EHR-Oriented Knowledge Graph System: Toward Efficient Utilization of Non-Used Information Buried in Routine Clinical Practice

Yong Shang, Yu Tian, Min Zhou, Tianshu Zhou, Kewei Lyu, Zhixiao Wang, Ran Xin, Tingbo Liang, Shiqiang Zhu, and Jingsong Li

***Abstract*—Non-used clinical information has negative implications on healthcare quality. Clinicians pay priority attention to clinical information relevant to their specialties during routine clinical practices but may be insensitive or less concerned about information showing disease risks beyond their specialties, resulting in delayed and missed diagnoses or improper management. In this study, we introduced an electronic health record (EHR)-oriented knowledge graph system to efficiently utilize non-used information buried in EHRs. EHR data were transformed into a semantic patient-centralized information model under the ontology structure of a knowledge graph. The knowledge graph then creates an EHR data trajectory and performs reasoning through semantic rules to identify important clinical findings within EHR data. A graphical reasoning pathway illustrates the reasoning footage and explains the clinical significance for clinicians to better understand the neglected information. An application study was performed to evaluate unconsidered chronic kidney disease (CKD) reminding for non-nephrology clinicians to identify important neglected information. The study covered 71,679 patients in non-nephrology departments. The system identified 2,774 patients meeting CKD diagnosis criteria and 10,377 patients requiring high attention. A follow-up study of 5,439 patients showed that 82.1% of patients who met the diagnosis criteria and 61.4% of patients requiring high attention were confirmed to be CKD positive during follow-up research. The application demonstrated that the proposed approach is feasible and effective in clinical information utilization. Additionally, it's valuable as an explainable artificial intelligence to provide interpretable recommendations for specialist physicians to understand the importance of non-used data and make comprehensive decisions.**

***Index Terms*—Knowledge graph, electronic health record, ontology, non-used information.**

Manuscript received August 3, 2020; revised November 5, 2020, April 8, 2021, and May 18, 2021; accepted May 23, 2021. Date of publication May 31, 2021; date of current version July 20, 2021. This work was supported in part by the Major Scientific Project of Zhejiang Lab under Grant 2020ND8AD01, in part by the National Natural Science Foundation of China under Grants 81771936 and 81801796, in part by the National Key Research and Development Program of China under Grant 2018YFC0116901. *(Yong Shang and Yu Tian are co-first authors.) (Corresponding authors: Jingsong Li; Shiqiang Zhu; Tingbo Liang).*

Yong Shang, Yu Tian, Tianshu Zhou, Kewei Lyu, and Zhixiao Wang are with the Engineering Research Center of EMR and Intelligent Expert System, Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou 310027, China (e-mail: yshang6618@126.com; ty.1987823@163.com; alexzhou@163.com; kwlyu@zju.edu.cn; zxwang6890@gmail.com).

Min Zhou and Tingbo Liang are with First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310027, China (e-mail: zyzm1993@163.com; liangtingbo@zju.edu.cn).

Ran Xin and Shiqiang Zhu are with Zhejiang Lab, Hangzhou 310027, China (e-mail: xinran@zhejianglab.com; zhusq@zhejianglab.com).

Jingsong Li is with the Engineering Research Center of EMR and Intelligent Expert System, Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou 310027, China, and also with Zhejiang Lab, Hangzhou 311100, China (e-mail: ljs@zju.edu.cn).

Digital Object Identifier 10.1109/JBHI.2021.3085003

I. Introduction

THE OMISSION of clinical information usage or non-used clinical information has negative implications on healthcare quality [1]. In routine clinical practice, clinicians focus mainly on disease-related information relevant to their specialties. Practitioners focus on the diagnosis and examinations of diseases that belong to their medical department [2] but may be insensitive or less concerned about information showing disease risks beyond their specialties, resulting in delayed, missed, and incorrect diagnoses or improper treatment [1], [3]–[5]. Patients with risks of cross-departmental diseases cannot be identified in time for treatment to be effective, worsening healthcare quality and creating financial burdens.

The probable causes of the omission of clinical information usage are as follows: (1) Specialist physicians lack sufficient knowledge of cross-departmental diseases, causing obstacles in information use. Studies have shown that clinicians may lack the knowledge of diseases related to other departments, leading to improper decisions or failure to diagnose [6], [7]. For example, clinicians in non-nephrology departments with deficient chronic kidney disease (CKD) knowledge may rarely identify abnormal kidney function data and fail to manage CKD progression promptly [8]. Doctors have access to longitudinal records; however, without proper attention or sufficient knowledge, this valuable information is buried in data, creating waste [9]. (2) Another influential factor is that massive, long-term clinical data are not practical for clinicians to comprehensively review due to heavy workloads and limited time [10], [11]. The electronic health record (EHR) data of a patient could comprise hundreds of diagnoses, examinations, and prescriptions, with a duration as long as 10 years. The problem lists and key clinical findings are usually not properly maintained, leading to omitted information if past clinical notes are not completely reviewed [12]. Additionally, extracting important information among numerous clinical data is challenging, particularly finding overall relevant data on a certain disease.

Thus, due to limited cross-departmental disease knowledge, limited time and heavy workload, the utilization of this omitted information in clinical practice by doctors is insufficient. Additionally, providing methods to efficiently utilize non-used information is crucial. Furthermore, to arouse the attention of clinicians concerned about this information beyond their specialties, explainable illustrations of the important clinical findings in data are necessary to provide comprehensive and convincing details for better understanding and acceptance.

Explainable artificial intelligence (AI) is valuable in the medical domain because it provides contextual explanations and makes results retraceable and comprehensible [13]. Classic machine learning (ML) methods induce accurate models but still lack explicit declarative knowledge. The results defy interpretation by clinicians and cause trust problems. Currently, many studies have worked on making AI models explainable through feature interaction effects, causality models, data visualization and knowledge integration [14], [15]. However, the application of explainable AI in medical domains requires more doctor-oriented tools and methods to render the model interpretation and provide explanations of decision-making.

As the latest achievement in semantic web research, knowledge graphs are now valuable methods for information utilization and are applied in many fields [16], [17]. Current studies have shown several advantages of knowledge graphs for the valuable utilization of EHR data and the provision of explainable results. Knowledge graphs represent knowledge and data entities in a standard ontological structure so that the medical concepts in the knowledge graph are unambiguous [18]. Entities are connected through semantic relationships, so the recommendation-related knowledge and data are retraceable. In addition, knowledge graphs use deduction reasoning by integrating knowledge and real-world data. The reasoning pathway can be traced, and the evidence-based approach is more acceptable and interpretable for clinicians [19]. The knowledge graph method uses deduction reasoning and medical evidence to illustrate and explain the clinical significance of the decision support results. Meanwhile, the ML algorithm uses data induction to generate probability, and difficulties arise in interpreting the relationship between the results and various input data [20].

In this study, we introduce an EHR-oriented knowledge graph for the efficient utilization of non-used clinical information. The knowledge graph is created using a 2-level ontology structure based on the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) and clinical practice guidelines (CPGs). Using the combination of clinical knowledge and EHR data, the system provides an approach to help clinicians better recognize important clinical information neglected in practice through semantic reasoning and a graphical explanation of significant findings.

An application study with follow-up evaluation to assess EHR knowledge graphs was performed to help non-nephrology clinicians identify unconsidered CKD patients. Notably, the study aimed to find potential failed diagnoses of CKD in patients with abnormal results buried in EHRs of which non-nephrology clinicians were unaware, and it was not a prediction tool.

II. Related Works

Currently, many approaches have utilized EHR data to improve healthcare quality. Studies have shown the advantages of semantic representation on EHR data and knowledge graphs for effective clinical decision support (CDS) [23], [24], [28]–[30]. Different from these approaches, the proposed method in this study focuses on knowledge barriers between departments and utilizing non-used medical information for improvements in healthcare quality. The architecture of the system is also designed to cover most aspects of EHR data while having a global semantic structure so that the system is able to utilize a wide spectrum of medical information and to generalize to other clinical domains. The related works and limitations compared to the proposed EHR knowledge graph system are summarized in Table I.

III. EHR Knowledge Graph System

A. System Architecture

This study aimed to create an EHR knowledge graph system for utilizing non-used structured clinical data through explainable semantic reasoning to provide explainable and visualized support during routine practices. The proposed EHR knowledge graph in this study is based on the OMOP CDM, which is a standard data schema for normalizing heterogeneous EHR datasets to achieve multicenter collaborative research [31]. The system uses EHR data that underwent the extraction, transformation, and loading (ETL) process into the OMOP CDM format. The system converts EHR data into resource description framework (RDF)-type triples, creating semantic relationships between data elements, forming a patient information model to contain the medical pathway of the patient's EHR and performing semantic reasoning to analyze non-used clinical information. The system functions automatically. The construction phase requires the participation of medical experts and domain experts for accuracy and clinical practicality. The architecture of the EHR knowledge graph system is presented in Fig. 1. The design and functionalities of each component are as follows:

- The EHR Knowledge Graph Domain Ontology Module (Part 1) defines the semantic structure and reasoning logics of the knowledge graph. A 2-level knowledge graph ontology structure was created to hold medical knowledge and RDF-type EHR data suitable for semantic reasoning and clinical functionality. A top-level ontology defines the global structure for containing medical concepts, while the disease local ontology fills the specific medical knowledge. The knowledge graph system uses the ontology structure to consistently store various medical concepts, classifications and EHR data relationship. *The OMOP EHR Data Conversion Module* refers to the ontology structure to analyze EHR databases and perform RDF

TABLE I
RELATED WORKS COMPARISON

Topics	References	Application Domain	Methodology	Limitations
Semantic representation of EHR data	EI Fadly A., et al. [21] C. Tao, et al. [22]	EHR data repurposing Semantic representation of EHR data	Semantic architecture creation for EHR data representation as a clinical model	The self-defined data models face adoption problems for heterogeneous EHR datasets. The studies lack application on data utilization other than consistency checks and classification.
Ontology approaches for clinical decision support	R. Zhang, et al. [23] N.I. Cole, et al. [24]	Drug-drug interaction Chronic kidney disease	Utilizing ontologies to identify certain clinical concepts in EHR datasets	These studies use ontologies as a vocabulary. The methods did not perform semantic analysis on patients' EHR status. The knowledge and the real-world data are loosely connected.
Using knowledge graph on EHR data	M. Wang, et al. [25] Y. Shen, et al. [26] M. Rotemansch, et al. [27]	Drug recommendation Drug similarity Knowledge graph completion	Knowledge graph embeddings, similarity calculation, naïve Bayes	These studies focus on knowledge engineering instead of clinical functionalities. In addition, they are hard to implement in routine practices.
	A. Gyrard, et al. [28] A. Sheth, et al. [29] O Seneviratne, et al. [30]	Health internet of things Disease self-management Breast cancer characterization	Hybrid reasoning of probability and deduction Rule reasoning	These methods covered partial aspects of one's health information and did not structure the whole EHR data domains, which may raise issues when transferring the system to other clinical domains.

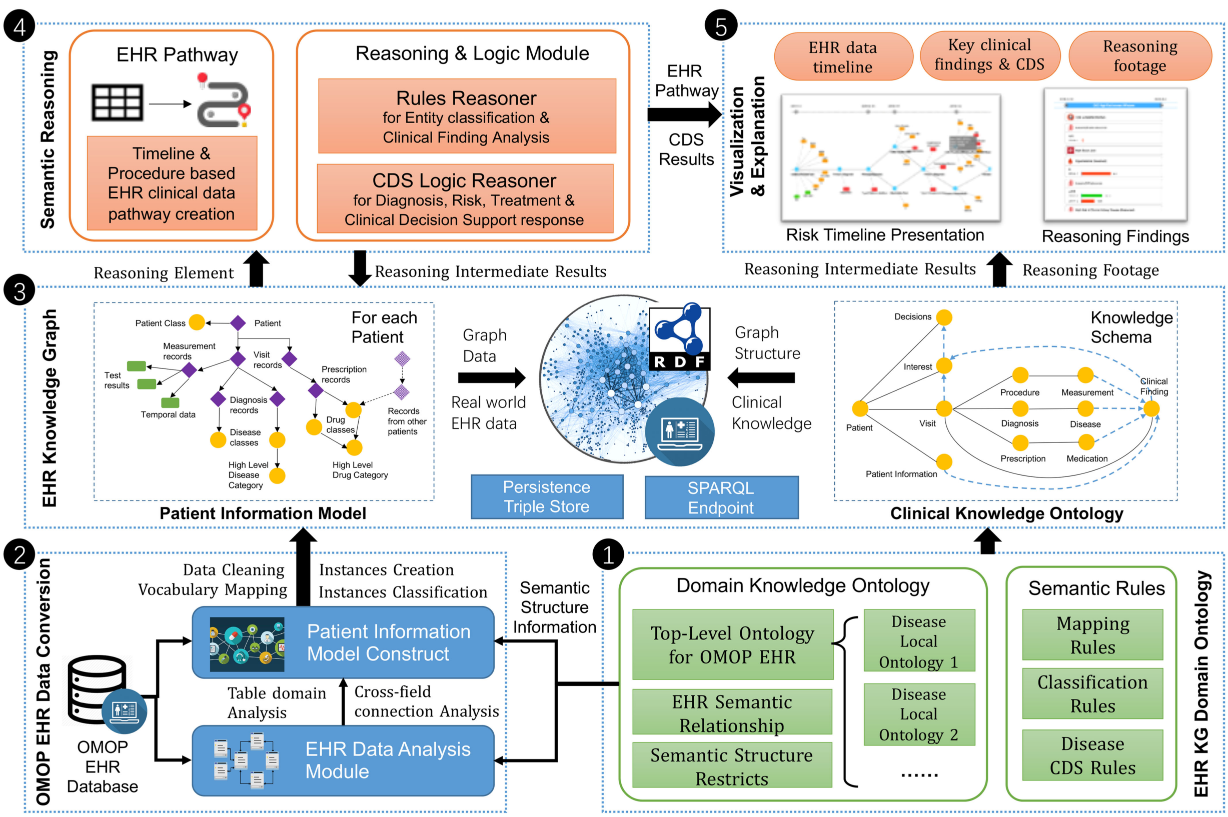

Fig. 1. Knowledge Graph System Architecture. (1) The EHR KG domain ontology module defines the semantic structure of the system and the reasoning logics. (2) Analysis of EHR data and transformation into an RDF-type patient information graph. (3) Storage of medical knowledge and EHR data in the knowledge graph. (4) Performing semantic reasoning and creating CDS results. (5) Visualization of the reasoning footage and EHR data for result interpretation and explanation.

transformation of clinical data in a table format into a semantic EHR data graph. *The Semantic Reasoning Module* utilizes reasoning rules to achieve functionalities.

- The EHR Data Conversion Module (Part 2) analyzes the OMOP EHR database and transforms the data in table format into an RDF-type semantic graph for further utilization. The module matches table domains and concept IDs with top-level ontology entities. The module automatically transforms EHR records into ontology web language (OWL) individuals, loads numerical values and creates semantic relationships according to mapping rules. The complete transformation is then loaded into *the EHR Knowledge Graph* for storage and further reasoning. This module utilizes D2RQ and the Apache Jena-based program to perform the transformation process.
- The EHR Knowledge Graph (Part 3) is the core to store medical knowledge and RDF-type EHR data. A persistent triple store holds the graph data and interacts with other

components through a SPARQL endpoint. The knowledge graph consists of medical knowledge and EHR data. Population-level triple queries and patient-level clinical information queries are available since all of the EHR data nodes are linked with certain medical knowledge entities. All of the modules query and update information through the *EHR knowledge graph*.

- The Semantic Reasoning Module (Part 4) provides reasoning functionalities to complete the semantic relationships of EHR data nodes, evaluate semantic consistency and achieve rule-based reasoning for CDS recommendation. This module analyzes patients' RDF-type data nodes and performs detailed classification. Additionally, the temporal and clinical sequences of each record are reasoned, forming a medical pathway or each patient's records. The reasoning process functions automatically based on reasoning rules and CDS logics. The reasoning module queries the RDF graph from *the EHR Knowledge Graph* and generates clinical findings and CDSs. Intermediate reasoning results are created as OWL entities to record the reasoning footage as well as related data nodes for visualization, evidence tracing and explanation.
- The Visualization and Explanation Module (Part 5) is the window to communicate with clinicians, providing visualization of important neglected medical information and CDS recommendations generated by the knowledge graph system. This module summarizes the reasoning results and query reasoning-related data nodes (e.g., related clinical findings, exam results, risks of certain disease, etc.) from the *EHR Knowledge Graph* and visualizes the information through a graph-like timeline. The process requires no manual input to visualize the timeline for clinicians to review the related information.

B. System Components

This section introduces the detailed design of each module and the core contribution of the EHR knowledge graph system.

1) Knowledge Graph Ontology Structure: A 2-level ontology structure formed the core of the EHR Knowledge Graph Domain Ontology Module. It defines the semantic structure of the whole knowledge graph system for knowledge and clinical data preservation.

a) Top-level ontology: The design of the top-level ontology refers to previous ontology designs from OMOP CDM semantic mapping studies [32], [33], but adjustments and additions to the ontology structure have also been made according to the design of the OMOP CDM standard clinical data structure [31] and the need to perform CDS reasoning on EHR data. Top-level ontology focuses on the core information of patient clinical data, medical knowledge representation and semantic reasoning. It is comprised of the most important domains of the OMOP CDM for EHR data. The key classes of the ontology are listed in Table II.

- The ontology classes of Patient, Visit, Diagnosis, Procedure and Prescription are for basic OMOP EHR data element storage. These classes structured the basic framework of the RDF nodes of the clinical pathway.
- The Disease, Measurement, and Medication classes are hierarchies containing specific subclasses for diseases, tests and drugs as well as OWL reasoning rules for classification. These classes provide medical knowledge for clinical information in the EHR database to be stored and provide support for reasoning.
- The Clinical Finding class and the Interest class are used for semantic reasoning. The clinical finding class is comprised of intermediate reasoning findings such as abnormalities in measurement or interaction between medications. The clinical finding class acts as basic blocks for further decision support and reasoning pathway tracing. The interest class records the regions of interest of a patient's EHR data. The knowledge graph locates the region of interest and then finds all abnormalities, risks, and relevant clinical findings nearby to provide decision support.

b) Disease local ontology: The local ontology is a disease-specific ontology based on top-level ontology. Domain experts review CPGs, collect clinicians' experiences and identify all required medical concepts and rules related to diagnosis criteria, classification criteria and recommendations. The resulting knowledge will be created as OWL entities and reasoning rules for CDS. The local ontology expands the knowledge graph, providing clinical knowledge to utilize RDF-type EHR data for reasoning and CDS generation. A demonstration of a local ontology for CKD is presented in Fig. 2. Clinical concepts are disambiguated according to standard concepts in OMOP vocabulary and coded with OMOP concept IDs to facilitate transformation from EHR data to an EHR knowledge graph.

2) EHR Data Conversion Into a Knowledge Graph:

a) Patient information model: The patient information model defines the semantic structure and relationship of RDF-type EHR data records. It is a 3-level patient-visit-treatment structure modeling each patient's clinical records into a semantic clinical trajectory. The EHR Data Conversion Module transforms EHR data in a table format into a semantic graph according to the structure of the patient information model. The patient-centralized RDF structure connects isolated individuals in the knowledge graph for semantic reasoning to be effective for patient-level clinical recommendations. The structure of the patient information model is shown in Fig. 3.

b) The transformation process:

- The system first analyzes patients' records and visit records in EHR databases. Individuals of these records are created, annotated with record IDs and connected through the object property based on EHR data. Numerical values such as the age and visit date and time are loaded into individuals using the datatype property. The knowledge graph then analyzes the visit date and time of each patient, sorting the sequence and labeling each visit individual with series order numbers.

TABLE II
TOP-LEVEL ONTOLOGY CLASSES OF THE EHR KNOWLEDGE GRAPH

Class Entity Name	Main Content	Subclass Examples*	Statement
ckded:ClinicalFinding (SNOMED CT)	Chief complaint Symptom Medical examination findings Intermediate reasoned findings	LaboratoryFinding ElectrolyteFinding PotassiumAbnormal KidneyFunctionFinding SerumCreatinineAbnormal UricAcidAbnormal Symptom Oliguria Edema	• Individuals of symptoms and intermediate reasoning findings. • The intermediate reasoning results from the diagnosis, measurement, prescription, etc., and will add to the clinical finding class, labeling the abnormalities and findings from the EHR ontology of a patient. Individuals in the clinical finding class provide the basics for decision support and reasoning footage tracing.
ckded:Diagnosis (SNOMED CT)	Disease Risk of disease	AcuteKidneyDisease CardiovascularDisease Atherosclerosis MyocardialInfarction HepatitisB	• Hierarchy of diseases. Diagnosis individuals and the Visit individuals who have a diagnosis will be classified into corresponding disease subclasses.
ckded:Procedure (SNOMED CT & CPT4)	Examination Surgery Nursing	Examination LabExam FullBloodCount KidneyFunctionTest Surgery Ablation Amputation	• Individuals of medical procedure records. • Records of examination, surgery and nursing will be created as individuals and classified into certain subclasses. The results of examination will be saved separately in an individual belonging to ckded:Measurement and linked to ckded:Procedure.
ckded:Interest	Region of interest in patients' clinical records entities	CKDHighRiskInterest CKDLowRiskInterest AKIInterest	• The interest class has individuals linked to certain visits that contain key findings of disease risks to set a region in ontology data for the knowledge graph to perform reasoning contrapuntally.
ckded:Measurement (LOINC)	Physical examination results Laboratory examination results Image examination results	LabResults FullBloodCountResults KidneyFunctionResults UrineRoutineResults ImageResults UltrasoundResults	• Individuals of examination results. • Measurement individuals use owl:DatatypeProperty and owl:ObjectProperty to save medical examination results. Measurements are linked to Procedures.
ckded:Prescription (OMOP defined)	Prescriptions		• Individuals of prescriptions records. • Prescription individuals save time and type and use rdf:type to indicate prescribed medications.
ckded:Medication (RxNorm & VA Class)	Medication hierarchy	ACE-I Biguanides Metformin Macrolides Statin	• Hierarchy of medications. Prescription individuals with a specific medication will be classified into corresponding medication subclasses.
ckded:Patient (OMOP defined)	Patient individuals		• Patient individuals.
ckded:PatientInfo (SNOMED CT & OMOP defined)	Patient biological information Family history Behavior information	Sex FamilyHistory DailyBehavior	• Patient biological information, family history and daily behavior such as smoking, exercise rate, drinking frequency, etc.
ckded:Values	Concept type of values		• Values such as *"Moderate"* and *"2 times a week"*.
ckded:Visit (OMOP defined)	Hospital visit record individuals	InPatient OutPatient	• Individuals of visit occurrence. All diagnoses, procedures, and medication individuals during one hospital visit are linked to a specific visit individual.

∗ Only presents part of the example subclasses to illustrate the function of each top-level class.

- Diagnosis, measurement, procedure, and prescription records are then analyzed and created as individuals. Each treatment individual is connected with related visit individuals according to the visit occurrence ID through the object property. Treatment individuals are then classified into certain subclasses according to diagnosed diseases, measurement items and procedure items.
- Numerical values such as test results, the date and time, and provider IDs are loaded into individuals using the datatype property. Each numerical entity has its own datatype property labeled with OMOP concept IDs. The knowledge graph compares the concept ID and chooses the correct property and numerical format to load the values into certain OWL instances. The disease local ontology uses the unified code for units of measure (UCUM) as unit terminology, the same as the OMOP CDM. The UCUM converter API is used to standardize test result units before loading into the knowledge graph.

The knowledge graph analyzes treatment individuals during each visit and sequences the instances based on the date and time, labeling each individual with a series number within the same visit. Then, according to the predefined clinical pathway order illustrated in Appendix A, the knowledge graph contacts all of the treatments and constructs the clinical trajectory of a patient to model the sequence of a clinical visit.

Through the above process, 3-level patient information models are created for each patient's EHR data records. Moreover,

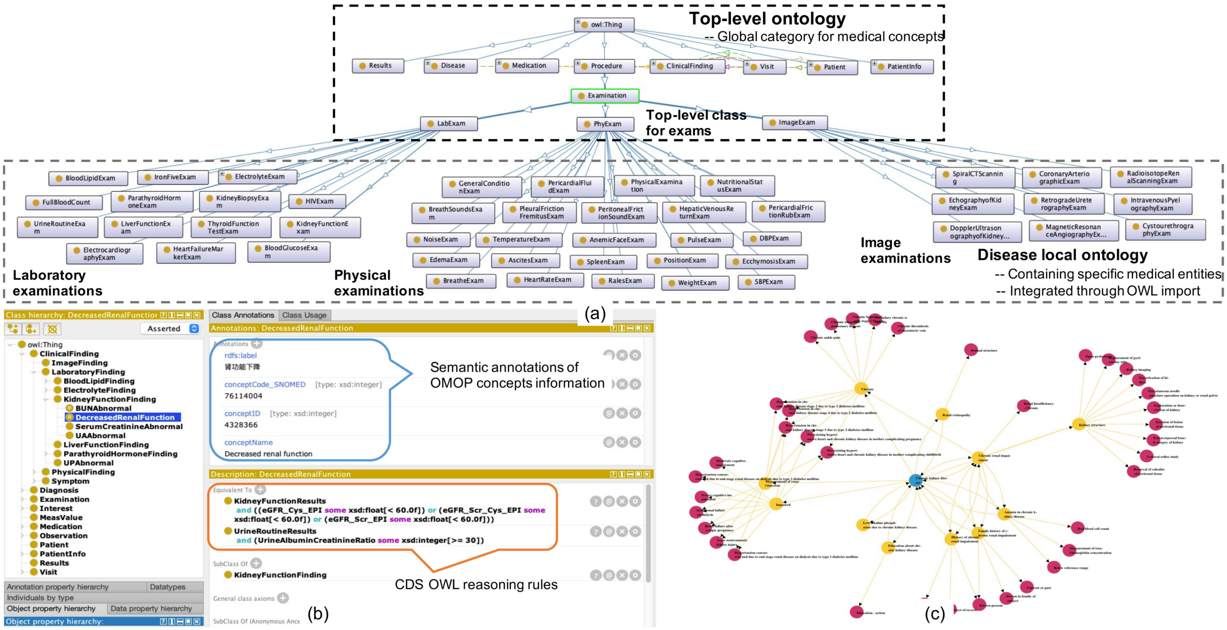

Fig. 2. Examples of local ontology for chronic kidney disease. (a) Disease-specific detailed subclasses added under the top-level ontology structure. The elements in disease-specific local ontology are added under the top-level ontology structure for detailed knowledge representation. (b) The annotation and OWL rules are added to certain elements. The annotation uses OMOP concept information for unified identification and data mapping. (c) Example of the knowledge element network in the knowledge graph system.

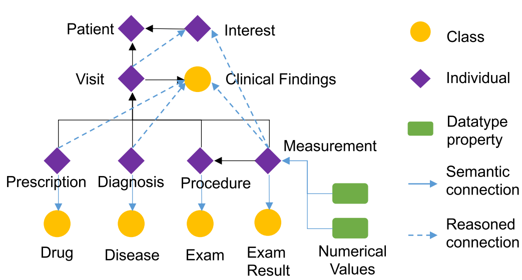

Fig. 3. Patient information model structure.

the knowledge graph contains both a population-level interconnected graph structure and a patient-level medical pathway for personalized reasoning. A demonstration of the patient information model network is shown in Fig. 4.

3) Semantic Reasoning and CDS Logic: The Reasoning and Logic Module performs semantic reasoning on RDF-type patient data according to semantic rules and CDS logic. This module performs classification, trajectory construction, consistency tests and CDS reasoning to identify significant information and obtain clinical recommendations.

a) Clinical trajectory: The reasoning module creates an EHR data trajectory for each patient, forming a medical pathway as illustrated in the patient information model. The trajectory is based on the date of each element and the standard clinical sequence. The trajectory sequence model is shown in Appendix A. Clinical individuals are linked through the object property to indicate the procedure sequence and labeled with series numbers to indicate the sequence order.

b) Semantic rules: Domain experts create reasoning rules by reviewing CPGs and collecting clinicians' opinions. The reasoning rules comprise data classification, clinical findings, risk identification, measurement analysis, and CDS recommendation. The reasoner performs EHR data categorization, clinical data analysis, independent clinical finding generation and summary CDS recommendation according to these rules. The rules are encoded as OWL restrictions or Jena rules for different uses. Fig. 5. shows an example of the reasoning rules.

The Reasoning and Logic Module uses the HermiT reasoner [32] and Apache Jena for semantic reasoning, analyzing RDF-type patient EHR graphs and providing CDS recommendations. There are 3 steps in the reasoning procedure, as shown in Fig. 6:(1) analyze clinical entities and generate independent clinical findings; (2) identify disease-related risks from clinical findings and set regions of interest for further reasoning; and (3) summarize the clinical findings and obtain reasoning results in each region of interest.

4) Reasoning Visualization and Explanation of the Results: To provide a clear illustration of important clinical information found by the EHR knowledge graph system, a reasoning Visualization and Explanation Module was created. A demonstration user interface is developed for the CDS response and reasoning pathway illustration, as shown in Fig. 7.

To provide detailed information on cross-departmental knowledge applications, a network graph based on D3.js for the clinical trajectory of a patient's EHR data is presented. The key elements, including the disease-related diagnosis, abnormal measurements and prescriptions, are listed using the force-direct graph, sorted in the order of clinical trajectory. The risk-related diagnosis and abnormal measurements are color-labeled for identification. The regions of interest are marked under the

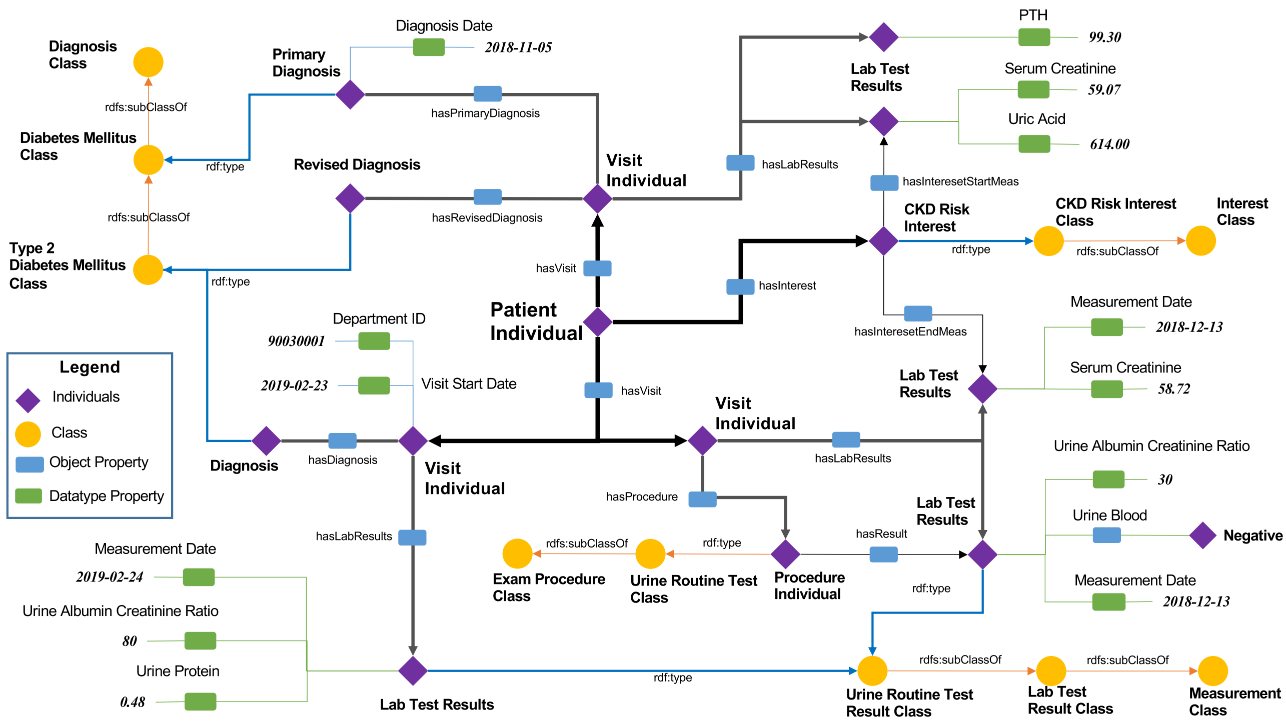

Fig. 4. Example of a patient information model (partial nodes listed). The patient individual links to multiple visit individuals. The visit individual consists of diagnosis, procedures and medical exams. Each clinical individual is identified through classification to a certain class, and numerical values are recorded as datatype properties to represent the clinical information.

```
[Hyperkalemia:(?x rdf:type ns:Visit)(?x ns:hasLabResults ?lr)
    (?lr ns:K ?k)greaterThan(?k,5.5)
    ->(?x rdf:type ns:Hyperkalemia)]

[Anemia2:(?p ns:hasVisit ?x)(?p rdf:type ns:Female)
    (?x ns:hasLabResults ?lr)(?lr ns:HGB ?hgb)lessThan(?hgb,110)
    ->(?x rdf:type ns:Anemia)(?x ns:ReasonerAdding '1'^^xsd:integer)]

[PTHAbnormal:(?lr rdf:type ns:ParathyroidHormoneResults)
    (?lr ns:PTH ?pth)greaterThan(?tg,65.0)
    ->(?lr rdf:type ns:PTHAbnormal)]

[AKIHis:(?p ns:hasVisit ?x)(?p ns:hasVisit ?y)
    (?x rdf:type ns:AcuteKidneyInjury)(?x ns:VisitOrder ?vor)
    (?y ns:VisitOrder ?vor2)greaterThan(?vor2,?vor)
    ->(?y rdf:type ns:PI_DH_AcuteKidneyInjury)]
```

Fig. 5. Example reasoning rules for the EHR knowledge graph system.

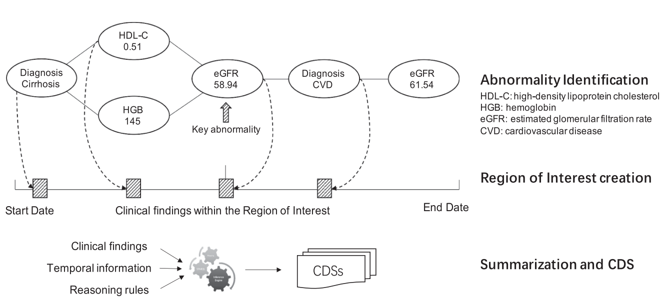

Fig. 6. Three-step reasoning procedures.

timeline, indicating the key reasoning areas and important clinical data information that should be emphasized with greater attention. In each region of interest, the derived clinical findings are listed, showing the reasoning footage of the diagnosis or warning. The graphical illustration helps the users understand why the decisions are given and how they are acquired.

5) Knowledge Graph Generation: The D2RQ tool and an Apache Jena program were used to convert OMOP EHR data into the knowledge graph system.

The D2RQ handles the ontology individual creation and classification. EHR data records in the OMOP tables are transformed into OWL instances, classified into related top-level classes and annotated with IDs. Mappings rules were created for D2RQ to perform the rdf-dump service.

An Apache Jena-based program was created to convert detailed clinical information into OWL instances to cover the deficiency of D2RQ in TBox generation. The Jena program scans the OMOP EHR database, reading the table domain, record ID, concept ID and related values of each record. Numerical values are added into corresponding OWL instances, and entity relationships are created through the object property.

The created knowledge graph was then persistently stored in the Apache Jena TDB2 triple store. A SPARQL endpoint using Apache Jena Fuseki2 provides a knowledge graph I/O.

IV. Knowledge Graph Application

A. Reminding Non-Nephrology Clinicians of Unconsidered CKD in Patients During Routine Clinical Practice

One potential application of the EHR knowledge graph for cross-departmental knowledge sharing and the valuable use of neglected EHR data for decision support is to identify patients with CKD-related risks who were ignored or missed by non-nephrology clinicians during routine practices.

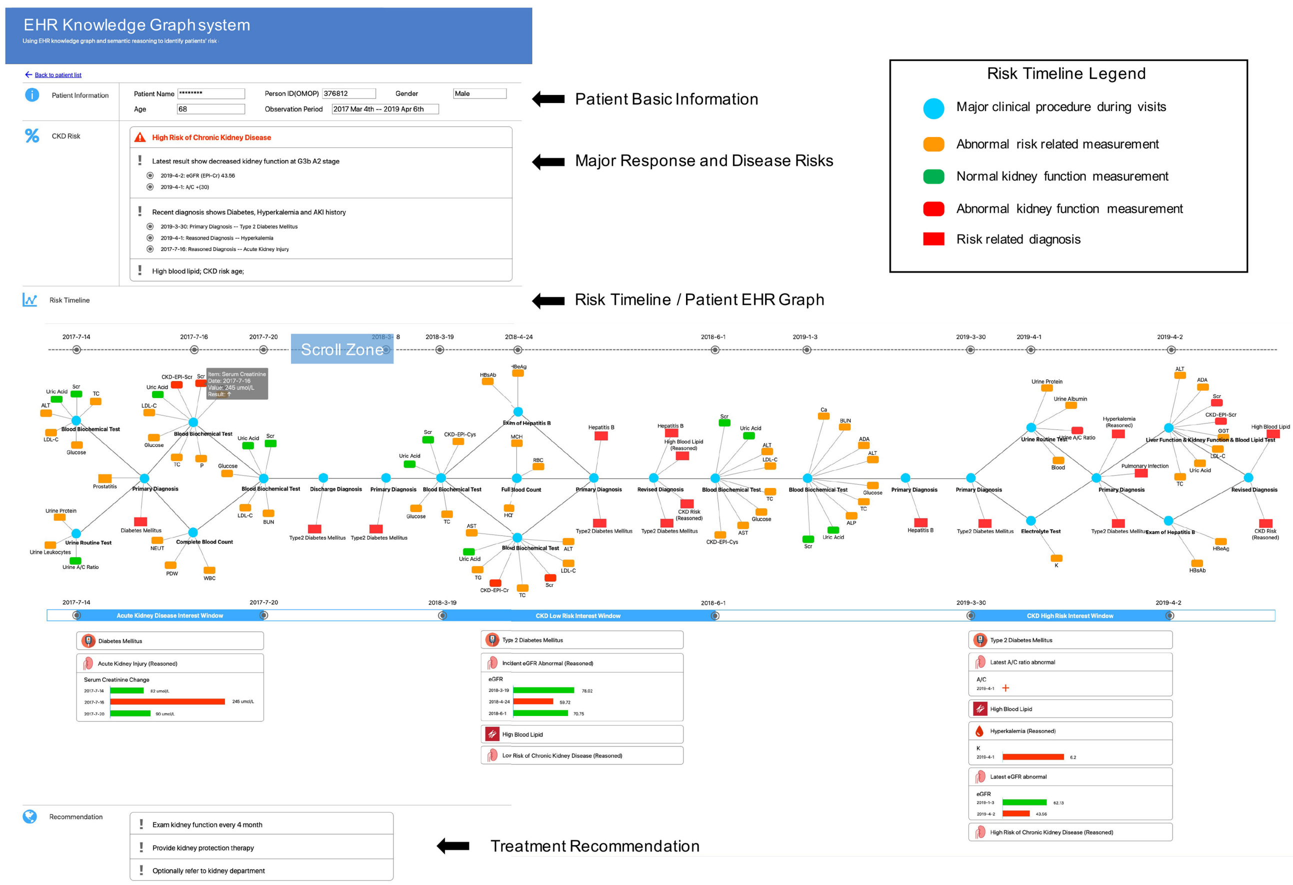

Fig. 7. Graphical illustration of the patient EHR data trajectory and knowledge graph reasoning footage explanation (chronic kidney disease as an example case). The risk timeline presents a trajectory of patients' EHR data used by the reasoning process. The system defines chronic kidney disease regions of interest as blue bars under the timeline, indicating potential risks during that time window. The related clinical findings generated by the system are illustrated under the regions of interest to interpret the evidence for chronic kidney disease. This would help clinicians to better understand the importance of the data and the reasons of the CDS results.

CKD is a common disorder that increases the risk of cardiovascular disease, kidney failure, and other complications. Epidemiological research has shown that the prevalence of CKD in China is 10.8%, while only 12.5% of patients are aware of the disease [35]. A major reason for the low awareness is that CKD and risks are often not identified in time for treatment to be effective. The duty falls on doctors in non-nephrology departments to identify abnormal kidney function and CKD risks because CKD shows no symptoms during the early stage [36]. Patients often undergo medical exam test batteries that contain tests of renal function or risks for kidney abnormality during clinical visits. Moreover, some diagnoses could indicate CKD risk factors. However, the valuable information buried in EHRs is hardly recognized by non-nephrology clinicians because of their lack of clinical knowledge of CKD, leading to failure to consider potential CKD in patients showing risks or abnormal test results and failure to follow CPGs for CKD management [8], [37], [38].

Thus, an EHR knowledge graph system utilizing neglected EHR data and cross-departmental knowledge sharing is helpful to remind non-nephrology clinicians to consider CKD. The current CKD-related CDS is mainly focused on providing progression predictions and clinical examination recommendations for CKD patients [39]–[41] rather than helping clinicians in non-nephrology departments identify possible CKD patients. Few studies have used semantic technology for CKD identification, and semantic usage is mostly for terminology recognition [24]. To achieve cross-departmental informing of unconsidered CKD, it is important to combine both EHR data and CKD knowledge for CDSs to be effective.

B. Application Study Design

The purpose of this application is to remind non-nephrology clinicians of unconsidered CKD risks through buried information in EHR data. Notably, the study is not for prediction of CKD progress. A local ontology and semantic reasoning rules were created based on the CPGs of CKD and acute kidney disease and reviews of CKD management [42]–[46]. Furthermore, the clinical experiences of clinicians from the kidney department at the First Affiliated Hospital, College of Medicine, Zhejiang University (FAHZU) were collected.

The system will identify patients with abnormal kidney function test results or CKD-related clinical risks who have neither visited the nephrology department nor have diagnosis records of any kidney diseases. The EHR knowledge graph first identifies abnormal kidney function examination results and then analyzes related diagnoses, measurements, clinical findings, prescriptions and kidney function trends. The knowledge graph delineates the CKD risk interest region and performs semantic reasoning to classify patients into 3 groups:

1) *Meeting CKD diagnosis criteria*: Patients with test results and clinical findings meeting the CKD diagnosis criteria of CPGs (3 months of continued abnormal kidney function). These patients were considered to have a missed diagnosis that should be identified immediately.
2) *Require high attention*: Patients not meeting CKD diagnosis criteria but have several CKD-related risks (either risk factors for CKD development or consequences of CKD) and decreased kidney function trends. These patients require high attention from non-nephrology clinicians, and kidney function should be tracked.
3) *Require low attention*: Patients with occasional abnormal kidney function test results and few CKD-related risks. These patients were considered to require less attention to avoid over-warning occurrences because the kidney function abnormalities are mostly acute and incidental.

The reasoning workflow and the risk factors considered by the knowledge graph are shown in Appendix B.

The application study uses anonymized EHR data between March 2007 and May 2019 at FAHZU. To evaluate the application result, a prospective study was performed to verify the accuracy of the classification by following up on some of the patients with CKD risks found by the knowledge graph. The follow-up study was performed between June 2019 and December 2019, and the diagnosis of CKD and test results of kidney function were analyzed. The discovery lead time of patients with a positive CKD diagnosis during the follow-up study was calculated to indicate the extent of time that the knowledge graph system identified the CKD risks before clinicians noticed during the follow-up study. The statistics of the EHR knowledge graph and the Visualization and Explanation Module were also evaluated. The study was approved by the Clinical Research Ethics Committee of FAHZU (No. 2020-330).

C. Results

The cohort identified and used by the knowledge graph contains patient data from March 2007 to May 2019. Patients with records of an estimated glomerular filtration rate (eGFR) lower than 60 mL/min or a urine albumin creatinine ratio higher than 30 mg/g were included. Patients with a kidney department visit history or a kidney disease diagnosis and patients who had undergone dialysis or kidney replacement were excluded. In total, 71679 patients were included in the target cohort. Table III shows the characteristics of the patients at cohort entry.

The EHR knowledge graph for this application was generated. We provide knowledge graph ontology elements, axioms, and the number of triples in Table IV.

The evaluation study results from the EHR knowledge graph are shown in Table V. Among the 71679 cohort patients, the knowledge graph found that 2774 patients met the CKD diagnosis criteria. In total, 10377 patients were grouped as requiring high attention because of continues kidney function decreases and several CKD-related risks. A total of 58528 patients were grouped as requiring low attention due to only occasional abnormal kidney function or few CKD-related risks.

A total of 5439 patients were followed up and underwent kidney function tests at FAHZU. The follow-up results of confirmed CKD or unconfirmed CKD ratios are shown in Table V. The results indicate that the proposed system effectively utilized non-used EHR data to identify neglected CKD patients in

TABLE III

CHARACTERISTICS OF THE STUDY COHORT. THE SUMMARY IS BASED ON THE EHR DATA AT COHORT ENTRY. THE MEASUREMENT SUMMARY USES THE FIRST EXAMINATION RESULT AT THE COHORT ENTRY VISIT. THE DIAGNOSIS SUMMARY USES THE PRIMARY AND REVISED DIAGNOSIS DURING THE WHOLE ENTRY VISIT. THE VISITED DEPARTMENT SUMMARY LISTS THE TOP 10 VISITED DEPARTMENTS

Characteristics		Number
Target Cohort		71,679
Age	18–59 years	26,307 (36.7%)
	Older than 60 years	45,372 (63.3%)
	Mean (years)	61.74 ± 16.99
Sex	Female	29,715 (41.5%)
	Male	41,964 (58.5%)
Measurement (Mean ± Std)	Serum Creatinine (μmol/L)	88.88 ± 52.03
	eGFR (mL/min)	77.07 ± 25.66
	BUN (mmol/L)	6.69 ± 3.37
	ACR (mg/g)	77.79 ± 87.78
	HDL-C (mmol/L)	1.07 ± 0.44
	LDL-C (mmol/L)	2.25 ± 0.96
	TC (mmol/L)	4.09 ± 1.28
Diagnosis	Diabetes Mellitus	7,590 (10.6%)
	Hypertension	19,775 (27.6%)
	CVD	1,348 (18.8%)
	Cirrhosis	2,745 (3.8%)
Visited Department (Top 10 most)	General Surgery	9,438 (13.2%)
	Department of Infectious Disease	8,546 (11.9%)
	Cardiology	5,822 (8.1%)
	Urology	2,941 (4.1%)
	Department of Respiratory Medicine	2,823 (3.9%)
	Hematology	2,414 (3.4%)
	Cardio-thoracic Surgery	2,059 (2.9%)
	Ophthalmology	2,044 (2.9%)
	Anorectal Surgery	1,750 (2.4%)
	Neurosurgery	1,427 (2.0%)

eGFR: estimated glomerular filtration rate; BUN: blood urea nitrogen; ACR: urine albumin creatinine ratio; HDL-C: high-density lipoprotein cholesterol; LDL-C: low-density lipoprotein cholesterol; TC: total cholesterol; CVD: cardiovascular disease.

TABLE IV

STATISTICS OF THE KNOWLEDGE GRAPH. THE KNOWLEDGE GRAPH CONTAINS THE TOP-LEVEL ONTOLOGY, A LOCAL ONTOLOGY FOR CHRONIC KIDNEY DISEASE AND EHR DATA FOR COHORT PATIENTS

	Variables	Number
Knowledge graph	Number of triples	24,296,458
	Entities	3,498,135
	Relations	238
Structure & ontology	Ontology class	422
	Object property	23
	Datatype property	191
	Annotation	2,222
Reasoning elements	OWL restrictions and Jena rules	981

TABLE V
EVALUATION STUDY RESULTS OF CROSS-DEPARTMENTAL UNCONSIDERED CKD REMINDING

Groups	Number of patients	Number of follow-up patients	Confirmed CKD during follow-up	Did not confirm CKD during follow-up
Full Cohort	71,679	5,439	--	--
Meeting CKD diagnosis criteria	2,774	223	**183 (82.1%)**	40
Require high attention	10,377	464	**285 (61.4%)**	179
Require low attention	58,528	4,752	**287 (6.0%)**	4,465

The follow-up results of positive CKD are shown in bold.

TABLE VI
DISCOVERY LEAD TIME OF THE CDS RESULTS

Variables	Meeting CKD diagnosis criteria	Require high attention
Mean (days)	884 ± 728	707 ± 651
Median (days)	640	420

The lead time results are based on a follow-up study and were calculated by the difference between the date of the identified risks occur and the date that on which CKD was confirmed during follow-up.

non-nephrology departments. A large number of patients in the group requiring high attention derived benefit from the system for in-time recommendation. Furthermore, patients showing negative CKD indications still need to be monitored at the time of reasoning to prevent chronic risks. The results of the group requiring low attention showed that the system avoided over-warning occurrences. The results in groups requiring high attention and low attention indicated an accuracy of 0.911, a recall of 0.498 and an F1 score of 0.550.

The low recall and F1 scores are mainly due to data imbalance and uncertainty of the chronic development of abnormal kidney function. In the follow-up study, the patients in the group requiring low attention outnumbered those in the other group. A small fraction of these patients had CKD-positive indications in the follow-up study because of insufficient EHR data evidence during reasoning or because the time interval between the reasoning and the follow-up was long and the health status changed. However, the large number of patients in the group requiring low attention created a data imbalance and caused a low recall and F1 score. As the follow-up study continues to cover more patients and the implementation of the system grows, the data imbalance may improve. The study cohort reflected the real-world status of the risk-neglected patients, so implementing a scaling operation to equalize the proportions of the two groups would be irrational, although it would lead to a much higher recall and F1 score.

The lead times of detection for the group meeting the criteria and the group requiring high attention are shown in Table VI. The lead time results are based on follow-up patients, showing that the buried EHR data could identify potential unconsidered CKD patients long before the follow-up diagnosis.

D. CDS Reasoning Pathway Explanation for Clinicians

The CDS response and reasoning pathway are shown in Fig. 7. The patient's RDF EHR data are converted into a timeline relation graph using D3.js. Basic patient information and major CDS responses, including the CKD warning group and related risks, are listed at the top of the page. The risk timeline presents the patient's RDF information based on the knowledge graph, following the order of the clinical visit pathway, showing the diagnosis, kidney function test results and related measurements. The knowledge graph identifies each kidney function abnormality and analyzes clinical information related to the abnormality to create a CKD region of interest. A region of interest is a time window comprised of a series of continued clinical visits showing CKD risks. For example, the patient with ID 376812 shown in the risk timeline of Fig. 7 has 3 regions of interest. On April 24, 2018, the patient data showed abnormal kidney function, but the knowledge graph found that the abnormality was occasional and had few related risks, and the region was labeled as a low CKD risk. The abnormality on April 2, 2019, was the latest patient result, and kidney function had decreased to the G3b stage with risks such as an abnormal ACR, hyperkalemia, cirrhosis, diabetes and hyperlipidemia. Thus, the knowledge graph labeled this region as having a high CKD risk, requiring a high attention classification result.

V. DISCUSSION

In this study, we presented an EHR-oriented knowledge graph system for the efficient utilization of non-used information buried in EHR data. Different from explainable methods that focus on model interpretation, the proposed approach uses the advantages of a knowledge graph to provide evidence-based recommendations and interprets the importance of neglected clinical information for clinicians to understand. Instead of competing with clinicians at their own specialties, the proposed method helps clinicians make comprehensive decisions and improve healthcare quality. Under a standardized semantic structure, the system applies knowledge on cross-departmental EHR data. Through traceable reasoning and pathway visualization, the system presents explanations for reasoning results and helps clinicians make comprehensive decisions.

A 2-level knowledge graph ontology structure was created based on the OMOP CDM. The top-level ontology defines the overall semantic architecture for covering most aspects of EHR data, while the disease local ontology provides flexibility to add new knowledge and generalizes the system to various domains. The patient information model constructs semantic relationships between EHR data entities and forms a clinical trajectory for personalized reasoning. In addition, the system performs step-by-step reasoning, setting regions of interest and summarizing clinical findings for trackable CDS results.

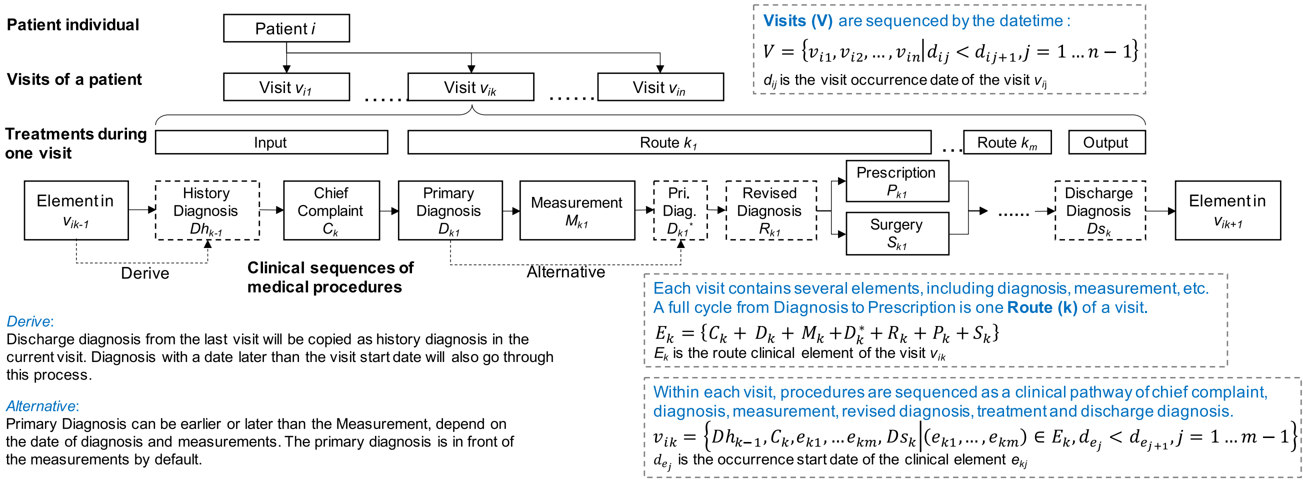

Fig. 8. Clinical trajectory for EHR data conversion. The EHR data of a patient is constructed as a patient-visit-treatment hierarchy style. The visits of a patient are sorted according to visit date. Within each visit, the clinical records are sequenced as a pathway of chief complaint, diagnosis, measurement, treatment procedures and discharge diagnosis. The date of each procedure is also considered. The system uses the object property to link the procedures and the datatype property to label the sequence number.

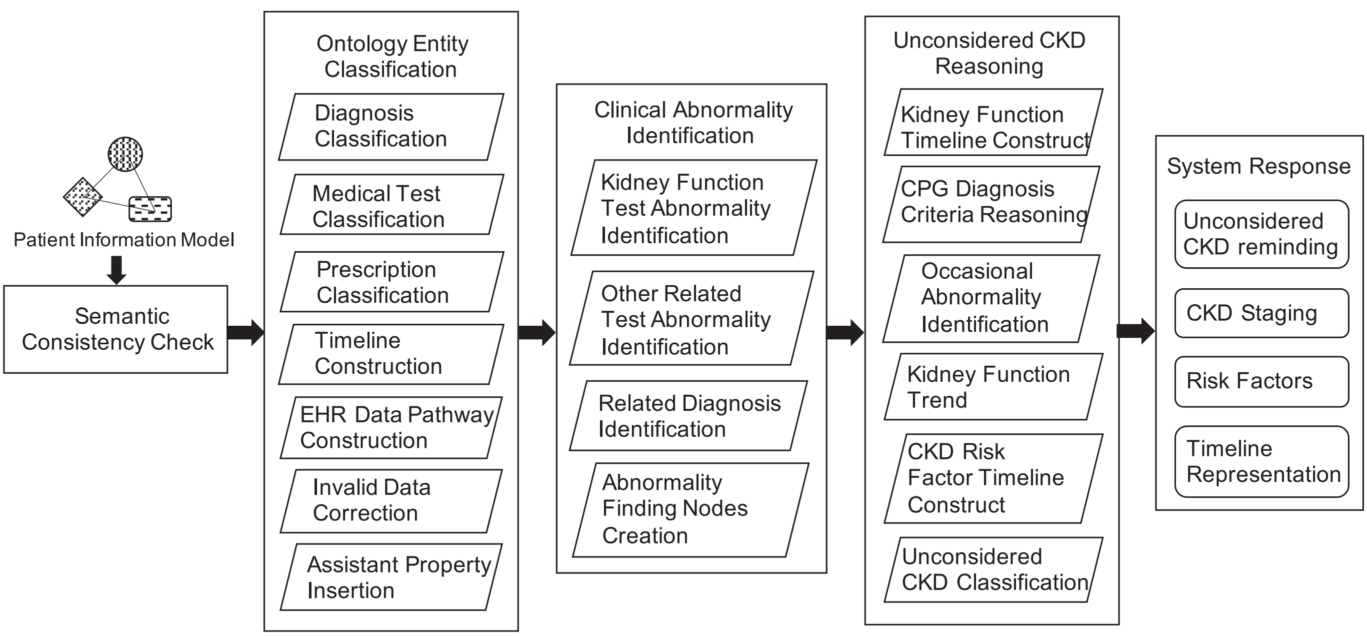

Fig. 9. Workflow for the cross-departmental unconsidered CKD reminding.

The application study showed that the proposed method is effective in utilizing non-used information and helping non-nephrology clinicians break the knowledge barrier to understand the importance of patients' CKD risks and take prompt action. The system identified 2774 patients meeting the diagnosis criteria who were neglected. For other patients, in the group requiring high attention, 61.4% of patients showed positive CKD indications during follow-up. This indicates that a large number of patients will benefit from the in-time CDS of the system. Patients showing negative CKD indications also require kidney function monitoring at the time of reasoning to rule out chronic risk. For patients in the group requiring low attention, only 6.0% showed positive CKD. The system avoided large occurrences of over-warning of patients with occasional decreased kidney function and few related risks so that the routine practice of clinicians is not interrupted. Furthermore, the system performs dynamic monitoring. When new data are added, the system will re-reason the information and give new recommendations if the CKD risk status has changed. Patients showing new evidence of urgency regarding kidney function monitoring will be informed, and the recovered patients will no longer be notified.

The proposed method is not a prediction model. The aim of the application study and the proposed system is to identify neglected information and patients missed by clinicians to achieve comprehensive decisions. The precision of each group is the priority concern of the study. This indicates a positive impact

TABLE VII
CLINICAL FINDINGS USED IN CROSS-DEPARTMENTAL REMINDERS OF UNCONSIDERED CKD

	Characteristics	Score
Kidney Function Condition	Occasional kidney function abnormal	-2
	Decreased kidney function trend	1-3
	Current kidney function lower than G3b	1
Risk factors for CKD	Age over 60 years	1
	Diabetes	1
	Hypertension	1
	Acute kidney injury history	1
	Cardiovascular disease	1
	Autoimmune disease	1
	Cirrhosis	1
	Urine acid abnormal	1
	Urine protein abnormal	1
	Hyperlipidemia	1
Consequences of CKD	Anemia	1
	Bicarbonate abnormal	1
	Parathyroid hormone abnormal	1
	Hyperkalemia	1
	Blood urine nitrogen abnormal	1

Occasional kidney function abnormality is defined as having normal kidney function results in adjoined visits. A total score higher than 4 will be considered as requiring high attention in patients not meeting CKD diagnosis criteria. The total score is counted within each region of interest of CKD risk.

on patients through the system output. Despite false positives, all of the patients in the group requiring high attention needed to be monitored according to the evidence. Furthermore, each of the positive CKD indications found is a major impact since these patients were originally missed by clinicians and might be neglected for a long time.

The EHR knowledge graph system proposed in this study is expandable and can be generalized to other clinical domains. The 2-level ontology structure gives the knowledge graph a global semantic structure for entities and the expansion ability for the addition of new knowledge. The top-level ontology defines the architecture of knowledge entities and RDF-type EHR data structure. The OMOP CDM-based design is able to cover most aspects of EHR data and is suitable for most diseases. Medical knowledge and EHR data from various diseases use a standard and unified semantic structure. Therefore, the knowledge graph system performs semantic queries and semantic reasoning based on a consistent graph data structure, and it is independent of specific diseases. By creating disease local ontologies for different diseases and integrating them into the knowledge graph, the system is able to be transferred to other clinical purposes and diseases.

The disease local ontology construction plays the role of expanding the system into other clinical domains. The main effort to transfer the system to other diseases is to create new disease local ontologies. Additional system APIs and user interfaces are also needed for system invocation and clinicians to review information. Domain experts and medical experts are needed for the construction to ensure accuracy and clinical practicality of the disease local ontology. An overview of the construction process is provided. (1) Medical experts pinpoint the clinical need and propose requirements on medical knowledge (diseases, drugs, examinations, etc.). (2) Proper concepts are selected through semantic mappings and are checked by domain experts. Medical knowledge is constructed as semantic entities and automatically classified into top-level classes based on concept information. (3) Medical experts provide diagnosis criteria, exam limits, risks and other analysis logics. Domain experts transform the knowledge into semantic relationships, OWL restricts and reasoning rules. (4) System APIs are created for functionality and evaluation. Although the reasoning rules require manual efforts, the standard semantic structure leads to a consistent building strategy and standard rule formation, which reduces the workload of the manual process. The experts and work hours needed for the construction phase vary from disease to disease. It would take more efforts to create a disease local ontology for complex diseases, such as cancer applications, because of the numerous omics concepts required and the complicated knowledge of diagnosis and treatment.

There are still some limitations to this study. Currently, the knowledge graph system focuses mainly on structured EHR data and lacks the usage of omics data and medical image data. The adoption of such information would lead to more complete data coverage. Additionally, the clinical information of a patient is usually separated between several hospitals, creating data fragmentation. The application of the knowledge graph system needs to utilize multicenter EHR data in collaborating conditions, which raises data security problems and cross-institute reasoning.

VI. CONCLUSION AND FUTURE WORK

The study demonstrated that the proposed EHR knowledge graph system is effective in utilizing neglected medical data during routine practice and is able to help clinicians make comprehensive decisions through explainable AI.

In future work, we will improve the system for multicenter implementation to overcome patient data fragmentation through collaborative reasoning while maintaining data security and privacy between different institutes. The system will also be implemented for other diseases, such as colorectal cancer detection, and pancreatic cancer risk detection among diabetes patients. The risk of these diseases may be neglected by clinicians during routine practices. Omics data and medical image data will be added to the system for more comprehensive decision support.

APPENDIX A

Fig. 8 presents an RDF-type EHR data trajectory model. The individuals are linked according to temporal and predefined clinical sequences.

APPENDIX B

Table VII presents the evidence considered by the knowledge graph. Fig. 9 presents the workflow for the cross-departmental unconsidered CKD reminders.

ACKNOWLEDGMENT

The authors thank all of the contributors, clinical experts, reviewers, and editors who helped to improve this work.

REFERENCES

[1] K. H. Jones, G. Laurie, L. Stevens, C. Dobbs, D. V. Ford, and N. Lea, “The other side of the coin: Harm due to the non-use of health-related data,” *Int. J. Med. Inform.*, vol. 97, pp. 43–51, 2017.

[2] F. Rutten, M. Cramer, J. Lammers, D. Grobbee, and A. Hoes, “Heart failure and chronic obstructive pulmonary disease: An ignored combination?,” *Eur. J. Heart Failure*, vol. 8, no. 7, pp. 706–711, 2006.

[3] T. Shaw *et al.*, “Attitudes of health professionals to using routinely collected clinical data for performance feedback and personalised professional development,” *Med. J. Aust.*, vol. 210, no. 6, pp. S17–S21, 2019.

[4] B. Tang, G. Hanna, and A. Cuschieri, “Analysis of errors enacted by surgical trainees during skills training courses,” *Surgery*, vol. 138, no. 1, pp. 14–20, 2005.

[5] P. Ramnarayan *et al.*, “Diagnostic omission errors in acute paediatric practice: Impact of a reminder system on decision-making,” *BMC Med. Inform. Decis. Mak.*, vol. 6, no. 1, 2006, Art. no. 37.

[6] M. Audétat, S. Laurin, V. Dory, B. Charlin, and M. Nendaz, “Diagnosis and management of clinical reasoning difficulties: Part II clinical reasoning difficulties: Management and remediation strategies,” *Med. Teacher*, vol. 39, no. 8, pp. 797–801, 2017.

[7] R. Rahav Koren, C. Suriu, O. Yakir, L. Akria, M. Barhoum, and A. Braester, “Physicians’ lack of knowledge - A possible reason for red blood cell transfusion overuse?,” *Isr. J. Health Policy Res.*, vol. 6, no. 1, 2017, Art. no. 49.

[8] L. E. Boulware, M. U. Troll, B. G. Jaar, D. I. Myers, and N. R. Powe, “Identification and referral of patients with progressive CKD: A national study,” *Amer. J. Kidney Dis.*, vol. 48, no. 2, pp. 192–204, 2006.

[9] D. G. Goetz, A. J. Kuzel, L. B. Feng, J. P. DeShazo, and L. E. Love, “EHRs in primary care practices: Benefits, challenges, and successful strategies” *Amer. J. Managed Care*, vol. 18, no. 2, pp. 48–54, 2012.

[10] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, “Doctor ai: Predicting clinical events via recurrent neural networks,” in *Proc. Mach. Learn. Healthcare Conf.*, Dec. 2016, pp. 301–318.

[11] R. L. Gardner *et al.*, “Physician stress and burnout: The impact of health information technology,” *J. Amer. Med. Inform. Assoc.*, vol. 26, no. 2, pp. 106–114, 2018.

[12] M. V. Devarakonda, N. Mehta, C.-H. Tsou, J. J. Liang, A. S. Nowacki, and J. E. Jelovsek, “Automated problem list generation and physicians perspective from a pilot study,” *Int. J. Med. Inform.*, vol. 105, pp. 121–129, 2017.

[13] A. Holzinger, “From machine learning to explainable aI,” in *Proc. World Symp. Digit. Intell. Syst. Machines*, Aug. 2018, pp. 55–66.

[14] R. Goebel *et al.*, “Explainable ai: The new 42?,” in *Proc. Int. Cross-Domain Conf. Mach. Learn. Knowl. Extract.*, Springer, Cham, 2018, pp. 295–303.

[15] H. Hagras. “Toward human-understandable, explainable AI.” *Computer*, vol. 51, no. 9, pp. 28–36, 2018.

[16] X. Chen, S. Jia, and Y. Xiang, “A review: Knowledge reasoning over knowledge graph,” *Expert Syst. Appl.*, vol. 141, 2020, Art. no. 112948.

[17] Q. Wang, Z. Mao, B. Wang, and L. Guo, “Knowledge graph embedding: A survey of approaches and applications,” *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 12, pp. 2724–2743, Dec. 2017.

[18] D. Fensel *et al.*, *Knowledge Graphs: Methodology, Tools and Selected Use Cases*. Cham Switzerland: Springer, 2020, pp. 2–3.

[19] R. B. Haynes, D. L. Sackett, W. S. Richardson, W. Rosenberg, and G. R. Langley, “Evidence-based medicine: How to practice and teach EBM ,” *Can. Med. Assoc.*, vol. 157, no. 6, 1997, Art. no. 788.

[20] F. Wang, L. P. Casalino, and D. Khullar, “Deep learning in medicine—Promise, progress, and challenges,” *JAMA Intern. Med.*, vol. 179, no. 3, pp. 293–294, 2019.

[21] A. El Fadly *et al.*, “Integrating clinical research with the healthcare enterprise: From the RE-USE project to the EHR4CR platform,” *J. Biomed. Inform.*, vol. 44, pp. S94–S102, 2011.

[22] C. Tao *et al.*, “A semantic-web oriented representation of the clinical element model for secondary use of electronic health records data,” *J. Amer. Med. Informat. Assoc.*, vol. 20, no. 3, pp. 554–562, 2012.

[23] R. Zhang *et al.*, “Using semantic predications to uncover drug–drug interactions in clinical data,” *J. Biomed. Inform.*, vol. 49, pp. 134–147, 2014.

[24] N. I. Cole *et al.*, “An ontological approach to identifying cases of chronic kidney disease from routine primary care data: A cross-sectional study,” *BMC Nephrol.*, vol. 19, no. 1, 2018, Art. no. 85.

[25] M. Wang, M. Liu, J. Liu, S. Wang, G. Long, and B. Qian, “Safe medicine recommendation via medical knowledge graph embedding,” 2017, *arXiv:1710.05980*.

[26] Y. Shen, K. Yuan, J. Dai, B. Tang, M. Yang, and K. Lei, “KGDDS: A system for drug-drug similarity measure in therapeutic substitution based on knowledge graph curation,” *J. Med. Syst.*, vol. 43, no. 4, 2019, Art. no. 92.

[27] M. Rotmensch, Y. Halpern, A. Tlimat, S. Horng, and D. Sontag, “Learning a health knowledge graph from electronic medical records,” *Sci. Rep.*, vol. 7, no. 1, 2017, Art. no. 5994.

[28] A. Gyrard, M. Gaur, S. Shekarpour, K. Thirunarayan, and A. Sheth, “Personalized Health Knowledge Graph,” Dayton, Ohio, United States, 2018. [Online]. Available: https://corescholar.libraries.wright.edu/knoesis/1173.

[29] A. Sheth, P. Anantharam, and K. Thirunarayan, “kHealth: Proactive personalized actionable information for better healthcare,” in *Proc. Workshop Pers. Data Anal. Internet Things*, Hangzhou, China, Sep. 2014. [Online]. Available: https://www.researchgate.net/publication/267866120_kHealth_Proactive_Personalized_Actionable_Information_for_Better_Healthcare

[30] O. Seneviratne *et al.*, “Knowledge integration for disease characterization: A breast cancer example,” in *Proc. Int. Semantic Web Conf.*. Springer, Cham, 2018, pp. 223–238.

[31] “*The book of OHDSI*,” 2019. Accessed: Oct, 17 2019. [Online]. Available: https://ohdsi.github.io/TheBookOfOhdsi/

[32] J. G. Klann, L. C. Phillips, C. Herrick, M. A. H. Joss, K. B. Wagholikar, and S. N. Murphy, “Web services for data warehouses: OMOP and PCORnet on I2B2,” *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1331–1338, 2018.

[33] A. Pacaci, S. Gonul, A. A. Sinaci, M. Yuksel, and G. B. L. Erturkmen, “A semantic transformation methodology for the secondary use of observational healthcare data in postmarketing safety studies,” *Front. Pharmacol.*, vol. 9, 2018, Art. no. 435.

[34] B. Glimm, I. Horrocks, B. Motik, G. Stoilos, and Z. Wang, “HermiT: An OWL 2 reasoner,” *J. Autom. Reasoning*, vol. 53, no. 3, pp. 245–269, 2014.

[35] L. Zhang *et al.*, “Prevalence of chronic kidney disease in China: A cross-sectional survey,” *Lancet*, vol. 379, no. 9818, pp. 815–822, 2012.

[36] A. S. Allen, J. P. Forman, E. J. Orav, D. W. Bates, B. M. Denker, and T. D. Sequist, “Primary care management of chronic kidney disease,” *J. Gener. Intern. Med.*, vol. 26, no. 4, pp. 386–392, 2011.

[37] C. H. Fox, A. Brooks, L. E. Zayas, W. Mcclellan, and B. Murray, “Primary care physicians knowledge and practice patterns in the treatment of chronic kidney disease: An upstate new york Practice-based research network (UNYNET) study,” *J. Amer. Board Fam. Med.*, vol. 19, no. 1, pp. 54–61, 2006.

[38] R. K. Israni, J. A. Shea, M. M. Joffe, and H. I. Feldman, “Physician characteristics and knowledge of CKD management,” *Amer. J. Kidney Dis.*, vol. 54, no. 2, pp. 238–247, 2009.

[39] M. Makino *et al.*, “Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning,” *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 11862.

[40] S. J. Barbour *et al.*, “Evaluating a new international risk-prediction tool in IgA nephropathy,” *JAMA Intern. Med.*, vol. 179, no. 7, 2019, Art. no. 942.

[41] A. Perotte, R. Ranganath, J. S. Hirsch, and N. Elhadad, “Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis,” *J. Amer. Med. Inform. Assoc.*, vol. 22, no. 4, pp. 872–880, 2015.

[42] A. Levin *et al.*, “Kidney disease: Improving global outcomes (KDIGO) CKD work group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease,” *Kidney Int. Supplements*, vol. 3, no. 1, pp. 1–150, 2013.

[43] J. A. Kellum *et al.*, “Kidney disease: Improving global outcomes (KDIGO) acute kidney injury work group. KDIGO clinical practice guideline for acute kidney injury,” *Kidney Int. Supplements*, vol. 2, no. 1, pp. 1–138, 2012.

[44] J. A. Vassalotti, R. Centor, B. J. Turner, R. C. Greer, M. Choi, and T. D. Sequist, “Practical approach to detection and management of chronic kidney disease for the primary care clinician,” *Amer. J. Med.*, vol. 129, no. 2, 2016, pp. 153–162.e7.

[45] P. Romagnani *et al.*, “Chronic kidney disease,” *Nature Rev. Dis. Primers*, vol. 3, no. 1, pp. 1–24, 2017.

[46] D. W. Johnson *et al.*, “KHA-CARI guideline: Early chronic kidney disease: Detection, prevention and management,” *Nephrology*, vol. 18, no. 5, pp. 340–350, 2013.