, one major hurdle that limits the application of black-box models in clinical practice is the lack of the interpretability of these approaches [78–80]. Research taxonomy of the proposed hybrid case-based approach is provided in Figure 2.1.

Alternatively, the proposed hybrid case-based approach provides interpretable medication selection and dosage adjustment recommendations given the small amount of clinical data with reasonably acceptable accuracy. Therefore, in this section, we focus on those aspects of the CDSS that are within the scope of the proposed methodology, such as expert knowledge acquisition, medication selection and dosing adjustment.

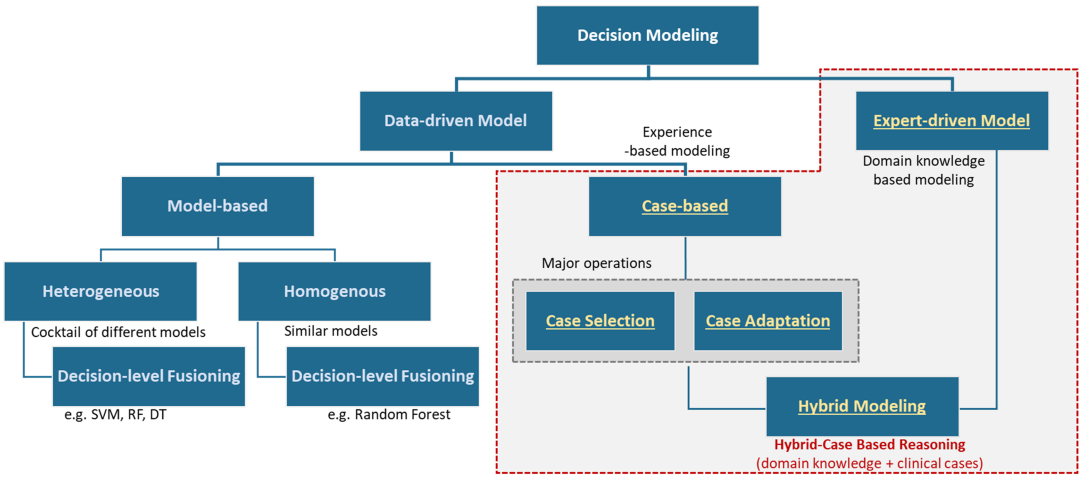


Figure 2.1: Research Taxonomy - Hybrid knowledge modeling for case selection and adaptation

2.2 Hybrid-Case Based Reasoning: Knowledge Acquisition

A mind-map-based knowledge acquisition process is proposed by Yu et al. [81] for the treatment of thyroid nodules. The authors proposed a consultative process between domain experts and knowledge engineers in which a domain model is produced. A number of clinical practice guidelines (CPGs) pertaining to the thyroid nodule treatment are analyzed by the domain experts and, subsequently, an iterative decision tree (DT) model is generated by the knowledge engineers for automating the decision-making process. The CDSS was evaluated using retrospective medical records of 483 patients. The authors reported 78.9% concordance between the CDSS recommendations and routine clinical practice. A similar modeling approach is adapted by Choi et al. [82], in which a CDSS for heart failure diagnosis is proposed. The authors proposed a hybrid knowledge modeling approach in which both expert-driven and data-driven models are consolidated into a single model. In this regard, the Classification and Regression Tree (CART) model is used to build a decision tree from patients' medical records. Moreover, the resulting model is combined with the DT model produced by the domain expert. The authors reported higher accuracy of the combined model as compared with both the expert-driven model and the data-driven model. Hussain et al. [83] proposed a knowledge validation and verification approach for such cases when multiple stakeholders are involved in the knowledge modeling process and diverse sources are consulted. A hybrid approach was used which consists of both the domain expert

knowledge as well as patients' medical records. The resulting CDSS is used for the treatment of oral cavity cancer patients. It was observed that knowledge verification is an important aspect of expert-based knowledge modeling to address issues raised due to various inconsistencies, e.g., non-standard terminologies. The authors evaluated four different knowledge acquisition scenarios and reported higher classification accuracy for the hybrid approach with formal knowledge verification. A knowledge based CDSS was proposed by Afzal et al. [84] for treating cancer patients. In this regard, an automated knowledge acquisition approach was proposed to acquire relevant data from head and neck cancer patients' unstructured documents. Finally, a CART model was used for treatment regimen prediction. The authors reported 69.0% accuracy in correctly selecting the treatment recommendation with respect to routine clinical practice. Bach et al. [85] proposed a clinical dashboard to facilitate co-decision making in the management of non-specific low back pain patients. The system collects data from questionnaires and wearable devices to make predictions about the course of non-specific low back pain treatment. A case-based approach is used to provide personalized recommendations for patients. The knowledge acquisition and recommendation process are primarily based on pain management guidelines, consultation with clinicians, and past patient cases. Ali et al. [86] proposed a multi-modal-based interactive authoring environment for expert knowledge acquisition that is also shareable. A case study on oral cavity lesions treatment plan generation was presented in which expert-based knowledge in the form of a mind-map was converted into a set of medical logic modules. In this study, the authors attempted to automate the process for shareable knowledge creation in a user-friendly manner. The proposed system was evaluated from both system-oriented and user-oriented aspects. Wit et al. [87] evaluated clinical rules in a standalone pharmacy-based CDSS for hospitalized and nursing home patients. The authors investigated the utility of clinical rules for reducing prescription errors. The knowledge acquisition process for creating clinical rules was based on guidelines that are developed by both pharmacists and physicians. The main objective of the study was to evaluate the clinical significance of automated alerts in routine clinical practice. In this regard, the relevance was determined whether or not the pharmaceutics contacted the physician for each alert. The authors reported that the average efficiency of the CDSS was low, whereas a few clinical rules have an efficiency of greater than 10%. A number of factors contributed to the low efficiency of the system, such as

alert fatigue and the daily recurrence of previously evaluated alerts, etc.

The aforementioned studies performed knowledge acquisition for developing the expert-driven model. The main focus was on the completeness of the model, i.e., the developed expert knowledge model is sufficient for providing a recommendation for a given task, e.g., treatment selection. The main advantage of the proposed hybridization approach is that it can synthesize partial domain models, i.e., the domain model is used only to provide abstract level generic recommendations. The abstract recommendation is refined using clinical cases that are relatively easy to acquire as compared to rigorously constructing a detailed domain model for complex recommendations such as medication selection and dosing adjustment. Moreover, the proposed approach of synthesizing a partial domain model with clinical cases is also more practical where codified domain knowledge, e.g., clinical practice guidelines, are not sufficiently available for the task at hand. For example, in the domain of CKD–MBD management, the leading guidelines do not provide a detailed recommendation model for dealing with medication selection and dosing prescription. Therefore, general pointers are extracted through domain experts to construct a generic model as per the recommendations of the guideline, while the specialization of the recommendation is aided through the clinical practice of the clinicians in the form of specific cases. Therefore, the main role of the proposed approach is to combine the abstract domain model with that of specific clinical cases for final multi-factor recommendation generation. Table 2.1 provides a summarized comparison of related techniques for knowledge acquisition for domain model construction.

2.3 Hybrid-Case Based Reasoning: Medication Selection

A CDSS for the management of CKD–MBD in patients with ESRD who receive maintenance hemodialysis has the potential to improve different stages of prescription, such as medication initiation, modification, monitoring, or discontinuation [56, 88, 89]. Furthermore, it is reported that the usage of CDSSs improves overall adherence to clinical practice guidelines and streamlines the decision-making process of clinicians [90]. Vogel et al. [91] compared the effectiveness of an outpatient renal dose adjustment alert through a computerized provider order entry (CPOE) CDSS and a CDSS providing alerts to pharmacists. The authors concluded that both types of CDSS resulted in low rates of potential medication errors. In prescriber-based CDSSs, a pre-defined pro-

Table 2.1: Literature comparison of related knowledge acquisition techniques for clinical decision support systems

Reference	Area of Application	Objective	Characteristics	Limitations
[39]	Thyroid nodules	Treatment	* Knowledge-based system modeling * Expert driven domain model * Retrospective evaluation	*Complete knowledge model is difficult to acquire * Model evolution requires domain expert involvement
[40]	Heart disease	Diagnosis	*Hybrid knowledge model * Interpretable decision making * Retrospective and pilot study	*Difficult to express domain consensus for complex cases * Combined model is prone to overfitting
[41]	Oral cavity cancer	Diagnosis	*Hybrid knowledge model * Model consistency evaluation through formal methods * Retrospective evaluation	*Complete domain model is difficult to acquire for complex decision tasks *Combined model is prone to overfitting
[42]	Head and neck cancer	Diagnosis	*Automated knowledge acquisition from documents * Interpretable decision model * Offline and online evaluation	*Domain expert involvement is required for data quality validation * Resulting model does not incorporate domain knowledge that is not reflected in selected data
[43]	Low back pain	Treatment	*Co-decision making model * Implicit knowledge modeling through case-based framework * Reference group selection based on positive outcome	*Clinical guidelines are not integral part of the case-based model * Data acquisition through wearable devices is unreliable, and self-reporting data are subjective
[44]	General healthcare	Wellness management	*Framework for domain model enrichment * Wellness concept model for health management * Model evaluated using nominal group technique	*Only SNOMED CT is used for standard terminology harmonization * Model evolution requires domain expert involvement
[45]	Standard medical care	Treatment	*CDSS based on clinical rules for pharmacy application * Automated alerts for prescription error reduction * Retrospective evaluation	*Difficult to express domain consensus for complex cases * Model evolution requires domain expert involvement

cess map is used to aid in the decision-making of medication prescription and dosing. Hellden et al. [92] evaluated the impact of a CDSS on the general practitioners' (GPs) experience of drug dosing. The information-gathering process included a questionnaire and a focus group discussion. The study presented favorable evaluations by GPs in terms of ease of use and overall usefulness in medication dosing. Furthermore, primary care physicians have reported higher acceptance of simple graphical user interfaces, along with task-oriented clear navigation and concise advice. Pirnejad et al. [93] proposed a methodology for appropriate drug therapy recommendations for kidney transplant patients based on clinical knowledge as well as international recommendations. Niazkhani et al. [94] proposed a context-aware CDSS for managing drug-laboratory interactions in order to reduce medication errors. The main focus of the study was to develop a user-friendly CDSS to accommodate drug-laboratory interactions (DLIs) while reducing the alert fatigue of clinicians. The knowledge base was based on DLI-rules that were extracted from pharmacology references and clinicians' direct input. The overall efficacy of the system was evaluated using the "Questionnaire for User Interface Satisfaction". Shemeikka et al. [95] proposed a CDSS to support prescriptions of pharmaceutical drugs in patients with reduced renal function. The proposed system was integrated with an electronic health record system (EHR) used in both hospitals and outpatient facilitates. The evaluation of the system was based on a usability questionnaire and the frequency of system logging. The main focus of this research was to integrate the CDSS in the Janus toolbar for appropriate drugs therapy recommendations. Awdishu et al. [96] proposed a CDSS for supporting medication prescription for CKD patients. The system targeted 20 medications and aided clinicians in the drug therapy discontinuation or dosage adjustment for adult patients with impaired renal function. Medication alterations were based on reviewing primary literature and CPGs, among other resources. The authors reported that the proposed CDSS achieved favorable results in providing guidance on new prescriptions. Ting et al. [56] proposed a hybrid case-based reasoning approach for medication prescription recommendations. In this regard, the proposed approach combined results from case-based modeling and Bayesian reasoning using a set of heuristic rules. Highly recommended medications were those that were selected by both models. The main focus was on utilizing the clinical experience of physicians along with modeling clinical knowledge in the form of a Bayesian network. Medication prescription recommendation

is a non-trivial task that includes the selection of medication from among a number of alternate medications. Furthermore, medication dosage selection adds to the complexity of the task. For such complex scenarios, clinical practice guidelines (CPGs) do not sufficiently capture the wide range of suitable recommendations. Therefore, most of the studies utilizing domain knowledge can only provide medication selection recommendations. In the proposed hybridization pipeline, we demonstrate an approach which involves combining an abstract domain model with that of clinical cases for medication dosing estimation. In this regard, the proposed approach combines an expert-based model for medication selection and a statistical technique, such as an interquartile range (IQR), for medication dosage estimation. Table 2.2 provides a summarized comparison of related techniques for medication prescription.

Table 2.2: Literature comparison of related medication prescription techniques for clinical decision support system

Reference	Area of Application	Objective	Characteristics	Limitations
[50]	Kidney disease	Medication Selection	*CDSS based on pre-defined process map * Outpatient renal dose adjustment using CDSS * Retrospective evaluation	*Difficult to express domain consensus for complex cases * Model maintenance requires domain expert involvement
[51]	Primary healthcare	Medication Selection	*A two-step drug recommendation through CDSS * Integrated into Janus web solution * Evaluation through questionnaire responses and focus group	*Complete domain model is difficult to acquire for complex decision tasks with multiple preferences * Clinical experience of different clinicians for dosing recommendation is not integral part of the CDSS
[52]	Kidney patients	Medication Selection	*CDSS for potential drug–drug interactions (pDDIs) recommendation * Knowledge base construction for pDDIs alert generation * Prospective evaluation	*The knowledge does not provide medication dosing recommendation * Difficult to express domain consensus for complex cases
[53]	Kidney patients	Medication Selection	*Context-aware CDSS for drug–laboratory interactions (DLIs) * Knowledge base for DLIs recommendations * Prospective cross-sectional evaluation using real clinical patient data	*Complete domain model is difficult to acquire for complex decision tasks with multiple preferences * Difficult to maintain complex rule-based models, e.g., medication adjustment
[54]	Kidney patients	Medication Selection	*Drug prescription for reduced renal function patients * CDSS is integrated in Janus toolbar * Evaluation using questionnaire technique	*Clinical experience of different clinicians for dosing recommendation is not integral part of the CDSS * Clinical experience of medication selection is not reflected in the model
[55]	Kidney patients	Medication Selection	*CDSS for drug therapy selection/discontinuation * Different alerts are designed based on multiple domain sources * Prospective evaluation using randomized control trial	*The CDSS does not provide medication dosing recommendation * Knowledge maintenance for new medications would pose a major challenge
[46]	Standard medical care	Medication Selection	*Data-driven hybrid model using case-based reasoning and Bayesian reasoning, execute in parallel * Heuristic rules to combine results from both models	*The CDSS does not provide medication dosing recommendation * Complete domain model is difficult to acquire for complex decision tasks reflecting multiple preferences

Chapter 3

Proposed Methodology for Hybrid Case-based Modeling

This chapter deals with elaboration of our proposed hybrid knowledge modeling approaching for case selection and adaptation. In this regard, first we provide an overview of the proposed methodology. We specify the steps required to construct a domain model and also touch upon our strategy of incorporating the domain model into the case-based reasoning framework. Afterwards, we elaborate on the three key steps of our methodology i.e. active case base partitioning, multi-level case selection and hybrid case adaptation. Lastly, we discuss details of the CKD-MBD CDSS that is designed on the basis of our methodology.

3.1 Overview of Hybrid Case-based Reasoning

3.1.1 Extract, Transform, Load (ETL) pipeline for data acquisition

Relevant data are generally available in disparate form i.e. data in electronic medical records, clinical notes, medical images, etc. In this regard, all the relevant data are brought from multi-sources and forms into a staging area where it can be processed and transformed into a standard form. Extract operation, retrieves data from multiple sources while transform converts into a standard form. Finally, load operation brings the data from a staging area into a centralized repository. The proposed CDSS takes into account the laboratory and imaging test results of patients and assists clinicians in selecting an appropriate treatment regimen. Treatment recommendations in terms of medication selection and dosage adjustment are based on similar patients and domain knowledge. The proposed hybrid approach is illustrated in Figure 3.2. The expert knowledge is codified into a hierarchical structure such as a DT and it is utilized for partitioning the past clinical cases into multiple groups. Each case is composed of two components, i.e., problem component and solution

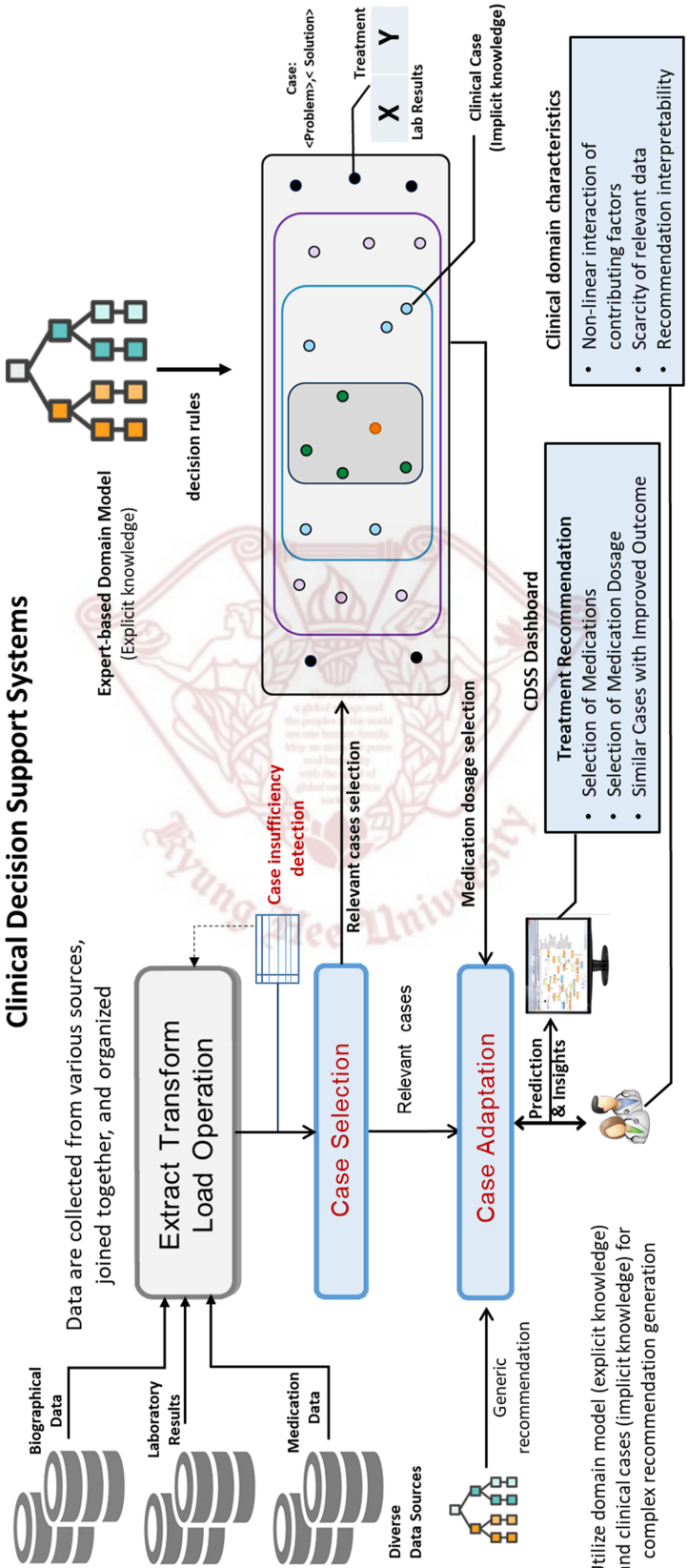


Figure 3.1: Proposed methodology at a glance.

component, represented by X and Y. The problem component represents measurements for multiple laboratory test results such as PTH, phosphate, and albumin-corrected calcium levels along with the status of vascular calcification in the body. The solution component specifies different prescribed medications along with the dosage. A new patient encounter, Xt, is treated as a test case and assigned one of the recommendation groups.

Moreover, each medication recommendation group is denoted by Ti, where i refers to the number of pre-specified partitions by the domain experts, and k represents any specific partition that contains Xt. In the proposed CKD–MBD CDSS, the entire case base is divided into 33 mutually exclusive partitions, starting from T1 up to T33 (refer to Table 4.3). X0 and Y0 represent cases from partition Tk that are treated as similar cases for the given Xt. Moreover, a subset of reference cases, m, are selected from the Tk partition containing n cases where m ¡ n. All similar cases are assigned an outcome value based on the PII and only those cases are selected as reference cases that have a PII > 0. Reference cases along with domain knowledge-based adaptation rules are used for prescription recommendation denoted by ˆY. ˆY represents a set of selected medications along with their dosage range, e.g., Medication <Cinacalcet>: = 25 mg/day–50 mg/day. The IQR is used for estimating dosage ranges for multiple selected medications. The prescription recommendation is provided to the clinician that may further refine the dosage. Afterwards, a final prescription, Y", is provided to the new patient case, Xt.

A comprehensive recommendation scenario depicting key stages of the proposed recommendation system is depicted in Figure **??**. A patient's laboratory and imaging tests are evaluated through the domain model, and an abstract recommendation is subsequently generated based on the patient's type, i.e., negative for cardiovascular calcification (type-II), and patient's group, i.e., T1. A set of similar cases are acquired from the case-base pertaining to both type-II and T1 patients. Each retrieved case is assigned a case outcome score using the PII. A set of references cases are selected i.e., cases having a PII > 0. Prescribed medications of the selected cases are processed using both adaptation rules acquired from domain knowledge, such as "start or increase medication class A", "maintain medication class B", and "decrease or stop medication class C", etc. IQR is applied on prescribed medication dosages when generating lower and upper bounds, e.g., medication A1: 25 mg/day–50 mg/day, where A1 is one of the medications in medication

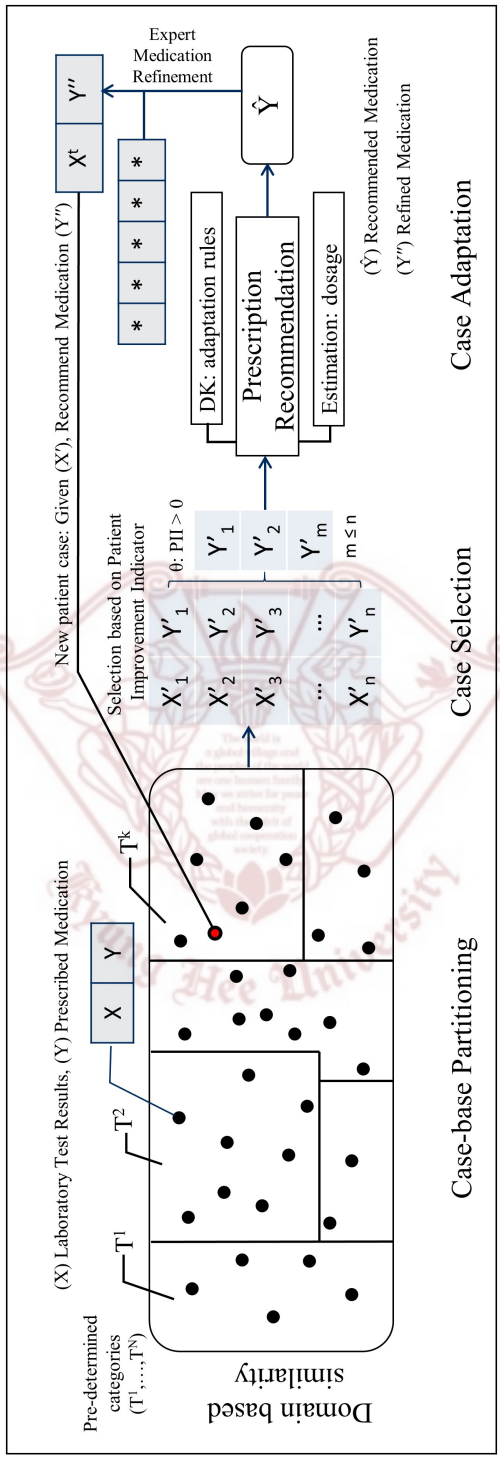


Figure 3.2: Schematic representation of the proposed hybrid knowledge modeling approach.

class A. Adaptation rules are used for medication class level recommendations, i.e., initiation/-continuation/discontinuation of a medication class, while clinical cases are used for estimating the lower and upper bounds of dosages for specific medications.

The major advantages of the proposed approach are as follows:

1. Domain theory-based patient categorization enhances the confidence of clinicians
2. Each group is denoted by a variable-sized neighborhood
3. Easily identifiable patient groups that have an insufficient number of associated clinical cases
4. Easily identifiable treatment regimens that are effective for similar patient cases
5. Medication dosage adjustment support based on domain theory along with evidence from past clinical cases
6. The enhanced interpretability of the medication selection and dosage adjustment recommendation by clinicians

A multi-level data flow diagram (DFD) of the proposed CKD–MBD CDSS is shown in Figure 3.3. The proposed CDSS is composed of three main tasks, i.e., case-base partitioning, case selection, and case adaptation. An abstract domain model is used to partition the case-base in a pre-determined set of groups, i.e., abstract recommendations. One major advantage of eager partitioning is the active identification of those partitions that lack a sufficient number of clinical cases in the case repository. Case selection deals with retrieving similar cases for a given test case and selecting a set of reference cases from the selected partition. Finally, case adaptation is applied to a set of reference cases with the help of adaptation rules. Adaptation rules are based on domain knowledge. The IQR is used as a measure of statistical dispersion of the selected medication dosages. In this regard, a final recommendation is generated specifying the lower and upper bounds for the dosage values.

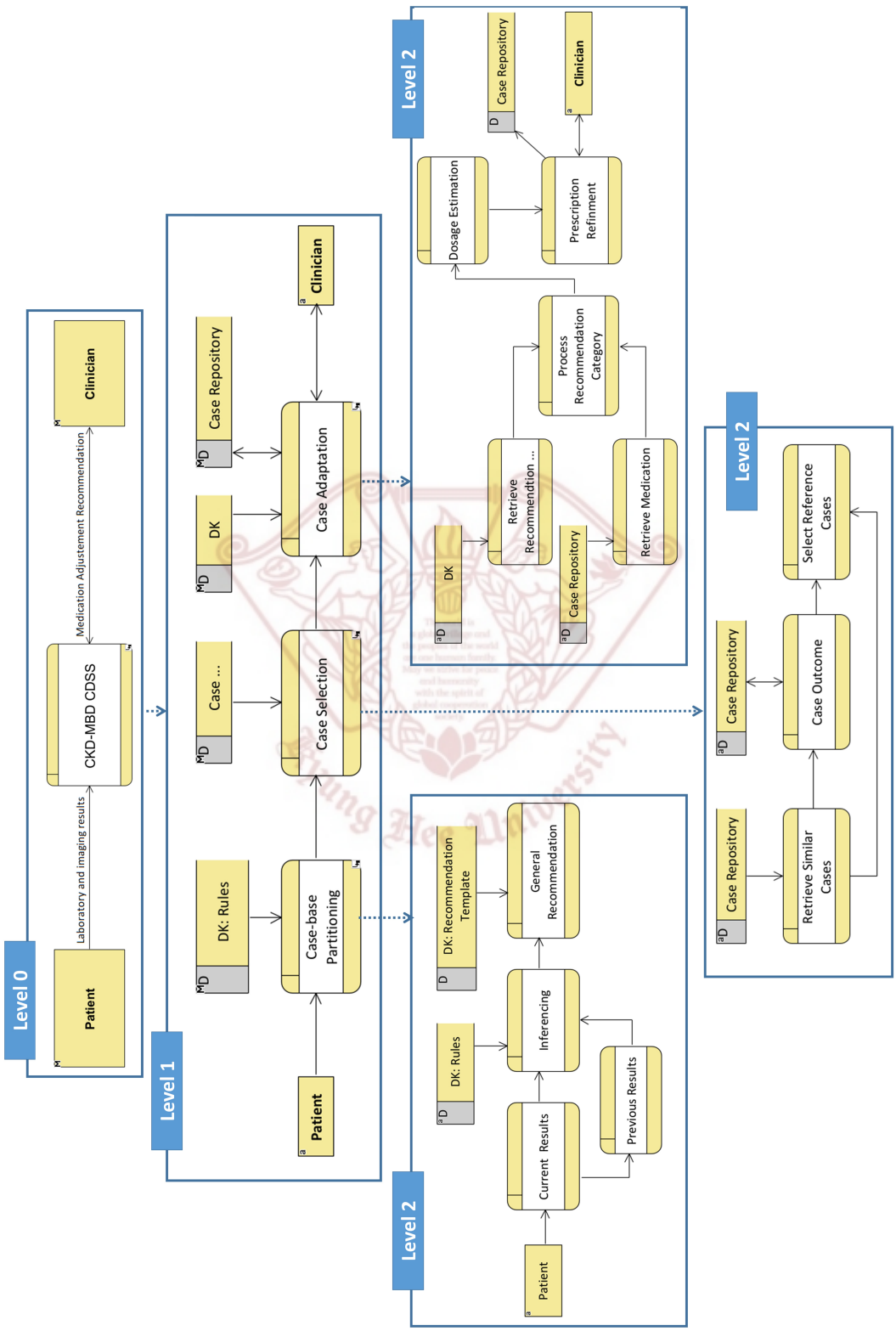


Figure 3.3: Data flow diagram depicting the relationship between processes and data..

3.1.2 Active Case Partitioning through Domain Knowledge

Domain knowledge (DK) plays a critical role in the development of CDSSs. It primarily deals with specifying key concepts and the relationships among the concepts. In the proposed system, DK is used for a priori partitioning of the case-base into multiple patient groups. The process of DK acquisition and codifying it into a domain model is depicted in Figure 3.4. DK is also used for generating generic medication intake recommendations that are also used in case adaptation operations for medication selection. The benefits derived from active case partitioning include the variable size of each partition (i.e., neighborhood size is not defined a priori) and, as the relevant cases are localized, this therefore reduces the run-time processing for retrieving similar cases every time a new test case is received. Algorithm 1 specifies the main steps for active case-base partitioning, provided as follows:

Algorithm 1: Active Case-base Partitioning

Input : $KB-$ Rules , $CB-$ Case Base
Output: Partition $P = T_1, T_2, T_3, ..., T_n$

```
for ∀ c ∈ CB do
    rule = reasoing(c);
    rule = rule[1].code;
    p[rec] = c;
end
for ∀ pt ∈ P do
    if pt.length <= θ then
        flag(pt)
    end
    rule = rule[1].code;
    p[rec] = c;
end
return P
```

The final output of the knowledge acquisition process is a domain-decision model that is similar to a DT structure. Over the course of multiple consultations, the domain experts develop a domain-decision model based for the most part on KDIGO CKD–MBD guidelines [12]. The domain model provides sufficient knowledge to group patient casebase into multiple categories. The domain model is converted into production rules of the form IF-THEN. Figure 3.4 illustrates a mind-map depicting a domain model for CKD–MBD patients. As recommended in the KDIGO guidelines, all hemodialysis patients are subject to lateral abdominal radiographs and echocar-

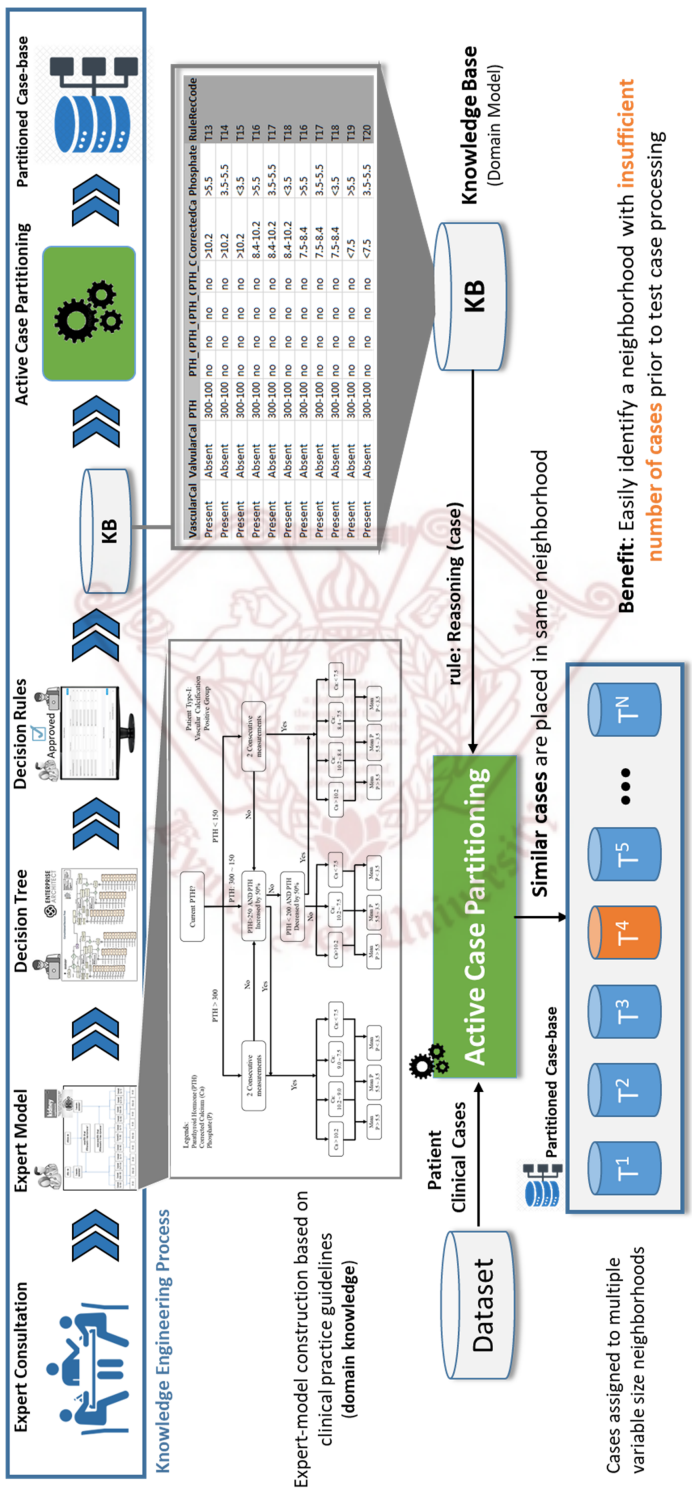


Figure 3.4: Knowledge engineering process of domain model construction and active case-base partitioning.

Table 3.1: Relevant clinical parameters and their target ranges.

Clinical Parameter	**Target Range**
PTH (type-I patient)	150 - 300 pg/mL
PTH (type-II patient)	130 - 600 pg/mL
Phosphate	3.5 - 5.5 mg/dL
Albumin-corrected Calcium	7.5 - 10.2 mg/dL

diography in order to evaluate vascular and valvular calcification, respectively. The severity of vascular calcification is graded on the abdominal aorta by a validated method [97], while valvular calcification is assessed in a dichotomous manner, i.e., its presence or absence. In the proposed CDSS for CKD–MBD management, hemodialysis patients are broadly categorized into two types based on the degree of ectopic calcification (as shown in Table 3.1): type-I patients who have valvular calcification or at least a moderate degree of vascular calcification (calcification score $>$ 5 out of 24), and type-II patients who are negative for valvular calcification and have a mild degree of vascular calcification (calcification score$<=$ 5) at most. The novel approach of the proposed CDSS is that a strict target range of PTH is set for type-I patients, whereas a relatively lenient target range of PTH recommended by KDIGO is set for type-II patients. Patient type categorization is performed by domain experts through the consultative method as mentioned in Section 2.1 as a part of the domain knowledge acquisition process.

In this regard, the resultant domain model accommodates both types of patients. There are three key attributes to the domain model, i.e., PTH, albumin-corrected calcium, and phosphate levels in the body (Figure 3.5). Furthermore, there are in total 33 patient groups identified by the domain experts. Each group is associated with a generic recommendation. A template for the multi-factor generic recommendation is provided in Table 3.2. 'Dialysate Calcium Concentration' is a non-medication factor that can be modified according to the partition to which the patient belongs. The recommendation against each factor is provided in general terms, such as whether to initiate (or increase) a particular medication/dialysate calcium concentration or discontinue (or decrease) the medication/dialysate calcium concentration.

Table 3.3 provides a generic recommendation template. As it can be seen from the table, each factor can take on one of the available treatment options.

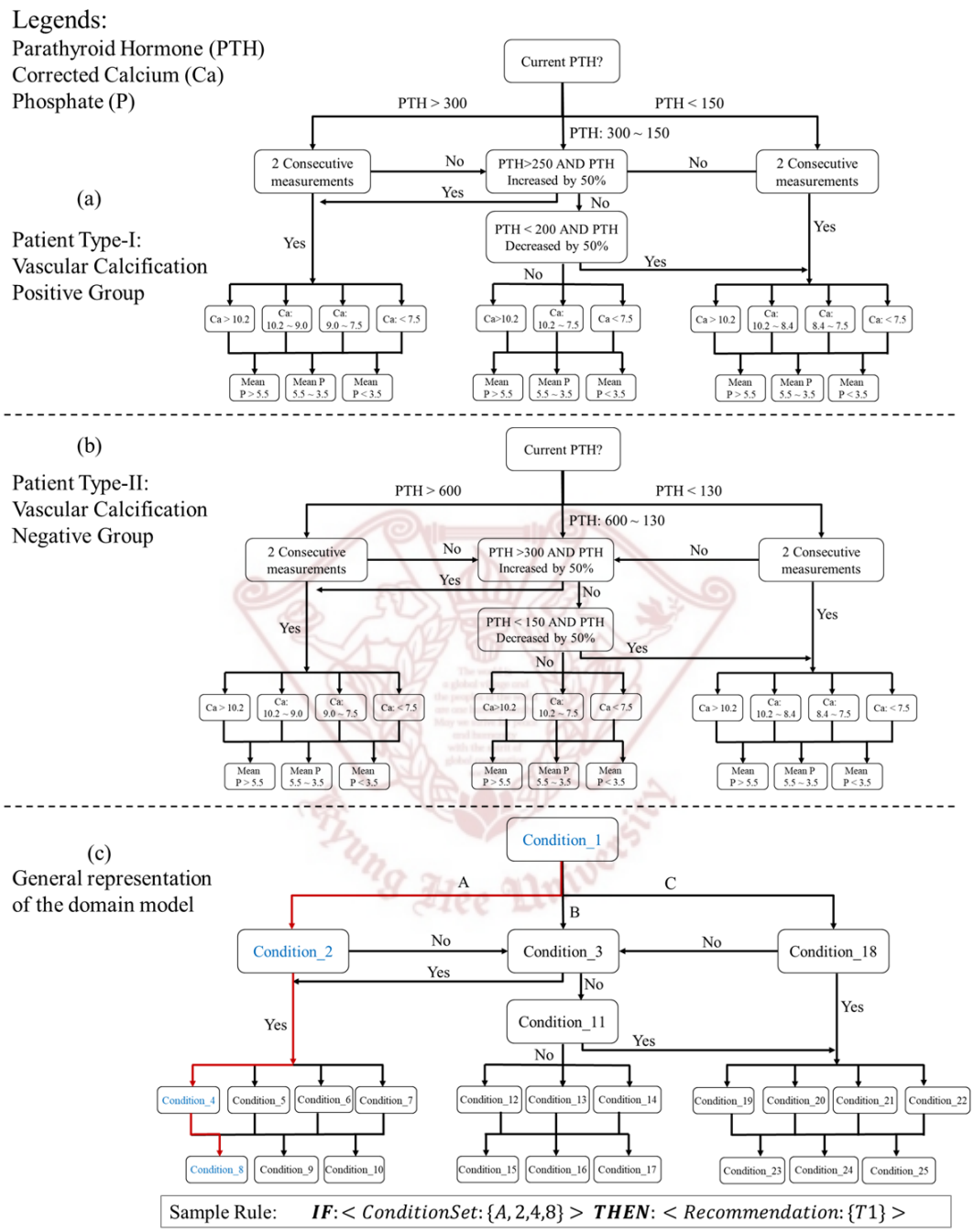


Figure 3.5: Mind-maps for expert-based domain models for (a) type-I and (b) type-II CKD–MBD patients along with (c) a sample mind-map structure for representing a CPG-based domain-model (Ca refers to albumin-corrected calcium).

Table 3.2: A sample generic recommendation.

Management Class	**Treatment Options**
Calcimimetics	Start or Increase Cincalcet
Calcitriol	Stop Calcitriol
Vitamin D and Analogs	Stop Vitamin D and Analogs
Calcium-based Phosphate Binder	Stop Calcium-based Phosphate Binder
Non-Calcium-based Phosphate Binder	Start or Increase Non-Calcium-based Phosphate Binder
Dialysate calcium concentration	Reduce dialysate calcium concentration by 0.25 mmol/L

Table 3.3: Generic Recommendation Template.

Management Class	**Available Options**
Calcimimetics	[Start or Increase Cinacalcet], [Decrease Cinacalcet], [Stop or Decrease Cinacalcet], [As it is]
Calcitriol	[Start or Increase Calcitriol], [Stop Calcitriol], [Decrease or Stop Calcitriol], [Consider Calcitriol], [As it is]
Vitamin D and Analogs	[Consider Vitamin D Analogs], [Decrease or Stop Vitamin D Analogs], [As it is]
Calcium-based Phosphate Binder	[Start or Increase CPB], [Stop CPB], [Decrease or Stop CPB], [As it is]
Non-Calcium-based Phosphate Binder	[Start or Increase NCPB], [Stop NCPB], [Decrease or Stop NCPB], [As it is]
Dialysate calcium concentration	[Increase by 0.25 mmol/L], [Reduce by 0.25 mmol/L], [Maintain Current Calcium Concentration]

Active case partitioning through domain knowledge serves two purposes, i.e., it partitions cases into multiple groups, and it also provides generic medication intake recommendations for each category (refer to Table 4.3). It is also important to note that one of the main objectives of the domain model construction is to include CKD–MBD guidelines in the decision-making process. The overall domain model for type-I and type-II patients has resulted in 432 production rules (see Figure 3.5). The production rules are useful in automating the reasoning process and maintaining the knowledge base.

3.1.3 Reference Case Selection Using the PII

One of the important contributions made in this paper is the development of the PII as a case scoring function. The main objective of the PII is to provide a summarized view to the clinician regarding the overall health status of the patient, i.e., patient-important outcome. The PII is comprised of three individual factors i.e., PTH, calcium, and phosphate levels in the body. A patient may visit multiple times over the period of treatment, and at each visit the aforementioned three clinical measurements are used to calculate the PII. The operation of PII computation is depicted in Figure 3.6 and the formula for its calculation is provided in Equation (1).

$$PII = \frac{\sum_i^m C_i}{m} \tag{3.1}$$

where, m refers to the total number of clinical measurements, i refers to the i-th measurement, C refers to a Boolean value, i.e., either 0 or 1. Each Ci value refers to a binary decision, i.e., whether the given test results are within a target range or not. For example, for patients with at least a moderate degree of vascular calcification (i.e., patient type-I), the ideal PTH level is between 150 300 pg/mL [57] [98], while target ranges of phosphate and albumin-corrected calcium are 3.5 5.5 and 7.5 10.2 mg/dL, respectively.

The PII is bounded between 0 and 1. PII values closer to 1 indicate a better patient important outcome, as depicted in Figure 3.6. The PII differs for different patients: seeing as the normal range of PTH varies with the type of patient, PII is therefore calculated accordingly. All the patient cases are assigned their respective PII values, except for corner cases, such as a patient having only a single encounter or the latest encounter of the patient. These cases are treated as corner cases because no subsequent patient encounter is available to calculate the PII value. It is important to note that for evaluating the efficacy of medication dosage prescribed on encounter i, the laboratory test results from the subsequent patient encounter, i + 1, are required, as shown in Figure 3.6. Algorithm 2 provides main steps for case selection through feedback as follows:

The PII is used to differentiate cases based on their outcome, i.e., whether the patient's condition (indicated by laboratory test results) improved given a certain prescription or not. The main

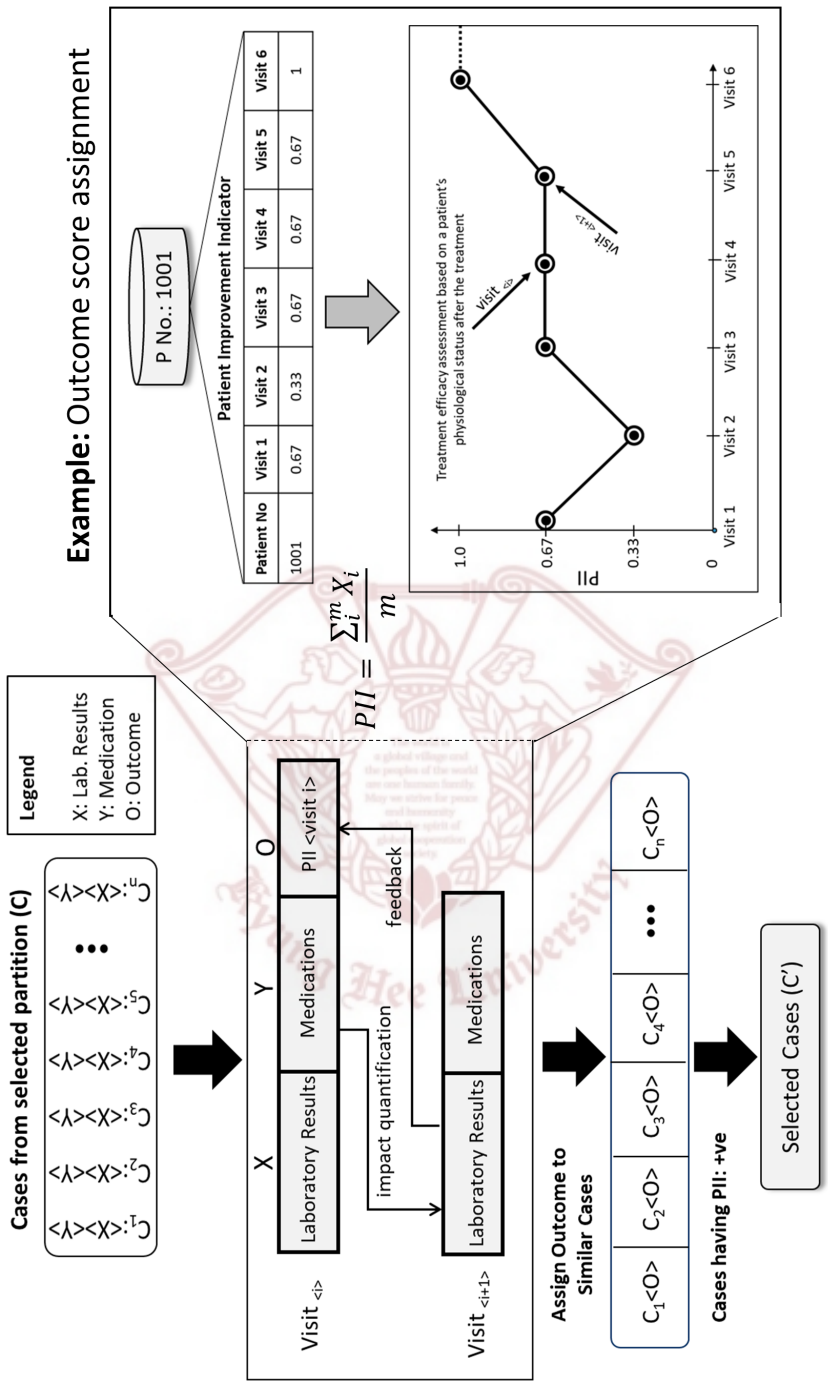


Figure 3.6: The Patient Improvement Indicator (PII) for selecting reference cases.

purpose is to select a set of reference cases that have a positive impact on the outcome, i.e., improvement is recorded in the patient's laboratory test results. Equation (1) is used to assign an

Algorithm 2: Case Selection based on feedback

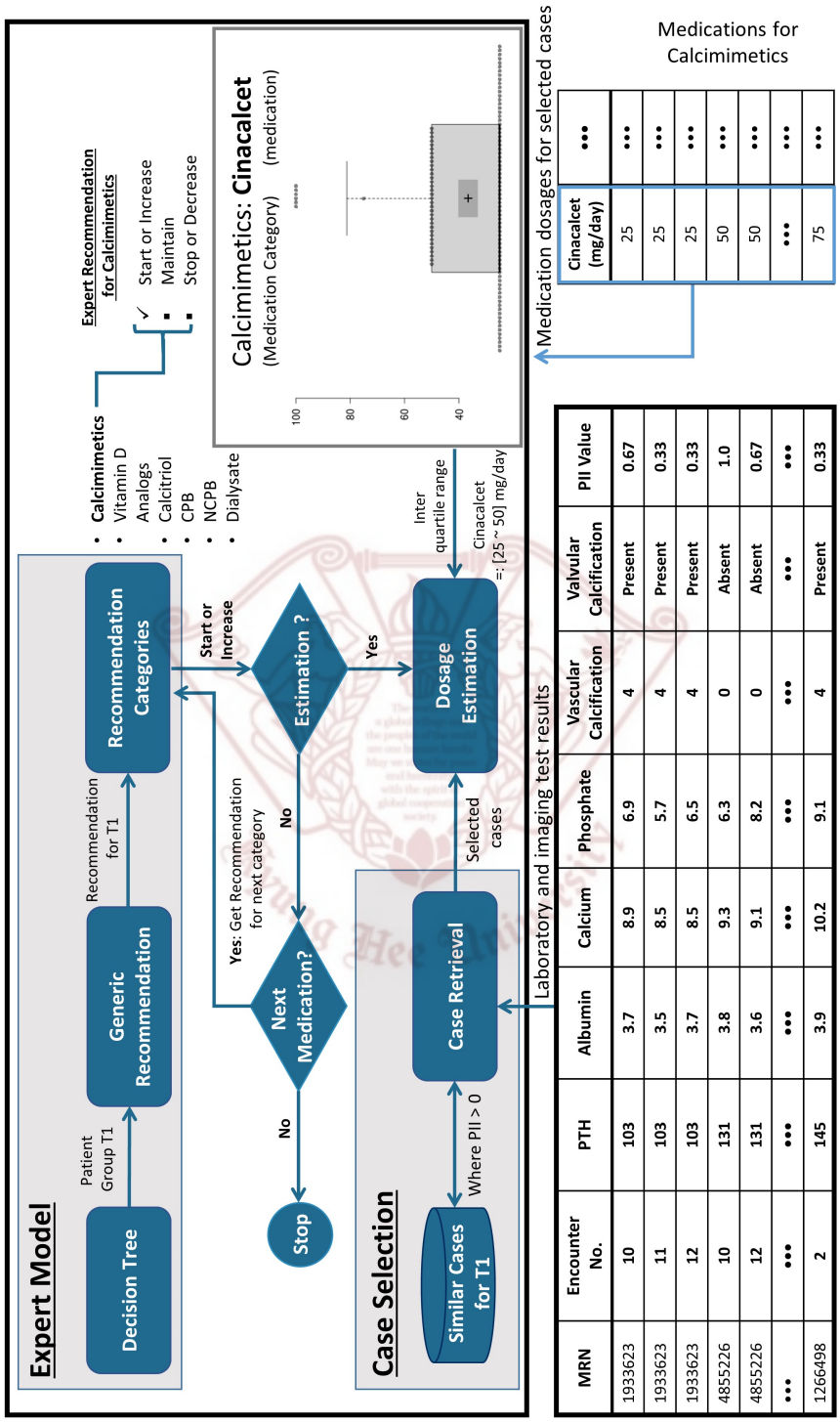
Input : $C-$ Partitioned Cases
Output: $C^{'}$ Referenced Cases

for $\forall\, c \in C$ **do**
  if $c_{outcome}! = Null$ **then**
    $c_{outcome} \leftarrow fetch_{PII}(c)$
  else
    if $t! = Null$ **then**
      $indicator = eval(ct, Ref)$
    end
  end
  $result+ = indicator$
end
$cout = result/len(result)$;
$cprev = fetch_pii(c)$;
$Cselected+ = c$;
return $Cselected$

outcome score to a clinical case. An important contribution of the PII is to refine case selection operation, i.e., select cases among similar cases based on their outcome. The refined selection provides a set of reference cases for medication dosage estimation. A similar approach was adopted by Bach et al. [85], whereby the authors initially retrieved a set of similar patients in the domain of low back pain therapy recommendations, and later a reference patient group was identified for further recommendation tuning. The reference group was comprised of patients with positive outcomes, such as decreased pain, improved pain self-efficacy, and better mood. In the case of the CKD–MBD domain, treatment prescription (i.e., medication and dosage selection) for a similar set of patients may vary as per the clinicians' decisions. The PII is therefore used to qualify different prescribed past treatments according to their efficacy.

3.1.4 Case Adaptation through Domain Knowledge and Clinical Cases

The main objective of the case adaptation operation is to provide medication intake recommendations to the physician. In this regard, domain knowledge and reference cases (i.e. cases selected after identifying their positive impact on patients' psychological measurements) are used for generating recommendations as shown in Figure 3.7. As stated earlier, domain knowledge is used for both case-base partitioning as well as generic recommendation generation. Case adaptation re-



MRN	Encounter No.	PTH	Albumin	Calcium	Phosphate	Vascular Calcification	Valvular Calcification	PII Value
1933623	10	103	3.7	8.9	6.9	4	Present	0.67
1933623	11	103	3.5	8.5	5.7	4	Present	0.33
1933623	12	103	3.7	8.5	6.5	4	Present	0.33
4855226	10	131	3.8	9.3	6.3	0	Absent	1.0
4855226	12	131	3.6	9.1	8.2	0	Absent	0.67
•••	•••	•••	•••	•••	•••	•••	•••	•••
1266498	2	145	3.9	10.2	9.1	4	Present	0.33

Figure 3.7: Medication dosage selection and dosage adjustment based on domain knowledge and interquartile range (IQR).

fines the expert-based generic recommendation through processing similar cases and statistically analyzing the co-occurrence of medication dosages. Case adaptation is the final step in the proposed methodology, through which a more precise treatment recommendation is generated that deals with both medication selection and dosage recommendation. Table **??** shows an example of a sample relationship between the generic recommendation and dosage recommendation.

Algorithm 3: Case Adaptation using Hybrid Approach

Input : $DTcode-$ Decision Tree, $C-$ Selected Cases
Output: R Multi-Attribute Recommendation

$Rec - gen$ = $fetchRec(DTcode)$;
for $\forall\ cat \in Rec - gen$ **do**
  $action - cat = process(cat)$;
  if $ation - cat ==' Start - or - Increase'$ **then**
    $med - dosage = fetch(Cm)$;
  end
  $med - size = len(med - dosage)$;
  $m = sort(med - dosage)$;
  $mid - index = m(m, 0, mid - index)$;
  $rangeL = m[median(m, 0, mid - index)]$;
  $rangeH = m[mid - index + median(m, 0, mid - index)]$;
  $R.append(rangeL, rangeH)$;
  if $t! = Null$ **then**
    $indicator = eval(ct, Ref)$
  end
  ; $result+ = indicator$
end

return R

As it can be seen that dosage recommendation relies on directions from generic recommendations, i.e., whether to select a particular class of medication or not. Moreover, the final dosage recommendation is based on the most frequent medication and its dosage among high prospect similar cases. It is important to note that, when using IQR statistics, situations in which similar cases take on different medication dosages for a given medicine are reflected in the final recommendation as a dosage range having both lower bound and upper bound values, as shown in Figure 3.7.

3.1.5 CKD–MBD CDSS Execution Process

The execution process workflow pertaining to the medication prescription is depicted in Figure 3.8. Patients are assigned a unique Medical Record Number (MRN) at the registration stage. Afterwards, both current and previous laboratory tests are acquired for patient type selection as well as patient group identification (through the domain-decision model). The medical laboratory tests include measurements for phosphate, calcium, albumin, and PTH. A set of reference cases is selected based on the PII of similar cases. Furthermore, if only a single case is available in the selected case set, then it is provided to the physician without any modification.

When there are multiple selected cases in the set, then a case adaptation operation takes place that generates a single medication prescription recommendation based on the multiple reference cases. Before persisting with the medication dosage, the clinician may choose to modify the contents of the recommendation, such as adjusting the medication dosage from the recommended one. The system automatically logs the concordance between the generated recommendation and the clinician's prescription.

The categorization of patients has two aspects, i.e., patient type selection and patient group selection. Patient type selection requires medical imaging results such as lateral abdominal radiography and echocardiography in order to determine the degree of ectopic calcification. The aforementioned imaging tests are conducted once every year. Patients are divided into two types, i.e., positive for vascular calcification and negative for vascular calcification. Patient group selection, on the other hand, is performed using the domain model, as shown in Figure 3.10. The group selection decision is taken every month, i.e., at each encounter with the patient. Furthermore, PTH laboratory medical results are conducted every three months, and both albumin-corrected calcium and phosphate tests are performed every month. Current and previous laboratory test results are required for patient group selection through the domain model. There are 33 different patient groups identified by clinicians within the scope of CKD–MBD management.

3.1.6 An extended CDSS

The proposed approach is geared towards treatment recommendation. The CDSS can be extended to include patient diagnosis screening as well in which case the extended CDSS would have two

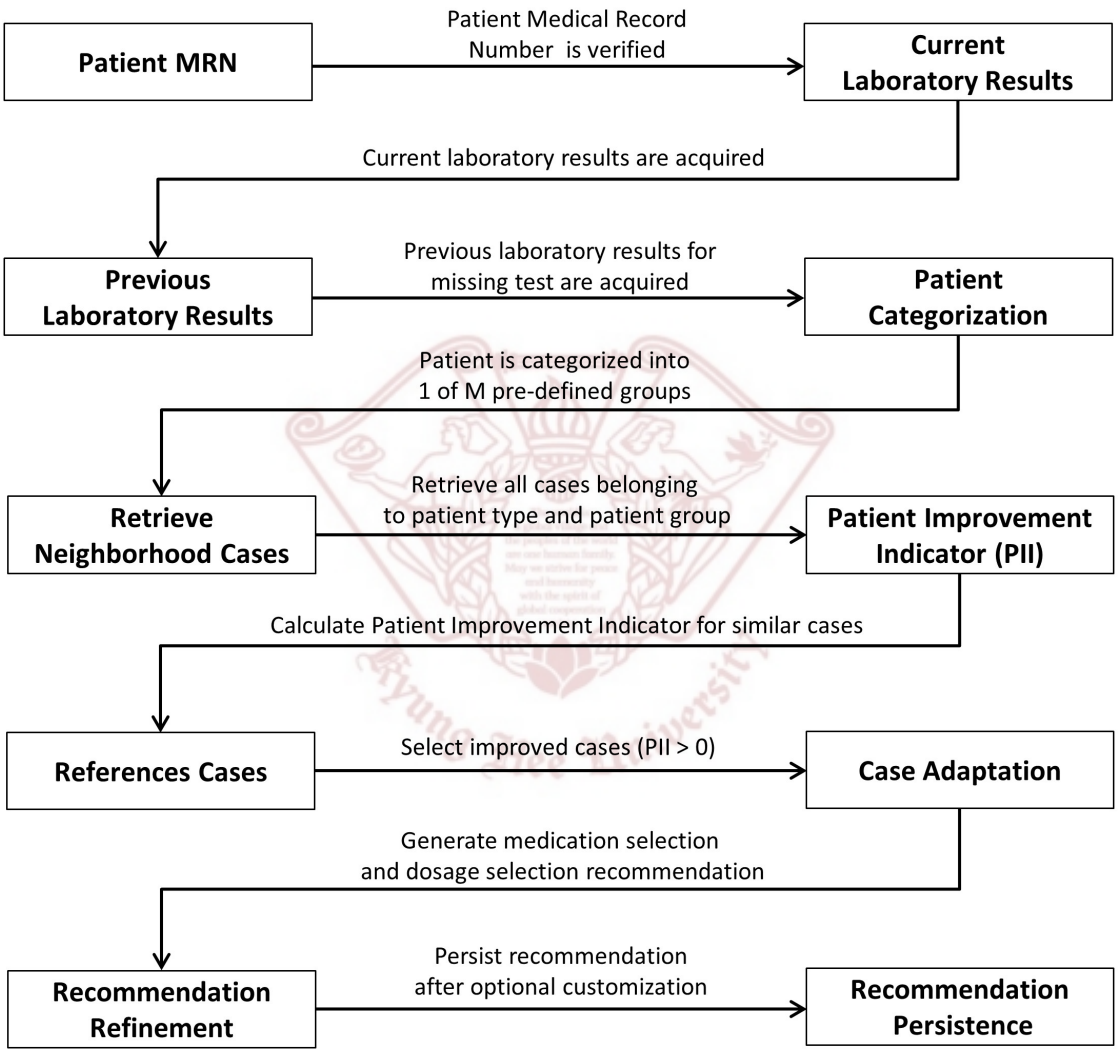


Figure 3.8: Process flow for CKD–MBD CDSS pertaining to the treatment regimen selection.

main tasks i.e. diagnosis screening to identify potential CKD patients. and after a subsequent patient diagnosis by the physician, treatment module may support physician in selecting a suitable treatment regimen for the patient. Figure 3.10 depicts an overall concept of a twin-system that independently provides services for interpretable decision making.

One key benefit of the extended CDSS is that the openly available diagnosis datasets can be used to extract knowledge from structured (i.e. tabular) data in order to provide interpreable decision making support in the form of decision rules. In this regard, the source data set is pre-processed to enhance the quality of data. Moreover, feature selection also needs to be performed in order to select a subset of most informative features that can be used for constructing decision rules through machine learning models e.g. decision tree. Once a model is learned then rule extraction can be performed to get decision rules in the form of *IF* $< X >$ *THEN*$< Y >$. A number of other considerations may also be incorporated in decision modeling such as the operating cost of the features being utilized in the decision making.

Automated medical diagnosis is one of the important machine learning applications in the domain of healthcare. In this regard, most of the approaches primarily focus on optimizing the accuracy of classification models. In the domain of medical data mining, several intelligent clinical decision support systems are designed which tend to automate the diagnosis process [6,7]. These decision systems employ machine learning techniques that assist physicians in the diagnosis and treatment of CKD in an efficient manner [99–101]. Based on a number of important indicators such as blood pressure, albumin levels, blood and urea tests, potassium, and other co-morbidities, e.g., diabetes, cardiovascular disease, etc., a patient is comprehensively assessed for CKD and its progression. As the earlier diagnosis of the disease onset can improve the chances of patients to favorably respond to treatment, therefore, most of the automated systems are optimized for enhancing the overall accuracy of the model [80, 100].

The medical decision support systems that solely focus on predictive performance are far from the field reality and hence are not unanimously approved by physicians [79]. In this regard, the interpretability of the classification model is stipulated as one of the important requirements among others for a successful medical decision system [79, 80]. Similarly, the cost factor as a practicability concern for medical decision systems recently gained traction in the medical data mining

community. Therefore, one of the key research directions pursued by the research community is to design decision systems that are accurate, interpretable, and cost-effective.

In a number of studies performed on CKD diagnosis, decision tree models consistently produced results with high predictive accuracy [100–102]. Hence, the main impetus for using tree-based models in an ensemble technique is two-fold. Firstly, tree models are easy to interpret by the domain experts, therefore, in domains such as medical diagnosis, it is desirable to assess the validity of the classification model through visual inspection [79, 80, 103]. Secondly, tree-models that are based on bagging and boosting techniques tend to produce highly accurate classifiers on small to medium datasets [61, 99–101]. Hence, tree models are suitable approaches for considering in an ensemble for a CKD dataset, as they can cater to both types of requirements i.e., interpretability and accuracy.

Moreover, feature selection is becoming an essential task in building classification models where the objective is to select a subset of useful features [61, 99, 100, 104, 105]. The notion of usefulness is based on the worth of a feature in a dataset in terms of its relevancy and redundancy. There are generally three approaches for feature selection i.e., filter-based approach, wrapper-based approach, and embedded approach [60, 104, 106]. In the case of filter approaches, the worth of a feature is evaluated through univariate statistical approaches such as Chi-Square, Gini index, information gain, etc. Therefore, feature ranking techniques fall into the filter category. On the other hand, wrapper approaches generally, construct a set of candidate feature subsets that are evaluated on a classifier [106]. Embedded techniques are implicitly used by some of the classifiers, such as decision trees, while constructing a model. A number of studies demonstrated that ensemble-based feature selection techniques generally perform better than non-ensemble techniques [100, 104, 105, 107, 108]. Ensemble feature selection approaches are composed of multiple evaluation functions for quantifying the worth of a feature or a subset of features. In this regard, multiple types of feature evaluation functions can be used such as univariate techniques, classification models, or a set of mixed techniques from the aforementioned categories [100]. Ensembles can be comprised of both homogeneous and heterogeneous configurations. In this regard, for a homogeneous configuration, a dataset is horizontally partitioned into multiple subsets where a single type of the feature evaluation function is executed on each partition [104, 108]. On the other hand,

for a heterogeneous configuration, multiple evaluation functions are executed on the dataset in parallel, and later their results are combined [104, 105, 108]. Similarly, ensemble feature ranking approaches can be arranged in either a homogeneous configuration or a heterogeneous manner. In both cases, a global ranked list of features is obtained based on multiple feature lists produced by the individual feature ranking functions. One key challenge in this regard is to select a threshold value which divides the global ranked list into a set of retained and removed features [107].

Most of the studies in the CKD domain assume that the cost of data acquisition is negligible, therefore, the cost factor associated with each feature is generally ignored [99,102,109]. However, this assumption may not hold in many real-world medical applications where a patient is required to undergo multiple tests such as urine analysis, electrocardiogram, blood culture, etc., and the tests may vary in terms of incurred cost. Therefore, feature selection methods for CKD diagnosis applications may take into account the cost factor as well. Recent studies reported significant scholarly work on developing chronic kidney disease diagnosis and management systems. Recent studies reported significant scholarly work on developing chronic kidney disease diagnosis and management systems [58, 60, 61, 99–101, 110–116]. Please refer to [117, 118] for details on cost-sensitive model construction and knowledge extraction from structured data for CKD diagnosis.

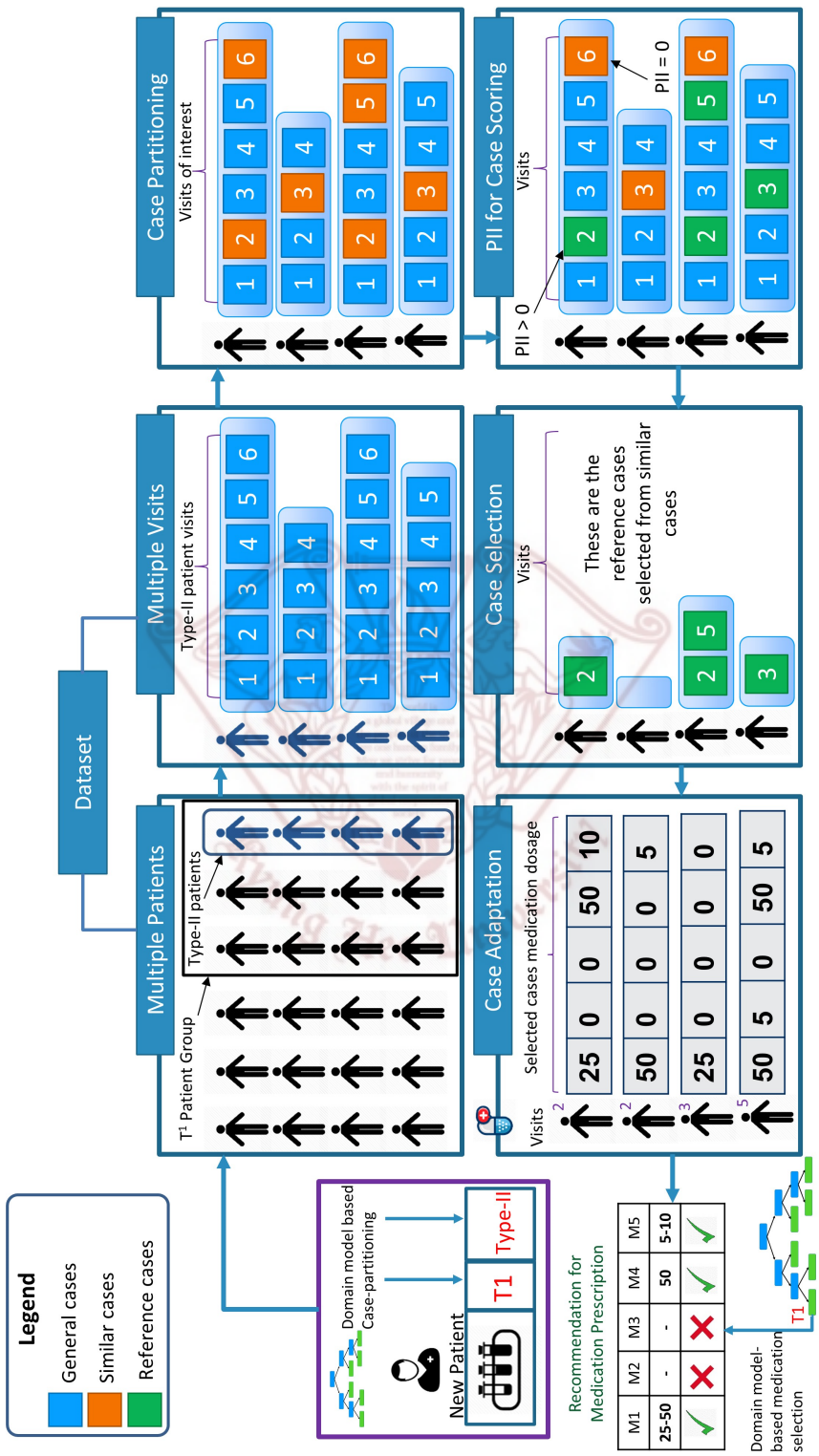


Figure 3.9: Medication selection and dosage adjustment scenario based on the proposed approach.

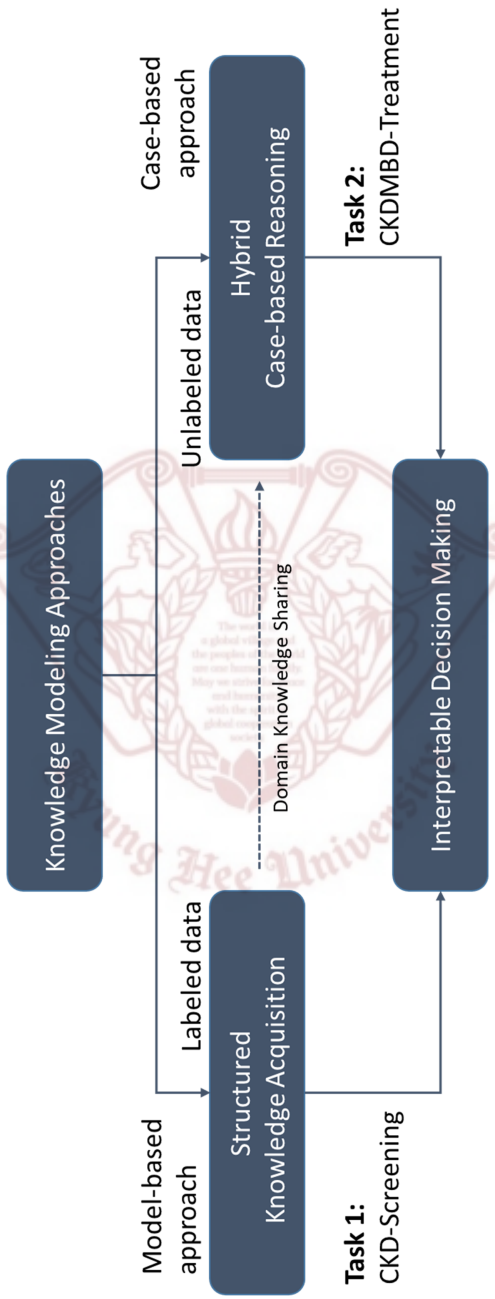


Figure 3.10: Extended CDSS for patient screening and medication recommendation

Chapter 4

Experiments and Discussion

4.1 Experiments Overview of Hybrid Case-based Reasoning

In this chapter we elaborate extended the chronic kidney disease case study for group-based feature cost and unlike model-based feature scoring, we consider filter measures such as symmetric uncertainty, chi-squared and Relief for feature weight-age calculation.

The CKD–MBD CDSS is evaluated using two perspectives, i.e., system perspective and user perspective. In the case of the system perspective, the evaluation is performed in terms of compliance between the CKD–MBD CDSS medication recommendation and routine clinical practice. The usability aspects of the proposed system are evaluated in terms of recommendation generation, assistance in preventing accidental dosage errors, and serial trend visualization of key measurements such as PTH, phosphate, and albumin-corrected calcium.

4.1.1 System-Centric Evaluation

To validate the system, we performed an experiment in which we established a concordance between the CDSS generated recommendations and that of physician's prescribed medication. We have 850 clinical cases extracted from 66 patients (each patient had at most 13 encounters) from Kyung Hee University Hospital, Seoul, South Korea. The gender ratio of the patients was 70:30, where 70% of the patients are male. Furthermore, the distribution of clinical cases between type-I and type-II patients are 374 and 476, respectively.

Domain Model Compliance

The domain model is primarily based on KDIGO CKD–MBD guidelines. As mentioned earlier, the generic recommendation is based on the domain model; therefore, it is worthwhile evaluating the compliance between the routine practice and the domain model. The recommendation consists of general directions for clinicians regarding the initiation, modification, or discontinuation of a certain medication class, as indicated in Table 3.2. The evaluation results, as provided in Figure 4.1, show the overall compliance between the clinical cases and the generic recommendation. It can be seen in Figure 4.1 that in general most of the recommendation factors have complied with the routine clinical practice as well. Therefore, the domain model-based generic recommendation plays an important role in the dosage estimation task.

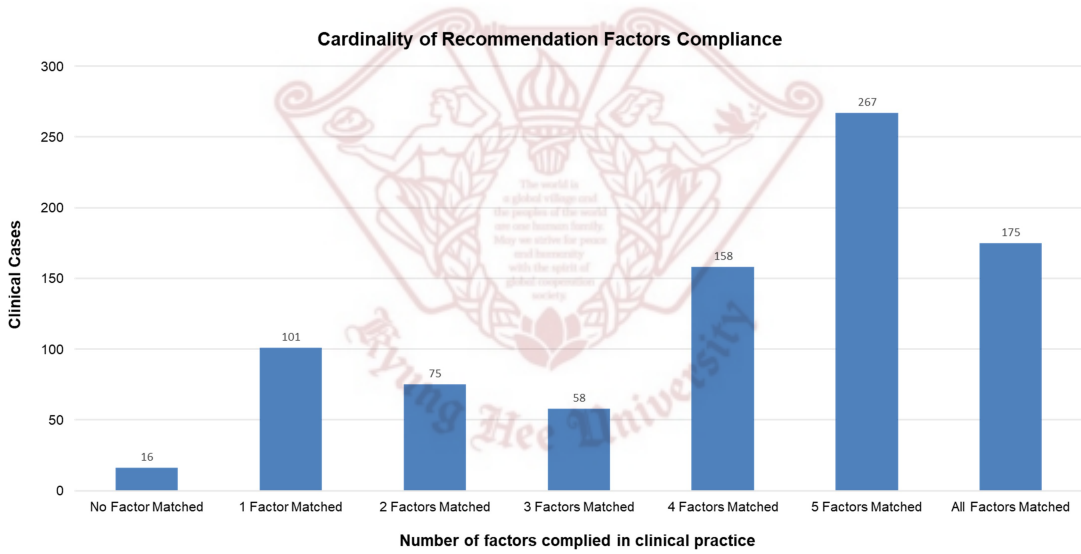


Figure 4.1: Cardinality of compliance among domain model and routine clinical practice for multifactor recommendations.

Figure 4.2 shows a breakdown of the compliance rate of six medication classes that are part of the overall recommendation. In the non-compliant cases, "decrease" slightly dominated, e.g., the system recommended to "maintain" while the clinician decreased the dosage. A confusion matrix based on the compliance evaluation between domain model and routine clinical practice is provided in Figure 4.3. "Maintain" remained the most dominant label in the recommendation across the medication classes. The average discrepancy across all the medication management

classes for the “start/increase”, “maintain”, and “stop/decrease” out of 850 cases is 97.50 cases, 42.16 cases, 82.83 cases, respectively. Moreover, NCPB had major discrepancies among all the medication classes, specifically in the “start/increase” recommendation.

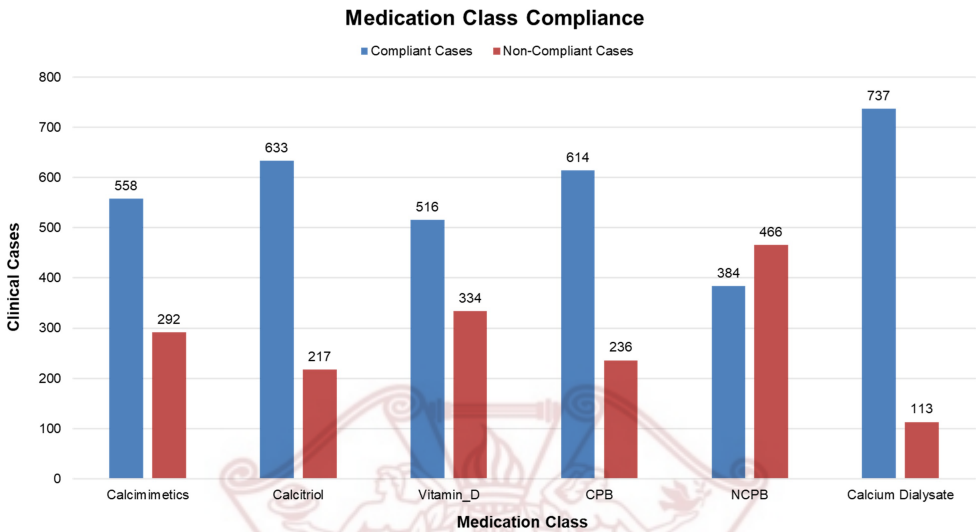


Figure 4.2: Compliance among different medication management classes along with dialysate calcium concentration.

Calcimimatics

Actual \ Predicted	Start/Increase	Maintain	Stop/Decrease
Start/Increase	167	132	17
Maintain	29	334	23
Stop/Decrease	19	72	57

Calcitriol

Actual \ Predicted	Start/Increase	Maintain	Stop/Decrease
Start/Increase	173	31	13
Maintain	18	448	11
Stop/Decrease	14	74	68

Vitamin D & Analogs

Actual \ Predicted	Start/Increase	Maintain	Stop/Decrease
Start/Increase	154	82	12
Maintain	17	450	20
Stop/Decrease	14	34	67

Calcium-based Phosphate Binders

Actual \ Predicted	Start/Increase	Maintain	Stop/Decrease
Start/Increase	140	66	8
Maintain	16	488	19
Stop/Decrease	3	41	69

Non-Calcium-based Phosphate Binders

Actual \ Predicted	Start/Increase	Maintain	Stop/Decrease
Start/Increase	71	101	19
Maintain	49	259	52
Stop/Decrease	9	187	103

Calcium Dialysate Concentration

Actual \ Predicted	Start/Increase	Maintain	Stop/Decrease
Start/Increase	45	7	0
Maintain	2	766	0
Stop/Decrease	0	8	22

Figure 4.3: Confusion matrix indicating compliance between domain model and routine clinical practice.

Evaluation for Dosage Recommendation

The medication dosage recommendation is the main objective of the CKD–MBD CDSS. In this regard, both domain knowledge and similar past cases are used to assist clinicians in dosage prescription. The clinical case-base consists of 600 cases whereas 250 cases are used to evaluate the recommendation system's efficacy with respect to the routine clinical practice of clinicians. There are 107 cases for type-I patients and 143 cases for type-II patients in the test dataset. Table 4.1 provides evaluation results based on the test data, indicating concordance between the dosage recommendation and clinical practice.

$$Concordance = \frac{\sum_{i}^{j}(System \cap Clinician)}{j} \tag{4.1}$$

The evaluation procedure is based on comparing the recommended dosage with that of the clinician's prescription using Equation (2), where i starts from 1 and j is the total number of factors in the recommendation, i.e., 10. Seeing as the recommended dosage is based on the IQR, i.e., 1st quartile and 3rd quartile, in most of the cases the recommendation is therefore in the form of a range of values, i.e., lower bound of the dosage and upper bound of the dosage. In such cases, the evaluation is based on whether the prescribed medication dosage is within the recommended dosage range or not. "In-Range" cases are those in which the prescribed medication is within the recommended range; otherwise they are regarded as "Out-of-Range" cases. Furthermore, not all medications are present in all the cases, i.e., Cinacalcet is present in 49 cases out of a total of 250 test cases, and so on. Concordance for Cinacalcet is 85.71%, Calcitriol (po) is 81.81%, Calcitriol (iv) is 66.66%, Paricalcitol (iv) is 82.24%, Calcium Carbonate is 76.47%, Calcium Acetate is 81.81%,Sevelamer is 76.12%, Lanthanum is 55%, and Dialysate calcium concentration is 98.40%. As Alfacalcidol does not include any case in the test set, it is therefore not part of the average concordance calculation. The average concordance of the medication dosage recommendation, as reported in Table 4.1, is 78.27%.

Table 4.1: Concordance evaluation for the medication dosage recommendation.

Management Class	**Total Cases**	**Present Cases**	**In-Range Cases**	**Out-of-Range Cases**
Cinacalcet	250	49	42	7
Calcitriol, po	250	11	9	2
Calcitriol, iv	250	15	10	5
Paricalcitol, iv	250	148	122	26
Alfacalcidol	250	0	0	0
Calcium Carbonate	250	34	26	8
Calcium Acetate	250	11	9	2
Sevelamer	250	155	118	37
Lanthanum	250	20	11	9
Dialysate Calcium Concentration		250	246	4

4.1.2 User-Centric Evaluation

The usability of the system is yet another important consideration apart from its efficacy. Systems that have bad user experiences, such as unnecessary complexity, workflow inconsistency, and distraction, lead to cognitive burdens on the user and results in limited usability. In this paper, we have also evaluate the usability aspect of the proposed system. The system is evaluated by 11 participants with different experience levels and expertise with healthcare applications. The system is evaluated in an end-to-end manner including tasks such as patient registration to the final recommendation generation and prescription persistence. The CDSS features under evaluation include user interfaces for recommendation generation, consistency of the user interfaces, timeliness of the relevant information, visualization of the clinical parameters, medication dosage-related pop-ups, among others. Participants' responses are acquired through a widely popular user experience evaluation questionnaire. Questionnaires are widely used as a research instrument for effective user experience evaluation. The User Experience Questionnaire (UEQ) compares the level of experience and assessed scale means of participants with a benchmark dataset of 4818 people across 163 studies on various services [85]. Figure 4.4 lists a number of key items describing a distinct quality aspect of an interactive product identified by usability experts. UEQ contains six user experience (UX) aspect scales with 26 items. Items belonging to a specific group are similar in meaning but represent different aspects of the system for a given aspect scale. The Cronbach's

alpha coefficient is a well-known metric for determining the average value per item [119]. Figure 4.4 demonstrates that the 50 percent mean values are more than or equal to 1.5, confirming the proposed system's substantial positive impact on the UX of the participants.

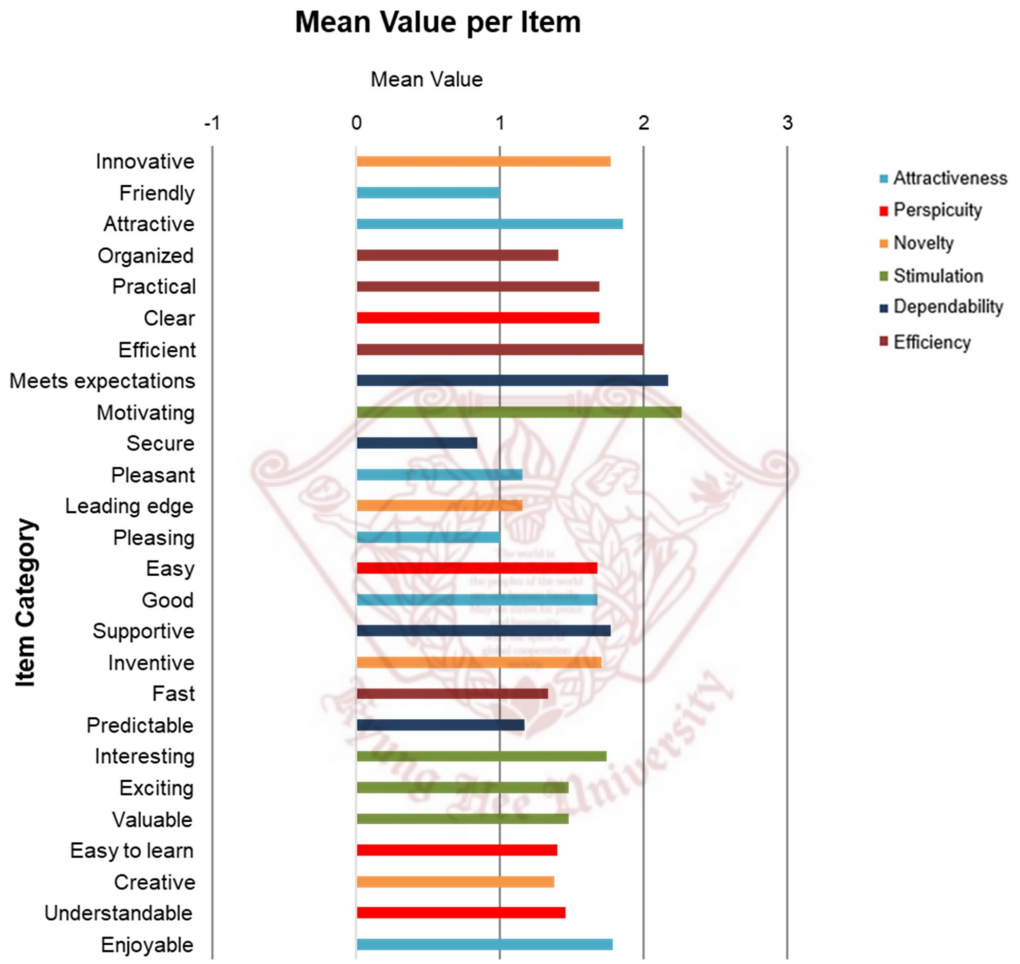


Figure 4.4: Scale mean value per item for multi-aspect user experience (UX) evaluation.

The overall six scales are attractiveness, perspicuity, efficiency, stimulation, and novelty. In this regard, attractiveness is a pure valence dimension. Furthermore, perspicuity, efficiency, and dependability are pragmatic quality aspects (goal-directed), while stimulation and novelty are hedonic quality aspects (not goal-directed). Attractiveness represents an overall impression of the system. Perspicuity characterizes ease of familiarity with the system, efficiency represents whether users can solve their task without unnecessary effort. The dependability aspect denotes if user feels

in control of the interaction. Stimulation represents whether it is an exciting and motivating product to use or not. And finally, novelty characterizes whether the system catches the interest of users or not? As shown in Figure 4.5, the analysis of UEQ support is used to determine the means of stimulation, attractiveness, perspicuity, dependability, efficiency, and novelty scales [119, 120] in the 0 to 2 range. The value of the dependability scale is close to 2.0, [121] indicating that the proposed system induces confidence in the decision-making of the participants.The 95% confidence intervals for the UEQ scale mean are used to evaluate the confidence interval (a measure of the precision of mean estimation) [122]. The mean confidence scores calculated are 1.452, 1.529, 1.475, 1.581, 1.456, 1.512 for attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty, respectively, as shown in Figure 4.5. The UEQ tool compares the UX of the proposed system with that of other services [122]. As indicated in Figure 4.6, the system provides higher dependability due to the transparency of its decision-making along with the inculcating of domain knowledge. Moreover, other aspects such as perspicuity, efficiency, and stimulation are also in the "Good" range, indicating a general acceptance across the participants. The attractiveness aspect of the system is "Above Average", while the novelty aspect is also reasonably high, indicating participants' interest in using the system [123].

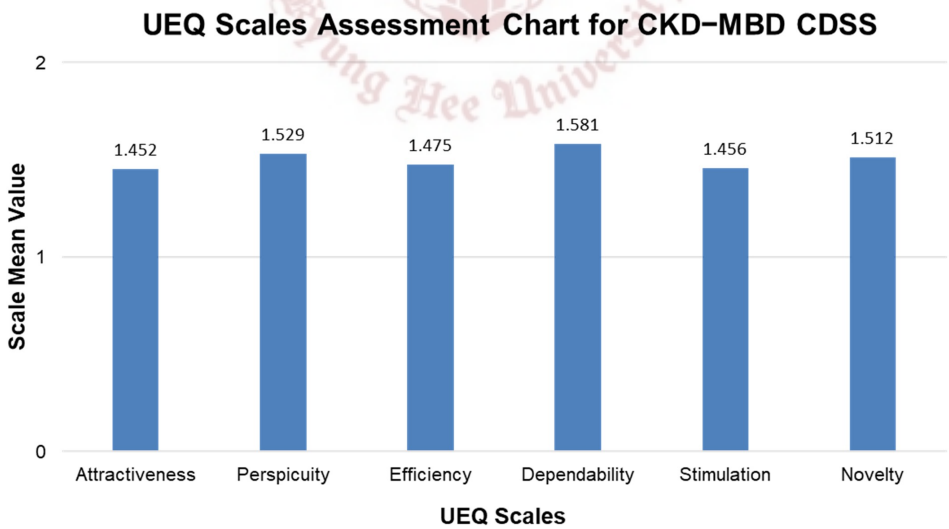


Figure 4.5: User Experience Questionnaire (UEQ) scale values for key 6 aspect dimensions.

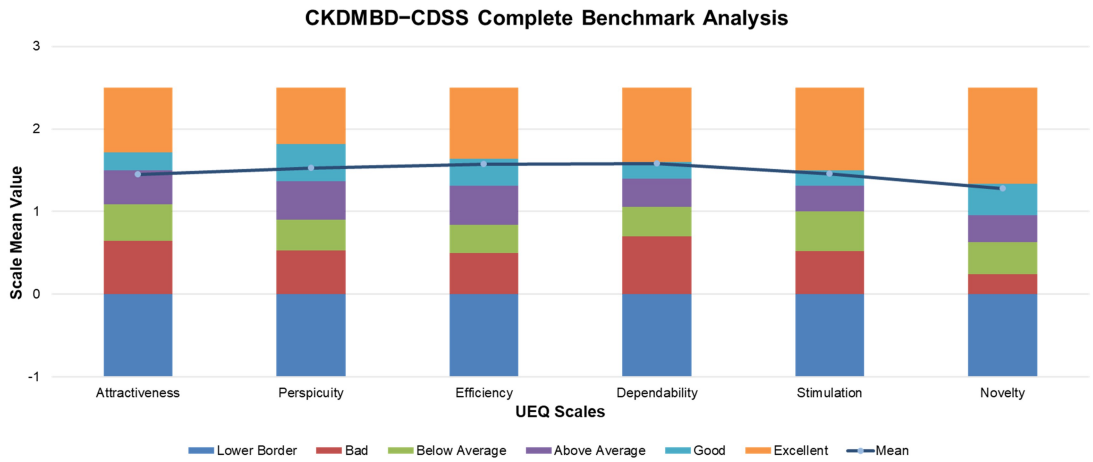


Figure 4.6: CKD–MBD CDSS Benchmark Analysis.

4.2 Discussion

The domain model plays a critical role in the proposed methodology as it guides the case adaptation operation along with providing it with a subset of relevant cases for estimating the medication dosage. The compliance between the clinical practice and the domain model depends on the overall compliance rate, which is generally high due to the fact that some of the highly frequent recommendations have a high compliance rate. In this regard, it is observed that among 33 different generic recommendations, only a few of the recommendations are more frequent, as shown in Figure 4.7. The medication dosage is mostly kept consistent, avoiding abrupt changes from one encounter to another, which corroborates long-term treatment regimens. Therefore, both T16 and T17 recommend to "maintain" the medication dosage in general, while suggesting little changes in the medication selection and dosing. The aforementioned observation also explains the relatively higher frequency of these two recommendations. The striking gap between real-world practice and algorithm-directed recommendation lies in non-calcium-based phosphate binders (NCPB). In most cases, physicians did not increase the dosage of NCPB despite elevated serum phosphate levels. The side effects of NCPB, which frequently causes nausea, vomiting, and abdominal discomfort, may be behind this lack of increases in dosage. It could also be the case thay physicians were likely to be reluctant to actively prescribe NCPB due to pill burden, since more than six tablets a day are required to meet recommendations in some cases. Our results reflect the practi-

cal difficulties of lowering elevated phosphate levels which are encountered by most physicians. Moreover, in some cases, the serum phosphate levels remained slightly higher than the upper limit of the target range, prompting the system to increase NCPB dosage while the clinician opted to maintain the current dosage. This behavior is due to the crisp nature of the rule-base, with little tolerance for on-the-edge cases.

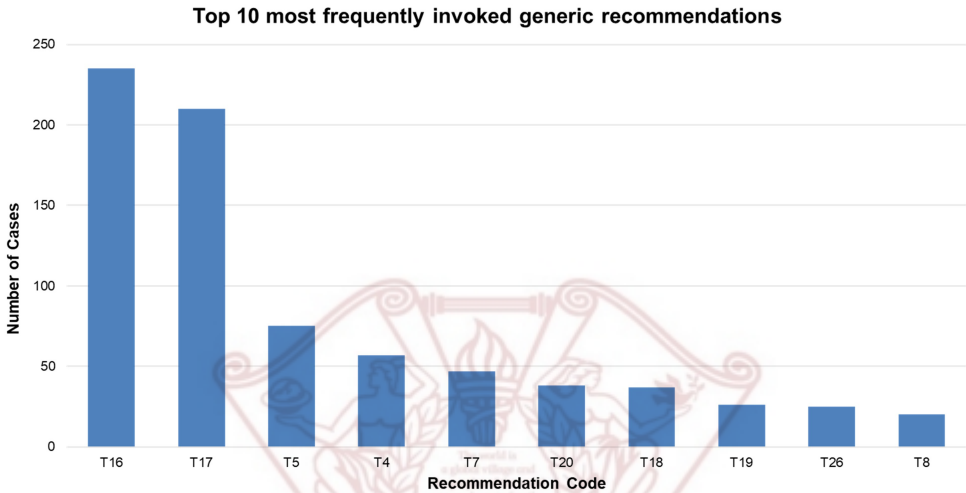


Figure 4.7: Top 10 frequent generic recommendations.

As recommended by the KDIGO guidelines, dialysate calcium concentration in most cases was 1.25 mmol/L, and “maintain” was the most frequent recommendation. Since dialysate calcium concentration was usually unchanged, the overall compliance rate was the highest. In a few exceptional cases of severely low blood calcium levels, “increase” was provided as a last resort. In the case of medication dosage recommendations, a high level of concordance is found for Cinacalcet. This can be explained by Cinacalcet being a single medication option available for prescription in the class of Calcimimetics within the scope of the proposed system. In all other medication classes, there are at least two medication options available, e.g., both Sevelamer and Lanthanum fall under the medication class of NCPB, whereas the CPB medication class includes medications such as Calcium Carbonate and Calcium Acetate. In terms of user-centric evaluation, the proposed system obtained a high score on dependability. This can be attributed to several reasons, such as adopting domain knowledge in decision making, indicating the selected relevant cases, quantifying patient-important outcomes in the form of the PII, and evaluating the system with real-world

patient data. The perspicuity aspect of the system, on the other hand, is also underscored by these results, as the system is user-friendly to navigate and the required information for decision making is readily available. The attractiveness of the system can be enhanced by improving the user interfaces (UI), such as de-cluttering the UI elements from the recommendation panel and demarking clear boundaries when multiple information panels are displayed in close proximity, such as in the cases of laboratory test results, prescribed medications, and recommended medications. In terms of limitations, the proposed system heavily relies on domain knowledge. Acquiring accurate and consensus-based knowledge from domain experts would therefore pose a challenge when adopting the proposed methodology where CPGs are not readily available. Furthermore, in the context medication dosage estimation, all the pre-defined partitions must have member cases associated with them. In the case of a partition that does not contain any clinical case, the dosage estimation operation cannot be efficiently performed.

4.2.1 Merits and demerits of the proposed methodology

The proposed methodology is based on a hybrid approach that combines two partial models to generate a complex domain knowledge compliant recommendation. One model captures the general level of consensus among the decision makers while the second model provides operational details of the recommendation generation based on decision makers' subjective experience. Therefore, the proposed approach is based on the synergy of two incompatible models that capture different aspects of the recommendation. In this regard, following are some of the operational aspects along with merits and demerits of the proposed solutions:

Active case-base partitioning

1. Knowledge acquisition from domain experts is a time consuming task. Without an accurate domain model the consensus of the decision makers is difficult to obtain. Therefore, it is one of the critical considerations for the accurate and domain compliant recommendation generation.

2. Major benefit of the active-case partitioning is to eagerly identify case partitions that have insufficient number of cases for recommendation generation. Therefore, it guides the case acquisition process in order to build the initial case base.

3. Another major benefit of active case-base partitioning is to generate variable size distinct partitions without specifying partition size by the user. In this manner, a set of candidate cases are identified where the selection is grounded on domain knowledge and clinical consensus.

Case selection based on feedback

1. Case selection is an important step within the proposed approach. It is specifically designed to capture longitudinal information available within the case base. A set of reference cases are selected from candidate cases for recommendation adaptation. One demerit of this approach is insufficient reference cases where partition size is already small.

2. Due to the inherent interpretability of the recommendation generation process, the decision maker clearly inspects whether the generated recommendation is based on reference cases or candidate cases.

3. Reference cases capture with clinical experience of clinicians where the prescribed treatment had positive impact on the patient's medical condition. This change is measured in terms of whether the obtained laboratory medical test results are within the reference ranges or not as compared to the previous encounter?

4. Patient improvement indicator quantifies the efficacy of a prescribed treatment. Along with case selection, it can also be used for data analytics to show longitudinal impact of various treatments over a period of time.

Case adaptation based on hybrid approach

1. Case adaptation consolidates multiple recommendations acquired from selected similar cases (i.e. candidate or referenced cases) and generates a unified recommendation that is compliant with the overall domain knowledge and also contains the necessary operational details for actionable recommendation.

2. One major demerit of the proposed case adaptation solution is imprecise recommendation. For such scenarios where the selected cases are sparse or contain a wide variation in the solution component, the resultant recommendation becomes general in nature.

Table 4.2: Merits and demerits of the proposed contributions

Contributions	Merits	Demerits
Active case-base partitioning	1. Assist in case acquisition for constructing an initial case-base 2. Determine status of different partitions and identify case insufficiency 3. Variable size partitions without user-specified parameters	1. Involvement of domain experts is time consuming 2. Multiple alternative solutions are difficult to capture in a decision tree model or data structure 3. Inference time may increase with large number of rules
Multi-level case selection	1. Enhances interpretability of the recommendation 2. Identify positive and negative experiences of the decision makers to guide future decisions 3. Capture longitudinal trends based on recommendations' efficacy i.e. positive or negative impact of treatment	1. Fewer data points for statistical operations e.g. point estimation using regression techniques
Hybrid case adaptation	1. Incorporate domain knowledge in recommendation generation along with operational details 2. Easily customizable with suitable statistical techniques based on the dataset 3. Preserve interpretability of the final recommendation i.e. decision maker can inspect decision making process	1. Imprecise recommendation due to fewer reference cases 2. Multiple alternative candidate recommendations are difficult to reflect in the final consolidated recommendation.

3. The major benefit of the proposed solution is that the generated solution reflects characteristics of both the consensus-based general framework for the given scenario and operations details to make the recommendation more actionable.

4. Furthermore, based on the nature of dataset, the case adaptation can easily extend the proposed solution by incorporating other related techniques for point estimation and/or range estimation such as regression models or deep learning models.

4.2.2 Evaluation of the proposed methodology

The proposed methodology is evaluated with real world data of chronic kidney disease - mineral and bone disorder patients i.e. longitudinal records over a 13 month period.The dataset contained both physiological measurements of the patients and prescribed treatments. A subset of the dataset i.e. 30%, is used to evaluate the concordance of the proposed methodology with that of routine clinical practice. In this regard, the generated recommendation of 250 cases are provided to clinicians to determine whether the recommendations by the proposed methodology are correct or not? Therefore, a retrospective evaluation approach is used to determine the efficacy of the proposed recommendations with an active involvement of the clinicians. Moreover, in order to determine the impact of the proposed decision support system, 11 clinicians participated in the study and operated the CKD-MBD CDSS for over a month (as a stand along application). In this evaluation, the objective is to quantify the overall user experience of clinicians. Although the participants

have shown willingness to integrate the CDSS into their clinical practice, the user interface of the CDSS and its seamless integration into the hospital management system remained important aspects to improve. The interpretability of the CDSS played a major role in attaining trust of the clinicians, therefore, it is identified as an other important avenue for further enhancement.

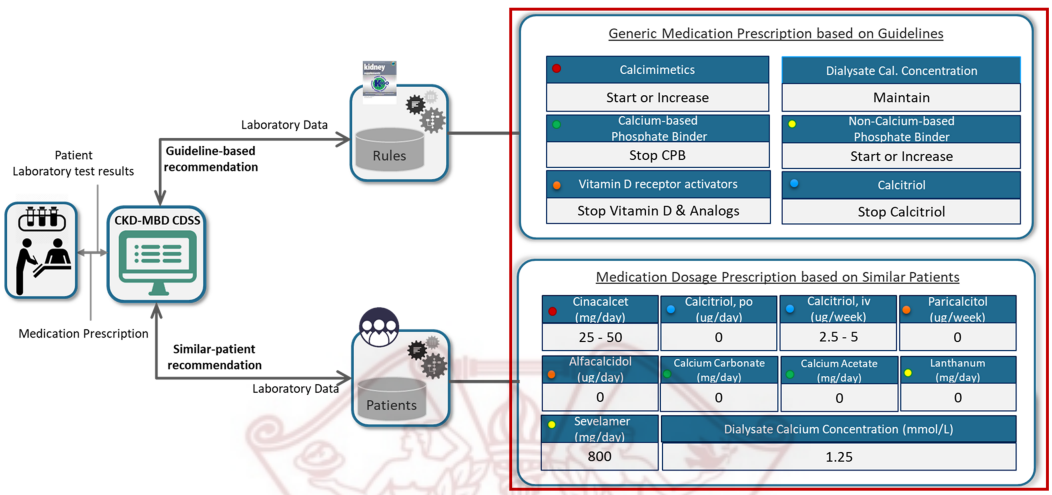


Figure 4.8: A sample multi-factor recommendation based on domain model and referenced cases.

4.2.3 Assistance to clinicians in clinical settings

Figure 4.8 depicts two types of recommendations that are provided to the clinician. First one is recommendation based on domain knowledge. This recommendation is generic in nature and provides general guidelines to clinicians. Second, recommendation is more specific and provides operational details for realizing the guideline-based general recommendation based on positive past experiences as depicted in Figure 4.8. Moreover, Table 4.3 provides a list of 33 categories against which domain model provides generic recommendations. These categories are identified through referencing clinical guidelines.

It is important to note that the dosage recommendation is calculated based on the direction from generic recommendations and high prospect similar cases. In this regard, as indicated in Table 4.3, the generic recommendation for Calcimimetics is “Start or Increase”, Cinacalcet is therefore recommended between 25 mg/day and 50 mg/day. Dosage range estimation is based on IQR of high prospect similar cases.

Table 4.3: A list of 33 generic recommendations

Rcode	Calcimimetics	Calcitriol	Vitamin D & Analogs	CPB	NCPB	Dialysate Calcium Concentration
T1	Start or Increase	Stop	Stop	Stop	Start or Increase	Reduce by 0.25 mmol/L If more than 1.5 mmol/L
T2	Start or Increase	Stop	Stop	Stop	As it is	Reduce by 0.25 mmol/L If more than 1.5 mmol/L
T3	Start or Increase	Stop	Consider Vitmain D Analogs	Stop	Decrease or Stop	Reduce by 0.25 mmol/L If more than 1.5 mmol/L
T4	Start or Increase	Stop	Consider Vitmain D Analogs	Stop	Start or Increase	Maintain current dialysate calcium concentration
T5	Start or Increase	As it is	Consider Vitmain D Analogs	As it is	As it is	Maintain current dialysate calcium concentration
T6	Start or Increase	As it is	Consider Vitmain D Analogs	Stop	Decrease or Stop	Maintain current dialysate calcium concentration
T7	As it is	As it is	Consider Vitmain D Analogs	As it is	Start or Increase	Maintain current dialysate calcium concentration
T8	As it is	Consider Calcitriol	Consider Vitmain D Analogs	As it is	As it is	Maintain current dialysate calcium concentration
T9	As it is	Consider Calcitriol	Consider Vitmain D Analogs	Decrease or Stop	Decrease or Stop	Maintain current dialysate calcium concentration
T10	Decrease	Consider Calcitriol	Consider Vitmain D Analogs	Start or Increase	As it is	Increase by 0.25 mmol/L
T11	Decrease	Start or Increase	Consider Vitmain D Analogs	Start or Increase	As it is	Increase by 0.25 mmol/L
T12	Decrease	Start or Increase	Consider Vitmain D Analogs	As it is	Decrease or Stop	Increase by 0.25 mmol/L
T13	As it is	Stop	Stop Vitamin D Analogs	Stop	Start or Increase	Reduce by 0.25 mmol/L If more than 1.5 mmol/L
T14	As it is	Stop	Stop Vitamin D Analogs	Stop	As it is	Reduce by 0.25 mmol/L If more than 1.5 mmol/L
T15	As it is	Stop	Stop Vitamin D Analogs	Stop	Decrease	Reduce by 0.25 mmol/L If more than 1.5 mmol/L
T16	As it is	As it is	As it is	As it is	Start or Increase	Maintain current dialysate calcium concentration
T17	As it is	As it is	As it is	As it is	As it is	Maintain current dialysate calcium concentration
T18	Stop or Decrease	As it is	As it is	As it is	Decrease	Maintain current dialysate calcium concentration
T19	Stop or Decrease	As it is	As it is	Start or Increase	Start or Increase	Increase by 0.25 mmol/L
T20	Stop or Decrease	Start or Increase	As it is	Start or Increase	Decrease or Stop	Increase by 0.25 mmol/L
T21	Stop or Decrease	Start or Increase	As it is	As it is	Stop	Increase by 0.25 mmol/L
T22	Decrease or Stop	Stop	Stop	Stop	Start or Increase	Reduce by 0.25 mmol/L If more than 1.5 mmol/L
T23	Decrease or Stop	Stop	Stop	Stop	As it is	Reduce by 0.25 mmol/L If more than 1.5 mmol/L
T24	Decrease or Stop	Stop	Stop	Stop	Decrease	Reduce by 0.25 mmol/L If more than 1.5 mmol/L
T25	Decrease or Stop	Decrease or Stop	Decrease or Stop	Stop	Start or Increase	Maintain current dialysate calcium concentration
T26	Decrease or Stop	As it is	Decrease or Stop	As it is	Decrease or Stop	Maintain current dialysate calcium concentration
T27	Decrease or Stop	As it is	Decrease or Stop	Stop	Decrease or Stop	Maintain current dialysate calcium concentration
T28	Decrease or Stop	As it is	Decrease or Stop	As it is	Start or Increase	Maintain current dialysate calcium concentration
T29	Decrease or Stop	Decrease or Stop	Decrease or Stop	As it is	As it is	Maintain current dialysate calcium concentration
T30	Decrease or Stop	As it is	Decrease or Stop	Stop	Decrease or Stop	Maintain current dialysate calcium concentration
T31	Decrease or Stop	Decrease or Stop	Decrease or Stop	Start or Increase	Start or Increase	Increase by 0.25 mmol/L
T32	Decrease or Stop	Decrease or Stop	Decrease or Stop	Start or Increase	As it is	Increase by 0.25 mmol/L
T33	Decrease or Stop	Start or Increase	Decrease or Stop	Decrease or Stop	Decrease or Stop	Increase by 0.25 mmol/L

Table 4.4: A sample medication dosage recommendation with respect to generic recommendation.

Management Class	**Dosage Recommendation**	**Reference Dosage Range**
Calcimimetics: Start or Increase	Cinacalcet: 25 mg/day–50 mg/day	Cinacalcet: 0∼100 mg/day
Calcitriol: Stop Calcitriol	Calcitriol, po: 0 ug/day	Calcitriol, po: 0∼2.0 ug/day Calcitriol, iv: 0∼10 ug/week
Vitamin D and Analogs: Stop Vitamin D and Analogs	Paricalcitol, iv: 0 ug/week Alfacalcidol: 0 ug/day	Paricalcitrol, iv: 0∼50 ug/week Alfacalcidol: 0∼3 ug/day
Calcium-based Phosphate Binder: Stop CPB	Calcium Carbonate: 0 mg/day Calcium Acetate: 0 mg/day	Calcium Carbonate: 0∼3750 mg/day Calcium Acetate: 0∼6000 mg/day
Non-Calcium-based Phosphate Binder: Start of Increase NCPB	Sevelamer: 800 mg/day Lanthanum: 0 mg/day	Sevelamer: 0∼13,000 mg/day Lanthanum: 0∼3750 mg/day
Dialysate calcium concentration: Maintain current dialysate calcium concentration	Dialysate Calcium Concentration: 1.25 mmol/L	Dialysate Calcium Concentration: 1.25∼1.75 mmol/L

The main objective of the proposed methodology is to align clinical cases with that of domain knowledge. In practice, the clinicians may not fully adhere to direction given in the clinical guidelines due to a number of reasons such as insurance coverage of the patient, commodities that are not covered in the scope of the guidelines, unavailability or shortage of certain medications, etc. These operational aspects that are not within the scope of clinical guidelines are not addressed in the proposed methodology. Although the methodology can be extended to generate multiple alternative recommendations based on positive past experiences that show domain-model compliant and non-compliant results. As the stated objective is to align two complementary models on different level of granularity, the alternative recommendation generation aspect is not explored in this thesis.

Chapter 5

Conclusion

This chapter concludes the thesis and provides future directions for extending the proposed methodology.

5.1 Conclusion and Future Research Directions

CDSSs assist clinicians and healthcare providers in both complex decision-making and addressing routine healthcare tasks. CDSSs process and analyze healthcare data, e.g., laboratory and imaging test results, in addition to medication history in order to provide prompts and reminders at the point of care. Applied to CKD–MBD management, CDSSs can assist clinicians in the selection of appropriate treatment protocols and tailored recommendations based on the status of vascular calcification. In this study, a hybrid knowledge modeling approach is proposed that incorporates both domain knowledge and patients' clinical cases for complex decision making, such as appropriate initiation, modification, monitoring, or discontinuation of the medication. Furthermore, we propose a PII which provides an overall summary of the patient record over a period of time. The PII is helpful in identifying past similar cases that have positive patient-important outcomes, e.g., patient laboratory tests that have improved with the prescribed medication regimen, so that similar patients may also be recommended the same medication regimens. Medication dosage estimation is performed on reference cases (acquired from similar patient cases) using the IQR to assist clinicians in selecting appropriate dosing. The proposed system is evaluated based on 250 clinical cases from hemodialysis patients and the overall concordance is found at 78.27recommendations and the routine clinical practice. A widely used user experience evaluation tool, UEQ, is used to evaluate the proposed systems' usability aspects with respect to clinicians. The usability assessment is based on clinicians who have independently evaluated the system. The dependabil-

ity and perspicuity of the system scored highly, while its attractiveness remained relatively low across the participants. This shows that the system provides useful recommendations along with initiative workflows that seamlessly align with clinical workflows, whereas information displaying panels can be further improved to de-clutter the user interface. We intend to expand the automated decision-making framework to other comorbidities of CKD–MBD, such as cardiovascular disease, osteoporosis, diabetes, among others. Moreover, patient data from multiple medical centers will be acquired to reflect sufficient diversity of different treatment approaches adopted by clinicians. Bayesian reasoning along with deep learning approaches are some of the candidate approaches that will be evaluated for hybridization along with domain knowledge.

This line of research can be extended in a number of directions such as augmenting analysis of both improved and unimproved similar cases for medication selection as it will provide the clinician with more information as to what treatment options are not productive in past. Moreover, multiple alternative recommendations can also be explored for a given case. Lastly, apart from treatment services, disease diagnosis based on interpretable model along with counterfactual scenarios augmentation can be useful to comprehensively cover the different facets of the CDSS services.

Bibliography

[1] L. Souza-Pereira, N. Pombo, S. Ouhbi, V. Felizardo, and N. Garcia, "Clinical decision support systems for chronic diseases: A systematic literature review," *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105565, 2020.

[2] N. Garcelon, A. Burgun, R. Salomon, and A. Neuraz, "Electronic health records for the diagnosis of rare diseases," *Kidney International*, vol. 97, no. 4, pp. 676–686, 2020.

[3] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, "An overview of clinical decision support systems: benefits, risks, and strategies for success," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–10, 2020.

[4] M. Laka, A. Milazzo, and T. Merlin, "Factors that impact the adoption of clinical decision support systems (cdss) for antibiotic management," *International journal of environmental research and public health*, vol. 18, no. 4, p. 1901, 2021.

[5] T. N. Poly, M. M. Islam, M. S. Muhtar, H.-C. Yang, P. A. A. Nguyen, and Y.-C. J. Li, "Machine learning approach to reduce alert fatigue using a disease medication–related clinical decision support system: model development and validation," *JMIR medical informatics*, vol. 8, no. 11, p. e19489, 2020.

[6] T. N. Poly, M. M. Islam, H.-C. Yang, and Y.-C. J. Li, "Appropriateness of overridden alerts in computerized physician order entry: systematic review," *JMIR medical informatics*, vol. 8, no. 7, p. e15653, 2020.

[7] J. Yoo, J. Lee, P.-L. Rhee, D. K. Chang, M. Kang, J. S. Choi, D. W. Bates, W. C. Cha *et al.*, "Alert override patterns with a medication clinical decision support system in an academic

emergency department: retrospective descriptive study," *JMIR Medical Informatics*, vol. 8, no. 11, p. e23351, 2020.

[8] A. Levin, G. Bakris, M. Molitch, M. Smulders, J. Tian, L. Williams, and D. Andress, "Prevalence of abnormal serum vitamin d, pth, calcium, and phosphorus in patients with chronic kidney disease: results of the study to evaluate early kidney disease," *Kidney international*, vol. 71, no. 1, pp. 31–38, 2007.

[9] A. M. Alem, D. J. Sherrard, D. L. Gillen, N. S. Weiss, S. A. Beresford, S. R. Heckbert, C. Wong, and C. Stehman-Breen, "Increased risk of hip fracture among patients with end-stage renal disease," *Kidney international*, vol. 58, no. 1, pp. 396–399, 2000.

[10] A. Pimentel, P. Ureña-Torres, M. C. Zillikens, J. Bover, and M. Cohen-Solal, "Fractures in patients with ckd—diagnosis, treatment, and prevention: a review by members of the european calcified tissue society and the european renal association of nephrology dialysis and transplantation," *Kidney international*, vol. 92, no. 6, pp. 1343–1355, 2017.

[11] G. A. Block, R. D. Kilpatrick, K. A. Lowe, W. Wang, and M. D. Danese, "Ckd–mineral and bone disorder and risk of death and cardiovascular hospitalization in patients on hemodialysis," *Clinical Journal of the American Society of Nephrology*, vol. 8, no. 12, pp. 2132–2140, 2013.

[12] M. Ketteler, G. A. Block, P. Evenepoel, M. Fukagawa, C. A. Herzog, L. McCann, S. M. Moe, R. Shroff, M. A. Tonelli, N. D. Toussaint *et al.*, "Diagnosis, evaluation, prevention, and treatment of chronic kidney disease–mineral and bone disorder: Synopsis of the kidney disease: improving global outcomes 2017 clinical practice guideline update," *Annals of internal medicine*, vol. 168, no. 6, pp. 422–430, 2018.

[13] S. Fishbane, W. B. Shapiro, D. B. Corry, S. L. Vicks, M. Roppolo, K. Rappaport, X. Ling, W. G. Goodman, S. Turner, and C. Charytan, "Cinacalcet hcl and concurrent low-dose vitamin d improves treatment of secondary hyperparathyroidism in dialysis patients compared with vitamin d alone: the achieve study results," *Clinical Journal of the American Society of Nephrology*, vol. 3, no. 6, pp. 1718–1725, 2008.

[14] F. Tentori, M. J. Blayney, J. M. Albert, B. W. Gillespie, P. G. Kerr, J. Bommer, E. W. Young, T. Akizawa, T. Akiba, R. L. Pisoni *et al.*, "Mortality risk for dialysis patients with different levels of serum calcium, phosphorus, and pth: the dialysis outcomes and practice patterns study (dopps)," *American Journal of Kidney Diseases*, vol. 52, no. 3, pp. 519–530, 2008.

[15] Y. Slinin, R. N. Foley, and A. J. Collins, "Calcium, phosphorus, parathyroid hormone, and cardiovascular disease in hemodialysis patients: the usrds waves 1, 3, and 4 study," *Journal of the American Society of Nephrology*, vol. 16, no. 6, pp. 1788–1793, 2005.

[16] J. Kolodner, *Case-based reasoning*. Morgan Kaufmann, 2014.

[17] M. M. Richter and R. O. Weber, *Case-based reasoning*. Springer, 2016.

[18] I. Watson and F. Marir, "Case-based reasoning: A review," *The knowledge engineering review*, vol. 9, no. 4, pp. 327–354, 1994.

[19] J. L. Kolodner, "An introduction to case-based reasoning," *Artificial intelligence review*, vol. 6, no. 1, pp. 3–34, 1992.

[20] C. K. Riesbeck and R. C. Schank, *Inside case-based reasoning*. Psychology Press, 2013.

[21] I. Bichindaritz and C. Marling, "Case-based reasoning in the health sciences: What's next?" *Artificial intelligence in medicine*, vol. 36, no. 2, pp. 127–135, 2006.

[22] M. M. Richter and A. Aamodt, "Case-based reasoning foundations," *The Knowledge Engineering Review*, vol. 20, no. 3, pp. 203–207, 2005.

[23] Y. Shen, J. Colloc, A. Jacquet-Andrieu, and K. Lei, "Emerging medical informatics with case-based reasoning for aiding clinical decision in multi-agent system," *Journal of biomedical informatics*, vol. 56, pp. 307–317, 2015.

[24] R. Miotto and C. Weng, "Case-based reasoning using electronic health records efficiently identifies eligible patients for clinical trials," *Journal of the American Medical Informatics Association*, vol. 22, no. e1, pp. e141–e150, 2015.

[25] K. A. Kumar, Y. Singh, and S. Sanyal, "Hybrid approach using case-based reasoning and rule-based reasoning for domain independent clinical decision support in icu," *Expert Systems with Applications*, vol. 36, no. 1, pp. 65–71, 2009.

[26] I. Bichindaritz, "Memoire: A framework for semantic interoperability of case-based reasoning systems in biology and medicine," *Artificial intelligence in medicine*, vol. 36, no. 2, pp. 177–192, 2006.

[27] M. A. Mohammed, M. K. Abd Ghani, N. a. Arunkumar, O. I. Obaid, S. A. Mostafa, M. M. Jaber, M. Burhanuddin, B. M. Matar, D. A. Ibrahim *et al.*, "Genetic case-based reasoning for improved mobile phone faults diagnosis," *Computers & Electrical Engineering*, vol. 71, pp. 212–222, 2018.

[28] O. N. Oyelade and A. E. Ezugwu, "A case-based reasoning framework for early detection and diagnosis of novel coronavirus," *Informatics in Medicine Unlocked*, vol. 20, p. 100395, 2020.

[29] D. Rossille, J.-F. Laurent, and A. Burgun, "Modelling a decision-support system for oncology using rule-based and case-based reasoning methodologies," *International journal of medical informatics*, vol. 74, no. 2-4, pp. 299–306, 2005.

[30] J. Bajo, J. F. de Paz, Y. de Paz, and J. M. Corchado, "Integrating case-based planning and rptw neural networks to construct an intelligent environment for health care," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5844–5858, 2009.

[31] J. M. Corchado, J. Bajo, and A. Abraham, "Gerami: Improving healthcare delivery in geriatric residences," *IEEE Intelligent Systems*, vol. 23, no. 2, pp. 19–25, 2008.

[32] J. A. Recio-García, P. A. González-Calero, and B. Díaz-Agudo, "jcolibri2: A framework for building case-based reasoning systems," *Science of Computer Programming*, vol. 79, pp. 126–145, 2014.

[33] S. Montani, L. C. Jain *et al.*, *Successful case-based reasoning applications*. Springer, 2010, vol. 62.

[34] R. L. De Mantaras, D. McSherry, D. Bridge, D. Leake, B. Smyth, S. Craw, B. Faltings, M. L. Maher, M. T COX, K. Forbus *et al.*, "Retrieval, reuse, revision and retention in case-based reasoning," *The Knowledge Engineering Review*, vol. 20, no. 3, pp. 215–240, 2005.

[35] O. Okudan, C. Budayan, and I. Dikmen, "A knowledge-based risk management tool for construction projects using case-based reasoning," *Expert Systems with Applications*, vol. 173, p. 114776, 2021.

[36] S. Chattopadhyay, S. Banerjee, F. A. Rabhi, and U. R. Acharya, "A case-based reasoning system for complex medical diagnosis," *Expert Systems*, vol. 30, no. 1, pp. 12–20, 2013.

[37] B. López, "Case-based reasoning: a concise introduction," *Synthesis lectures on artificial intelligence and machine learning*, vol. 7, no. 1, pp. 1–103, 2013.

[38] X. Blanco, S. Rodríguez, J. M. Corchado, and C. Zato, "Case-based reasoning applied to medical diagnosis and treatment," in *distributed computing and artificial intelligence*. Springer, 2013, pp. 137–146.

[39] G.-N. Zhu, J. Hu, J. Qi, J. Ma, and Y.-H. Peng, "An integrated feature selection and cluster analysis techniques for case-based reasoning," *Engineering Applications of Artificial Intelligence*, vol. 39, pp. 14–22, 2015.

[40] D. Gu, C. Liang, and H. Zhao, "A case-based reasoning system based on weighted heterogeneous value distance metric for breast cancer diagnosis," *Artificial intelligence in medicine*, vol. 77, pp. 31–47, 2017.

[41] M. Smeulers, A. T. Onderwater, M. C. van Zwieten, and H. Vermeulen, "Nurses' experiences and perspectives on medication safety practices: an explorative qualitative study," *Journal of nursing management*, vol. 22, no. 3, pp. 276–285, 2014.

[42] F. Sartori, A. Mazzucchelli, and A. Di Gregorio, "Bankruptcy forecasting using case-based reasoning: The creperie approach," *Expert Systems with Applications*, vol. 64, pp. 400–411, 2016.

[43] A. González-Briones, J. Prieto, F. De La Prieta, E. Herrera-Viedma, and J. M. Corchado, "Energy optimization using a case-based reasoning strategy," *Sensors*, vol. 18, no. 3, p. 865, 2018.

[44] M. R. Khosravani and S. Nasiri, "Injection molding manufacturing process: Review of case-based reasoning applications," *Journal of Intelligent Manufacturing*, vol. 31, no. 4, pp. 847–864, 2020.

[45] E. Hernandez-Nieves, G. Hernández, A. B. Gil-González, S. Rodríguez-González, and J. M. Corchado, "Cebra: A case-based reasoning application to recommend banking products," *Engineering Applications of Artificial Intelligence*, vol. 104, p. 104327, 2021.

[46] S. K. Biswas, D. Devi, and M. Chakraborty, "A hybrid case based reasoning model for classification in internet of things (iot) environment," *Journal of Organizational and End User Computing (JOEUC)*, vol. 30, no. 4, pp. 104–122, 2018.

[47] I.-C. Yeh and T.-K. Hsu, "Building real estate valuation models with comparative approach through case-based reasoning," *Applied Soft Computing*, vol. 65, pp. 260–271, 2018.

[48] N. Choudhury and S. A. Begum, "A survey on case-based reasoning in medicine," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 8, 2016.

[49] J.-B. Lamy, B. Sekar, G. Guezennec, J. Bouaud, and B. Séroussi, "Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach," *Artificial intelligence in medicine*, vol. 94, pp. 42–53, 2019.

[50] B. Perez, C. Lang, J. Henriet, L. Philippe, and F. Auber, "Risk prediction in surgery using case-based reasoning and agent-based modelization," *Computers in Biology and Medicine*, vol. 128, p. 104040, 2021.

[51] E. Guerrero, M.-H. Lu, H.-P. Yueh, and H. Lindgren, "Designing and evaluating an intelligent augmented reality system for assisting older adults' medication management," *Cognitive Systems Research*, vol. 58, pp. 278–291, 2019.

[52] C.-Y. Fan, P.-C. Chang, J.-J. Lin, and J. Hsieh, "A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification," *Applied Soft Computing*, vol. 11, no. 1, pp. 632–644, 2011.

[53] H. Ahn and K.-j. Kim, "Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach," *Applied soft computing*, vol. 9, no. 2, pp. 599–607, 2009.

[54] P.-C. Chang, C.-H. Liu, J.-L. Lin, C.-Y. Fan, and C. S. Ng, "A neural network with a case based dynamic window for stock trading prediction," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6889–6898, 2009.

[55] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[56] S. Ting, S. Kwok, A. H. Tsang, and W. B. Lee, "A hybrid knowledge-based approach to supporting the medical prescription for general practitioners: Real case in a hong kong medical center," *Knowledge-Based Systems*, vol. 24, no. 3, pp. 444–456, 2011.

[57] M. J. Khan, H. Hayat, and I. Awan, "Hybrid case-base maintenance approach for modeling large scale case-based reasoning systems," *Human-centric Computing and Information Sciences*, vol. 9, no. 1, pp. 1–25, 2019.

[58] N. A. Almansour, H. F. Syed, N. R. Khayat, R. K. Altheeb, R. E. Juri, J. Alhiyafi, S. Alrashed, and S. O. Olatunji, "Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study," *Computers in biology and medicine*, vol. 109, pp. 101–111, 2019.

[59] X. Song, L. R. Waitman, Y. Hu, A. S. Yu, D. Robins, and M. Liu, "Robust clinical marker identification for diabetic kidney disease with ensemble feature selection," *Journal of the American Medical Informatics Association*, vol. 26, no. 3, pp. 242–253, 2019.

[60] H. Polat, H. Danaei Mehr, and A. Cetin, "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods," *Journal of medical systems*, vol. 41, no. 4, pp. 1–11, 2017.

[61] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, "A machine learning methodology for diagnosing chronic kidney disease," *IEEE Access*, vol. 8, pp. 20 991–21 002, 2019.

[62] S. I. Ali, H. S. M. Bilal, M. Hussain, J. Hussain, F. A. Satti, M. Hussain, G. H. Park, T. Chung, and S. Lee, "Ensemble feature ranking for cost-based non-overlapping groups: A case study of chronic kidney disease diagnosis in developing countries," *IEEE Access*, vol. 8, pp. 215 623–215 648, 2020.

[63] N. Larios Delgado, N. Usuyama, A. K. Hall, R. J. Hazen, M. Ma, S. Sahu, and J. Lundin, "Fast and accurate medication identification," *NPJ digital medicine*, vol. 2, no. 1, pp. 1–9, 2019.

[64] S. I. Ali, S. W. Jung, H. S. M. Bilal, S.-H. Lee, J. Hussain, M. Afzal, M. Hussain, T. Ali, T. Chung, and S. Lee, "Clinical decision support system based on hybrid knowledge modeling: A case study of chronic kidney disease-mineral and bone disorder treatment," *International Journal of Environmental Research and Public Health*, vol. 19, no. 1, p. 226, 2021.

[65] C. A. Lester, J. Li, Y. Ding, B. Rowell, J. Yang, and R. A. Kontar, "Performance evaluation of a prescription medication image classification model: an observational cohort," *NPJ Digital Medicine*, vol. 4, no. 1, pp. 1–8, 2021.

[66] T. T. Tran, A. J. Richardson, V. M. Chen, and K. Y. Lin, "Fast and accurate ophthalmic medication bottle identification using deep learning on a smartphone device," *Ophthalmology Glaucoma*, 2021.

[67] D. Khongorzul, S.-M. Lee, Y.-K. Kim, and M.-H. Kim, "Image denoising methods based on daecnn for medication prescriptions," *Journal of the Korea Convergence Society*, vol. 10, no. 5, pp. 17–26, 2019.

[68] J. Jouffroy, S. F. Feldman, I. Lerner, B. Rance, A. Burgun, and A. Neuraz, "Medext: combining expert knowledge and deep learning for medication extraction from french clinical texts," *Published online January*, vol. 23, 2020.

[69] O. Solarte-Pabón, A. Blazquez-Herranz, M. Torrente, A. Rodríguez-Gonzalez, M. Provencio, and E. Menasalvas, "Extracting cancer treatments from clinical text written in spanish: A deep learning approach," in *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2021, pp. 1–6.

[70] B. Guzman, I. Metzger, Y. Aphinyanaphongs, H. Grover *et al.*, "Assessment of amazon comprehend medical: Medication information extraction," *arXiv preprint arXiv:2002.00481*, 2020.

[71] M. A. Al-Garadi, Y.-C. Yang, H. Cai, Y. Ruan, K. O'Connor, G.-H. Graciela, J. Perrone, and A. Sarker, "Text classification models for the automatic detection of nonmedical prescription medication use from social media," *BMC medical informatics and decision making*, vol. 21, no. 1, pp. 1–13, 2021.

[72] J. Xie, D. D. Zeng, and Z. A. Marcum, "Using deep learning to improve medication safety: the untapped potential of social media," pp. 375–377, 2017.

[73] B. Fan, W. Fan, C. Smith *et al.*, "Adverse drug event detection and extraction from open data: A deep learning approach," *Information Processing & Management*, vol. 57, no. 1, p. 102131, 2020.

[74] J. Y. Ryu, H. U. Kim, and S. Y. Lee, "Deep learning improves prediction of drug–drug and drug–food interactions," *Proceedings of the National Academy of Sciences*, vol. 115, no. 18, pp. E4304–E4311, 2018.

[75] D. Demner-Fushman, J. G. Mork, W. J. Rogers, S. E. Shooshan, L. Rodriguez, and A. R. Aronson, "Finding medication doses in the liteature," in *AMIA Annual Symposium Proceedings*, vol. 2018. American Medical Informatics Association, 2018, p. 368.

[76] D. Mahajan, J. J. Liang, and C.-H. Tsou, "Extracting daily dosage from medication instructions in ehrs: An automated approach and lessons learned," *arXiv preprint arXiv:2005.10899*, 2020.

[77] Y. Yang, P. Xie, X. Gao, C. Cheng, C. Li, H. Zhang, and E. Xing, "Predicting discharge medications at admission time based on deep learning," *arXiv preprint arXiv:1711.01386*, 2017.

[78] V. Osmani, L. Li, M. Danieletto, B. Glicksberg, J. Dudley, and O. Mayora, "Processing of electronic health records using deep learning: a review," *arXiv preprint arXiv:1804.01758*, 2018.

[79] S. Itani, F. Lecron, and P. Fortemps, "Specifics of medical data mining for diagnosis aid: A survey," *Expert systems with applications*, vol. 118, pp. 300–314, 2019.

[80] K. J. Cios and G. W. Moore, "Uniqueness of medical data mining," *Artificial intelligence in medicine*, vol. 26, no. 1-2, pp. 1–24, 2002.

[81] H. W. Yu, M. Hussain, M. Afzal, T. Ali, J. Y. Choi, H.-S. Han, and S. Lee, "Use of mind maps and iterative decision trees to develop a guideline-based clinical decision support system for routine surgical practice: case study in thyroid nodules," *Journal of the American Medical Informatics Association*, vol. 26, no. 6, pp. 524–536, 2019.

[82] D.-J. Choi, J. J. Park, T. Ali, and S. Lee, "Artificial intelligence for the diagnosis of heart failure," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–6, 2020.

[83] M. Hussain, M. Afzal, K. M. Malik, T. Ali, W. A. Khan, M. Irfan, A. Jamshed, and S. Lee, "Acquiring guideline-enabled data driven clinical knowledge model using formally verified refined knowledge acquisition method," *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105701, 2020.

[84] M. Afzal, M. Hussain, W. A. Khan, T. Ali, S. Lee, E.-N. Huh, H. F. Ahmad, A. Jamshed, H. Iqbal, M. Irfan *et al.*, "Comprehensible knowledge model creation for cancer treatment decision making," *Computers in biology and medicine*, vol. 82, pp. 119–129, 2017.

[85] K. Bach, C. Marling, P. J. Mork, A. Aamodt, F. S. Mair, and B. I. Nicholl, "Design of a clinician dashboard to facilitate co-decision making in the management of non-specific low back pain," *Journal of Intelligent Information Systems*, vol. 52, no. 2, pp. 269–284, 2019.

[86] T. Ali, M. Hussain, W. A. Khan, M. Afzal, J. Hussain, R. Ali, W. Hassan, A. Jamshed, B. H. Kang, and S. Lee, "Multi-model-based interactive authoring environment for creating shareable medical knowledge," *Computer methods and programs in biomedicine*, vol. 150, pp. 41–72, 2017.

[87] H. A. de Wit, C. M. Gonzalvo, J. Cardenas, H. J. Derijks, R. Janknegt, P.-H. M. van der Kuy, B. Winkens, and J. M. Schols, "Evaluation of clinical rules in a standalone pharmacy based clinical decision support system for hospitalized and nursing home patients," *International Journal of Medical Informatics*, vol. 84, no. 6, pp. 396–405, 2015.

[88] I. A. Scott, P. I. Pillans, M. Barras, and C. Morris, "Using emr-enabled computerized decision support systems to reduce prescribing of potentially inappropriate medications: a narrative review," *Therapeutic Advances in Drug Safety*, vol. 9, no. 9, pp. 559–573, 2018.

[89] D. Tawadrous, S. Z. Shariff, R. B. Haynes, A. V. Iansavichus, A. K. Jain, and A. X. Garg, "Use of clinical decision support systems for kidney-related drug prescribing: a systematic review," *American journal of kidney diseases*, vol. 58, no. 6, pp. 903–914, 2011.

[90] J. Ennis, D. Gillen, A. Rubenstein, E. Worcester, M. E. Brecher, J. Asplin, and F. Coe, "Clinical decision support improves physician guideline adherence for laboratory monitoring of chronic kidney disease: a matched cohort study," *BMC nephrology*, vol. 16, no. 1, pp. 1–11, 2015.

[91] E. A. Vogel, S. J. Billups, S. J. Herner, and T. Delate, "Renal drug dosing," *Applied Clinical Informatics*, vol. 7, no. 03, pp. 731–744, 2016.

[92] A. Helldén, F. Al-Aieshy, P. Bastholm-Rahmner, U. Bergman, L. L. Gustafsson, H. Höök, S. Sjöviker, A. Söderström, and I. Odar-Cederlöf, "Development of a computerised decisions support system for renal risk drugs targeting primary healthcare," *BMJ open*, vol. 5, no. 7, p. e006775, 2015.

[93] H. Pirnejad, P. Amiri, Z. Niazkhani, A. Shiva, K. Makhdoomi, S. Abkhiz, H. van der Sijs, and R. Bal, "Preventing potential drug-drug interactions through alerting decision support

systems: a clinical context based methodology," *International journal of medical informatics*, vol. 127, pp. 18–26, 2019.

[94] Z. Niazkhani, M. Fereidoni, P. Rashidi Khazaee, A. Shiva, K. Makhdoomi, A. Georgiou, and H. Pirnejad, "Translation of evidence into kidney transplant clinical practice: managing drug-lab interactions by a context-aware clinical decision support system," *BMC medical informatics and decision making*, vol. 20, no. 1, pp. 1–13, 2020.

[95] T. Shemeikka, P. Bastholm-Rahmner, C.-G. Elinder, A. Vég, E. Törnqvist, B. Cornelius, and S. Korkmaz, "A health record integrated clinical decision support system to support prescriptions of pharmaceutical drugs in patients with reduced renal function: design, development and proof of concept," *International Journal of Medical Informatics*, vol. 84, no. 6, pp. 387–395, 2015.

[96] L. Awdishu, C. R. Coates, A. Lyddane, K. Tran, C. E. Daniels, J. Lee, and R. El-Kareh, "The impact of real-time alerting on appropriate prescribing in kidney disease: a cluster randomized controlled trial," *Journal of the American Medical Informatics Association*, vol. 23, no. 3, pp. 609–616, 2016.

[97] E. Honkanen, L. Kauppila, B. Wikström, P. L. Rensma, J.-M. Krzesinski, K. Aasarod, F. Verbeke, P. B. Jensen, P. Mattelaer, B. Volck *et al.*, "Abdominal aortic calcification in dialysis patients: results of the cord study," *Nephrology Dialysis Transplantation*, vol. 23, no. 12, pp. 4009–4015, 2008.

[98] N. K. Foundation, "K/doqi clinical practice guidelines for bone metabolism and disease in chronic kidney disease," *American journal of kidney diseases: the official journal of the National Kidney Foundation*, vol. 42, no. 4 Suppl 3, pp. S1–S201, 2003.

[99] A. Sobrinho, A. C. D. S. Queiroz, L. D. Da Silva, E. D. B. Costa, M. E. Pinheiro, and A. Perkusich, "Computer-aided diagnosis of chronic kidney disease in developing countries: A comparative analysis of machine learning techniques," *IEEE Access*, vol. 8, pp. 25 407–25 419, 2020.

[100] A. Ogunleye and Q.-G. Wang, "Xgboost model for chronic kidney disease diagnosis," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 17, no. 6, pp. 2131–2140, 2019.

[101] B. Khan, R. Naseem, F. Muhammad, G. Abbas, and S. Kim, "An empirical evaluation of machine learning techniques for chronic kidney disease prophecy," *IEEE Access*, vol. 8, pp. 55 012–55 022, 2020.

[102] Q. Zhou, H. Zhou, and T. Li, "Cost-sensitive feature selection using random forest: Selecting low-cost subsets of informative features," *Knowledge-based systems*, vol. 95, pp. 1–11, 2016.

[103] G. R. Vásquez-Morales, S. M. Martinez-Monterrubio, P. Moreno-Ger, and J. A. Recio-Garcia, "Explainable prediction of chronic renal disease in the colombian population using neural networks and case-based reasoning," *Ieee Access*, vol. 7, pp. 152 900–152 910, 2019.

[104] V. Bolón-Canedo and A. Alonso-Betanzos, "Ensembles for feature selection: A review and future trends," *Information Fusion*, vol. 52, pp. 1–12, 2019.

[105] C.-F. Tsai and Y.-C. Hsiao, "Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches," *Decision support systems*, vol. 50, no. 1, pp. 258–269, 2010.

[106] M. Taradeh, M. Mafarja, A. A. Heidari, H. Faris, I. Aljarah, S. Mirjalili, and H. Fujita, "An evolutionary gravitational search-based feature selection," *Information Sciences*, vol. 497, pp. 219–239, 2019.

[107] B. Seijo-Pardo, V. Bolón-Canedo, and A. Alonso-Betanzos, "On developing an automatic threshold applied to feature selection ensembles," *Information Fusion*, vol. 45, pp. 227–245, 2019.

[108] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "An ensemble of filters and classifiers for microarray data classification," *Pattern Recognition*, vol. 45, no. 1, pp. 531–539, 2012.

[109] F. Min and Q. Liu, "A hierarchical model for test-cost-sensitive decision systems," *Information Sciences*, vol. 179, no. 14, pp. 2442–2452, 2009.

[110] A. Salekin and J. Stankovic, "Detection of chronic kidney disease and selecting important predictive attributes," in *2016 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 2016, pp. 262–270.

[111] Z. Chen, X. Zhang, and Z. Zhang, "Clinical risk assessment of patients with chronic kidney disease by using clinical data and multivariate models," *International urology and nephrology*, vol. 48, no. 12, pp. 2069–2075, 2016.

[112] A. A. Serpen, "Diagnosis rule extraction from patient data for chronic kidney disease using machine learning," *International Journal of Biomedical and Clinical Engineering (IJBCE)*, vol. 5, no. 2, pp. 64–72, 2016.

[113] A. Y. Al-Hyari, A. M. Al-Taee, and M. A. Al-Taee, "Clinical decision support system for diagnosis and management of chronic renal failure," in *2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*. IEEE, 2013, pp. 1–6.

[114] R. Ani, G. Sasi, U. R. Sankar, and O. Deepa, "Decision support system for diagnosis and prediction of chronic renal failure using random subspace classification," in *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2016, pp. 1287–1292.

[115] N. Tazin, S. A. Sabab, and M. T. Chowdhury, "Diagnosis of chronic kidney disease using effective classification and feature selection technique," in *2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec)*. IEEE, 2016, pp. 1–6.

[116] N. Cahyani and M. Muslim, "Increasing accuracy of c4. 5 algorithm by applying discretization and correlation-based feature selection for chronic kidney disease diagnosis," *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 12, no. 1, pp. 25–32, 2020.

[117] M. Hussain, F. A. Satti, S. I. Ali, J. Hussain, T. Ali, H.-S. Kim, K.-H. Yoon, T. Chung, and S. Lee, "Intelligent knowledge consolidation: From data to wisdom," *Knowledge-Based Systems*, vol. 234, p. 107578, 2021.

[118] S. Imran Ali, B. Ali, J. Hussain, M. Hussain, F. A. Satti, G. H. Park, and S. Lee, "Cost-sensitive ensemble feature ranking and automatic threshold selection for chronic kidney disease diagnosis," *Applied Sciences*, vol. 10, no. 16, p. 5663, 2020.

[119] H. B. Santoso, S. Fadhilah, I. Nurrohmah, and W. H. Goodridge, "The usability and user experience evaluation of web-based online self-monitoring tool: Case study human-computer interaction course," in *2016 4th International Conference on User Science and Engineering (i-USEr)*. IEEE, 2016, pp. 122–127.

[120] M. Schrepp, J. Thomaschewski, and A. Hinderks, "Construction of a benchmark for the user experience questionnaire (ueq)," 2017.

[121] H. B. Santoso, M. Schrepp, R. Isal, A. Y. Utomo, and B. Priyogi, "Measuring user experience of the student-centered e-learning environment." *Journal of Educators Online*, vol. 13, no. 1, pp. 58–79, 2016.

[122] M. Schrepp, "User experience questionnaire handbook," *All you need to know to apply the UEQ successfully in your project*, 2015.

[123] H. S. M. Bilal, M. B. Amin, J. Hussain, S. I. Ali, S. Hussain, M. Sadiq, M. A. Razzaq, A. Abbas, C. Choi, and S. Lee, "On computing critical factors based healthy behavior index for behavior assessment," *International Journal of Medical Informatics*, vol. 141, p. 104181, 2020.

Appendix A

List of Acronyms

Acronyms

A list of acronyms used in the manuscript:

CKD-MBD Chronic Kidney Disease-Mineral and Bone Disorder

ESRD End-Stage Renal Disease

CDSS Clinical Decision Support System

PTH Parathyroid Hormone

Ca Calcium

P Phosphate

CPB Calcium-based Phosphate Binder

NCPB Non-Calcium-based Phosphate Binder

UX User Experience

UEQ User Experience Questionnaire

DK Domain Knowledge

CPG Clinical Practice Guidelines

UI User Interface

CART Classification and Regression Tree

CPOE Computerized Provider Order Entry

GP General Practitioner

DLI Drug Laboratory Interactions

EHR Electronic Health Records

DT Decision Tree

IQR Interquartile Range

MRN Medical Record Number

ATT Attractiveness

PQ Pragmatic Quality

HQ Hedonic Quality

PII Patient Improvement Indicator

ANN Artificial Neural Networks

CART Classification And Regression Trees

CKD Chronic Kidney Disease

DFS-CT Direct Feature Selection - Combine Threshold

DFS-TC Direct Feature Selection - Threshold Combine

FSF Feature Scoring Function

GA Genetic Algorithm

GFR Glomerular Filtration Rate

KFRS Kernelized Fuzzy Rough Sets

KNN K-Nearest Neighbor

LASSO Least Absolute Shrinkage and Selection Operator

LG Linear Regression

NB Naive Bayes

PCA Principal Component Analysis

PKR Pakistani Rupees

RF Random Forest

SU Symmetric Uncertainty

SVM Support Vector Machine

TG Test Group

UCI University of California at Irvine

XGBoost Extreme Gradient Boosting

Appendix B

List of Publications

B.1 International Journal Papers [11]

1 **Syed Imran Ali**, Su Woong Jung, Hafiz Syed Muhammad Bilal, Sang-Ho Lee, Jamil Hussain, Muhammad Afzal, Maqbool Hussain, Taqdir Ali, Taechoong Chung, and Sungyoung Lee. ”Clinical Decision Support System Based on Hybrid Knowledge Modeling: A Case Study of Chronic Kidney Disease-Mineral and Bone Disorder Treatment.” International Journal of Environmental Research and Public Health 19, no. 1 (2022): 226.

2 **Syed Imran Ali**, Bilal Ali, Jamil Hussain, Musarrat Hussain, Fahad Ahmed Satti, Gwang Hoon Park, and Sungyoung Lee. ”Cost-sensitive ensemble feature ranking and automatic threshold selection for chronic kidney disease diagnosis.” Applied Sciences 10, no. 16 (2020): 5663.

3 **Syed Imran Ali**, Hafiz Syed Muhammad Bilal, Musarrat Hussain, Jamil Hussain, Fahad Ahmed Satti, Maqbool Hussain, Gwang Hoon Park, Taechoong Chung, and Sungyoung Lee. ”Ensemble feature ranking for cost-based non-overlapping groups: A case study of chronic kidney disease diagnosis in developing countries.” IEEE Access 8 (2020): 215623-215648.

4 Hussain, Musarrat, Fahad Ahmed Satti, **Syed Imran Ali**, Jamil Hussain, Taqdir Ali, Hun-Sung Kim, Kun-Ho Yoon, TaeChoong Chung, and Sungyoung Lee. ”Intelligent knowledge consolidation: From data to wisdom.” Knowledge-Based Systems 234 (2021): 107578.

5 Hussain, Musarrat, Fahad Ahmed Satti, Jamil Hussain, Taqdir Ali, **Syed Imran Ali**, Hafiz Syed Muhammad Bilal, Gwang Hoon Park, Sungyoung Lee, and TaeChoong Chung. ”A

practical approach towards causality mining in clinical text using active transfer learning." Journal of Biomedical Informatics 123 (2021): 103932.

6 Satti, Fahad Ahmed, Musarrat Hussain, Jamil Hussain, **Syed Imran Ali**, Taqdir Ali, Hafiz Syed Muhammad Bilal, Taechoong Chung, and Sungyoung Lee. "Unsupervised Semantic Mapping for Healthcare Data Storage Schema." IEEE Access 9 (2021): 107267-107278.

7 Hussain, Musarrat, Jamil Hussain, Taqdir Ali, **Syed Imran Ali**, Hafiz Syed Muhammad Bilal, Sungyoung Lee, and Taechoong Chung. "Text Classification in Clinical Practice Guidelines Using Machine-Learning Assisted Pattern-Based Approach." Applied Sciences 11, no. 8 (2021): 3296.

8 Afzal, Muhammad, **Syed Imran Ali**, Rahman Ali, Maqbool Hussain, Taqdir Ali, Wajahat Ali Khan, Muhammad Bilal Amin, Byeong Ho Kang, and Sungyoung Lee. "Personalization of wellness recommendations using contextual interpretation." Expert Systems with Applications 96 (2018): 506-521.

9 Ali, Maqbool, **Syed Imran Ali**, Dohyeong Kim, Taeho Hur, Jaehun Bang, Sungyoung Lee, Byeong Ho Kang, and Maqbool Hussain. "uEFS: An efficient and comprehensive ensemble-based feature selection methodology to select informative features." PloS one 13, no. 8 (2018): e0202705.

10 Bilal, Hafiz Syed Muhammad, Muhammad Bilal Amin, Jamil Hussain, **Syed Imran Ali**, Shujaat Hussain, Muhammad Sadiq, Muhammad Asif Razzaq, Asim Abbas, Chunho Choi, and Sungyoung Lee. "On computing critical factors based healthy behavior index for behavior assessment." International journal of medical informatics 141 (2020): 104181.

11 Bilal, Hafiz Syed Muhammad, Muhammad Bilal Amin, Jamil Hussain, **Syed Imran Ali**, Muhammad Asif Razzaq, Musarrat Hussain, Asim Abbas Turi, Gwang Hoon Park, Sun Moo Kang, and Sungyoung Lee. "Towards user-centric intervention adaptiveness: influencing behavior-context based healthy lifestyle interventions." IEEE Access 8 (2020): 177156-177179.

B.2 International Conference Papers [7]

12 **S. I. Ali** and S. Lee, "Ensemble based Cost-Sensitive Feature Selection for Consolidated Knowledge Base Creation," 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM), Taichung, Taiwan, 2020, pp. 1-7, doi: 10.1109/IMCOM48794.2020.9001751.

13 **S. I. Ali**, W. Ali Khan, S. Lee and S. Lee, "Cost-Sensitive Feature Selection using Particle Swarm Optimization: A Post-Processing Approach," 2020 International Conference on Information Networking (ICOIN), Barcelona, Spain, 2020, pp. 97-101, doi: 10.1109/ICOIN48656.2020.9016554.

14 **Ali S.I.**, Lee S. (2019) The Comparative Analysis of Single-Objective and Multi-objective Evolutionary Feature Selection Methods. In: Lee S., Ismail R., Choo H. (eds) Proceedings of the 13th International Conference on Ubiquitous Information Management and Communication (IMCOM) 2019. IMCOM 2019. Advances in Intelligent Systems and Computing, vol 935. Springer, Cham. https://doi.org/10.1007/978-3-030-19063-7-76

15 **Ali S.I.**, Amin M.B., Kim S., Lee S. (2018) A Hybrid Framework for a Comprehensive Physical Activity and Diet Recommendation System. In: Mokhtari M., Abdulrazak B., Aloulou H. (eds) Smart Homes and Health Telematics, Designing a Better Future: Urban Assisted Living. ICOST 2018. Lecture Notes in Computer Science, vol 10898. Springer, Cham. https://doi.org/10.1007/978-3-319-94523-1-9

16 **Ali S.I.**, Kang B.H., Lee S. (2016) Application of Feature Subset Selection Methods on Classifiers Comprehensibility for Bio-Medical Datasets. In: García C., Caballero-Gil P., Burmester M., Quesada-Arencibia A. (eds) Ubiquitous Computing and Ambient Intelligence. UCAmI 2016. Lecture Notes in Computer Science, vol 10069. Springer, Cham. https://doi.org/10.1007/978-3-319-48746-5-4

17 **Ali Syed Imran**, Muhammad Sadiq, Muhammad Afzal and Sungyoung Lee, "Service Curation Framework for Context-Aware Personalzied Recommendations", 2016 International

Symposium on Perception, Action and Cognitive Systems (PACS), Seoul, Korea, Oct 27-28, 2016

18 M. Sadiq, **S. I. Ali**, M. Bilal Amin and S. Lee, ”A Vertex Matcher for Entity Resolution on Graphs,” 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM), Taichung, Taiwan, 2020, pp. 1-4, doi: 10.1109/IMCOM48794.2020.9001799.

B.3 Patents [2]

19 Sungyoung Lee, **Syed Imran Ali**, “METHOD AND SYSTEM FOR ADAPTIVE RECOMMENDATION SELECTION FOR CONTEXT-AWARE PHYSICAL ACTIVITIES”, *Korean Intellectual Property Office*, Applied on: 2020-11-12, Application No. 10-2020-0151240. (Registered)

20 Sungyoung Lee, Sang-Ho Lee, **Syed Imran Ali**, “METHOD AND SYSTEM FOR DOMAIN-BASED ACTIVE CASE PARTITIONING AND CASE ADAPTATION FOR MEDICATION DECISION MODELING”, *Korean Intellectual Property Office*, Applied on: 2021-12-27, Application No. 10-2021-0188378. (Applied)