

A fuzzy-ontology-oriented case-based reasoning framework for semantic diabetes diagnosis

Shaker El-Sappagh ^a, Mohammed Elmogy ^{b,*}, A.M. Riad ^c



^a Department of Mathematics, College of Science, King Saud University, PO 2455, Riyadh, Saudi Arabia

^b Information Technology Department, Faculty of Computers & Information, Mansoura University, PO 35516, Mansoura, Egypt

^c Information Systems Department, Faculty of Computers & Information, Mansoura University, PO 35516, Mansoura, Egypt

ARTICLE INFO

Article history:

Received 30 October 2014

Received in revised form 2 June 2015

Accepted 5 August 2015

Keywords:

Case-based reasoning
Knowledge based system
Fuzzy ontology
Semantic retrieval
Diabetes diagnosis
Standard SNOMED CT terminology

ABSTRACT

Objective: Case-based reasoning (CBR) is a problem-solving paradigm that uses past knowledge to interpret or solve new problems. It is suitable for experience-based and theory-less problems. Building a semantically intelligent CBR that mimic the expert thinking can solve many problems especially medical ones.

Methods: Knowledge-intensive CBR using formal ontologies is an evolution of this paradigm. Ontologies can be used for case representation and storage, and it can be used as a background knowledge. Using standard medical ontologies, such as SNOMED CT, enhances the interoperability and integration with the health care systems. Moreover, utilizing vague or imprecise knowledge further improves the CBR semantic effectiveness. This paper proposes a fuzzy ontology-based CBR framework. It proposes a fuzzy case-base OWL2 ontology, and a fuzzy semantic retrieval algorithm that handles many feature types.

Material: This framework is implemented and tested on the diabetes diagnosis problem. The fuzzy ontology is populated with 60 real diabetic cases. The effectiveness of the proposed approach is illustrated with a set of experiments and case studies.

Results: The resulting system can answer complex medical queries related to semantic understanding of medical concepts and handling of vague terms. The resulting fuzzy case-base ontology has 63 concepts, 54 (fuzzy) object properties, 138 (fuzzy) datatype properties, 105 fuzzy datatypes, and 2640 instances. The system achieves an accuracy of 97.67%. We compare our framework with existing CBR systems and a set of five machine-learning classifiers; our system outperforms all of these systems.

Conclusion: Building an integrated CBR system can improve its performance. Representing CBR knowledge using the fuzzy ontology and building a case retrieval algorithm that treats different features differently improves the accuracy of the resulting systems.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Diabetes is a complex, chronic illness requiring continuous medical care with multifactorial risk-reduction strategies beyond glycemic control. According to World Health Organization (WHO), diabetes will be the seventh leading cause of death in 2030 [1]. Globally, about 336 million people are living with type 2 diabetes mellitus, and this figure is set to rise to over 552 million by 2030 [2]. In 2014, 9% of adults 18 years and older had diabetes [1]. There are three main types of diabetes. The first type is type 1 diabetes mellitus or insulin dependent diabetes mellitus; this type occurs when the pancreas cannot produce sufficient insulin. The second type is

type 2 diabetes mellitus or insulin-independent diabetes mellitus; this type occurs when the body cannot effectively use the produced insulin. The third type is gestational diabetes, which occurs in pregnant women. A patient of diabetes symptoms but not really diabetic is called a pre-diabetes patient.

The early diagnosis of diabetes is critical in its care process because the early care can prevent long-term microvascular complications such as retinopathy, nephropathy and neuropathy, and cardiovascular disease. Moreover, the early diagnosis can prevent the pre-diabetes patient to become a diabetic. At present, the results for early detection of diabetes are not highly accurate. Therefore, there is a need to develop a diagnosis system for diabetes that has better accuracy. Clinical decision support systems (CDSS) can help in this regard. Existing rule-based diagnose diabetes systems are mainly based on the A1C criteria or plasma glucose criteria, either the fasting plasma glucose (FPG) or the 2-h plasma glucose

* Corresponding author. Tel.: +0020 1098889791; fax: +0020 502223754.
E-mail address: melmogy@mans.edu.eg (M. Elmogy).

(2-h PG) value after a 75-g oral glucose tolerance test (OGTT). For example, they take decisions using rules such as if ($A1C \geq 6.5\%$ or $FPG \geq 126 \text{ mg/dL}$ or $2\text{-h PG} \geq 200 \text{ mg/dL}$) then the patient is diabetic [3]. However, diabetes diagnosis is more complicated than these direct decisions. Diabetes is related to other diseases including renal diseases, heart diseases, foot diseases, etc. Moreover, it has symptoms related to hyperglycemia or hypoglycemia. The *true or false* decisions about these symptoms, e.g. *thirst=true*, is not enough.

Diabetes diagnosis is a theory-less and unstructured problem, and it depends on the physician's experience. For experience-based problem solving, case based reasoning (CBR) is one of the most suitable AI techniques for decision support [4]. CBR imitates human reasoning, and it is suitable when we cannot formulate a problem in a set of generalized rules. It is appropriate in a medical context where symptoms represent the problem, and diagnosis and treatment represent the solution. The CBR paradigm has been successfully used in various medical fields from lung disease and eating disorders to diabetes and Alzheimer's disease [5]. Many pieces of research utilized CBR for diabetes diagnosis [6–9]. Although any CBR system relies on a set of specific previous experiences, its reasoning power can be improved by general knowledge about the domain [10]. Ontologies can enhance the capabilities of CBR by creating knowledge intensive-CBR (KI-CBR) systems [11]. It can play many roles in CBR such as background domain ontology, case-base ontology, semantic similarity measurement, and others [12]. Ontology can enhance CBR systems in many dimensions, as shown in Fig. 1. In this figure, we suggest three types of KI-CBRs paradigms. In part (a) of Fig. 1, the case-base is stored in a traditional database, and the domain knowledge is stored in an ontology. In part (b), the case-base is stored in a crisp ontology, and the domain knowledge is stored in an ontology. In part (c), the case-base is stored in a fuzzy ontology, and the domain knowledge is stored in an ontology. We have selected the most complicated and recent approach (part c). For diabetes diagnosis, researchers made efforts toward diabetes ontology development [13]. Nevertheless, the literature of ontology-based CBR for diabetes is not rich with studies [7,8].

The most critical steps in CBR paradigm are the *case representation* and *case retrieval*. We concentrate on these two main steps to improve the performance of medical CBR. The case base building process reduces the efforts and time to build the system's knowledge base compared to rule-based systems. No generalized knowledge is required to build a successful CBR system. However, the collection of cases for patients requires the integration between the CDSS system and the distributed electronic health record (EHR) environment. As a result, the standardization of CBR knowledge and data is critical to achieving interoperability. Interoperability between EHR systems and CDSS facilitates the automatic collection of knowledge from patients' EHRs, supports the integration of CDSS in the healthcare environment, and eases the physician's querying process. EHR uses standards as Health Level 7s reference information model (HL7 RIM) [14] and systematized nomenclature of medicine-clinical terms (SNOMED-CT) [15], SCT for short,

ontology for data storage and exchange, which can be utilized in CBR. RIM can be used as a standard case-base structure, and SCT can be used as background knowledge to enhance semantic retrieval [16,17]. El-Sappagh et al. [9] proposed a standard data model for diabetes case-base. SCT is a huge ontology, which affects the performance of the CBR retrieval algorithm. Creating a reference set from SCT for diabetes is required. El-Sappagh et al. [18] proposed a diabetes diagnosis OWL2 standard ontology from an SCT reference set. As far as we know, there are no studies utilize SCT reference sets in CBR systems for diabetes diagnosis, which is considered as a required issue for semantic retrieval and integration of CDSS in EHR environment. Using the created SCT-based OWL2 for semantic retrieval requires the encoding of the case-base unstructured knowledge with the same code. The encoding process is not a straightforward process, and it requires a methodology. El-Sappagh et al. [19] proposed an encoding methodology and utilized it to encode the case-base contents.

Physicians often describe patients using imperfect and linguistic data, and their knowledge and natural language have a great deal of imprecision and vagueness. As Zadeh [20] argued much of the knowledge that humans acquire through experience is perception-based and thus subject to imprecision and inaccuracy. Such knowledge, when not treated in some suitable way that can consider and convey its inherent imprecision, usually leads to the poor effectiveness of the knowledge-based systems that use it. As a result, KI-CBR paradigm must handle the imprecise knowledge representation and reasoning [21]. The existing fuzzy CBR systems utilize imprecise knowledge through the use of fuzzy logic for case representation and relevant fuzzy pattern matching techniques for similarity assessment [22]. A survey of existing systems of fuzzy CBR in diabetes diagnosis indicates that there are few works in this field. However, the lack of representation of this knowledge in ontological restricts the effectiveness of these systems because they did not take advantage of the reasoning capabilities that ontologies provide. The fuzzy ontology focuses on assigning a meaning to the fuzziness of the ontology's components. It is an important characteristic as it makes the fuzzy ontology's imprecision explicit, thus facilitating more efficient knowledge acquisition and ontology reuse. Moreover, it enables the definition of more effective semantic similarity measures, which facilitate case retrieval. For diabetes, the existing fuzzy CBR systems have not used fuzzy ontology or even crisp ontology as background domain knowledge or case-base ontologies [8]. On the other hand, ontologies and fuzzy logic have been utilized in diabetes in other reasoning methods such as rule-based expert systems [23].

In this paper, we present a fuzzy KI-CBR framework that handles and exploits imprecise knowledge through the effective integration of fuzzy logic in the ontology-based CBR paradigm. Fuzzy case-base ontology and a fuzzy semantic retrieval algorithm are proposed and integrated to build an intelligent CBR for diabetes diagnosis. This approach introduces fuzzy semantics to CBR in two places. The first is the representation of imprecise knowledge itself, and the second is case retrieval. In particular, our proposed framework

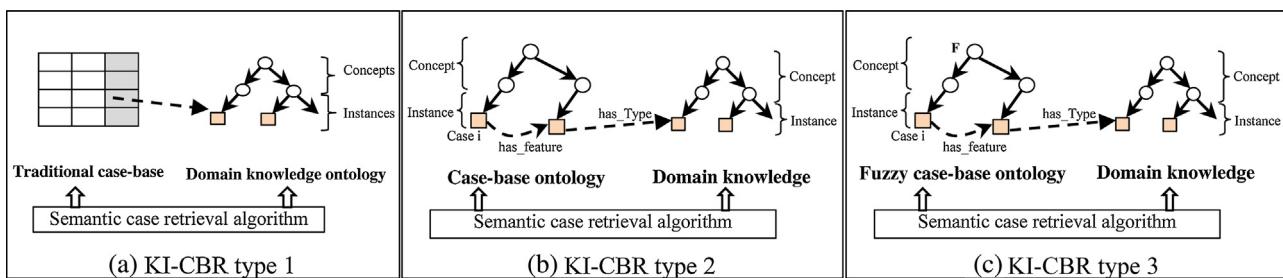


Fig. 1. KI-CBR frameworks.

is built using a fuzzy ontology that supports the representation of imprecise case-specific knowledge while the retrieval of cases is enabled by proposing a highly customizable fuzzy semantic similarity framework. As most of the CBR studies did not implement the entire cycle [12,24], we concentrate on the most critical and most related steps (i.e., case representation and retrieval). Case adaptation, reuse, retention, and case-base maintenance will be handled in other works.

Importantly, our system is implemented in six modules: Case source preparation, case base ontology engineering, terminology server, fuzzy case-base ontology population, case retrieval engine, and case query parser. We implement and test the proposed framework on a real case-base. The system has a user-friendly interface; it supports the selection of standard medical concepts from an SCT dialog, and it implements the clinical distance in the case retrieval process.

As a result, the system achieves a high-level performance compared to the traditional CBR systems, other CBR systems in the literature, and machine learning algorithms. The system's accuracy is 97.67%. Therefore, it is highly accurate and can be applied in a real medical environment.

To this end, the remainder of the paper is organized as follows: Section 2 provides studies related to KI-CBR, especially for diabetes, and show its limitations. Section 3 is a set of preliminaries including our dataset description. Section 4 illustrates the research methodology used in the study. Section 5 is the proposed CBR framework. Implementation and evaluation are discussed in Section 6. Finally, Section 7 concludes the paper and highlights future work directions.

2. Related work

The physician can depend on clinical practice guidelines (CPG) to diagnose diabetes. However, CPGs are long plaintext documents. Some languages such as Arden syntax can be used for representing and sharing this medical knowledge. It can convert CPGs into actionable rules to implement rule-based CDSS systems. Samwald et al. [25] proposed a development environment including a compiler and rule engine for Arden Syntax rules. However, diabetes diagnosis is an ill-formed, theory-less, and experience based problems. Depending on rules, is not suitable because there will be many exceptional cases. Rules results often require adaptations by a physician. Rules cannot be customized for specific patient conditions. It is time-consuming to build and maintain a large rule-base. CBR is one of the most suitable AI technique for the experience based problems because it is easier for an expert physician to formulate specific cases than to formulate generalized rules. Traditional CBR has been used for diabetes diagnosis in many studies [4–7]. An evolution of this paradigm is the ontology-based-CBR [21]. This approach is generally more effective in retrieving similar cases than traditional ones [10]. Ontology plays many roles to enhance CBR semantics ranging from case storage and representation to case adaptation and reuse [11]. Moreover, case semantic retrieval algorithms can be improved by using case-base and domain background knowledge in the form of ontologies [26,27].

2.1. Regarding the role of ontology in diabetes management

In the diabetes domain, ontology has been used in many CDSSs [13,23,28–30]. For example, Chen et al. [13] introduced an ontology for diabetes drugs and an ontology for patients' symptoms. These ontologies utilized semantic web rule language (SWRL) and Java expert system shell (JESS) to determine potential prescriptions for the patients. Rahimi et al. [28] developed a type 2 diabetes mellitus (T2DM) ontology (DMO) to diagnose and manage patients

with diabetes, and they proposed an algorithm to query the ePBRN data repository to diagnose T2DM. Sherimon et al. [29] proposed a dynamic adaptive questionnaire ontology for gathering the diabetic patient's medical history. Hayuhardhika et al. [30] developed an ontology for diabetes disease and used a weighted tree similarity algorithm for diagnosis. However, regarding diabetes diagnosis, none of these ontologies is designed for CBR, and few studies have used ontology in CBR [6,8]. In diabetes diagnosis systems, ontologies have not been utilized in neither case-base nor background knowledge nor case retrieval. Jaya and Uma [7] have listed the roles of ontology in a diabetes diagnosis CBR. El-Sappagh et al. [31] proposed a case-base ontology engineering methodology, and they proposed a diabetes case-base ontology. However, there are no decision support capabilities provided in the study. The resulting OWL2 ontology can be utilized in the current study to store and retrieve cases semantically. In addition, this ontology is crisp and cannot handle the existed vagueness in diabetes diagnosis environment [20].

2.2. Regarding the encoding of medical data

Some of medical knowledge is stored in the unstructured form. This knowledge is not suitable for CBR. To enhance the semantic intelligence of a CBR system, the case-base textual contents have to be encoded in a formal way. Samwald et al. [32] asserted that the building CDSS system requires the encoding of clinical data by using ontologies. They developed a CDSS for pharmacogenomic knowledge representation and reasoning based on an OWL2 ontology [33]. However, using standard medical ontologies, such as SCT, supports the implementation of semantically intelligent case retrieval algorithms [34], enhances the interoperability and seamlessly integration between CDSS and EHR environment [16], and supports the creation of standard encoded case-base knowledge [35]. As a result, the unstructured medical data of EHR are standardized into a unified form, which facilitate the automatic collection of cases knowledge of the distributed EHR environments. Moreover, the CBR system becomes more intelligent by interpreting the meaning of medical concepts. In addition, case retrieval algorithm can calculate the clinical distance between patients rather than geometric or semantic distances. To the best of our knowledge, standard medical ontologies such as SCT have not been used in diabetes diagnosis CBR systems. El-Sappagh et al. [18] proposed an OWL2 ontology for SCT to be used as background domain knowledge with CBR. In addition, this ontology can be used to encode the diabetes case-base unstructured knowledge into a formal and standard form [19].

2.3. Regarding the fuzzification of medical data

Diagnosis of diabetes depends on the physician's experience and the patient's description of his case. Most medical data are described using vague terms (i.e., partially known) [36]. Vagueness can be handled using fuzzy logic (FL) [20]. FL is useful for CBR because CBR is fundamentally analogical reasoning, which can operate with linguistic expressions. FL facilitates the knowledge elicitation from a domain expert, eases the transfer of knowledge between domains, and enhances the similarity measurement. Fuzzy logic has been integrated with CBR in hybrid systems [37,38] and used for calculating the fuzzy similarity between cases [22]. However, there are no real studies in the literature for fuzzy-CBR systems for diabetes diagnosis. Thirugnanam et al. [39] built a hybrid CDSS system for diabetes diagnosis using a neural network, fuzzy, and CBR. This study used the fuzzy and CBR reasoning mechanisms separately, and no fuzziness has been added to enhance the CBR functionality. Most CBR systems in the literature utilized FL in case retrieval step only. Building a fuzzy case-base knowledge is required to support fuzziness in CBR systems. However, these

hybrid systems have not benefited from fuzzy ontology reasoning capabilities in CBR system.

2.4. Regarding the role of fuzzy ontology in CDSS

As crisp ontology has proved its roles in CBR, fuzzy ontologies can extend the capabilities of crisp ontologies [40]. Crisp ontologies are not suitable to address imprecise and vague knowledge, which is inherent in real world domains [41]. Fuzzy ontology can come from two sources: mapping of a fuzzy database [42] or as an extension of crisp ontology [40]. Fuzzy ontology has been used in medical and non-medical systems [41,43–47]. Ali et al. [47] proposed T2FOBOMIE; this system is an opinion mining system based on a type-2 fuzzy rough ontology. Rodríguez et al. [41] proposed a fuzzy ontology-based system for modeling human behavior. Torschizi et al. [43] utilized fuzzy ontology to build an intelligent rule-based system to determine the severity of Benign Prostatic Hyperplasia and recommend the appropriate therapies. Carlsson et al. [44] discussed the capabilities of fuzzy ontology over crisp one and utilized it in a knowledge mobilization application. Mezei et al. [45] asserted that fuzzy ontology is critical to building actionable knowledge to aid complex decisions, and they proposed a fuzzy wine ontology. Molinera et al. [46] proposed a decision support system for recommending smartphones using fuzzy ontologies. Lee and Wang [23] used fuzzy ontology to build a diabetes diagnosis CDSS system. This system is based on rule-based reasoning paradigm. It used the freely available Pima Indians dataset, which is not diabetes representative data. It achieved the accuracy of 91.2%.

2.5. Regarding the role of fuzzy ontology in CBR

For CBR systems, many studies have utilized fuzzy ontology for case base representation and fuzzy retrieval processes [21,48]. Alexopoulos et al. [21] proposed a fuzzy ontology-based CBR system using fuzzy algebra. Ali et al. [48] proposed a type-2 fuzzy ontology-based CBR system for collision avoidance of autonomous underwater vehicles. Fuzzy ontology can enhance CBR in different ways such as physician can more easily define experience cases using natural-like language, cases can be indexed more efficiently, and fuzzy-semantic retrieval algorithms can be implemented. Diabetes has utilized fuzzy ontologies in many fields [49]; however, there is no fuzzy ontology-based CBR for diabetes management. CBR effectiveness is further improved if ontology-based CBR systems can utilize vague or imprecise knowledge [21]. We argue that there is a difference between ontology-based fuzzy CBR [50] and fuzzy-ontology based CBR [21]. The former builds a fuzzy CBR system and uses crisp ontology to enhance its functionality, but the later builds a fuzzy ontology for its case-base. Alexopoulos et al. [21] have concentrated only on fuzzy properties using fuzzy algebra. Fuzzy Ontologies can extend query cases.

Fuzzy-ontology-based KI-CBR is a yet unstudied topic, especially in the medical domains such as diabetes diagnosis. Moreover, there are no studies on diabetes diagnosis, which incorporate subsets of standard ontologies such as SCT, unified medical language system (UMLS), gene ontology (GO), international classification of diseases (ICD), disease ontology, or logical observation identifiers names and codes (LOINC) as the background domain knowledge [51]. In addition, in our study, we are the first to separate case-base ontology from background knowledge ontology. This separation has a great role in the medical domain because the case base and domain ontologies are usually huge; moreover, many standard ontologies can be simultaneously utilized in the same CBR system.

As shown in Fig. 2, the purpose of this paper is to propose, implement, and test a fuzzy KI-CBR framework using characteristics of ontology, fuzzy logic, and standard medical terminology (i.e., SCT). To accomplish this purpose, the major contributions to performing this research can be summarized as follows:

- We propose an integrated fuzzy knowledge-intensive CBR framework. This system (shown in Fig. 5) is distinctive in its novel architecture and can be applied in the development of a variety of CDSS systems.
- We introduce an efficient way to develop the case-base fuzzy ontology, which is the backbone of the proposed system. This ontology is built based on our previously published crisp ontology [31] and the top-level CBR crisp ontology namely CBROnto proposed by [52]. The step-by-step tutorial on the fuzzy ontology development process can be helpful for interested readers to conduct experiments. The proposed fuzzy ontology is the first in the medical domain.
- We propose a fuzzy semantic retrieval algorithm for retrieving cases from the fuzzy ontology according to the physician new coming problems. This hybrid algorithm is accurate and takes into account the types of patient's features including numerical, fuzzy, ordinal, lexical, and semantic types. Moreover, the fuzzy types are represented in a fuzzy ontology, and the semantic types are based on a standard diabetes diagnosis SCT ontology.
- To perform the case study, we develop a JAVA-based prototype based on the most popular CBR APIs (i.e. JCQLIBRI). The internal intelligent processes of the prototype control the query processes. The physician enters the patient description data in a new case Q_v . The system converts the query case crisp vector into a fuzzy semantic vector Q_{FSV} . The Q_{FSV} is passed to the retrieval engine, which retrieves the most similar cases to the Q_{FSV} case based on the clinical distances between patients. The experimental results that are generated by utilizing this prototype advocate the efficiency of the proposed architecture.

The proposed framework utilizes our previously proposed ontologies such as the crisp case-base ontology [31] and the

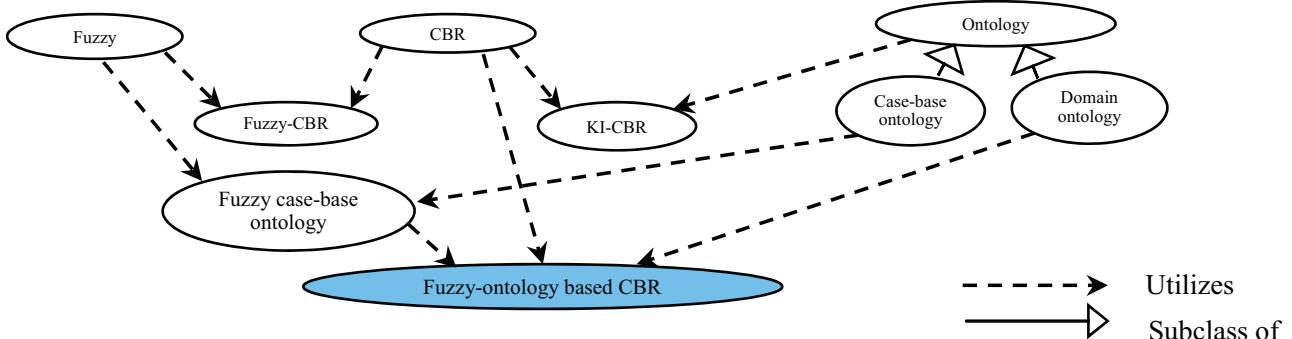


Fig. 2. Our research focus.

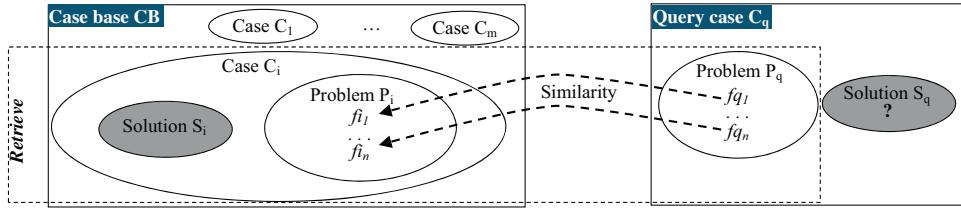


Fig. 3. The correspondence between stored cases and query case.

diabetes standard ontology from SCT taxonomy [18]. In addition, to prepare our case-base contents, we utilize the case-base standard data model [9], use the pre-processing step to handle noisy data, select relevant features, and calculate the weight vector [35]. We utilized our encoding methodology to encode the case-base unstructured knowledge into standard form [19]. Moreover, we utilize our fuzzification methodology to fuzzify the case-base vague concepts.

3. Preliminaries

To make the article self-contained, in this section we define some concepts, definitions, and terminologies before discussing the proposed framework.

3.1. Case base reasoning

Generally, CBR is an AI technique for solving a problem by remembering similar past experiences. For example, physicians look for groups of known symptoms and engineers take many of their ideas from previously successful solutions. The main concept of CBR is “similar problems have similar solutions.” CBR knowledge is formed in a case-base of previous experiences (either success or failure). It does not depend on the explicit model of the problem as in rule base reasoning for the inference process, but it simply utilizes the experience captured, in the same way, the expert usually inputs and processes it. The newly solved problems can be added as a new experience in the CBR system’s experience-base (case-base), which supports the auto-learning process. The CBR can be defined as a cyclic process named “the four Rs” [54]: (i) *Retrieve* the most similar cases, (ii) *Reuse* the cases that might solve the problem, (iii) *Revise* the proposed solution if necessary, and (iv) *Retain* the new solution as part of a new case. The most important aspects of CBR system are the case-base knowledge representation and the case retrieval algorithm, and these are our contributions in the current paper.

Definition 1. A case-base CB is a finite set of cases $\{C_1, C_2, \dots, C_m\}$, where m is the number of cases in the CB.

Definition 2. A case is a case is a contextualized piece of knowledge representing an experience. The i th experience case $C_i \in CB$ is formally defined as $C_i = \langle P_i, S_i \rangle$, where P_i and S_i respectively represent the case problem description and the case solution features.

Definition 3. A case retrieval algorithm is an algorithm that takes as input (query case C_q , case base CB, and features weighting vector W); it calculates the level of similarity between C_q and every case in CB; and finally it returns the solution of the most similar cases. The k -nearest neighbour (k -NN) is the most applicable retrieval algorithm. More formally, let $C_q = \langle P_q, X \rangle$ be a query case, where P_q is the query case’s problem and X denote is its solution. It should be mentioned that the main objective of the CBR system is to determine the value of X , which is unknown before the execution of the case retrieval process. In general, multiple features describe the problem situations of both the case-base historical cases and the target case. Let $N = \{1, 2, \dots, n\}$, is the total number of attributes.

Let $f = \{f_1, f_2, \dots, f_n\}$ be a finite set of n features concerning the problem situations of both the historical cases and the target case, where f_j denotes the j th attribute, $j \in N$. Let $\bar{W} = (w_1, w_2, \dots, w_n)^T$ be a weights vector of case features which determine the features importance, where w_j denotes the weight or the importance degree of attribute f_j , such that $\sum_{j=1}^n w_j = 1$ and $0 \leq w_j \leq 1$, $j \in N$. Let $\rightarrow C_i = (f_{i1}, f_{i2}, \dots, f_{in})^T$ be a vector of feature values for the problem situation of historical case C_i , where f_{ij} denotes the consequence of historical problem situation C_i concerning attribute f_j , $i \in M$, $j \in N$. Let $\rightarrow C_q = (f_{q1}, f_{q2}, \dots, f_{qn})^T$ be a vector of feature values for the problem situation of target case C_q , where f_{qj} denotes the consequence of current problem situation C_q concerning attribute f_j , $j \in N$. As shown in Fig. 3, the correspondence between query case’ and the historical cases’ features can be easily defined.

The case retrieval algorithm depends on the level of similarity between two cases $SIM(C_i, C_q)$, i.e. the global similarity, where $SIM(C_i, C_q) \in [0, 1]$. The similarity function $SIM(C_i, C_q)$ is the collection of feature-level similarities $sim(f_{ij}, f_{qj})$, the local similarity, where $sim(f_{ij}, f_{qj}) \in [0, 1]$. Many studies of existing CBR assume that all features are of the same datatype (e.g. numerical) and provide as single local similarity function $sim(f_{ij}, f_{qj})$ to measure the similarity between f_{ij} and f_{qj} . This is not the normal case [55]. In our study, we propose one of the most complete similarity measure, which takes into account the numerical, nominal, ordinal, fuzzy, and semantic feature types, as shown in Fig. 4.

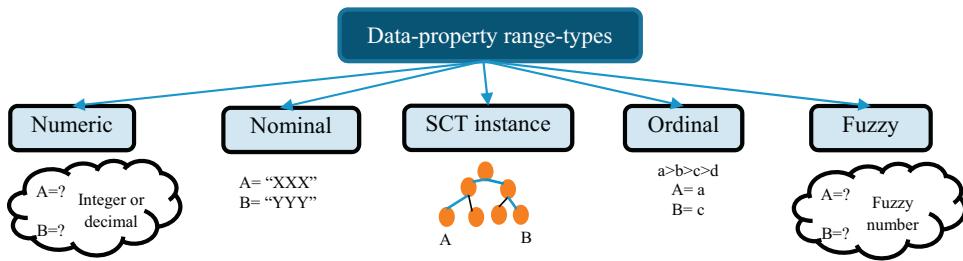
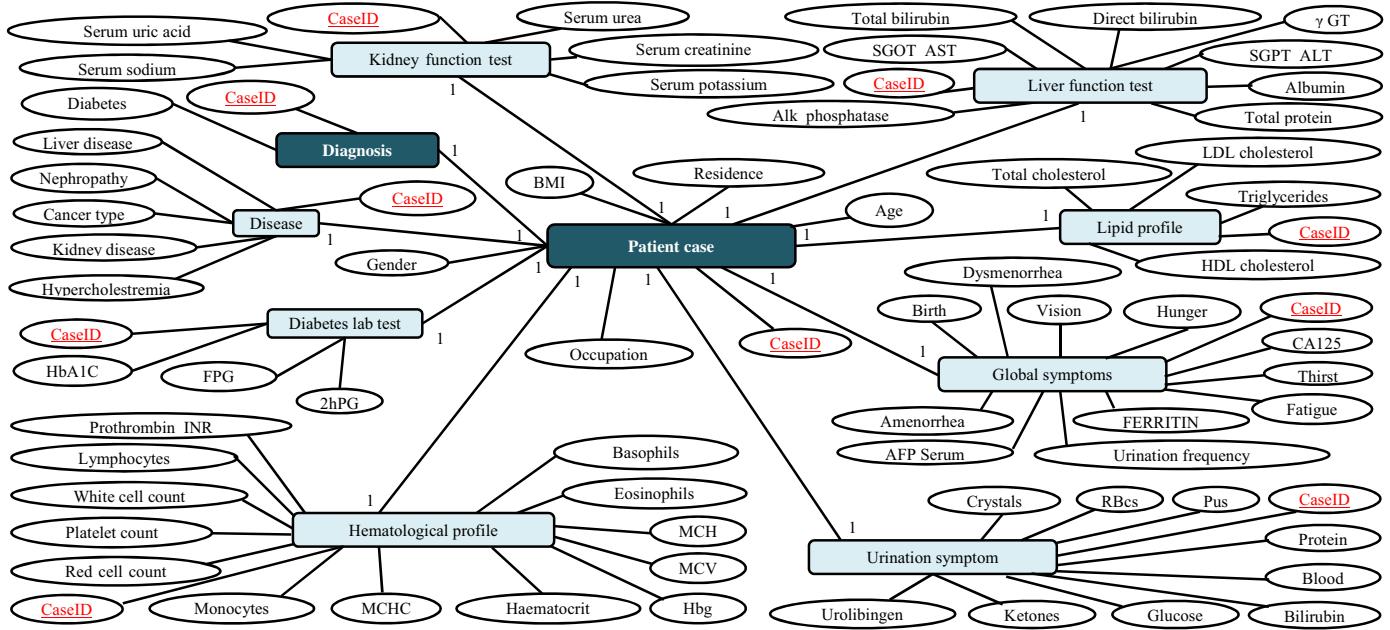
The global similarity between the two cases $SIM(C_i, C_q)$ can be defined by a distance method. The most widely used measures are the Euclidean distance or Hamming distance, as shown in the following equation:

$$sim(C_i, C_q) = \begin{cases} 1 - \sqrt{\sum_j w_j^2 \times dist^2(f_{ij}, f_{qj})} & \text{if the Euclidean distance used,} \\ 1 - \sum_j w_j \times dist(f_{ij}, f_{qj}) & \text{if the Hamming distance used.} \end{cases} \quad (1)$$

where $sim(f_{ij}, f_{qj})$ function is defined in terms of the function $dist(f_{ij}, f_{qj})$.

3.2. Case representation

The contents of a case-base must be defined in the first beginning of a CBR system. These contents determine all of the subsequent steps such as case-base ontology, case base fuzzy ontology, and case retrieval. After checking with the domain experts, CPGs, and handbooks of case histories in diabetes diagnosis domain, our case will contain the features described in Table 1. The data have been obtained and managed by the hospitals of Mansoura University, Mansoura, Egypt. All the features that affect the diabetes diagnosis have been collected by our domain experts. Some data are collected from a diagnostic biochemical lab (AutoLab, Mansoura, Egypt). The used data set was collected from January 2010 through

**Fig. 4.** The case feature types.**Fig. 5.** Diabetes diagnosis and other related complaints case base data model.

August 2013. There are 67 eligible patients, who enrolled in this study. However, seven control subjects were excluded due to limited blood samples for testing AFP. Our data set contains 70 features for describing diabetic patients and for linking diabetes with other disorders such as cancer, kidney diseases, and liver diseases. The data set is distributed as 33.3% pre-diabetic patients, 53% diabetic patients, and 13.7% normal patients. **Table 1** shows descriptions of considered features in this study.

3.3. The structure of a diabetes diagnosis case

Fig. 5 shows an Entity Relationship (ER) model for all entities and attributes used in our data set. This data model is compatible with HL7 RIM [56]. This compatibility facilitates the integration with EHR and supports the auto collection of cases. Moreover, this data model has been fuzzified with our proposed fuzzification methodology into a fuzzy ER model, then converted to a fuzzy case-base database, which was the source of instances for our proposed fuzzy case-base ontology. These entities and attributes were enriched by entities and attributes in diabetes diagnosis CPGs as in the National Guidelines Clearing House¹. Entities and features related to diabetes treatment, medications, and drugs are out of scope.

Definition 4. Diabetes diagnosis cases are defined according to our data model. A case $C = \langle P, S \rangle$ is defined as follows: $P = \{LFT, GS, A, B, R, G, O, KFT, LT, US, HP, DI\}$ where $LFT = \text{liver function tests}$, $LP = \text{lipid profile}$, $GS = \text{global symptoms}$, $A = \text{age}$, $B = \text{BMI}$, $R = \text{residence}$, $G = \text{gender}$, $O = \text{occupation}$, $KFT = \text{kidney function tests}$, $LT = \text{lab tests}$, $US = \text{urination symptoms}$, $HP = \text{haematological profile}$, and $DI = \{L + N + C + H\}$ where $L = \text{probable liver problem}$, $N = \text{probable nephropathy problem}$, $C = \text{probable cancer type}$, and $H = \text{probable hypercholesterolemia problem}$. $S(P)$ is the solution part describes the diagnosis of diabetes including diabetic, prediabetic, gestational-diabetic, and prediabetic-gestational.

$S = \{DD\}$

where $DD = \text{diabetes diagnosis}$. Our diagnostic features can be numerical features (e.g., age, lab tests, BMI and so on), ordinal features (e.g., features in *Global_symptoms* table in Fig. 5), and text features (e.g., sex, occupation, etc.). All these features have not been encoded in SCT concepts because their coding will not enhance the semantic retrieval algorithm of CBR. On the other hand, patient disorders are instance features, and we have mapped it to standard SCT concepts in another work [18]. We concentrated on the CBR semantic retrieval aspect, not sharing and interoperability issues. For example, if feature $HbA1c = 6.4$ is encoded in SCT as |43396009: Hemoglobin A1c measurement| = 6.4, this code enhances semantic interoperability but does not enhance semantic retrieval process in CBR. On the other hand, if the patient has a disorder such as nephropathy, this concept has a long sub-tree of disorders (e.g., caliectasis, amyloid nephropathy, calyceal fistula, and so on), which can be described by different physicians. The semantic similarity of

¹ <http://www.guideline.gov/>.

Table 1

The patient attributes used to describe cases. Data type = {P = primitive, I = instance of SCT concept, N = numerical, C = categorical, F = fuzzy, O = ordinal}.

Feature type	Feature name	Data type	Normal range	UoM	Min-mean-max	F. No.
Demographics	Residence	P, C	{Urban, Rural}	–	–	1
	Occupation	P, C	{Farmer, Police}	–	–	2
	Gender	P, C	{Male, Female}	–	–	3
	Age	P, N, F	–	Year	29–48.117–74	4
Diabetes lab tests	BMI	P, N, F	18.5–25	kg/m ²	20–33.117–45	5
	HbA1C	P, N, F	≤5	mmol/L	5–6.373–7.4	6
	2h PG	P, N, F	<139	mg/dL	165–202.733–235	7
	FPG	P, N, F	<99	mg/dL	96–129.633–156	8
Haematological profile	Prothrombin INR	P, N, F	0–1	%	1–1.16–1.4	9
	Red cell count	P, N, F	4.2–5.4	10 ⁶ /cmm	3.8–5.194–5.88	10
	Hbg	P, N, F	12–16	g/dL	9.8–12.332–13.4	11
	Haematocrit (PCV)	P, N, F	37–47	vol%	31.1–35.215–36.8	12
	MCV	P, N, F	80–90	fL	26.8–71.908–76.4	13
	MCH	P, N, F	27–32	pg	3.3–25.47–29.4	14
	MCHC	P, N, F	30–37	%	1.8–35.465–41.7	15
	Platelet count	P, N, F	150–400	10 ³ /cmm	135–316.183–2000	16
	White cell count	P, N, F	4–11	10 ³ /cmm	6–8.055–9.2	17
	Basophils	P, N, F	0–1	%	0–1.013–5	18
	Lymphocytes	P, N, F	20–45	%	21.2–25.768–29	19
	Monocytes	P, N, F	2–10	%	1.7–2.942–4	20
	Eosinophils	P, N, F	1–4	%	1–1.897–3.4	21
Symptoms	Urination frequency	O	–	–	–	22
	Vision	O	–	–	–	23
	Thirst	O	–	–	–	24
	Hunger	O	–	–	–	25
	Fatigue	O	–	–	–	26
Kidney Function Lab tests	Serum potassium	P, N, F	3.5–5.3	mEq/L	2.4–3.767–4.3	27
	Serum urea	P, N, F	5–50	mg/dL	17–31.56–67	28
	Serum Uric acid	P, N, F	3.0–7.0	mg/dL	3–4.237–7.9	29
	Serum creatinine	P, N, F	0.7–1.4	mg/dL	0.9–1.35–3.6	30
Lipid profile	Serum sodium	P, N, F	135–150	mEq/L	134–137.833–158	31
	LDL cholesterol	P, N, F	0–130	mg/dL	50–94.917–170	32
	Total cholesterol	P, N, F	0–200	mg/dL	158–209.367–275	33
	Triglycerides	P, N, F	60–160	mg/dL	78–144.767–189	34
Tumor markers	HDL cholesterol	P, N, F	45–65	mg/dL	30–55.533–65	35
	FERRITIN	P, C	28–397	ng/mL	–	36
	AFP Serum	P, C	0.5–5.5	IU/ml	–	37
Urine analysis	CA-125	P, C	1.9–16.3	U/mL	–	38
	Chemical examination	Protein	O	–	–	39
		Blood	O	–	–	40
		Bilirubin	O	–	–	41
		Glucose	O	–	–	42
		Ketones	O	–	–	43
		Urobilinogen	O	–	–	44
	Microscopic examination	Pus	O	–	–	45
		RBcs	O	–	–	46
		Crystals	O	–	–	47
Liver function tests	S. albumin	P, N, F	3.5–5.0	g/dL	1.9–4.082–5.4	48
	Total bilirubin	P, N, F	0.0–1.0	mg/dL	0.8–1.317–3	49
	Direct bilirubin	P, N, F	0.0–0.3	mg/dL	0.3–0.533–1.6	50
	SGOT (AST)	P, N, F	0–40	U/L	35–54.567–165	51
	SGPT (ALT)	P, N, F	0–45	U/L	35–57.317–183	52
	Alk. phosphatase	P, N, F	64–306	U/L	170–214.2–360	53
	γ GT	P, N, F	7–32	U/L	18–35.833–98	54
	Total protein	P, N, F	6.0–8.7	g/dL	3.1–4.858–8.7	55
Females history	Amenorrhea	I	–	–	–	56
	Birth	I	–	–	–	57
	Dysmenorrhea	I	–	–	–	58
Diagnosis	Diabetes type	P, C	–	–	–	59
	Nephropathy	I	–	–	–	60
	Lipid disease	I	–	–	–	61
	Cancer type	I	–	–	–	62
	Liver disease	I	–	–	–	63
	Radiological examination	I	–	–	–	64

these concepts is critical in KI-CBR retrieval engine. Moreover, the case solution features are not encoded because these features do not participate in measuring similarity between cases.

3.4. Ontology

Ontology is a formal, explicit specification of a shared conceptualization. It is a unified view of a domain, which describes its instances, concepts, and relationships between them [57]. The main

advantage of ontology usage is that it supports the sharing and reusing of formally represented knowledge by explicitly stating the concepts, relationships and axioms of a domain. Ontology is defined by a particular language. OWL2 is the most recent ontology representation language defined by W3C². In addition, ontology mainly

² <http://www.w3.org/TR/owl2-overview/>.

depends on a specific description logic (DL). For example, OWL2 is based on the SROIQ (D) DL. DL is a formal logic that can enhance the reasoning capabilities of CBR systems. Ontology can be formally defined as

Definition 5. Ontology is defined as

$$O = \{I, C, R, A\}$$

where O is an ontology for a domain of interest; C is a set of concepts in a domain; I is a set of individuals or instances in a domain; R is a set of relations among concepts including object and data relationships; A is a set of axioms holding among concepts, relations, or individual. Axioms provide explicit logical assertions about these three elements.

3.5. Fuzzy sets

The backbone of the proposed framework is the case-base fuzzy ontology. This ontology is defined by the combination of fuzzy sets theory with crisp ontology. Fuzzy set theory was introduced by Zadeh [58] to address vague and imprecise concepts. Classical sets are defined by characteristic functions:

Definition 6. Let χ be a set and A be a subset of χ ($A \subseteq \chi$). Then the function in the following equation:

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \quad (2)$$

is called the characteristic function of the set A in χ . Fuzzy sets introduce the concept of partial membership where an element can be a member of a set with a certain degree in $[0,1]$ other than $\{0,1\}$ in crisp sets. As a result, it allows the reasoning by linguistic terms. A fuzzy set can be defined as follows:

Definition 7. A fuzzy set A over a universe of discourse X is defined by a membership function μ_A (or simply A) which maps each element x to a value between $[0,1]$, as shown in the following equation:

$$\mu_A(x) : X \rightarrow [0, 1] \quad (3)$$

where A is the fuzzy set, μ_A is the degree of membership, $x \in X$, and $\mu_A(x) \in [0, 1]$. A fuzzy set A can be defined as set of ordered pairs: $A = \{x/\mu_A(x)|x \in X\}$.

3.6. Fuzzy ontology

Vagueness is the vital part of any suitable medical diagnosis system. Fuzzy logic-based systems employ the classical fuzzy logic theory, which can handle vagueness at a certain level. After the successfulness of crisp ontology and the applicability of fuzzy logic in case representation and retrieval of CBR, the integration of these two technologies (in a fuzzy ontology) will surely enhance the performance of CBR systems. Up to 17 formal definitions can be found in fuzzy ontology [59]. One definition is *an ontology that uses fuzzy logic to provide a natural representation of imprecise and vague knowledge and eases reasoning over it*. Formally speaking, a fuzzy ontology can be defined as follows:

Definition 8. A Fuzzy OWL ontology FO consists of a fuzzy ontology structure FO_S and a fuzzy ontology instances FO_I , so $FO = (FO_S, FO_I)$ [42]:

$FO_S = FID_0 \cup FAxiom_0$, where $FID_0 = FCID_0 \cup FDRID_0 \cup FOPID_0 \cup FDPIP_0$ is a set of fuzzy class descriptions, and $FAxiom_0$ is a set of fuzzy class and property axioms defined over FID_0 :

- $FCID_0$ is a set of fuzzy classes or concepts. Each fuzzy class may be a user-defined fuzzy class, or one of two predefined fuzzy classes owl: Thing and owl: Nothing.
- $FDRID_0$ is a set of fuzzy datatypes. Each fuzzy data type may be a predefined XML Schema fuzzy datatype.
- $FOPID_0$ is a set of fuzzy object properties.
- $FDPIP_0$ is a set of fuzzy data properties.
- $FAxiom_0$ is a set of fuzzy class and property axioms defined over FID_0 .

$FO_I = FIID_0 \cup FAxiom_0$, where $FIID_0$ is a set of individuals, and $FAxiom_0$ is a set of fuzzy individual axioms.

FO = fuzzy ABOX A + fuzzy TBOX T . A Fuzzy TBOX is a finite set of fuzzy concept inclusion axioms of the form $(\alpha \bowtie n)$, and fuzzy role inclusion axioms of the form $(\alpha \bowtie n)$, where $n \in (0, 1]$ and α can be a concept inclusion axiom or a role inclusion axiom. A Fuzzy ABOX is a finite set of fuzzy concept and fuzzy role assertions axioms of the form $(\alpha \bowtie n)$, where $n \in (0, 1]$ and α is a role or concept assertion of the form $a:C \bowtie \alpha, (a,b):R \bowtie \alpha, a \neq b$, and $a = b$.

The main idea in fuzzy DLs is that concepts and roles are interpreted as fuzzy subsets of an interpretation's domain. In fuzzy DLs, axioms can occur with a certain degree of truth. The notion of satisfaction of a fuzzy axiom E by a fuzzy interpretation I , denoted $I \models E$, is defined in [60] as follows:

- $I \models (\tau \geq \alpha)$ iff $\tau^I \geq \alpha$
- $I \models (\text{trans } R)$ iff $\forall_{x,y \in \Delta^I}, R_I(x, y) \geq \sup_{z \in \Delta^I} R^I(x, z) \otimes R^I(z, y)$
- $I \models R_1 \subseteq R_2$ iff $\forall x, y \in \Delta^I, R_1^I(x, y) \leq R_2^I(x, y)$
- $I \models (\text{inv } R_1, R_2)$ iff $\forall x, y \in \Delta^I, R_1^I(x, y) - R_2^I(x, y) \geq 0$

Concept C is satisfiable iff there is an interpretation I , and an individual $X \in \Delta^I$ such that: $C^I(x) > 0$. For a set of axioms ε , we say that I satisfies ε iff I satisfies each element in ε . I is a model of E iff $I \models E$. I satisfies (is a model of) a fuzzy KB $K = (A, T)$, denoted $I \models K$, iff I is a model of each component A, T , respectively. An axiom E is a logical consequence of a knowledge base K , denoted $K \models E$ iff every model of K satisfies E .

3.7. The fuzzy-semantic case representation

Given a case base crisp ontology, elements that can be fuzzified include datatypes, object properties (through fuzzy modifiers), and data properties. Moreover, fuzzy case base ontology can include crisp assertions side-by-side with fuzzy assertions. Cases are stored in a fuzzy ontology as concept instances. As a result, a case-base CB is defined as:

$CB = \{0_1, 0_2 \dots 0_m\}$, where m is the number of cases and 0_k is the k 's case. Each case in the case-base ontology is defined as follows:

Definition 9. A case 0_k is a vector of conjunctive set of predicates of the form:

$$0_k = P_1 \cap P_2, \dots, P_n$$

where P_i is the i 's predicate of four forms: A (fuzzy) concept assertion $a:C_i \bowtie \alpha$, A (fuzzy) object property assertion $(a,b):R_i \bowtie \alpha$, or A (fuzzy) data property assertion $(a,v):T_i \bowtie \alpha$, for a, b as abstract individuals and v as a literal value. A (fuzzy) data property assertion $(a,v):T_i$, for v as a fuzzy linguistic term defined using a fuzzy datatype. By converting the physician query into a semantic query of the form $i_Q \equiv \sim P_{Q1} \cap P_{Q2}, \dots, P_{Qn}$, the similarity calculation between these predicates becomes straightforward. This similarity depends on the inference capabilities of the utilized ontology reasoners. The querying process will be detailed in subsequent sections.

As we assert, (fuzzy) ontology has many roles in CBR. It can play many roles in every phase of CBR including case representation, indexing, retrieval, adaptation, and maintenance. For the case representation and retrieval steps, these roles include the following:

It considerably reduces the knowledge acquisition bottleneck [11]. It allows knowledge engineers to use knowledge already acquired, conceptualized, and implemented in a formal language, such as DLs based languages.

It supports persistence of cases and indexes using individuals or concepts that are embedded in the ontology [31].

It can be used as a vocabulary to define the case structure, either if the cases are embedded as individuals in the ontology itself, or if the cases are stored in a different persistent media such as a database [10].

It can play the role of terminology to define the query vocabulary [31]. The user can better express his requirements if he can use a richer vocabulary to define the query.

During the similarity computation, the ontology allows the user to bridge the semantic gap between the query terminology and the case base terminology [18].

It supports dynamic case storage where features can be added, updated, or deleted from the case base.

It preserves storage space in which many cases can point to the same feature values.

It can define a semantic index of cases for an in-memory case base [61].

Ontology's description logic reasoners such as Pellet and FaCT++ [62] can check the case base consistency, redundancy, and adequacy, which is not possible in a regular database environment [42]. Moreover, reasoners significantly enhance the effectiveness of the case retrieval process.

Domain's background knowledge such as SCT medical terminology can be integrated with case base ontology to create a KI-CBR [11]. For active CDSS, ontology supports interoperability between CBR-based CDSS and EHR system database.

Ontology provides a common understanding of a domain. As a result, it supports the implementation of distributed CBR systems [52].

By using an ontology, complex relations between case features can be created. For example, the relationship between diabetes symptoms and disorders can be used for inference values of missing features.

Heterogeneous cases, which have no fixed structure, can be designed. They may have different structures with different types and numbers of features.

Cases can have relationships with each other such as Cause, ISA, Part.Of, Result.From.etc. These relationships can handle incomplete cases and allow default values (by inheritance) [11].

Compound features, which contain many other simple and compound features, can be defined.

Utilizing (fuzzy) ontology engineering methodologies can help making the CBR knowledge acquisition process more efficient.

In the medical domain, where there are many standard ontologies as SCT, GO.etc., ontologies reuse has many benefits such as standardization of the CDSS knowledge, interoperability, distribution of knowledge, and so on. Moreover, many ontologies can be integrated with the CBR, where each case feature can be semantically connected with an ontology. For example, the patient *disease feature* can be connected to Disease ontology; patient *gene feature* can be connected to GO ontology; patient *lab test features* can be connected to LOINC ontology.etc.

Ontology-based representation of cases enables reusing and adaptation in a variety of application scenarios [21].

Creating a fuzzy case-base ontology from a fuzzy case-base database is supported by methodologies [40], languages [63], tools [48], and reasoners [60]. These fuzzy ontologies add vagueness to the KI-CBR systems.

4. Research methodology

As shown in Fig. 6, we follow a specific methodology to finish this study. To accomplish the purpose of this study, we have utilized some existing technologies and studies. Moreover, we have utilized our previous research studies to complete some specific steps. In the figure, we make a clear cut between the current study goals and the other utilized works.

In the first step, the detailed understanding of the nature of diabetes mellitus disease and its diagnosis process requires deep interviews with the domain experts. The next step involves the collection of patients EHR records to implement the case-base fuzzy ontology. This dataset will determine the structure of the case-base ontology, and it will be used to populate the ontology. However, the collected medical data needed preparation processes including (pre-processing to enhance the quality of data and calculate the weight vector, coding to formalize the unstructured contents of medical data using a standard medical-ontology, and fuzzification to fuzzify some numerical features). Moreover, a standard ontology needs to be created from the huge SCT ontology to be used as the domain background knowledge in similarity calculation process. We utilize our previous studies to accomplish this

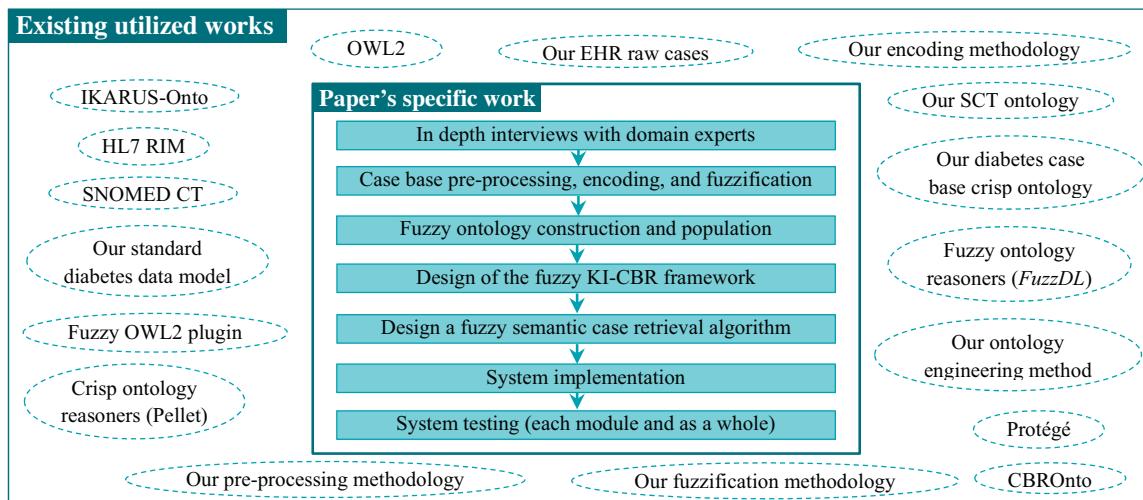


Fig. 6. Research structure.

step [9,18,19,35]. The next step involves the construction and population of the case-base fuzzy ontology. As this step is complex, we extend the previously proposed crisp ontology [40] using a high-level methodology [31] and create a fuzzy OWL2 ontology using protégé tool. As far as we know, there are no fuzzy case-base ontologies for medical CBR systems. In the next step, we propose a fuzzy KI-CBR framework. This framework is an integrated set of modules. Each module is for a specific purpose, and each one has inputs and outputs. The framework will be detailed in the next section. Next, for the fuzzy case base ontology, and for handling the supported feature types, we design a hybrid semantic retrieval algorithm. The next step is the implementation of our framework using JAVA programming language. Finally, we test the implemented system using case base of real diabetics.

5. The proposed fuzzy KI-CBR framework for diabetes diagnosis

This section provides a description of our proposed fuzzy-ontology based CBR system for diabetes diagnosis. The architecture of this system is shown in Fig. 7. It has six modules: Case source preparation, case base ontology engineering, terminology server, fuzzy case-base ontology population, case retrieval engine, and case query parser. The main steps of the framework are case-base preparation and case retrieval.

The case-base preparation step is achieved by the case source preparation, case-base ontology engineering, terminology server, and fuzzy case-base ontology population modules as follows:

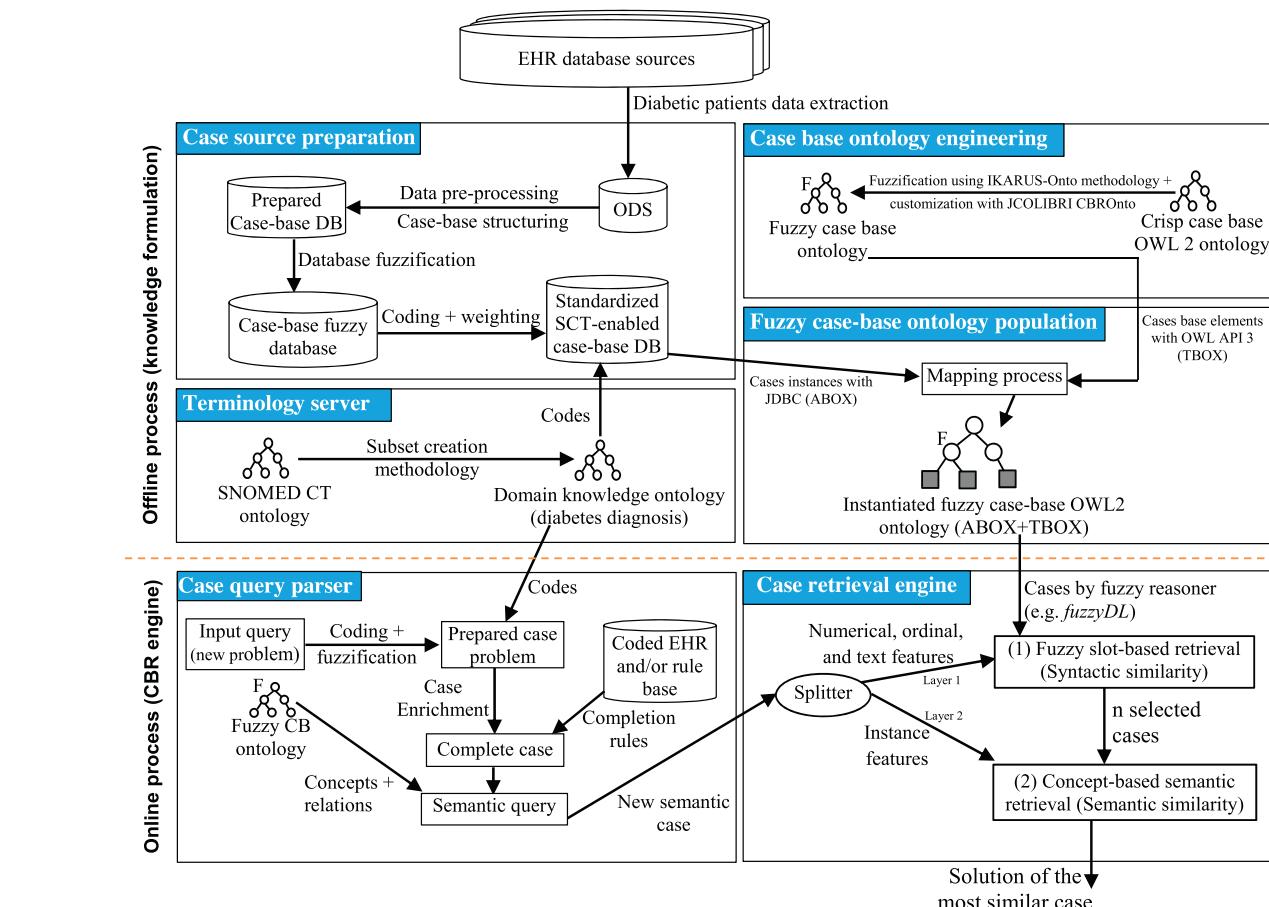


Fig. 7. The proposed CBR framework.

1. The **case-source preparation module** takes the EHR raw data and converts it into pre-processed, encoded, and fuzzified relational database.
2. The encoding process is based on SCT codes from the **terminology server module**.
3. The **case-base ontology-engineering module** builds the case-base crisp ontology and extends it to a fuzzy ontology.
4. The **fuzzy case-base ontology population module** populates the resulting fuzzy ontology in step 3 with the fuzzy relational database in step 1.

The **case retrieval step** is achieved by the terminology server, case retrieval engine, and case query parser modules.

1. The **case query-parser module** takes the user query vector and converts it to a semantic query vector according to the case base fuzzy ontology terminologies.
2. The **case retrieval-engine module** takes the created semantic query vector generated in step 1 and searches for the most similar k cases in the fuzzy case-base ontology.
3. The clinical similarity between medical concepts of semantic features is based on the SCT ontology in the **terminology server module**.

5.1. Case source preparation module

This module prepared the EHR raw data to a case-base structure and content. It collected the patient's features related to a diabetes diagnosis from distributed EHR systems and stored it in an operational data store (ODS). We have collected 60 cases, which describe

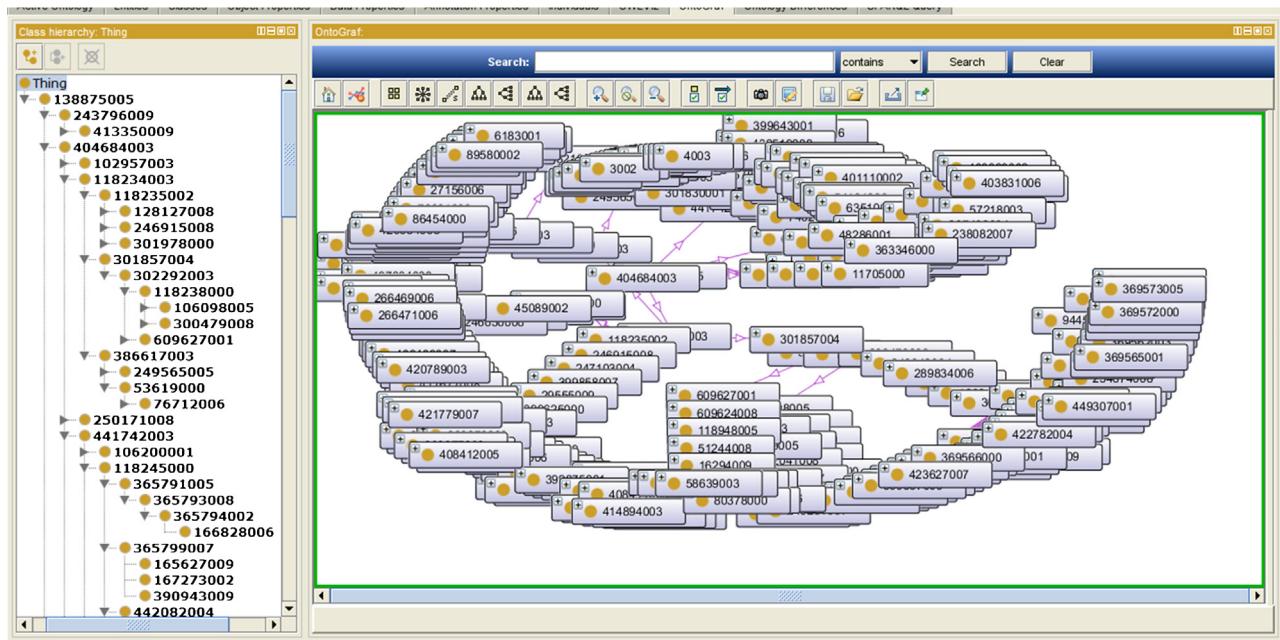


Fig. 8. The proposed SCT reference set for diabetes diagnosis.

diabetic patients, as shown in Table 1. These cases are descriptive of all types of cases as in [64], which used 47 cases only. Next, these data were anonymized, cleaned, and normalized. Features' weights were calculated using machine learning algorithms including genetic algorithm, decision tree, and others. El-Sappagh et al. [35] have proposed a case-base preparation process and applied it to the used case-base data. Moreover, the data were converted to a case base structure using our proposed standard data model [9]. In addition, the prepared case-base was coded according to SCT reference set that was created, which is specialized for diabetes diagnosis [18]. Finally, the encoded case-base was fuzzified in a fuzzy relational database using our proposed methodology in another work. The works of El-Sappagh et al. in [18,35,64] are utilized in this module to prepare the used EHR medical data. The resulting database is the source of instances (ABOX) for our proposed fuzzy case-base ontology.

5.2. Terminology server module

This module creates the domain background ontology. This knowledge is critical in two places: (1) in semantic similarity measurement, and (2) in query formulation. The domain knowledge ontology can be built locally, or it can depend on a standard medical ontology such as SCT [56]. Unfortunately, ontologies are typically created in an ad-hoc manner, which may influence the accuracy of the similarity calculations [64]. The second choice is better because clinical ontologies are mature, and they include all required medical concepts and relationships. Moreover, this standardization enhances the interoperability, reuse, sharing, and integration with the EHR environment. SCT was the terminology used in this study. Building a complete ontology is not realistic and using the whole SCT in CBR affects the retrieval algorithm because it is a very large ontology (i.e., it contains 361,800 concepts). We have collected all SCT concepts related to diabetes according to our proposed methodology [18], and built its OWL 2 ontology (TBOX), as shown in Fig. 8. This ontology only contains 550 concepts. When measuring semantic similarity with JCOLIBRI API, it is between concept instances; however, SCT contains only concepts. We have solved this problem by creating an instance for each selected concept with the same name (ABOX). Moreover, we have represented

the selected concepts using its conceptIDs. Fully specified names, symptoms, and preferred names can be added as annotations with their corresponding names. As shown in Fig. 8, this ontology is not user readable. We resolve this issue in our future work. Each concept name begins with the pattern "C_." to be readable by JCOLIBRI API³ as a concept and differentiate it from instances. The resulting ontology is a directed acyclic graph (DAG), which supports single inheritance only, but the whole SCT supports multiple inheritances. An ontology has a structured format with relationships between concepts. The "IS_A" relationship between a parent and a child is the core relationship, whereas other semantic relationships provide additional associations between terms (such as "part-of" or "active-ingredient-of"). Our ontology concentrates on the "IS_A" relationship only to form a taxonomy of concepts. Enriching the ontology with other relationships and axioms will be considered in future work.

5.3. Case-base ontology engineering module

This module converts our crisp case base ontology created in our previous work [31] into a fuzzy case-base ontology. We apply the procedural steps of IKARUS-Onto [40] methodology for converting a crisp ontology to a fuzzy ontology. The IKARUS-Onto is a high-level and abstract methodology to add a fuzzification aspects to a crisp ontology. We customize this methodology according to our requirements. It is the most accurate and complete methodology. Moreover, the resulting ontology is represented by Bobillo and Straccia syntax as OWL 2 ontology using Fuzzy OWL2 2.1.1 plug-in in Protégé 4.1 [63]. This syntax adds the fuzzy components as annotations for concepts and relationships (i.e., datatype and object properties). Moreover, it allows the creation of hedges and fuzzy data types. The default reasoners such as Pellet [62] and default modeling tools such as protégé can be used with the resulting ontology because all fuzzy aspects are coded as annotations (i.e., FuzzyLabel annotation). Every annotation is delimited by a start tag <fuzzyOwl2> and an end tag </fuzzyOwl2>, with an attribute fuzzyType specifying the fuzzy element being tagged.

³ <http://gaia.fdi.ucm.es/research/colibri/jcolibri>

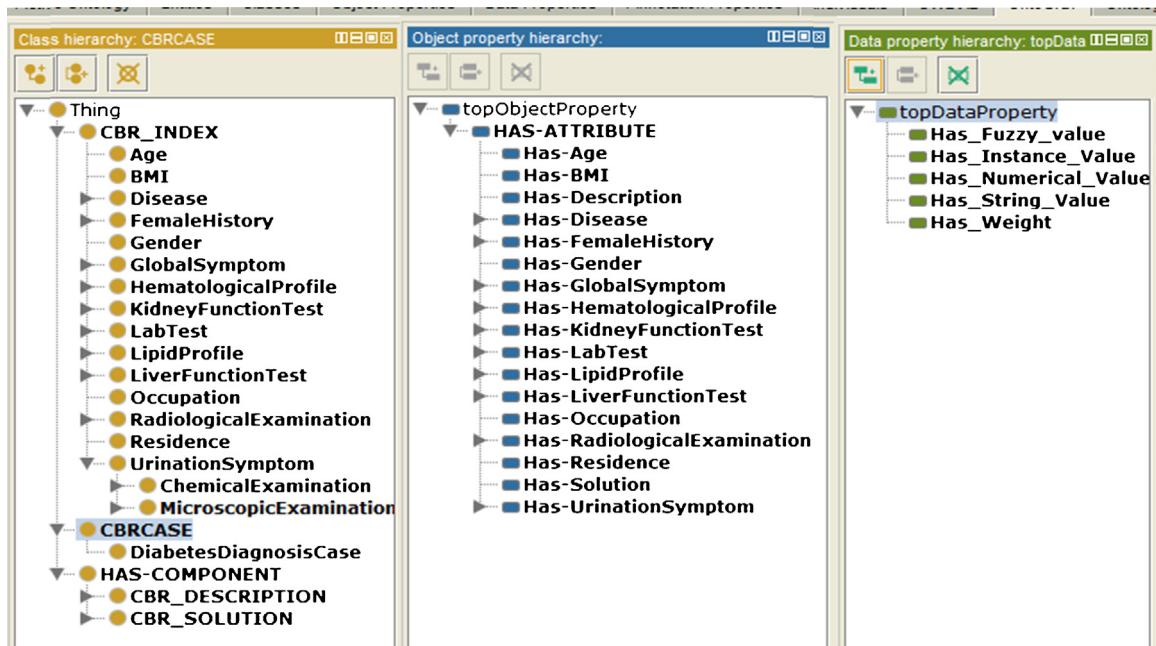


Fig. 9. The crisp case base ontology.

5.3.1. Crisp ontology customization

Before starting the fuzzification process, our previously created crisp case-base ontology [31] is customized according to our case-base's fuzzy database contents and the CBROnto standard ontology of JCOLIBRI 2 API [52]. Fig. 9 shows our crisp case base ontology after customization. This customization includes:

1. No outcome concept in our new ontology,
2. We have removed the *temporal aspect* because we do not provide the treatment plan for the diabetic patient, and our data set does not have multiple values over time for case features,
3. The *context* has been removed, and we will propose an indexing methodology in another work,
4. The *diagnoses* are Normal, Prediabetic, Prediabetic_Gestational, Diabetic, and Diabetic_Gestational only. Our data set cannot determine the type of diabetes (i.e., Type 1 or Type 2) and the type of pre-diabetes (i.e., IFG, IGT),
5. In our data set, many of the problem description features are new and not modeled in the previous ontology [31],
6. The hierarchy of the ontology is simplified as much as possible to be compatible with CBROnto,
7. Dealing with rules in the form of SWRL will be a future work to enhance the semantic of our case base.

As shown in Fig. 9, CASE_INDEX subsumes all of the case features, CBRCASE subsumes case instances, and HAS-COMPONENT subsumes the two parts of the case. This way, we utilize OntoBridge API of JCOLIBRI2 to address ontology storage, retrieval, and manipulation in a straightforward way [52]. In ontology-based CBR, cases are represented as concept instances and their attributes are represented as ontology relations or properties. The values that relation attributes may take are instances defined within some domain ontology. For example, consider a small fragment of our case base containing only age, gender, cancer, and labTest.

In Fig. 10, all of the case base data and structure are inside the case-base ontology. We may implement this ontology as two separate components: A case base structure stored in an OWL2 ontology, and instances of cases and features stored in a database, as shown in Fig. 11. Each choice has its advantages and limitations, and we chose the first one.

5.3.2. Case-base ontology fuzzification process

Our proposed fuzzy KI-CBR (FKI-CBR) framework operates on two axes, namely the ontology-based representation of imprecise knowledge and the utilization of this knowledge for effective case retrieval. For the first axis, a fuzzy ontology is proposed. For case retrieval, an algorithm that utilizes ontology and fuzzy is proposed.

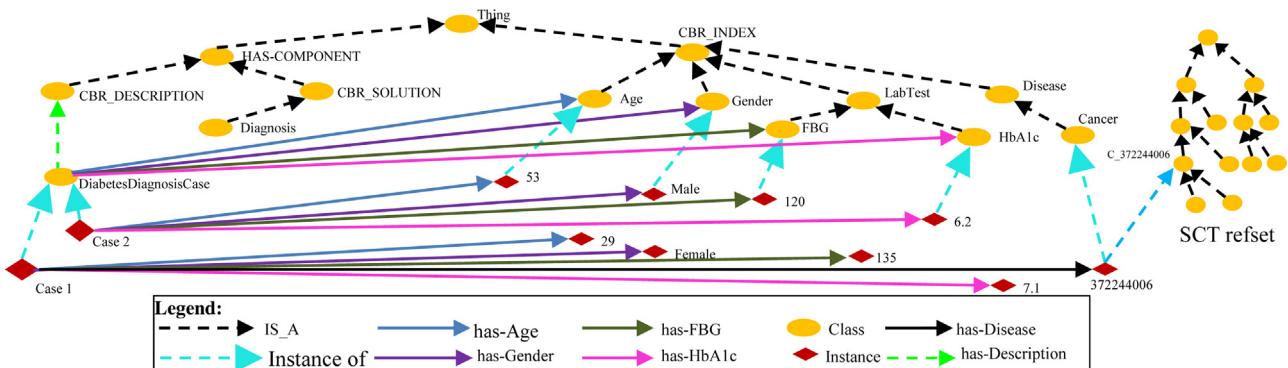


Fig. 10. A small fragment of crisp case base instantiation structure.

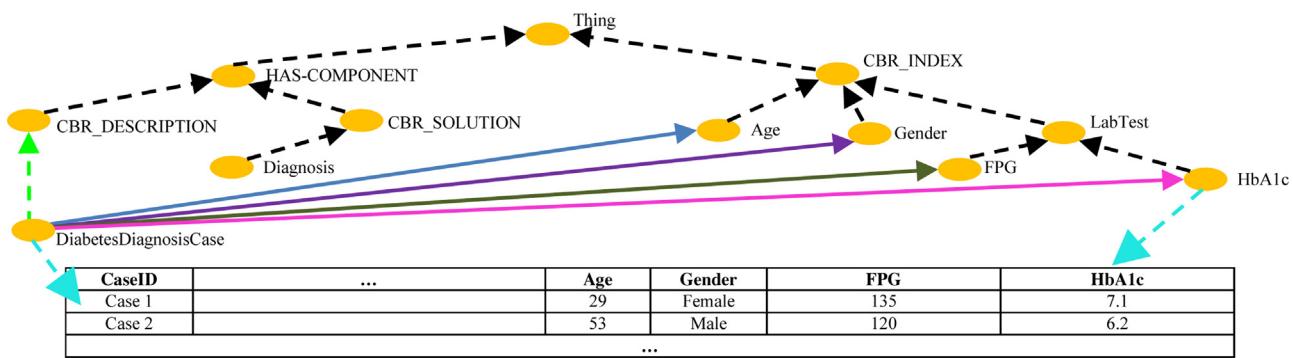


Fig. 11. Case instances stored in database.

An ontology may be defined as a set of concepts, instances, properties (data type properties) and relations (object properties). A concept represents a set or a class of entities within a domain while the entities that belong to a concept are called instances of this concept. A relation in turn links a concept instance to another instance while a property links an instance to a standard data type such as a string, integer, float, Boolean, etc.

5.3.2.1. The case-base fuzzy ontology design. In this paper, we handle only vagueness (i.e., imprecision), but uncertainty is not handled (i.e., probability, ambiguity, or inexactness). A fuzzy ontology may be informally defined as an ontology that expresses vague knowledge using a fuzzy set (fuzzy concept) namely *degree-vagueness* and fuzzy relation and properties namely *combinatory-vagueness* [40]. Because a crisp ontology is a special case of a fuzzy ontology, in which all relation and property degrees are equal to 1, fuzzy ontology-based CBR retains the characteristics of the traditional ontology-based CBR paradigm. Crisp elements that can be fuzzified include data types, object properties (through fuzzy modifiers), and data properties (through fuzzy modified data types). In other words, the fuzziness of ontology includes modeling of [40]:

1. Fuzzy concepts: concepts whose instances may belong to it in certain degrees, such as *youngPatient* are fuzzy concepts. Because *young* is a vague predicate, the concept is also vague and, therefore, can be represented as a fuzzy one; it allows the fuzzy concept assertions such as “*patient X is an instance of youngPatient at a degree of 0.7*.”
2. Fuzzy relations: there are two main types, (2.1) Fuzzy object relations, which link concept instances at a certain degree, and it allows fuzzy role assertions as “*patient X has-Disease Y at a degree of 0.8*.” (2.2) Fuzzy data type relations, which either assign literal value to concept instances at certain degrees (e.g., *patient X has-Residence “Rural” at a degree of 0.4*), or a fuzzy datatype is assigned to a concept instance (e.g., *patient X has-Fuzzy-Age young*), which includes the *age* fuzzy predicate.

There are many fuzzy ontology construction methodologies as IKARUS-Onto [40], UFOC [65], UPFON [66] and OntoMethodology [67]. Moreover, fuzzy ontology representation languages have been proposed in [63,68]. Fuzzy ontology reasoners include FuzzyDL, Fire, and DeLorean. Fuzzy reasoners use fuzzy description logics as fuzzy SROIQ (D), F-ALC, fuzzy SHIN, and fuzzy SHOID (D). As shown in Table 2, in our case, the fuzzy case-base ontology construction process, to store fuzzy cases about diabetic patient, used this IKARUS-Onto methodology, OWL 2 fuzzy extension [63], the

*FuzzyDL*⁴ reasoner using fuzzy DL SROIQ (D) [60], and a protégé tool with the *fuzzy OWL* plugin [69]. The plug-in does not translate fuzzy representations into OWL 2, but rather eases their representation by allowing specification of the type of fuzzy logic used, the definition of fuzzy data types, fuzzy modified concepts, weighted concepts, weighted sum concepts, fuzzy nominals, fuzzy modifiers, fuzzy modified roles and data types, and fuzzy axioms. Table 2 shows the execution steps of the IKARUS-Onto methodology in our case study.

5.3.2.2. The case-base fuzzy ontology implementation. For the fuzzification of our crisp case-base ontology, we use the Fuzzy OWL2 2.1.1 plug-in⁵ in Protégé 4.1.⁶ In the following, we detail fuzzy concepts, data types, relations, and data types. A fuzzy data type D is a pair Δ_D, ϕ_D where Δ_D is a concrete interpretation domain, and ϕ_D is a set of fuzzy concrete predicates d with an arity n and an interpretation $d^1 : \Delta_D^n \rightarrow [0, 1]$, which is an n -ary fuzzy relation over Δ_D [63]. For fuzzy data types, the functions allowed in Fuzzy OWL 2, defined over an interval $[k_1, k_2] \subseteq \mathbb{Q}$ are $d \rightarrow \{\text{left}(k_1, k_2, a, b)\}(\text{fig. 13c}), \text{right}(k_1, k_2, a, b)\}(\text{fig. 13d}), \text{Triangle}(k_1, k_2, a, b, c)\}(\text{fig. 13b}), \text{Trapezoidal}(k_1, k_2, a, b, c, d)\}(\text{fig. 13a}), \text{linear}(k_1, k_2, c)\}(\text{fig. 13e}), \text{mod}(d)\}$. The formalization of each element in the ontology is conducted as follows:

5.3.2.2.1. Fuzzy data types and fuzzy concrete roles (data properties). For each of the numerical features in our case base, our domain experts have defined their ranges, and fuzzy membership functions, their shapes, and parameters. For fuzzification of these values, we define two things: (1) a fuzzy data type, (2) a fuzzy concrete role. Because we have ~70 features, and most of them are numerical, we only give examples here. In cooperation with our domain experts, we have used MATLAB to define the fuzzy membership functions and their ranges, shapes, and equations, as shown in Fig. 12. Experience suggests that the overlap of triangle-to-triangle and trapezoid-to-triangle fuzzy regions averages somewhere between 25% and 50% of the fuzzy set base [70]. In our case, our domain expert has recommended fixing the normal ranges and overlapping low and high ranges by 50% to the normal range, see Fig. 12b. Considering HbA1c lab test values, let us assume its range is [71, 71] and its linguistic terms are *lowA1c* (left shoulder 5.7, 6.05), *normalA1c* (triangle (5.7, 6.05, 6.4)), and *highA1c* (right-shoulder (6.05, 6.4)). Firstly, we create a datatype *lowA1c* and then annotate it as fuzzy datatype. This action is repeated for every linguistic term in each

⁴ FuzzyDL Reasoner: <http://gaia.isti.cnr.it/straccia/software/fuzzyDL/fuzzyDL.html>.

⁵ Fuzzy OWL2 2.1.1 plug-in: <http://www.straccia.info/software/FuzzyOWL/>.

⁶ Protégé 4.1: <http://protege.stanford.edu>.

Table 2

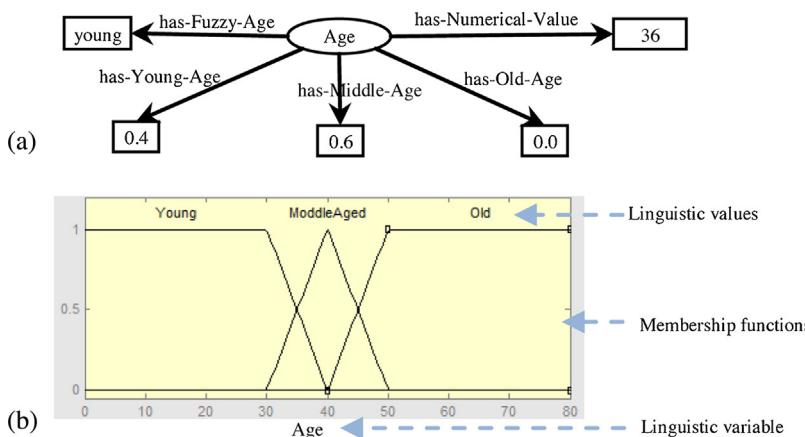
The fuzzy case-base ontology construction process.

IKARUS-Onto Step	Its application in our domain												
(1) Acquire crisp ontology	As shown in figure 9, we have developed our case base crisp ontology.												
(2) Establishment of the need for fuzziness	As our domain experts asserted, handling vagueness in patient's description is required in our diabetes diagnosis CDSS based on CBR. Both of stored cases in the knowledge base and physician queries are expressed to use vague terms.												
(3) Definition of fuzzy ontology elements	<p>(1) <i>Identification of vague concepts:</i> we have the option of considering each numerical feature, in our dataset (Table 1), as a fuzzy concept or fuzzy variable such as HbA1c, FPG, BMI, age, etc. Moreover, for each concept, we can create a fuzzy sub-concept for each linguistic value. For example, for BMI, we can create the fuzzy concepts <i>low-BMI</i>, <i>normal-BMI</i> and <i>obese-BMI</i> as fuzzy sub-concepts of BMI. These fuzzy concepts allow instances to be in certain concepts with certain degrees, such as <i>low-BMI</i> (20) with 0.6. However, this modeling complicates the resulting ontology, as each concept have at least three sub-concepts. As a result, the ontology population and the fuzzy semantic retrieval algorithm are complicated. These fuzzy aspects are modeled using fuzzy datatypes, fuzzy modifiers, and fuzzy concrete roles, as shown in section 5.3.2.</p> <p>(2) <i>Identification of vague relations and attributes:</i> In our case, we have no fuzzy relations between concepts. For fuzzy attributes, we have two fuzzy attributes (i.e. data type properties) namely, <i>has-Disease</i> and <i>lived-In</i> of type <i>Degree-vagueness</i>.</p> <table border="1"> <thead> <tr> <th>Attribute</th><th>Destinations</th><th>Vagueness nature</th><th>Degree interpretation</th></tr> </thead> <tbody> <tr> <td><i>has-Disease</i></td><td>(case/patient, string)</td><td><i>Degree-vagueness</i> of patient's disease</td><td>The extent to which the patient has a disease</td></tr> <tr> <td><i>lived-In</i></td><td>(case/patient, string)</td><td><i>Degree-vagueness</i> of living place</td><td>The extent to which lived in Rural and Urban</td></tr> </tbody> </table>	Attribute	Destinations	Vagueness nature	Degree interpretation	<i>has-Disease</i>	(case/patient, string)	<i>Degree-vagueness</i> of patient's disease	The extent to which the patient has a disease	<i>lived-In</i>	(case/patient, string)	<i>Degree-vagueness</i> of living place	The extent to which lived in Rural and Urban
Attribute	Destinations	Vagueness nature	Degree interpretation										
<i>has-Disease</i>	(case/patient, string)	<i>Degree-vagueness</i> of patient's disease	The extent to which the patient has a disease										
<i>lived-In</i>	(case/patient, string)	<i>Degree-vagueness</i> of living place	The extent to which lived in Rural and Urban										
	<p>For example, patient X has-Disease "Liver Cancer" with degree 0.3 is a fuzzy property. This degree may represent the severity of the disease or the time in years where this patient has the disease. Moreover, the patient may be lived in Rural and Urban for different periods (e.g. patient X lived-in "Rural" with degree 0.9, which means that the patient has lived in rural most of his life). The assignment of these degrees is populated from the case-base fuzzy relational database.</p> <p>Moreover, for each of our numerical features, we create an attribute <i>has-Fuzzy-X</i> (X=Age, BMI, FPG, HbA1c, etc.) which store the fuzzy value with the largest membership value. For example, in figure 12a, for Age=36, <i>young</i> has the largest membership value (i.e. $\mu_{middle-age} = 0.6$). As a result <i>has-Fuzzy-Age= middle-aged</i> is asserted. In addition, for each fuzzy value of numerical attributes, we create a fuzzy concrete role (e.g. for BMI feature, we create <i>has-Low-BMI</i>, <i>has-Normal-BMI</i>, and <i>has-Obesity</i>), as detailed in section 5.3.2. Figure 12b shows the fuzzy datatype properties for Age concept.</p> <p>(3) <i>Identification of fuzzy datatypes or vague terms:</i> All numerical features in Table 1 can be assigned vague terms. These terms enrich the querying of the case base, and similarity matrices can be used as a similarity measure. A fuzzy datatype is created for each fuzzy value to all numerical features in Table 1. These fuzzy datatypes are the range of the previously defined fuzzy concrete roles. Figure 12 (b) shows an example of fuzzy datatypes as <i>young</i>, and <i>old</i>.</p>												
(4) Formalization of fuzzy ontology elements	Our ontology is encoded in OWL2. We use the fuzzy extension to OWL2 proposed by [63]. This extension adds the fuzzy aspects as annotations in the crisp ontology, which allows using an ontology as crisp and fuzzy ontology at the same time, depending on the used reasoner. It supports the required fuzzy ontology elements as fuzzy datatypes. It has many reasoners (i.e. <i>fuzzyDL</i>) that support our needed fuzzy reasoning capabilities. Moreover, a protégé plugin exists to support modeling of fuzzy aspects. After modeling the fuzzy aspects, a parser exists for each reasoner to convert the ontology code, according to the selected reasoner's syntax. Details of this process are shown section 5.3.2.												
(5) Validation of fuzzy ontology	Ontology validation determines whether the built ontology meets certain application requirements. It includes the checking of its consistency, correctness, accuracy, and completeness. This step is detailed in section 6.2.1.												

fuzzy variable in our case base ontology. Next, for each numerical feature, we have defined a concrete role for each of its linguistic values. The previously defined fuzzy datatypes are used as ranges for these roles. Continuing with HbA1c, we define three fuzzy concrete roles *hasLowA1c*, *hasNormalA1c*, and *hasHighA1c*. For example, the

hasLowA1c is modeled as *hasLowA1c* (*HbA1c*, *lowA1c*) where *HbA1c* is a crisp concept and *lowA1c* is a fuzzy data type.

5.3.2.2. *Fuzzy modifiers, fuzzy modified data types, and fuzzy modified roles.* Modifiers can improve the expressiveness of the ontology and semantic queries. The degree of membership of fuzzy



$$\mu_{Young}(x) = \begin{cases} 0 & x > 40 \\ \frac{40-x}{10} & 30 \leq x \leq 40 \\ 1 & x < 30 \end{cases}$$

$$\mu_{MiddleAged}(x) = \begin{cases} 0 & x \leq 30 \\ \frac{x-30}{10} & 30 < x \leq 40 \\ 1 & 40 < x < 50 \\ 0 & x \geq 50 \end{cases}$$

$$\mu_{Old}(x) = \begin{cases} 0 & x < 40 \\ \frac{x-40}{10} & 40 \leq x \leq 50 \\ 1 & x > 50 \end{cases}$$

Fig. 12. An example of fuzzy datatypes.

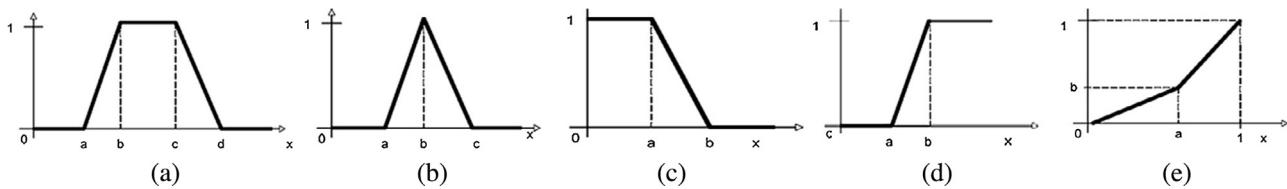


Fig. 13. Membership functions in [16] for fuzzy data type definition (i.e., fuzzy concrete domains) and fuzzy modifier functions: (a) trapezoidal function; (b) triangular function; (c) left-shoulder function; and (d) right-shoulder function.

data types may be changed using fuzzy modifiers. A fuzzy modifier is a function $f_{mod} : [0, 1] \rightarrow [0, 1]$, which applies to a fuzzy set to change its membership function, which can be linear (c) (Fig. 12e) or triangular (a, b and c) (Fig. 12b). By the help of domain expert, we defined modifies values as very, slightly, somewhat etc. with the help of a domain expert. For example, we have defined fuzzy modifiers very as linear (0.85). In our work, these modifiers have two purposes (Fig. 13):

To modify a data type, such as the new data type *veryLowA1c*, which is a modified version of *lowA1c* as shown next:

- <fuzzyOwl2 fuzzyType = "datatype">
- <Datatype type = "modified" modifier = "very" base = "lowA1c"/>
- </fuzzyOwl2>

To modify a fuzzy concrete role as shown next.

5.3.2.2.3. *Fuzzy modified data type properties.* The other type of fuzzy data type properties are Degree-vagueness as *has-Disease* and *lived-In* attributes as shown in Table 2, they are modeled as fuzzy modified roles. For example, *has-Disease* role can be modified by *very* modifier in a new role *very-has-Disease* as

- <fuzzyOwl2 fuzzyType = "role">
- <Role type = "modified" modifier = "very" base = "has-Disease"/>
- </fuzzyOwl2>

5.3.2.2.4. *Fuzzy logic of the ontology.* We have selected *Zadeh* fuzzy logic for our ontology where: $t\text{-Norm } \alpha \otimes \beta = \min \{\alpha, \beta\}$, $t\text{-Conorm } \alpha \oplus \beta = \max \{\alpha, \beta\}$, Negation $\neg \alpha = 1 - \alpha$, and Implication $\alpha \Rightarrow \beta = \max \{1 - \alpha, \beta\}$. This annotation is at the ontology level as

- <fuzzyOwl2 fuzzyType = "ontology">
- <FuzzyLogic logic = "zadeh"/>
- </fuzzyOwl2>

The resulting fuzzy ontology structure (TBOX) contains 63 classes, 54 object properties, 138 (fuzzy) datatype properties, 105 fuzzy datatypes. After creating the fuzzy ontology structure, the next step is to create the ontology instances. The instances of the cases and the instances of its describing features are populated from our fuzzy case base relational database. We populate the ontology with 60 real world diabetes diagnosis individual cases.

5.4. Fuzzy case-base ontology population module

Fuzzy ontology population from the fuzzy relational database has been studied [42]. Moreover, there are protégé plugins to automate the process such as FRDB2FOnto [42], which convert the fuzzy database schema and content to a fuzzy ontology structure and instance. On the other hand, for storage of large ontologies, fuzzy ontologies can be stored in semantic preserved databases [73]. We selected the first choice to be compatible with the JCOLIBRI2 framework. Inspired by ontology population approaches, we developed our procedure to fill the resulting case-base fuzzy ontology with cases (i.e., instances) from our previously modeled case-base fuzzy relational database. We show the process on a single fuzzy table *fAge*, which includes the fuzzy components of feature *age*. Case base crisp ER model in Fig. 5 has been previously fuzzified and implemented into a fuzzy relational database. As shown in Fig. 15, the *Age* feature in table *Patient.Case* (Fig. 15a) has been fuzzified into *fAge* table (Fig. 15b).

According to our resulting fuzzy ontology, we can map between fuzzy concrete properties as *has-Young-Age* and the attributes of *fAge* relation as *youngAge*. Moreover, the *has-Age* object property connects the instances from classes the *Case* and *Age* as: ClassAssertion (Case C1); ClassAssertion (Age A1); ObjectPropertyAssertion (has-Age C1 A1). The same process done in Fig. 10 for the whole case-base crisp ontology was performed for the fuzzy ontology. Our mapping rules of database instances to ontology instances are

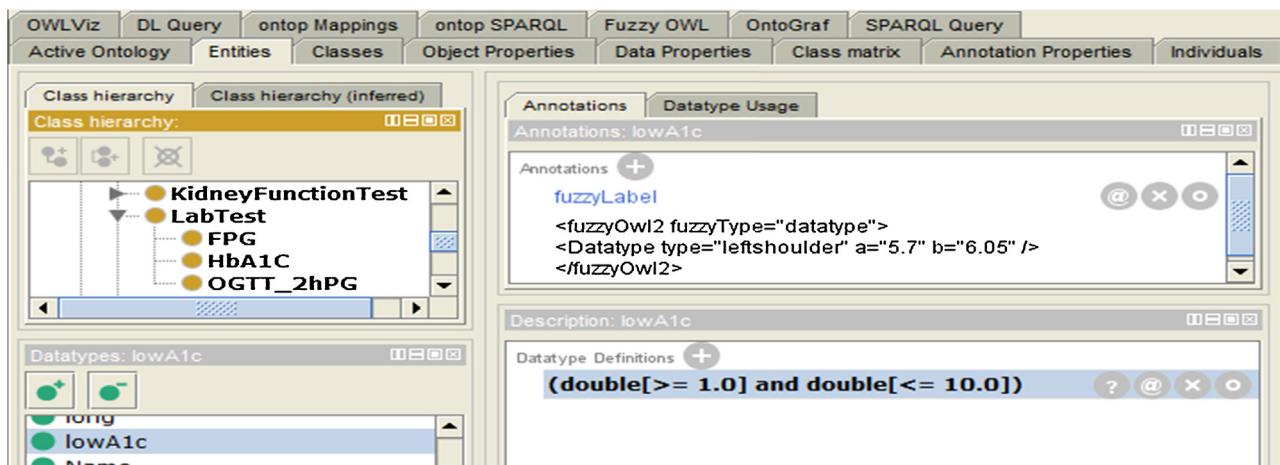


Fig. 14. An example of fuzzy data type definition.

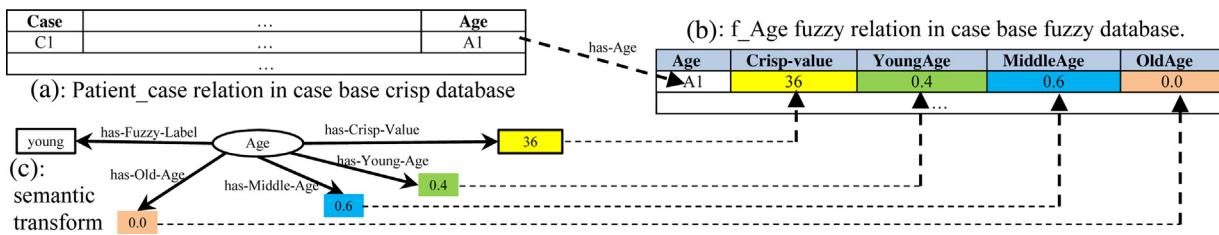


Fig. 15. Correspondence between fuzzy relational database and fuzzy ontology.

guided by W3C rules⁷ and the rules in Zhang et al. [74] by adapting their rules to work with the fuzzy relational database (e.g., table, tuple, attribute, primary key, and foreign key are mapped to concept, instance, data type property, axiom, and object property, respectively). The resulting ontology is now ready with the sets of TBOX and ABOX. After populating the ontology with 60 cases, it contains 2640 concept instances.

The resulting case-base is a collection of case instances in the ontology. Case attributes are represented as fuzzy data properties and fuzzy object properties, as follows:

$$CB = \bigcup_{i=1,n} CBRase_i, \text{ where } n \text{ is the number of cases,}$$

$$\begin{aligned} & (\exists Has_ID.string) \cap (\exists CASE_COMPONENT.CBR_DESCRIPTION) \cap (\exists has_Solution.CBR_Solution) \\ & \cap (\exists Has_Residence.string) \cap (\exists Has_Age.Age) \cap (\exists Has_BMI.BMI) \cap (\exists Has_Occupation.string) \\ & \cap (\exists Has_Disease.Disease) \cap (\exists Has_FemaleHistory.FemaleHistory) \\ & \cap (\exists Has_GlobalSymptom.GlobalSymptom) \cap (\exists Has_Hemato\log\icalProfile.Hemato\log\icalProfile) \\ CBRCase_i \equiv & \cap (\exists Has_KidneyFunctionTest.KidneyFunctionTest) \cap (\exists Has_LabTest.LabTest) \\ & \cap (\exists Has_LipidProfile.LipidProfile) \cap (\exists Has_LiverFunctionTest.LiverFunctionTest) \\ & \cap (\exists Has_Radio\log\icalExamination.Radio\log\icalExamination) \\ & \cap (\exists Has_UyinationSymptom.UyinationSymptom) \cap Has_Gender.\{"female", "male"\} \cap \dots \end{aligned}$$

Moreover, case diagnosis part is a nominal concept of the form DIAGNOSIS = {diabetic, preDiabetic, normal, diabeticGestational, prediabeticGestational}.

As a result, in the next module (case query parser module), the query case will be modeled in the same format as $CBRCase_i$, and in the case retrieval module (Section 5.6), a systematic comparison between the cases' predicates can calculate the similarity levels between cases.

5.5. Case query parser module

For a new patient diagnosis problem, the physician enters the new patient description in the query form; this forms the new case without a solution. We have asserted before that our cases have a homogeneous structure. Implementing heterogeneous case structure will be discussed in another study. As a result, all of the necessary patient features are known in advance, but the physician may not know all of the values of these features when describing the patient, and their entry may be time-consuming. Ontologies especially standard medical ontologies support the integration of a CBR system and EHR [18]. The query module can search the patient record for the necessary fields. Moreover, we can implement a rule base to link features and infer the missing ones. Next, the query is fuzzified and coded with the same methods used for the case-base ontology to facilitate similarity and mapping. The new problem structure is transformed into the fuzzy case-base

ontology vocabulary by some strategy; then, the semantic query is sent to the *Case Retrieval Engine* to compute the similarity between the query concepts and the concepts of the new semantic-query problem. The semantic query is a DL conjunctive query of the logic form $\hat{\wedge}_i(\emptyset_i) \bowtie \alpha$, where \emptyset_i is a conjunction of terms of the form $A(x), R(x,y)$, for atomic concept A , and atomic role R ; x, y are either individuals or variables names $\alpha \in \{0,1\}$, and $\bowtie \in \{>, \geq, \leq, <\}$.

To this end, let us take a semantic query example. After acquiring the query case Q from physician, it is represented as a vector $Q = \langle \text{attribute}_i = \text{value}_i \rangle$, for i is the number of features.

Our cases are represented with 70 features, so writing semantic queries using all of these features will create a long and complicated query. A very small fragment of these features is $Q = \langle \text{Age} = 38, \text{Residence} = \text{"Rural"}, \text{Fatigue} = \text{"++"}, \text{Gender} = \text{"Male"}, \text{disease} = \text{"Malignant tumor involving left ovary by direct extension from endometrium"} \dots \rangle$. This vector enters two main preparation steps: fuzzification of numerical data, and coding of unstructured data. After the fuzzification process, the vector is $Q = \langle \text{young} = 0.2, \text{middleAged} = 0.8, \text{old} = 0, \text{fuzzyLabel} = \text{middleAged}, \text{Age} = 38, \text{Residence} = \text{"Rural"}, \text{Fatigue} = \text{"++"}, \text{Gender} = \text{"Male"}, \text{disease} = \text{"Malignant tumor involving left ovary by direct extension from endometrium"} \dots \rangle$. After the encoding of the query by our SNOMED CT domain OWL2 ontology; this step encodes unstructured data into standard codes. The resulting vector is $Q = \langle \text{young} = 0.2, \text{middleAged} = 0.8, \text{old} = 0, \text{fuzzyLabel} = \text{middleAged}, \text{Age} = 38, \text{Residence} = \text{"Rural"}, \text{Fatigue} = \text{"++"}, \text{Gender} = \text{"Male"}, \text{disease} = \text{"369524001"} \dots \rangle$. The other ordinal and categorical features remain the same. The vector Q needs to be transformed into a semantic query. This query is a conjunction of a set of predicates as $P_1 \cap P_2 \cap \dots \cap P_n$ where P_i is a predicate of four forms: a (fuzzy) concept assertion $a:C_i \bowtie \alpha$, a (fuzzy) object property assertion $(a,b):R_i \bowtie \alpha$, a (fuzzy) data property assertion $(a,v):T_i \bowtie \alpha$, for a, b as abstract individuals and v as a literal value, or A (fuzzy) data property assertion $(a,v):T_i \bowtie \alpha$, for v as a fuzzy linguistic term defined using a fuzzy datatype.

According to the vocabulary of our fuzzy case-base ontology, the vector Q is transformed into a semantic query containing OWL individuals and property (i.e., data and object) instances of the form $\langle \text{concept instance}, \text{object property}, \text{concept instance}; \alpha \rangle$,

⁷ http://www.w3.org/2001/sw/rdb2rdf/wiki/Database-Instance-Only_and_Database-Instances-and-Schema-Mapping.

ClassAssertion (Patient_Case query1)	DataPropertyAssertion (has_Fuzzy_Label age2 "young")
ClassAssertion (q_desc CBR_DESCRIPTION)	DataPropertyAssertion (has_Crisp_Value age2 38)
ClassAssertion (Age age2)	DataPropertyAssertion (has_Gender q_desc "Male")
ClassAssertion (Disease ds2)	DataPropertyAssertion (has_Fatigue q_desc "+")
ObjectPropertyAssertion (has_Description query1 q_desc)	DataPropertyAssertion (has_disease q_desc ds2)
DataPropertyAssertion (has_Residence q_desc "Rural")	DataPropertyAssertion (has_disease_value ds2 "369524001")
ObjectPropertyAssertion (has_Age q_desc age2)	
DataPropertyAssertion (has_Young_Age age2 0.2)	
DataPropertyAssertion (has_middle_Aged age2 0.8)	
DataPropertyAssertion (has_Old_Age age2 0.0)	

(Continue for all object and datatype properties describing the new query-case 70 features)

Fig. 16. A part of the semantic representation of fuzzy query case.

```
PREFIX xmlns:owl=<"http://www.w3.org/2002/07/owl#">
PREFIX xsd=<"http://www.w3.org/2001/XMLSchema#">
PREFIX swrl=<"http://www.w3.org/2003/11/swrl#">
PREFIX swrlb=<"http://www.w3.org/2003/11/swrlb#">
PREFIX rdfs=<"http://www.w3.org/2000/01/rdf-schema#">
PREFIX case=<http://www.semanticweb.org/math/s/ontologies/casebase#>

SELECT DISTINCT ?sol
FROM <http://www.semanticweb.org/math/s/ontologies/casebase.owl>
WHERE {Class(?Ci), Class(?desc), Class(?sol), Class(?age), Class(?dis), Type(?C, case:Patient_Case), Type(?desc,CBR_DESCRIPTION), Type(?sol,CBR_SOLUTION), Type(?age,case:Age), Type(?dis,case:Disease), ObjectProperty(has_Description), PropertyValue(?desc,has_Disease,?dis), ObjectProperty(has_Disease), ObjectProperty(has_Solution), ObjectProperty(has_Age), DataProperty(?has_Fuzzy_Label), ObjectProperty(has_Part), ObjectProperty(has_Diagnosis), DataProperty(has_Crisp_Value), PropertyValue(?desc,has_Disease_Value,"369524001"), DataProperty(has_Disease_Value), DataProperty(has_Gender), DataProperty(has_Residence), DataProperty(has_Fatigue), PropertyValue(?Ci,has_Description,?desc), PropertyValue(?Ci,has_Solution,?sol), PropertyValue(?desc,has_Residence,"Rural"), PropertyValue(?desc,has_Age,?age), PropertyValue(?age,has_Young_Age,0.2), PropertyValue(?age,has_middle_Aged,0.8), PropertyValue(?age,has_Old_Age,0.0), PropertyValue(?age,has_Crisp_Value,38), PropertyValue(?age,has_Fuzzy_Label,"middleAged"), PropertyValue(?Ci,has_Gender,"Male"), PropertyValue(?Ci,has_Fatigue,"++"), ...has_Solution(?Ci,?sol)}{}
```

Fig. 17. A part of a semantic query over the case-base.

<concept instance, data property, fuzzy value; α >, <concept instance, data property, literal; α >, as shown in Fig. 15c. Fig. 16 shows an example using OWL2 Functional-Style Syntax⁸ for the previous query vector Q ; we concentrate on the fuzzy values representation and assume that $\alpha = 1$.

Up to this point, we have two options. We firstly search for an exact match between this query case and a case in the fuzzy ontology. In this case, an SPARQL-DL query can be used to query the case-base ontology and retrieve the diagnosis of the matched case, as shown in Fig. 17.

The other general option is the existence of partial similarity between the query case and all cases in the case-base. In this case, we use a set of APIs including JCOLIBRI, OntBridge, Pellet, and fuzzyDL APIs to implement a java project to implement the retrieval algorithm. The algorithm uses the proposed similarity function in the next section to retrieve the most suitable k cases in the fuzzy case-base ontology. The algorithm calculates the clinical similarity between the query case and all cases in the case-base according to the inference capabilities of the Pellet and fuzzyDL reasoners. The solutions of the most similar k cases are selected and retrieved to the physician to guide his decision process.

5.6. Case retrieval engine module

We can state that a case is equivalent to another case if both cases have exactly the same structure and attribute values. In crisp ontology-based CBR, the retrieval of cases involves the exploitation of the structure and the content of the ontology for computing the semanticsimilarity between the attribute values

and consequently, for the cases. There is some ontology-specific similarity functions that utilize ontological knowledge in a different manner [72]. None of these measures utilizes imprecise knowledge in any way. Case retrieval can be implemented with a neural network (NN), rule-based (RB), case indexing (CI), and a decision tree (DT). However, it is hard to determine the corresponding structure and parameters of NN and DT, in addition, extraction and the choice of rules and indexes are largely dependent on the experience of the knowledge engineers as well [75]. In this paper, we propose a case retrieval algorithm that involves combining the reasoning capabilities of classical ontologies (i.e., semantic similarity) with fuzzy similarity for numerical features in order to create a powerful hybrid reasoning mechanism. We assume that all case classes have a unique structure (i.e., the same set of attributes).

The performance of similarity measure totally depends on the type and the importance of features. We have used a set of machine learning algorithms to calculate feature weights, as in another study [35]. First, we calculated the local similarity of each feature according to its type [12]; next, we used a global similarity function based on a distance function as Euclidian or Minkowski. According to feature types, our proposed similarity algorithm had two stages of similarity. The first stage depends on syntactic features only to retrieve a set of potentially similar cases, and the second depends on the remaining semantic features to select the most similar case.

5.6.1. Similarity calculation's first stage

Consider a query case C_q , stored cases C_i for $i = 1, \dots, n$ and n is the number of cases in the case-base, and feature weights w_i . All instance features have weight $w_i = 0$. The first layer calculates $SIM_{\text{syntactic}}(C_q, C_i)$. This global similarity function $SIM_{\text{syntactic}}$ returns the most similar cases according to the similarity between C_q and C_i using syntactic similarity of syntactic features (fuzzy and

⁸ <http://www.w3.org/TR/owl2-syntax/>.

not fuzzy), see the following equation:

$$\begin{aligned} \text{SIM}_{\text{syntactic}}(C_q, C_i) &= \frac{\sum_{j=1}^n w_i \times \text{sim}(f_{qj}, f_{ij})}{\sum_{j=1}^n w_i} = \frac{\sum_{j=1}^m w_i \times \text{sim}_L(A_j, Z_j)}{\sum_{j=1}^n w_i} \\ &+ \frac{\sum_{j=m+1}^k w_i \times \text{sim}_O(B_j, Z_j)}{\sum_{j=1}^n w_i} + \frac{\sum_{j=k+1}^r w_i \times \text{sim}_F(C_j, Z_j)}{\sum_{j=1}^n w_i} \\ &+ \frac{\sum_{j=r+1}^n w_i \times \text{sim}_N(D_j, Z_j)}{\sum_{j=1}^n w_i} \end{aligned} \quad (4)$$

The A, B, C , and D are the sets of nominal, ordinal, fuzzy, and numerical features of the query case, and Z contains the corresponding features. Moreover, we can add importance for each type of features by using weights w_1, w_2, w_3, w_4 as shown in Eq. (5), where w_1, w_2, w_3 , and $w_4 \in (0, 1]$ and $w_1 + w_2 + w_3 + w_4 = 1$

$$\begin{aligned} \text{SIM}_{\text{syntactic}} &= w_1 \times \frac{\sum_{j=1}^m w_i \times \text{sim}_L(A_j, Z_j)}{\sum_{j=1}^n w_i} + w_2 \\ &\times \frac{\sum_{j=m+1}^k w_i \times \text{sim}_O(B_j, Z_j)}{\sum_{j=1}^n w_i} + w_3 \\ &\times \frac{\sum_{j=k+1}^r w_i \times \text{sim}_F(C_j, Z_j)}{\sum_{j=1}^n w_i} + w_4 \\ &\times \frac{\sum_{j=r+1}^n w_i \times \text{sim}_N(D_j, Z_j)}{\sum_{j=1}^n w_i} \end{aligned} \quad (5)$$

Depending on the type of a feature, the local similarity sim is selected as follows:

If the feature is nominal, the exact match is used as in the following equation:

$$\text{sim}_L(A_j, Z_j) = \begin{cases} 1 & \text{if } A_j = Z_j \\ 0 & \text{if } A_j \neq Z_j \end{cases} \quad (6)$$

If the feature is ordinal, our domain experts proposed a similarity matrix for each ordinal feature, and the similarity $\text{sim}_O(B_j, Z_j)$ is calculated based on this matrix. Due to space restrictions, we do not show matrices.

If the feature is fuzzy, we have two options:

(1) The feature value is numerical. Our proposed fuzzy similarity measure utilizes all of the fuzzy sets of compared features in calculating similarity. As the case-base fuzzy ontology store case with fuzzified features, the input query numerical features is fuzzified using the same fuzzy sets, and a comparison is conducted between stored and query fuzzy values. The normalized Euclidean distances between fuzzy sets of a feature are used to calculate similarity as in the following equation:

$$\text{Dist}_F(C_j, Z_j) = \frac{\sqrt{\sum_{k=1}^n (\mu_{cjk} - \mu_{zjk})^2}}{\sqrt{n}} \quad (7)$$

where C_j = crisp value of a feature in query, Z_j = crisp value of a feature in a case, n = number of fuzzy sets for feature f , μ_{cjk} and μ_{zjk} are k 's fuzzy values for query and stored cases' feature, respectively. The similarity is calculated using the following equation:

$$\text{sim}_F(C_j, Z_j) = 1 - \text{Dist}(C_j, Z_j) \quad (8)$$

After testing this function, we found it insensitive for extreme values because the membership functions are equal to zero except

one function, for example, for ages 60 and 70, $\text{Dist}_F(60, 74) = 0$ and $\text{sim}_F(60, 74) = 0$. To solve this problem, to calculate fuzzy similarity, we take the average of crisp similarity (Eq. (9)) and fuzzy one.

(2) The feature value is a vague term. A patient can be described using vague terms for numerical features (e.g., Age = young, BMI = obese, FPG = low). Our case-base fuzzy ontology supports all types of similarities. As shown in Fig. 15, the *has-Fuzzy-Age* data type property stores the linguistic term *young* for the numerical *age* = 36. When a patient is described by a linguistic term, proposed similarity matrices of our domain expert are used (see Table 3).

In addition, fuzzy hedges such as "very", "quite", "somewhat", "not", or "extremely" are possible in query case description. As shown in Section 4.3.1, it is possible to define hedges in the case base ontology. The stored and entered hedges can be compared using similarity matrices as proposed by our domain expert.

If the feature is simple numerical, then the similarity is calculated using the following equation:

$$\text{sim}_N(C_j, Z_j) = 1 - \frac{|D_j - Z_j|}{\text{Max} - \text{Min}} \quad (9)$$

5.6.2. Similarity calculation's second stage

Medical concepts similarity can be conducted non-semantically or lexically, or it can be done semantically using standard ontologies as SCT [16,64]. We selected the second choice to measure the similarity in meaning between concepts. The retrieved m cases from the first layer (i.e., by $\text{SIM}_{\text{syntactic}}(C_q, C_i)$) enter another evaluation based on the semantic similarity between the instance features. Lexical or exact similarity cannot be used to compare ontology concepts. All syntactic features have $w_i = 0$. The $\text{SIM}_{\text{syntactic}}(C_q, C_i)$ utilizes our proposed SCT domain ontology to calculate the semantic similarity between compared SCT concepts [18]. Instance features have the data type = I in Table 1. Not like relatedness, semantic similarity measures how similar the meaning of concepts are based on the IS-A relationship only [55]. These measures include edge-based, node-based (i.e., information content and features-based), and hybrid measures. Garla and Brandt [27] have provided a recent survey of all existing measures. Most these measures are suitable for WordNet nouns only. We do not utilize Information Content (IC), neither corpus nor intrinsic, because none of its calculation methods is applicable to SCT; its calculation is time consuming [55]; it is inaccurate due to shallow annotations [27]. The most popular methods for intrinsic IC are Seco et al. [76] using Eq. (10) and Sánchez et al. [77] using Eq. (11).

$$IC_{\text{Seco}}(u) = 1 - \frac{\log(D(u))}{\log |C|} \quad (10)$$

$$IC_{\text{Sanchez}} = -\log \left(\frac{\text{Leaves}(u)/|A(u)|^{+1}}{\text{Max_leaves} + 1} \right) \quad (11)$$

with $D(u) = \{v | v \sqsubseteq u\}$, C the set of all concepts in the ontology, $\text{leaves}(u)$ as the number of leaves subsumed by the concept u , and Max_Leaves as the number of terminal concepts of the ontology.

We propose a new hybrid measure based on path length and concept features. First, for path length, our similarity is based on

Table 3

The fuzzy sets similarity matrix for age feature.

Stored case	Query		
	Young	Middle-aged	Old
Young	1	0.5	0.1
Middle-aged	0.5	1	0.6
Old	0.1	0.6	1

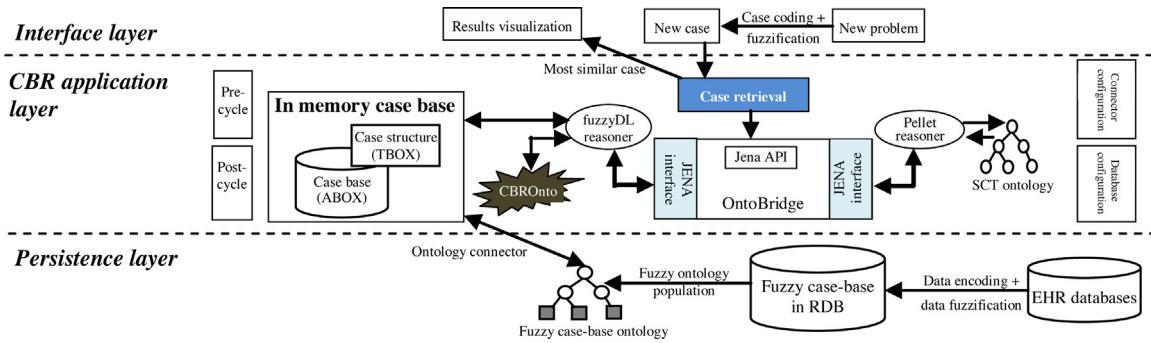


Fig. 18. The structure of the system implementation.

the *depth* of the Least Common Ancestor (LCA) of the two concepts and the *closeness level* of concepts to their LCA in our SCT sub-ontology is based on *IS-A* relationship only. In other words, (1) the deeper the LCA, the more specific it is considered and, thus, the more similar the compared concepts are assumed; (2) the closer the two concepts are to their LCA, the more similar they are. Second, to quantify similarity for concept features, the commonalities and differences between concepts must be considered [55]. JCOLIBRI API⁹ uses four semantic similarity measures: path-based such as *fdeep_basic* and *fdeep*, and feature-based such as *cosine*, and *detail* [52]. These measures have been tested in the API's tutorial; however, path-based measures do not take into account the depth of concepts from their LCA, and feature-based measures depend only on the commonalities between compared concepts. Our proposed measure overcomes these limitations and integrates path based and feature based approaches. The proposed composite similarity measure uses the equation the following equation:

$$SIM_{\text{Semantic}}(u, v) = w_1 sim_{\text{path}}(u, v) + w_2 sim_{\text{feature}}(u, v) \quad (12)$$

where $w_1, w_2 \in (0, 1]$ are weights for $w_1 + w_2 = 1$, and $sim_{\text{path}}(u, v)$ (Eq. (13)) is an adapted version of Wu and Palmer [69] (Eq. (14)) because $sim_{\text{wu and palmer}}(u, v) < 1$ which violates the Identity Of the Indiscernibles property (IOI) [55].

$$sim_{\text{path}}(u, v) = \begin{cases} 1 & \text{if } u = v \\ sim_{\text{wu and palmer}} & \text{otherwise} \end{cases} \quad (13)$$

$$sim_{\text{wu and palmer}}(u, v) = \frac{2 \times \text{depth}(lca(u, v))}{\text{shortest_path}(u, lca(u, v)) + \text{shortest_path}(v, lca(u, v)) + 2 \times \text{depth}(lca(u, v))} \quad (14)$$

In addition, $sim_{\text{Feature}}(u, v)$ is based on Batet et al. [26], Eqs. (15) and (16):

$$sim_{\text{Feature}}(u, v) = 1 - Dist_{\text{Batet}}(u, v) \quad (15)$$

$$Dist_{\text{Batet}}(u, v) = \log_2 \left(1 + \frac{|A(u) \setminus A(v)| + |A(v) \setminus A(u)|}{|A(u) \setminus A(v)| + |A(v) \setminus A(u)| + |A(u) \cap A(v)|} \right) \quad (16)$$

where $A(u)$ is the set of ancestors of u , i.e., $A(u) = \{v|uv\}$, $A(u)/A(v)$ is specificity of u , and $A(u) \cap A(v)$ is the commonality between u and v . We tried to calculate the clinical similarity between two concepts rather than the semantic distance. Clinical similarity is

influenced by the clinical granularity of concepts. For example, if we consider the hierarchy “*megacalycosis is-a caliectasis is-a kidney-disease*” from SCT, semantic similarity (*kidney disease, kidney disease*) = 1, but clinical similarity (*kidney disease, kidney disease*) < 1 because *kidney disease* is more general and abstract concept, which means other diseases as well. Moreover, clinical similarity (*caliectasis, caliectasis*) > clinical similarity (*kidney disease, kidney disease*) is decided. The main rule is that the deeper the concept, the more specific it is. Finally, the solution for the most similar case is suggested for the new problem.

6. Implementation and evaluation

6.1. System implementation

A CBR system was developed in Java by extending the APIs of the JCOLIBRI2 CBR framework [52]. As shown in Fig. 18, the proposed customization has three layers, and each layer has specific tasks. Due to space restrictions, we do not discuss this framework in detail. The persistence layer prepares the fuzzy case-base ontology. The CBR application layer is the core of the framework as it contains the whole CBR cycle. The interface layer accepts a query from the physician and returns the most similar case. We have implemented the case representation and retrieval steps only; case adaptation and retention are out of scope.

Due to space restrictions, we select only seven from our ~70 features to implement our system. These features are representative of the dataset because it includes fuzzy features as Age, HbA1c, and BMI; instance features as lipid disease, liver disease, and nephropathy; and nominal features as gender. The fuzzification of numerical features has been done in Matlab, and for space limitation, we will not discuss this process. Moreover, instance features have been encoded using our SCT reference set. Fig. 19 shows the query screen used to collect patient attributes. For instance features, the user selects an instance from shown ontology. Fig. 20 shows the similarity configuration window; It allows the dynamic selection of similarity functions and weights for each feature; the selection of the number of cases to retrieve (i.e., k). We have implemented all of the proposed similarity function in Section 5.5 including fuzzy and semantic. Spinner is used to let the user choose from a range of values to control the number of retrieved cases. The slider is used to set the weight. Fig. 21 shows the retrieved cases with their level of similarity.

6.2. Evaluation of the proposed CBR system

Each component of the proposed system is evaluated upon completion. These evaluations have provided proof of concept, illuminated system strengths, and weaknesses and guided system

⁹ <http://gaia.fdi.ucm.es/research/colibri/jcolibri>.

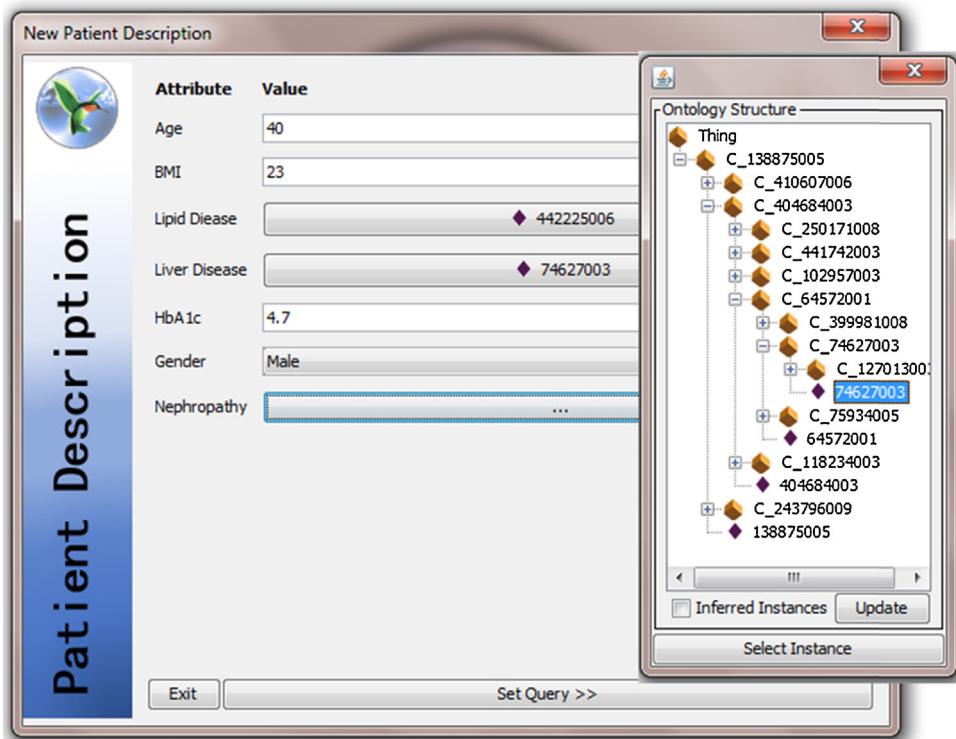


Fig. 19. A new case description.

development. The proposed framework is the first to integrate the capabilities of standard medical ontologies (i.e., SCT), fuzzy logic, ontology, and CBR in on the hybrid system. This combination has powerful benefits to CBR functionality. In the next sub-sections, we evaluate the proposed fuzzy case-base ontology (Section 6.2.1). Moreover, we evaluate the proposed semantic retrieval functions on a small fragment of the SCT medical ontology (Section 6.2.2).

In addition, the overall performance of the system is evaluated using the case-base ontology, the overall retrieval algorithm, and the domain standard ontology (Section 6.2.3).

6.2.1. Case-base fuzzy ontology evaluation

6.2.1.1. Our fuzzy ontology evaluation includes three dimensions. First, the ontology consistency has been checked using a set of

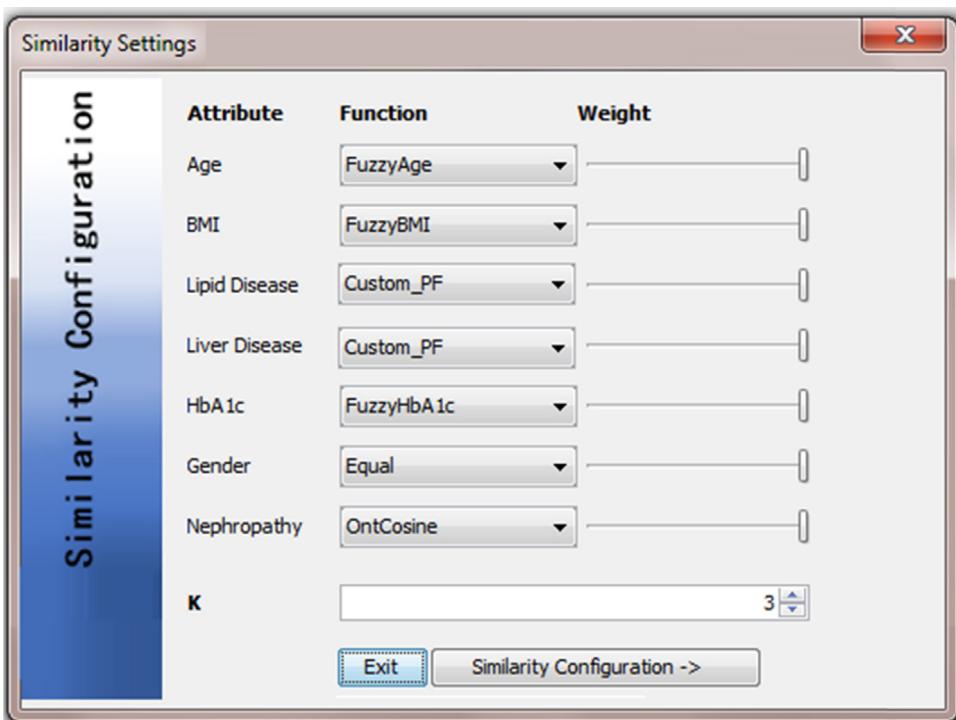


Fig. 20. Similarity measures setting.

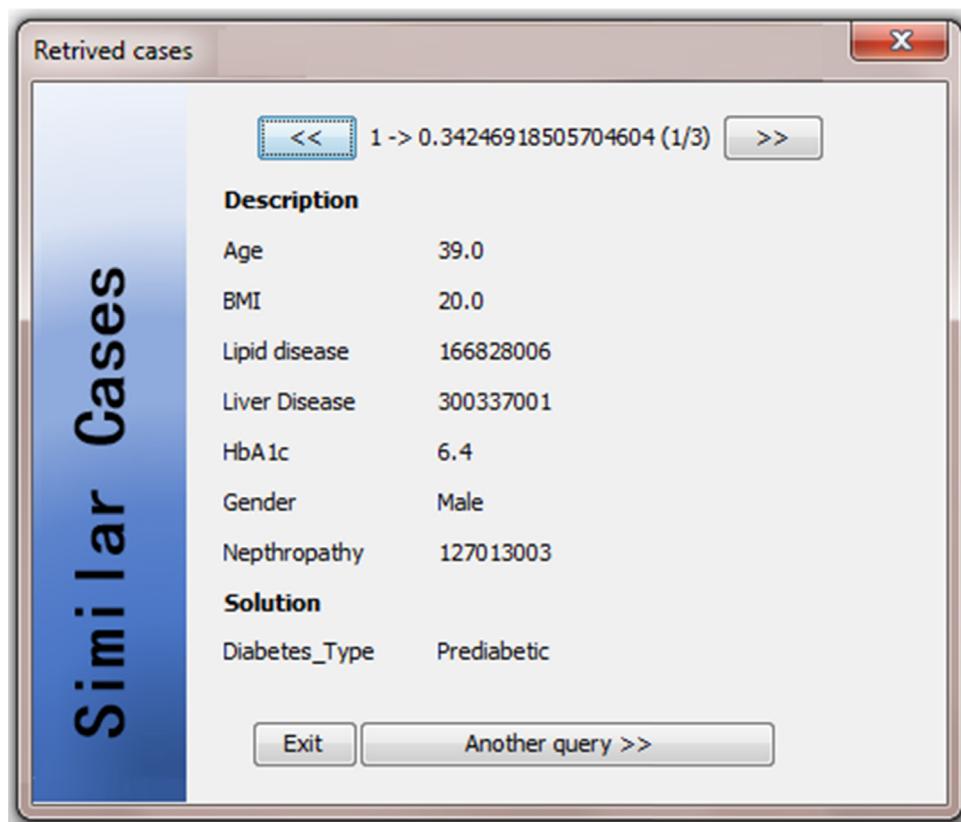


Fig. 21. Retrieved similar cases.

reasoners including HermiT 1.3.8, Fact++, and Pellet 2.3.0. Moreover, its fuzziness consistency has been checked by fuzzyDL 1.1. Pitfalls found in the ontology modeling process were detected using the OOPS! Pitfall Scanner [78] and carefully corrected.

Second, checking correctness, accuracy, and completeness is typically manual [48]. Our Domain experts have validated the correctness, accuracy, and completeness of the built case-base fuzzy ontology. Regarding the correctness, our two domain experts reviewed each fuzzy element (i.e. fuzzy datatype, fuzzy object property, and fuzzy data property) and asserted that these elements convey a meaning, which is indeed vague in diabetes diagnosis domain. Regarding the accuracy, our domain experts have unified the fuzzification process. They asserted that for each fuzzy variable such as HbA1c lab test its normal range is modeled using a triangular fuzzy set and the other fuzzy values such as low and high will be modeled using left and right shoulder functions; these shoulder functions are overlapped with the normal range by 50%. All these aspects have been reviewed by the domain experts, and they asserted that the vagueness has been done in an intuitively accurate way. Regarding the completeness, this fuzzy ontology is an extension of our crisp ontology [31]. First, the crisp ontology is complete because it has been tested using a set of competency questions and using all medical concepts of diabetes diagnosis domain. In other words, our domain experts have collected 328 medical terms from some diabetes diagnosis CPGs such as Canadian Diabetes Guideline, and they have tested the coverage of the ontology for all of these terms. The ontology has 100% concept coverage for all medical concepts required to describe diabetic patient cases. Second, domain experts have checked the completeness of vagueness. We used a set of SPARQL and protégé DL queries to verify the ability of the ontology to answer any fuzzy queries defined by domain experts; experts have verified that all the vagueness needed for diabetes domain has been represented in the ontology.

Third, we have evaluated our ontology using criteria-based and data-driven approaches. Brewster et al. [79] argued that precision and recall are not appropriate for ontology evaluation because they depend on a comparison between concepts of evaluated ontology and a standard one. There are no standard ontology evaluation mechanisms [80]. To measure the quality of our ontology, we can use criteria-based or data-driven evaluation mechanisms. *Regarding criteria-based evaluation mechanisms*, we need to compare it with other ontologies in the same domain. There are no other (fuzzy) case-base ontologies in the medical domain to compare our ontology with it. Alexopoulos et al. [21] have proposed a fuzzy case-base ontology for electricity market CBR system. On the other hand, there are some crisp case-base ontologies such as ArgCBROnto [81] for argumentation, [82] for mould design, and [83] for resource management. There are many proposed criteria to quantify the quality of ontologies [84]. Some of these criteria such as consistency can be successfully determined using semantic reasoners. Some criteria, such as clarity is difficult to evaluate as there are no means in place to determine them. Most of the proposed criteria are overlapped. From the set of criteria proposed in the literature, we depend on criteria proposed by Djedidi et al. [84], where each criterion can be measured by metrics. These criteria include:

1. *Complexity criterion*, which assesses structural and semantic links between ontology entities and the navigability in ontology structure,
2. *Cohesion criterion*, which takes into account the connected ontology components (i.e. classes),
3. *Conceptualization criterion*, which corresponds to design richness of the ontology content,
4. *Abstraction criterion*, which indicates class abstraction level (generalization/specialization) by measuring the depth of subsumption hierarchies,

Table 4
Ontology evaluation quality metrics.

Measure	Ontology	The proposed ontology	ArgCBROnto
Criteria	Metrics		
Complexity	The average number of paths to reach a class from the root	5	3
	Average number of semantic relations (object properties) per class	1.4	1.3
Abstraction	Average depth of the ontology	2	2
Cohesion	Average number of connected components (classes)	63	27
Conceptualization	<i>Semantic richness</i> : Ratio of the total number of semantic relations assigned to classes, divided by the total number of ontology relations (object properties and subsumption relations)	58/(58 + 59) = 0.495	38/(38 + 23) = 0.62
	<i>Attribute richness</i> : Ratio of the ontology's total number of attributes (i.e., the data properties), divided by the total number of classes	138/62 = 2.26	24/26 = 0.92
	<i>Inheritance richness</i> : Average number of subclasses per class	5.0	2.875
Comprehension	<i>Documentation of the properties</i>	2.04%	0.0%
	<i>Documentation of the classes</i>	88.71%	0.0%

Table 5
Ontology evaluation measures for three ontologies.

The measure	Knowledge coverage measures									
	Number of classes	Number of properties	Number of axioms	Maximum number of Parents	Documentation of the classes	Documentation of the properties	Properties with domain	Properties with range	Number of individuals	
The proposed ontology	62	196	1316	31	88.71%	2.04%	98.47%	98.98%	2640	
ArgCBROnto [81]	26	62	446	13	0%	0%	85.48%	77.41%	0	
Alexopoulos et al. [21]	10	18	N/A	N/A	N/A	N/A	N/A	N/A	N/A	

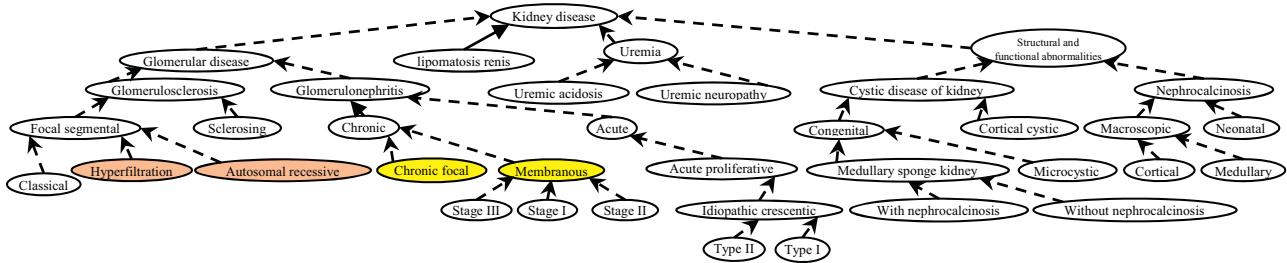


Fig. 22. A sub-graph of SCT for kidney disease.

5. *Completeness criterion*, which evaluates if the ontology covers domain relevant properties; This criterion has been evaluated previously by the ontology concept coverage,
6. *Comprehension criterion*, which assesses the facility of understanding ontology.

ArgCBROnto is the most complete ontology, and the other studies have no OWL2 ontologies. Table 4 represents a comparison between ArgCBROnto and our ontology regarding these metrics. For calculating metrics, we have used the equations proposed by Zhang et al. [85]. The ontology parameters used for metrics are calculated using a protégé 4.3 ontology editor's evaluation plugin (i.e., Ontology Evaluation¹⁰). Protégé has other automatic evaluation plugins such as OntoClean and AEON (Automatic Evaluation of ONtologies)¹¹.

As shown in the table, regarding *data-driven evaluation mechanisms*, Fernández et al. [86] have proposed another measure for data-driven evaluation. They measure the structure of the ontology including number of classes, properties, axioms, and individuals. Protégé Ontology Evaluation plugin calculates these metrics and others including naming conventions, class hierarchy, object properties hierarchy, data type properties hierarchy, documentation,

properties domain and range, disjointness restrictions, and lexically similar concepts and properties. The application of these measures is summarized in Table 5, where our ontology does overweight the compared ontologies.

6.2.2. Evaluation of the proposed semantic retrieval algorithm

The proposed retrieval algorithm supports five types of features including numerical, nominal, fuzzy, ordinal, and semantic. The last type (i.e. semantic type) measures the clinical distance between the compared SCT standard medical concepts. In this section, the proposed semantic similarity algorithm is evaluated by comparing it with the most popular semantic similarity algorithms in CBR (i.e. with JCQLIBRI2 [52]). As shown in Fig. 22, this is done by doing experiments using a sub-ontology from our SCT ontology for kidney diseases, assuming that w_1 and w_2 are 0.5 in Eq. (9).

We argue that there is a difference between the lexical, semantic, and clinical similarity. Lexical similarity depends on the level of textual similarity between the two concepts. Therefore, the lexical similarity Sim_{lexical} (Chronic focal, Membranous) is equal 0, and this is not accurate because both 197618004|chronic focal glomerulonephritis and 77182004|membranous glomerulonephritis are both 20917003|chronic glomerulonephritis. The semantic similarity adds some intelligence to this process. If we compare two patients P1 and P2 with diseases D1 = "kidney disease" and D2 = "renal disorder", then the semantic distance Sim_{Semantic} (D1, D2) = 1. Another example, if D1 = "autosomal dominant focal segmental

¹⁰ http://protegewiki.stanford.edu/wiki/Ontology_Evaluation.

¹¹ <http://code.google.com/p/aeon-project/>.

Table 6

The comparison between JCOLIBRI semantic similarity methods and our proposed one.

Similarity	Method					
	Fdeep_basic	Fdeep	Cosine	Detail	Proposed method	Case no.
Sim(Type I, Type I)	7/7	7/7	7/7	13/14	1	1
Sim("lipomatosis renis", uremia)	0	0	1	1/2	0.11	2
Sim(Cortical, Classical)	0	0	1/5	1/2	0.04	3
Sim(Chronic, Stage II)	4/7	4/6	2/√6	8	0.66	4
Sim(Glomerulosclerosis, Type I)	2/7	2/7	2/√21	4	0.47	5
Sim(Glomerulosclerosis, acute)	2/7	1/2	2/√12	4	0.383	6
Sim(Acute, chronic)	3/7	3/4	3/4	5/6	0.889	7
Sim(Type I, chronic)	3/7	3/7	3/√21	6	0.423	8

glomerulosclerosis" and D2="hyperfiltration focal segmental glomerulosclerosis" then $\text{Sim}_{\text{Semantic}}(\text{D1}, \text{D2})=1$. Semantic similarity depends on the ontology structure to infer the level of similarity between two concepts. However, the two patients in the second example are more similar than the first example. This is because while both patients with "kidney disease" from a semantic perspective have the same concept and therefore a semantic distance is zero, when applying these concepts to the patient case, "kidney disease" could mean many other disease entities including "Medullary sponge kidney", "medullary cystic disease OS", "caliectasis", "amyloid nephropathy", "hypertensive renal disease", and other. On the other hand, in the second example, 444977005|autosomal dominant focal segmental glomerulosclerosis and 236405006|hyperfiltration focal segmental glomerulosclerosis refer to a specific disease entity.

We propose to handle this issue by using the clinical similarity measurement. In clinical similarity, the $\text{Sim}_{\text{Clinical}}$ ("kidney disease", "kidney disease") < $\text{Sim}_{\text{Clinical}}$ ("autosomal dominant focal segmental glomerulosclerosis", "hyperfiltration focal segmental glomerulosclerosis"). As a result, the three similarities are not equal regarding accuracy, i.e. $\text{Sim}_{\text{Lexical}} \neq \text{Sim}_{\text{Clinical}} \neq \text{Sim}_{\text{Semantic}}$. Our proposed similarity measure takes into account the level of specificity of a concept that subsumes the two compared concepts and the level of commonality between the compared concepts. As a result, as shown in Table 6, the similarity $\text{Sim}(\text{Type I, Type I})=1$ because Type I and Type I are very specific in the ontology. The similarity $\text{Sim}(\text{Acute, Chronic})=0.889$ because these concepts are not specific; they contain many sub-concepts. Our algorithm is very sensitive to the level of similarity between the compared concepts.

As we can see in Table 6, Fdeep_basic and Fdeep do not take in to account the depth of concepts from their LCA (i.e., the closeness between concepts) as in cases 7, 8. Moreover, Cosine and Detail do not account for the differences between concepts such as cases 5, 6. What is more, there are distributed inefficiencies as Detail (Type I, Type I) ≠ 1, Cosine ("lipomatosis renis", Uremia)=1, etc. On the other hand, the proposed similarity measure provides logically consistent results for all types of problems because it accounts for into account the depth of the compared concepts from their LCA, and it takes the differences between compared concepts as well as commonalities. Eqs. (17)–(19) are the implementation of the semantic equations in the JCOLIBRI OntoBridge API environment, where (–) is the difference.

$$\text{Sim}_{\text{WU+Palmer}}(u, v) = \frac{2 \times \max \text{ProfLCS}(u, v)}{(\text{profConcept}(u) - \max \text{ProfLCS}(u, v)) + (\text{profConcept}(v) - \max \text{ProfLCS}(u, v)) + (2 \times \max \text{ProfLCS}(u, v))} \quad (17)$$

$$\text{Sim}_{\text{feature}}(u, v) = 1 - \frac{\text{Math.log}(1+x)}{\text{Math.log}(2)} \quad (18)$$

$$x = \left(\frac{|(\text{super}(u, \text{CN})) - (\text{super}(v, \text{CN}))| + |(\text{super}(v, \text{CN})) - (\text{super}(u, \text{CN}))|}{|(\text{super}(u, \text{CN})) - (\text{super}(v, \text{CN}))| + |(\text{super}(v, \text{CN})) - (\text{super}(u, \text{CN}))| + |(\text{super}(u, \text{CN})) - (\text{super}(v, \text{CN}))| + |(\text{super}(v, \text{CN})) - (\text{super}(u, \text{CN}))|} \right) \quad (19)$$

6.2.3. Performance of the proposed system

As domain experts knowledge are known to be the most relevant for evaluating the CDSS performance, one measure of the performance of our system is the extent to which the proposed

system decisions are matched with domain experts decisions [87]. After system development, our domain experts have conducted realistic experiments to test the accuracy, correctness, flexibility, applicability, and ease of use of the proposed diabetes diagnosis CDSS framework. The testing environment is the Mansura University Hospitals, and they have reported the results. The results show that our implemented CDSS is a realistic model of the real world of diabetes diagnosis.

Patient symptoms and tests are collected in real-time using crisp, fuzzy, text, and semantic values. If the crisp value of an attribute is not available, the domain expert can set a descriptive vague value according to the patient's description of these conditions. The determination of patient's current diseases (e.g., kidney, liver, cancer, etc.) is selected from the SCT ontology form. SCT provides the most comprehensive and standard interface for selecting concepts that describe patient diseases. These values form a new query case, and this case is formatted in the form of semantic query according to the case base fuzzy ontology. The CBR retrieval engine retrieves the most similar k cases. The value of k is selected by the domain expert. The system's produced decisions are compared with the experts' diagnoses of the case. After execution of the system for many times, domain experts have evaluated our system as 100% regarding flexibility, adequacy, and ease of use. Figs. 14 and 15 illustrate the screen shots of our prototype application in a testing scenario. We have applied this study on a case-base containing 60 cases from Mansura University Hospitals. Our method shows promising results. These results can be considered as a first step for real world testing of our proposed system. We did the evaluation of our system using a set of measures.

First, we used the leave-one-in evaluation technique to check the accuracy of our system to retrieve existing cases. Our system was 100% accurate when retrieving existing cases.

Second, we used the leave-one-out technique to measure the performance for non-existing cases. Namely, cases are taken out from the case-base one by one, and we have computed the similarity of this case with all the remaining cases in the case-base. It is a particular case of cross-validation. It has been used to evaluate many CBR systems including radiotherapy planning system [88], diabetes management system [89], and Fuzzy CBR systems [90]. The domain experts evaluate the performance of the implemented

framework by organizing a set of 43 experiments. The test cases are selected in a manner that allowed them to span the majority of topics and content represented in the case base. Each test query is fed into the system, and the corresponding response was recorded.

Table 7
The system performance evaluation.

K=3		
Test case	FKI-CBR decision	Expert decision
	Retrieved cases decision	Confidence (%)
Case 1	Diabetic	90.1
	Diabetic	88.2
	Diabetic	88
Case 2	Pre-diabetic	92
	Diabetic	91
	Pre-diabetic gestational	90
Case 3	Normal	98.1
	Normal	97.7
	Normal	95.7
Case 4	Normal	98.2
	Normal	97
	Normal	94.1
Case 5	Pre-diabetic	99.4
	Diabetic gestational	93
	Diabetic	91
Case 6	Pre-diabetic	99
	Diabetic gestational	92
	Diabetic	91
Case 7	Diabetic gestational	100
	Diabetic	92
	Pre-diabetic	91
Case 8	Diabetic	97
	Diabetic	96
	Diabetic	93
Case 9	Diabetic	95
	Diabetic	94
	Diabetic	92
Case 10	Diabetic	94
	Diabetic gestational	90
	Pre-diabetic	89
Case 11	Diabetic	97
	Diabetic	95
	Diabetic	94
Case 12	Diabetic	98
	Diabetic	92
	Diabetic	91
Case 13	Diabetic	98
	Pre-diabetic	95
	Diabetic	91.6
Case 14	Diabetic	94
	Diabetic	93
	Diabetic	93
Case 15	Diabetic	87
	Diabetic	85
	Diabetic	84
Case 16	Diabetic	86
	Diabetic	84
	Diabetic	83
Case 17	Diabetic	93
	Diabetic	91
	Normal	87
Case 18	Diabetic	93
	Diabetic	92.5
	Diabetic	91
Case 19	Pre-diabetic	84
	Diabetic	82
	Pre-diabetic	81
Case 20	Pre-diabetic	92
	Diabetic	91
	Diabetic	90
Case 21	Normal	91.5
	Normal	91
	Pre-diabetic	90
Case 22	Diabetic	94
	Diabetic gestational	92
	Pre-diabetic	91
Case 23	Diabetic	98.2
	Diabetic	93.5
	Diabetic	92.6
Case 24	Diabetic	96.5
	Pre-diabetic	94
	Diabetic gestational	94

Table 7 (Continued)

K=3			
Test case	FKI-CBR decision	Expert decision	
	Retrieved cases decision	Confidence (%)	
Case 25	Diabetic	92	Diabetic
Case 26	Diabetic	91	
	Diabetic	87	
Case 27	Diabetic	98	Diabetic
	Pre-diabetic	97	
Case 28	Diabetic	93	Diabetic
	Diabetic	92	
Case 29	Diabetic	91.9	
	Diabetic	95.43	Diabetic
Case 30	Diabetic	95.2	
	Diabetic	95	
Case 31	Diabetic	95	Diabetic
	Pre-diabetic	93.7	
Case 32	Diabetic	92.6	
	Normal	97.74	Normal
Case 33	Normal	97.6	
	Normal	94.6	
Case 34	Pre-diabetic	91.97	Pre-diabetic
	Diabetic	90.8	
Case 35	Diabetic	89.01	
	Normal	93.05	Normal
Case 36	Pre-diabetic	92.2	
	Pre-diabetic	91	
Case 37	Diabetic	87.7	Diabetic
	Diabetic	87.2	
Case 38	Normal	91.5	
	Normal	90.3	
Case 39	Diabetic	95.5	
	Diabetic	87.5	
Case 40	Diabetic	87.2	
	Normal	95.79	Normal
Case 41	Normal	94.63	
	Diabetic	94.1	
Case 42	Diabetic	97.52	Diabetic
	Diabetic	97.03	
Case 43	Diabetic	95.1	
	Normal	92.27	Normal
Case 44	Pre-diabetic	90.5	
	Normal	90.3	
Case 45	Diabetic	97.5	Diabetic
	Diabetic	94.5	
Case 46	Diabetic	93.8	

The proposed system's decisions are compared with the domain expert ones [21,71], and the "system's effectiveness" is referred to the amount of right answers, that is to say, the answers that verify what the expert had said. In other words, the accuracy is inversely proportional to the amount of the system's failures.

As shown in Table 7, our CDSS takes decisions similar to those of domain expert for all cases in the test set. The table contains three main columns the proposed system decision, the confidence of these decisions, and the corresponding domain expert decisions. This study testified the performance of the proposed CBR approach through experiment. The system results are contrasted with the domain expert decisions to determine if the results matched the

Table 8
The 2×2 ROC confusion matrix.

System decision	Domain expert decision	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

diagnosis expected by the expert or not. With these data, the accuracy, precision, recall, accuracy, and *f*-measure of the system could be measured. We have selected $k=3$ to assert the system behavior. For example, in case 1, the system has decided that this case has Diabetic diagnose for the three choices with similarities of 90.1, 88.2, and 88. The system performs right for most types of diagnosis, e.g. Pre-diabetes, Diabetic (cases 2, 3, 5, 10), and Normal. As we can see in Table 7, there is only one false decision in case 2, where the patient is diabetic, but the system diagnose it as per-diabetic. The semantic performance of the system is 97.67%, compared to 66% using Node Distance (ND) metrics only, 79% using IC similarity metric only, and 82% using combination of both IC and ND [91].

Based on results in Table 7, we use a 2×2 ROC confusion matrix to calculate the evaluation metrics of our system. For *Diabetic* decisions only, the values of TP, FP, FN, TN in Table 8 can be interpreted as:

- TP = the CBR system decides the diabetic case, and *domain expert* decides a diabetic case.
- FP = the CBR system decides a diabetic case, but the *domain expert* do not.
- FN = the CBR system decides not a diabetic case, but the *domain expert* decides it is diabetic.
- TN = the CBR system decides not a diabetic case, and the *expert decides* not a diabetic case.

The above parameters can be evaluated for Pre-diabetic and Normal as well. For space restrictions, we calculate Precision (*P*), Recall (*R*), Accuracy (*A*), Sensitivity (*S*), Effectiveness (*E*), and Negative Prediction Value (NPV) for *Diabetic* decisions only as follows. The metrics *E* and NPV are calculated using the following equations:

$$\text{Effectiveness} (E) = F - \text{Measure} (\text{Score}) = \frac{1}{(1/2P) + (1/2R)} \quad (20)$$

Table 9
Diabetic decision confusion matrix.

System decision	Domain expert decision	
	Positive	Negative
Positive	27	0
Negative	1	15

$$\text{Negative Prediction Value (NPV)} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (21)$$

From Table 7, we have calculated the values in Table 9 for the proposed systems.

The *P*, *R*, *A*, *S*, *E* regarding diabetic diagnosis are $P = \frac{27}{27+0} = 100\%$, $R = \frac{27}{27+1} = 96.43\%$, $A = \frac{27+15}{27+15+0+1} = 97.67\%$, $S = \frac{15}{15+0} = 100\%$, $E = \frac{1}{(1/2 \times (1)) + (1/2 \times (0.9643))} = 98.18\%$, and $\text{NPV} = \frac{15}{15+1} = 93.75\%$.

Although, the pre-diabetic and normal patients form less than half of the case-base, the proposed system accuracy for predicting them is 100%. The performance of our proposed system is enhanced because its similarity measures take into account the nature of all features.

6.2.4. A comparison between the proposed and other CBR systems

Most of the existing diabetes diagnosis CBR systems are traditional, and they did not provide adequate evaluations [6,8]. Fig. 23 shows a comparison with two diabetes diagnosis systems [92], and it asserts that our system has a better performance than these systems.

Montani et al. [93] proposed a traditional CBR system for diabetes care with the accuracy of 83%. The 4DSS hybrid CBR-RBR (Rule Based Reasoning) system proposed by Marling [89] has retrieval accuracy of 80%. Fuzzy case-based reasoning has not been used for diagnosis of diabetes before; however, it has been used for developing other medical systems as the diagnosis of stress [94]. The results of this system are Precision = 79.16% and Recall = 79.96%. Utilizing fuzzy ontology with a rule-based system for diabetes management in [23] has enhanced the accuracy to be 91.2%. However, Lee and Wang [23] used the rule-based reasoning technique, which is not suitable for experience-based problems such as diabetes diagnosis. Moreover, Lee and Wang's study used the Pima Indians Dataset, but we use real cases from Mansura University Hospitals in Egypt.

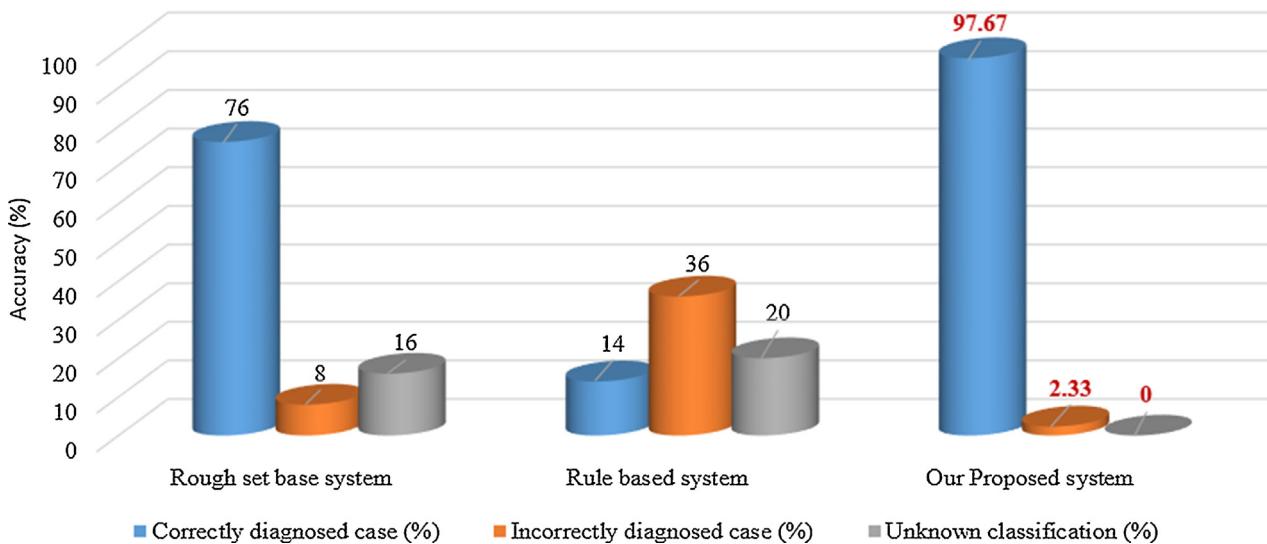


Fig. 23. A comparison between the proposed system and traditional ones.

Table 10

A comparison between the accuracy of proposed system and other studies.

Reasoning type	Domain	System name	Purpose	Accuracy (%)
Fuzzy CBR	Medical and semantic Medical	The proposed system	Diabetes diagnosis	97.67
		ConFuCiuS [95]	Diabetes diagnosis	75.53
		CBFDT [96]	Diagnosis of liver disorder	85
		Begum et al. [97]	Diagnosis of stress	80
		Petrovic et al. [88]	Radiotherapy planning	84.72
	Non-medical	Li et al. [98]	Financial application	92.36
		Arias-Aranda et al. [99]	Knowing the relationship between flexibility and operations strategy	89.23
		Khanum et al. [22]	Facial expression recognition	85
		Han et al. [100]	Endpoint prediction of basic oxygen furnace (BOF)	91.98
		Sushmita et al. [101]	Financial application	75
Traditional CBR	Medical	Xiong et al. [102]	Hybrid rule-CBR	93.25
		Martins-Bede et al. [103]	Classifying the prevalence of Schistosomiasis in the state of Minas Gerais in Brazil	71
		Jin et al. [104]	Customer-driven design	92
		System implemented on our dataset	Diabetes diagnosis	57.14
		T-IDDM [93]	Diabetes treatment and monitoring using conventional insulin therapy	84
		Marling et al. [105]	Type 1 diabetes management in insulin pump therapy	77.5
		Balakrishnan et al. [106]	Predictive system for diabetic retinopathy	85
		Bellazzi et al. [107]	Diabetes therapy	90
		Marling et al. [89]	4DSS system for diabetes diagnosis	80

Based on our case-base knowledge, we have implemented a CBR system without any semantic capabilities (i.e. neither case base fuzzy ontology nor domain standard ontology). The resulting system has achieved precision = 85.7%, recall = 42.85%, accuracy = 57.14%, specificity = 85.7%, effectiveness = 57.13%, and NPV = 42.9%. Our system has achieved a better performance, which explains the effects of case base knowledge preparation and semantic case retrieval algorithms. Moreover, our system has only one false case (Case 2 in Table 7). One of the most important features of CBR is the ability to retrieve k similar cases to the current problem. In our systems, if we just consider $k = 2$, then our system will have no false negative cases, and the accuracy will be 100%. To compare our system with previous studies, Table 10 compares our system's performance with a set of existing medical and non-medical CBR studies.

6.2.5. A comparison of the proposed system and machine learning classifiers

Shankaracharya et al. [108] presented a review of diabetes diagnosis techniques. Techniques such as artificial neural networks (ANN), support vector machines (SVMs), neuro-fuzzy systems and expert systems that developed by different authors have been discussed. Firstly, all these studies have lower performance than ours. However, these systems mostly depend on Pima Indians Dataset¹². To compare our system with these techniques, it is better to run these algorithms on our dataset. This dataset has been prepared before, and all noise and missing data have been handled [35]. For the comparing purpose, we apply some machine learning classifiers including C4.5, k -NN, SVM, Bayesian classifier, and ANN on our dataset and measure their performance. We use the 2-fold, 3-fold, 4-fold, 10-fold The cross-validation technique in the evaluation process. Cross-validation is a statistical technique useful in determining the robustness of a model. The n -fold cross validation divides the whole data set into n folds. The $n - 1$ folds are used for training, and one fold is used for testing. This process is continued until each fold from n is used for testing.

The overall performance of these algorithms is presented in Table 11. For the k -NN algorithm, we select $k = 3$ as done in our system; however, its performance is low. C4.5 achieves the best performance (about 89.19%) among machine-learning techniques; however, our system outperforms it. After testing the machine learning algorithms using from 2-fold to 10-fold cross-validation techniques, we calculate the average performance of each fold, and we make a comparison of different folds' results. Fig. 24 shows that the best performance is achieved with 5-fold cross validation.

We calculate the average precision, recall, accuracy, f -measure, and specificity for all folds. These averages are compared with the proposed system, the 5-fold cross validation, and the traditional (i.e. not fuzzy and not semantic) system, as shown in Fig. 25. Our findings show that the fuzzy KI-CBR can classify data more accurately than the other machine learning techniques and conventional CBR.

It can be seen in Fig. 25 that the machine learning classifiers have better performances than conventional CBR systems. This means that our study makes a high improvement in the CBR performance. The average accuracies of C4.5, conventional CBR, and proposed system are 88.88%, 57.14%, and 98.18%, respectively. The proposed approach demonstrates a major improvement than machine learning techniques and conventional CBR system.

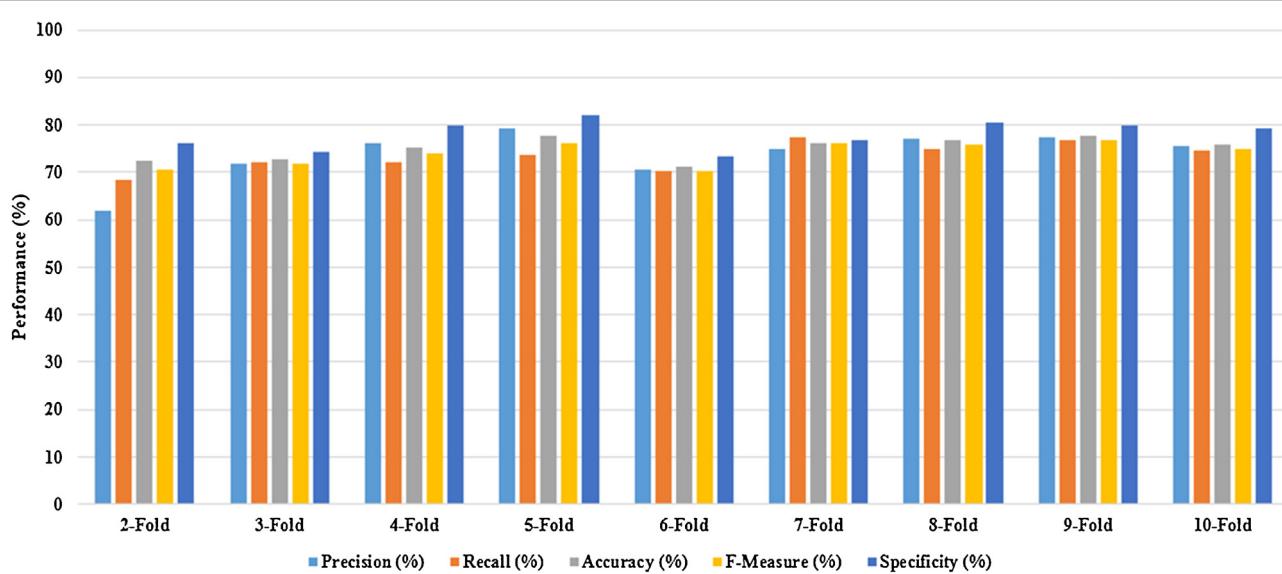
The results of this study clearly indicate that the hybridization of CBR with fuzzy ontology and medical ontologies is the most suitable technique for solving medical diagnosis problems. The enhanced performance of our system is a result of a couple of reasons. Firstly, the proposed CBR framework is integrated and complete. All components have been fully implemented and tested. The knowledge representation formalism using fuzzy ontology integrates the reasoning capabilities of fuzzy logic, description logic, and CBR. There are many studies, which use each of these reasoning mechanisms individually, but they have not achieved high accuracy. The second reason is the preparation of case-base data. These data have been pre-processed, fuzzified, and encoded before populated into the case-base knowledge. As a result, accurate data will produce accurate decisions. The third reason is the usage of a suitable weight vector for the used case features; the global similarity function has produced suitable similarities. The fourth reason is the proposed semantic retrieval algorithm. We have handled most of the possible datatypes, which appear in the medical domain. The fuzzy types

¹² <https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>.

Table 11

Performance of machine learning algorithms on our dataset.

Machine learning algorithms	Fold	Algorithm	Precision (%)	Recall (%)	Accuracy (%)	F-measure (%)	Specificity (%)
2-Fold	C4.5	93.1	93.1	93.33	93.1	93.1	93.54
	<i>k</i> -NN (<i>k</i> = 3)	63.3	63.3	63.33	63.3	64.5	64.5
	SVM	6.3	58.6	63.33	60.7	67.74	67.74
	Naive Bayes	81.8	62.1	75	70.6	87.09	
	ANN	65.5	65.5	66.66	65.5	67.74	67.74
	C4.5	90	93.1	91.66	91.5	90.32	
3-Fold	<i>k</i> -NN (<i>k</i> = 3)	60	60	60	59.9	64.51	
	SVM	71	75.9	73.33	73.3	70.96	
	Naive Bayes	65.4	58.6	65	61.8	70.96	
	ANN	72.4	72.4	73.33	72.4	74.19	
	C4.5	89.7	89.7	90	89.7	90.32	
	<i>k</i> -NN (<i>k</i> = 3)	68.7	68.3	68.33	68	77.41	
4-Fold	SVM	69	69	70	69	70.96	
	Naive Bayes	77.3	58.6	71.66	66.7	83.87	
	ANN	75.9	75.9	76.66	75.9	77.41	
	C4.5	92.9	89.7	91.66	91.2	93.54	
	<i>k</i> -NN (<i>k</i> = 3)	68.3	68.3	68.33	68.3	70.96	
	SVM	78.6	75.9	78.33	77.2	80.64	
5-Fold	Naive Bayes	77.3	58.6	71.66	66.7	83.87	
	ANN	75.9	75.9	76.66	75.9	77.41	
	C4.5	92.9	89.7	91.66	91.2	93.54	
	<i>k</i> -NN (<i>k</i> = 3)	68.3	68.3	68.33	68.3	70.96	
	SVM	78.6	75.9	78.33	77.2	80.64	
	Naive Bayes	77.3	58.6	71.66	66.7	83.87	
6-Fold	ANN	78.6	75.9	78.33	77.2	80.64	
	C4.5	89.3	86.2	88.33	87.7	90.32	
	<i>k</i> -NN (<i>k</i> = 3)	61.7	61.7	61.66	61.5	67.74	
	SVM	67.7	72.4	70	70	67.74	
	Naive Bayes	61.5	55.2	61.66	58.2	67.74	
	ANN	73.3	75.9	75	74.6	74.19	
7-Fold	C4.5	89.7	89.7	90	89.7	90.32	
	<i>k</i> -NN (<i>k</i> = 3)	73.6	73.3	73.33	73.2	80.64	
	SVM	69.7	79.3	73.33	74.2	67.74	
	Naive Bayes	70.4	65.5	70	67.9	74.19	
	ANN	71.9	79.3	75	75.4	70.96	
	C4.5	89.7	89.7	90	89.7	93	
8-Fold	<i>k</i> -NN (<i>k</i> = 3)	68.7	68.3	68.33	68	77.41	
	SVM	74.2	79.3	76.66	76.7	74.19	
	Naive Bayes	82.6	65.5	76.66	73.1	87.09	
	ANN	70	72.4	71.66	71.2	70.96	
	C4.5	89.3	86.2	88.33	87.7	90.32	
	<i>k</i> -NN (<i>k</i> = 3)	66.8	66.7	66.66	66.4	74.19	
9-Fold	SVM	75	82.8	78.33	78.7	74.19	
	Naive Bayes	79.2	65.5	75	71.7	83.87	
	ANN	77.4	82.8	80	80	77.41	
	C4.5	74.2	79.3	76.66	76.7	90.32	
	<i>k</i> -NN (<i>k</i> = 3)	73.1	65.5	71.66	69.1	70.96	
	SVM	77.4	82.8	80	80	77.41	
10-Fold	Naive Bayes	79.2	65.5	75	71.7	83.87	
	ANN	74.2	79.3	76.66	76.7	74.19	
	Average (%)	73.88	73.39	75.1	74.04	78.04	
	Conventional CBR system	85.7	42.85	57.14	57.13	85.7	
	Proposed fuzzy KI-CBR system	100	96.43	97.67	98.18	100	

**Fig. 24.** A comparison between the *n*-folds cross validation results.

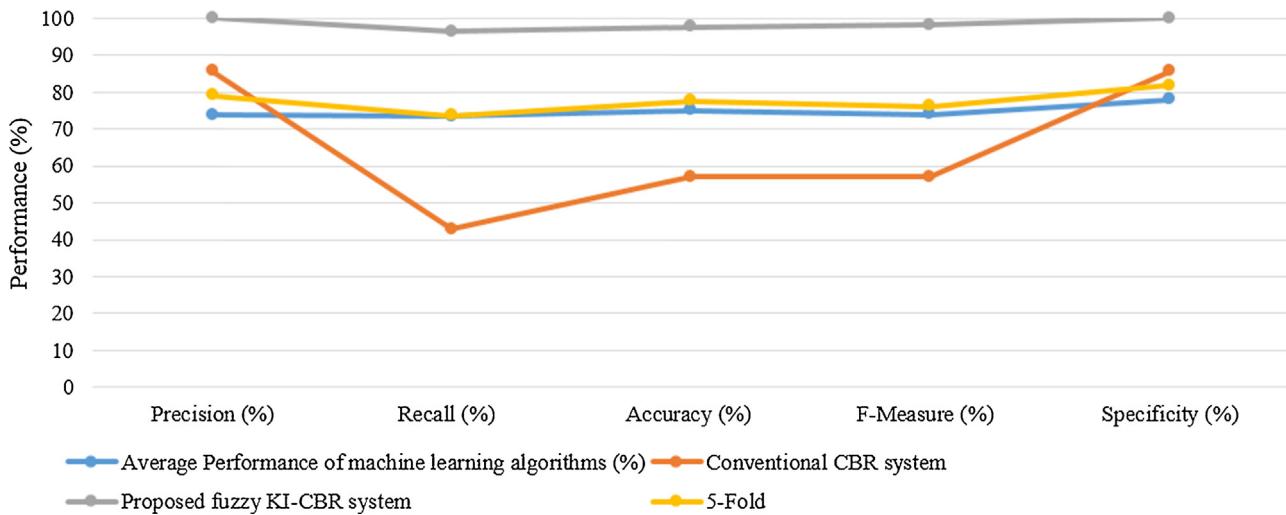


Fig. 25. Classification results comparison.

support the reasoning using linguistic terms and enhance the similarity calculation. Ordinal features' similarity is based on the expert domain knowledge in the form of similarity matrixes. Semantic features support the calculation of clinical similarities between SCT concepts.

In addition to its enhanced performance, the proposed system is tested for problems that are complex and cannot be solved by traditional systems. For example, If the case base contains a case $C_1 = \{age = 20, disease = "Acute proliferative", urination frequency = "+"\}$ and the query case is $\{age = young, disease = "Idiopathic crescentic", urination frequency = "Nil"\}$; in the traditional CBR systems, these cases are not similar and C_1 will not be returned. For fuzzy systems, the age is matched right as $age = 20$ is the same as $age = young$ (i.e., $\mu_{Young}(20) = 1$). However, the comparison of semantic and ordinal features fails to get the similarity. In semantic CBR systems, they fail to get the similarity of fuzzy and ordinal features. Due to these conditions, the results of these systems might prove to be not accurate. In our proposed system, we have proposed algorithms to handle all of these types.

7. Conclusion

This paper proposes a fuzzy ontology-based semantic CBR system and its implementation for a decision support system for diabetes diagnosis. This system enhances the decision maker efficiency in the diagnosing process. The proposed approach has many contributions and novelties: (1) It builds a case-base fuzzy ontology compatible with the most famous CBR framework, i.e. JCOLIBRI, (2) It builds and uses a standard medical terminology subset for diabetes diagnosis from SCT, which is the most complete medical ontology, and (3) It proposes a fuzzy-semantic similarity algorithm for case retrieval. Our implemented fuzzy ontology has followed a formal methodology, and it has represented using fuzzy OWL2 language. The proposed fuzzy-semantic retrieval algorithm outweighs all of the JCOLIBRI algorithms, and it covers their limitations. The integration of path-based similarity measures and feature-based measures enhances the accuracy of calculating clinical distances between concepts. Our system has achieved a performance of 97.67%. These results show that the proposed system has high accuracy, and physicians can consult it when diagnosing patients. In the future, we will implement the rest of the CBR steps especially the case adaptation process. We will utilize fuzzy ontology in the other steps of CBR as case adaptation, retention, and case-base maintenance. Moreover, we will try to

integrate multiple medical ontologies in our system because SCT has limitation in many aspects as lab tests and genes representation. Fortunately, there are many standard medical ontologies for these domains such as LOINC for lab tests and GO for genes representation. The integration of CBR with EHR environment will enhance the automation of the decision support process. Finally, we will benefit from the relational database for storing and querying the case base fuzzy ontology. The relational database supports storage of large case-base using a semantic preserving method.

Acknowledgments

This project was supported by King Saud University, Deanship of Scientific Research, College of Sciences, Research Centre. The authors would like to thank Dr. Farid Badria, Prof. of Pharmacognosy, Department and head of Liver Research Lab, Mansoura University, Egypt; and Dr. Hosam Zaghloul, Prof. at Clinical Pathology Department, Faculty of Medicine, Mansoura University, Egypt, for their efforts in this work.

References

- [1] World Health Organization (WHO). Diabetes; 2015. (<http://www.who.int/mediacentre/factsheets/fs312/en>) (accessed: 2 May 2015).
- [2] Ofori S, Unachukwu C. Holistic approach to prevention and management of type 2 diabetes mellitus in a family setting. *Diabetes Metab Syndr Obes* 2014;7:159–68.
- [3] Aljarullah A. Decision tree discovery for the diagnosis of type II diabetes. In: International conference on innovations in information technology. Abu Dhabi, UAE: IEEE; 2011. p. 303–7.
- [4] Begum S, Ahmed M, Funk P, Xiong N, Folke M. Case-based reasoning systems in the health sciences: a survey of recent trends and developments. *IEEE Trans Syst Man Cybernet, C* 2010;7(1):39–59.
- [5] Marlinga C, Montanib S, Bichindaritzc I, Funkd P. Synergistic case-based reasoning in medical domains. *Expert Syst Appl* 2014;41(2):249–59.
- [6] Jha M, Pakhira D, Chakraborty B. Diabetes detection and care applying CBR techniques. *Int J Soft Comput Eng (IJSCCE)* 2013;2(6):132–7.
- [7] Jaya A, Uma G. Role of ontology in case-based reasoning (CBR) for diagnosing diabetes. *J Inf Technol* 2009;5(3):17–23.
- [8] Chen J, Su S, Chang C. Diabetes care decision support system. In: The 2nd international conference on industrial and information systems (IIS). 2010. p. 323–6, 1.
- [9] El-Sappagh S, Elmogy M, Riad A. A CBR system for diabetes mellitus diagnosis: case-base standard data model. *Int J Med Eng Inf* 2015;7(3).
- [10] Dendani N, Khadir M, Guessoum S. Use a domain ontology to develop knowledge intensive CBR systems for fault diagnosis. In: International conference on information technology and e-Services (ICITeS). Sousse, Tunisia: IEEE; 2012. p. 1–6.
- [11] Diaz-Agudo B, Gonzalez-Calero P. An architecture for knowledge intensive CBR systems, advances in case-based reasoning, Enrico Blanzieri and Luigi Portinale, vol. 1898. Berlin, Heidelberg, Germany: Springer; 2000. p. 37–48.

- [12] Amailef K, Lu J. Ontology-supported case-based reasoning approach for intelligent m-Government emergency response services. *Decis Support Syst* 2013;55(1):79–97.
- [13] Chen R, Huang Y, Bau C, Chen S. A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection. *Expert Syst Appl* 2012;39(4):3995–4006.
- [14] Health Level Seven International (HL7), (<http://www.hl7.org/>) (accessed: 19 August 2015).
- [15] The International Health Terminology Standards Development Organization (IHTSDO), SNOMED CT: The Global Language of Healthcare, (<http://www.ihtsdo.org/snomed-ct>) (accessed: 19 August 2015).
- [16] Melton G, Parsons S. Inter-patient distance metrics using SNOMED CT defining relationships. *J Biomed Inf* 2006;39:697–705.
- [17] Jirathitkul P, Nithisanawadikul S, Tongphu S, Suntisrivaraporn B. A similarity measuring service for SNOMED-CT structural analysis of concepts in ontology. In: The 11th international conference on electrical engineering/electronics, computer, telecommunications and information technology (ECTI-CON). 2014. p. 1–6.
- [18] El-Sappagh S, Elmogy M, El-Masri S, Riad A. A diabetes diagnostic domain ontology for CBR system from the conceptual model of SNOMED CT. In: The second international conference on engineering and technology (ICET 2014). Cairo, Egypt: IEEE; 2014. p. 1–7.
- [19] El-Sappagh S, Elmogy M, Riad A, Zaghloul H, Badria F. A proposed SNOMED CT ontology-based encoding methodology for diabetes diagnosis case-base. In: The ninth international conference on computer engineering and systems (ICCES 2014). Cairo, Egypt: IEEE; 2014. p. 184–91.
- [20] Zadeh L. From search engines to question-answering systems the need for new tools, advances in web intelligence, Ernestina Menasalvas, Javier Segovia, and Piotr Szczepaniak, vol. 2663. Berlin, Heidelberg, Germany: Springer; 2003. p. 15–7.
- [21] Alexopoulos P, Wallace M, Kafentzis K, Askounis D. Utilizing imprecise knowledge in ontology-based CBR systems by means of fuzzy algebra. *Int J Fuzzy Syst* 2010;12(1):1–14.
- [22] Khanum A, Mufti M, Javed M, Shafiq M. Fuzzy case-based reasoning for facial expression recognition. *Fuzzy Sets Syst* 2009;160:231–50.
- [23] Lee C, Wang M. A fuzzy expert system for diabetes decision support application. *IEEE Trans Syst Man Cybernet, B: Cybernet* 2011;41(1):139–53.
- [24] Zhaoa J, Cui L, Zhao L, Qiu T, Chen B. Learning HAZOP expert system by case-based reasoning and ontology. *Comput Chem Eng* 2009;33:371–8.
- [25] Samwald M, Fehre K, Bruin J, Adlassnig K. The Arden syntax standard for clinical decision support: experiences and directions. *J Biomed Inf* 2012;45:711–8.
- [26] Sánchez D, Batet M, Isern D, Valls A. Ontology-based semantic similarity: a new feature-based approach. *Expert Syst Appl* 2012;39:7718–28.
- [27] Gan M, Dou X, Jiang R. From ontology to semantic similarity: calculation of ontology-based semantic similarity. *Sci World J* 2013;2013:1–11.
- [28] Rahimi A, Liaw S, Taggart J, Ray P, Yu H. Validating an ontology-based algorithm to identify patients with type 2 diabetes mellitus in electronic health records. *Int J Med Inf* 2014;83(10):768–78.
- [29] Sherimon P, Vinu P, Krishnan R, Takroni Y, AlKaabi Y, AlFars Y. Adaptive questionnaire ontology in gathering patient medical history in diabetes domain. Proceedings of the first international conference on advanced data and information engineering, Tutut Herawan, Mustafa Mat Deris, and Jemal Abawajy, vol. 285. Tannery Lane, Singapore: Springer; 2014. p. 453–60.
- [30] Sugiyanti Hayuhardhika W, Sarno R, Sidiq M. Weighted ontology and weighted tree similarity algorithm for diagnosing diabetes mellitus. In: International conference on computer, control, informatics and its applications (IC3INA). IEEE; 2013. p. 267–72.
- [31] El-Sappagh S, El-Masri S, Elmogy M, Riad AM, Saddik B. An ontological case base engineering methodology for diabetes management. *J Med Syst* 2014;38(8):67–81.
- [32] Samwald M, Stenzhorn H, Dumontier M, Marshall M, Luciano J, Adlassnig KP. Towards an interoperable information infrastructure providing decision support or genomic medicine. *Stud Health Technol Inform* 2011;169:165–9.
- [33] Samwald M, Antonio J, Giménez M, Boyce R, Freimuth R, Adlassnig KP, et al. Pharmacogenomic knowledge representation, reasoning and genome-based clinical decision support based on OWL 2 DL ontologies. *BMC Med Inf Decis Making* 2015;15:pp.12.
- [34] Yu H, Liaw S, Taggart J, Khorzoughi A. Using ontologies to identify patients with diabetes in electronic health records. In: International semantic web conference. Berlin, Heidelberg, Germany: Springer-Verlag; 2013. p. 77–80.
- [35] El-Sappagh S, Elmogy M, Riad A, Zaghloul H, Badria F. EHR data preparation for case based reasoning construction. The proceedings of the second international conference on advanced machine learning technologies and applications (AMLTA14), communications in computer and information science (CCIS), Aboul Ella Hassanien, Mohamed Tolba, and Ahmad Azar, vol. 488. Cham (ZG), Switzerland: Springer International Publishing; 2014. p. 483–97.
- [36] Adlassnig K. Fuzzy set theory in medical diagnosis. *IEEE Trans Syst Man Cybernet* 1986;16(2):260–5.
- [37] Abdul M, Muhammad A, Mustapha N, Muhammad S, Ahmad N. Database workload management through CBR and fuzzy based characterization. *Appl Soft Comput* 2014;22:605–21.
- [38] Ekong V, Inyang U, Onibere E. Intelligent decision support system for depression diagnosis based on neuro-fuzzy-CBR hybrid. *Modern Appl Sci* 2012;6(7):79–88.
- [39] Thirugnanam M, Kumar P, Srivatsan S, Nerlesh C. Improving the prediction rate of diabetes diagnosis using fuzzy, neural network, case based (FNC) approach. *Procedia Eng* 2012;38:1709–18.
- [40] Alexopoulos P, Wallace M, Kafentzis K, Askounis D. IKARUS-Onto: a methodology to develop fuzzy ontologies from crisp ones. *Knowl Inf Syst* 2012;32(3):667–95.
- [41] Rodríguez N, Cuéllar M, Lilius J, Calvo-Flores M. A fuzzy ontology for semantic modelling and recognition of human behaviour. *Knowl Based Syst* 2014;66:46–60.
- [42] Zhang F, Yan L, Ma Z. Reasoning of fuzzy relational databases with fuzzy ontologies. *Int J Intel Syst* 2012;27:613–34.
- [43] Torshizi A, Zarandi M, Torshizi G, Eghbali K. A hybrid fuzzy-ontology based intelligent system to determine level of severity and treatment recommendation for Benign Prostatic Hyperplasia. *Comput Methods Programs Biomed* 2014;113(1):301–13.
- [44] Carlsson C, Brunelli M, Mezei J. Decision making with a fuzzy ontology. *Soft Comput* 2012;16:1143–52.
- [45] Mezei J, Wikström R, Carlsson C. Aggregating linguistic expert knowledge in type-2 fuzzy ontologies. *Appl Soft Comput* 2015;(March), <http://dx.doi.org/10.1016/j.asoc.2015.03.023>. ISSN:1568-4946; pii:S1568494615001799.
- [46] Molinera J, Galvez I, Wikstrom R, Viedma E, Carlsson C. Designing a decision support system for recommending smartphones using fuzzy ontologies. *Intel Syst* 2015;323:323–34.
- [47] Ali F, Kim E, Kim Y. Type-2 fuzzy ontology-based opinion mining and information extraction: a proposal to automate the hotel reservation system. *Appl Intel* 2015;42(3):481–500.
- [48] Ali F, Kim E, Kim Y. Type-2 fuzzy ontology-based semantic knowledge for collision avoidance of autonomous underwater vehicles. *Inf Sci* 2015;295:441–64.
- [49] Lee C, Wang M, Hsu C, Chen Z. Type-2 fuzzy set and fuzzy ontology for diet application. *Adv Type-2 Fuzzy Sets Syst* 2013;301:237–56.
- [50] Park G, Benedictos J, Lee C, Wang M. Ontology-based fuzzy-CBR support system for ship's collision avoidance. International conference on machine learning and cybernetics, vol. 4. IEEE; 2007. p. 1845–50.
- [51] The Open Biological and Biomedical Ontologies, (<http://www.obofoundry.org/>) (accessed: 19 August 2015).
- [52] Recio-García J, Díaz-Agudo B, González-Calero P. In: Montani Stefania, Jain Lakhmi, editors. The COLIBRI platform: tools, features and working examples, successful case-based reasoning applications-2, vol. 494. Berlin, Heidelberg, Germany: Springer; 2014. p. 55–85.
- [54] Aamodt A, Plaza E. Case-based reasoning foundational issues, methodological variations, and system approaches. *J AI Commun* 1994;7(1):39–59.
- [55] Harispé S, Sanchez D, Ranwez S, Janaqi S, Montmain J. A framework for unifying ontology-based semantic similarity measures a study in the biomedical domain. *J Biomed Inf* 2014;48:38–53.
- [56] The SNOMED CT technical implementation guide, July 2014 International Release, International Health Terminology Standards Development Organization, (http://ihtsdo.org/fileadmin/user.upload/doc/download/doc.TechnicalImplementationGuide_Current-en-US_INT_20140731.pdf) (accessed: 19 August 2015).
- [57] Gruber T. Towards principles for the design of ontologies used for knowledge sharing. *Int J Hum Comput Stud* 1995;43(5–6):907–28.
- [58] Zadeh L. Fuzzy sets. *Inf Control* 1965;8:338–53.
- [59] Bobillo F. Managing vagueness in ontologies. Spain: University of Granada; 2008 (Ph. D. Thesis).
- [60] Bobillo F, Straccia U. fuzzyDL: an expressive fuzzy description logic reasoner. In: International conference on fuzzy systems (FUZZ-08). IEEE Computer Society; 2008. p. 923–30.
- [61] Recio-García J, Díaz-Agudo B. In: Ellis Richard, Allen Tony, Andrew Tuson, editors. Ontology based CBR with jCOLIBRI applications and innovations in intelligent systems XIV. London, WC1X 8HB, United Kingdom: Springer-Verlag; 2007. p. 149–62.
- [62] Description logics, (<http://dl.kr.org/>) (accessed: 19 August 2015).
- [63] Bobillo F, Straccia U. Fuzzy ontology representation using OWL 2. *Int J Approx Reason* 2011;52(7):1073–94.
- [64] Akmal S, Shih L, Batres R. Ontology-based similarity for product information retrieval. *Comput Ind* 2014;65(1):91–107.
- [65] Gu H, Wang X, Ling Y, Shi J. Building a fuzzy ontology of edutainment using OWL. *Comput Sci–ICCS* 2007;4489:591–4.
- [66] Baldarrago A, Santos M, Prado A. UPFON: unified process for building fuzzy ontology. In: The ninth international conference on fuzzy systems and knowledge discovery (FSKD). Chongqing, Sichuan, China: IEEE; 2012. p. 617–22.
- [67] Ghorbel H, Bahri A, Bouaziz R. Fuzzy ontologies building method: fuzzy ontomethodology. In: Annual meeting of the North American Fuzzy Information Processing Society (NAFIPS). Toronto, Canada: IEEE; 2010. p. 1–8.
- [68] Yagquinuma C, Santos M, Camargo H, Nicoletti M. Fuzz-onto: a meta-ontology for representing fuzzy elements and supporting fuzzy classification rules. In: The 12th international conference on intelligent systems design and applications (ISDA). Cochin, India: IEEE; 2012. p. 166–71.
- [69] Wu Z, Palmer M. Verb semantics and lexical selection. In: Proceedings of the 32nd annual meeting on association for computational linguistics. San Francisco, CA: Morgan Kaufmann Publishers; 1994. p. 133–8.
- [70] Ghanea-Hercock R. Applied evolutionary algorithms in Java. 2003 ed. New York, NY: Springer-Verlag; 2003.
- [71] Lin K, Shih L, Lu S, Lin Y. Strategy selection for product service systems using case-based reasoning. *Afr J Bus Manage* 2010;4(6):987–94.

- [72] Garla V, Brandt C. Semantic similarity in the biomedical domain: an evaluation across knowledge sources. *BMC Bioinf* 2012;13(1):pp. 261.
- [73] Ma Z, Zhang F, Yan L, Cheng J. Fuzzy ontology knowledge bases storage in fuzzy databases, fuzzy knowledge management for the semantic web. *Stud Fuzz Soft Comput* 2014;306:233–42.
- [74] Zhang F, Ma Z, Fan G, Wang X. automatic fuzzy semantic web ontology learning from fuzzy object-oriented database model. *Database Expert Syst Appl* 2010;6261:16–30.
- [75] Qi J, Hu J, Peng Y, Wang W, Zhang Z. A case retrieval method combined with similarity measurement and multi-criteria decision making for concurrent design. *Expert Syst Appl* 2009;36:10357–66.
- [76] Seco N, Veale T, Hayes J. An intrinsic information content metric for semantic similarity in WordNet. The 16th Eur Conf Artif Intell (ECAI), vol. 16. Amsterdam, Netherlands: IOS Press; 2004. p. 1089–90.
- [77] Sánchez D, Batet M, Isern D. Ontology-based information content computation. *Knowledge-Based Syst* 2011;24:297–303.
- [78] Poveda-Villalón M, Suárez-Figueroa M, Gómez-Pérez A. Validating ontologies with OOPS! *Knowl Eng Knowl Manage* 2012;7603:267–81.
- [79] Brewster C, Alani H, Dasmahapatra S, Wilks Y. Data driven ontology evaluation. In: Proceedings of the international conference on language resources and evaluation. 2004. p. 164–8.
- [80] Bright T, Furuya E, Kuperman G, Cimino J, Bakken S. Development and evaluation of an ontology for guiding appropriate antibiotic prescribing. *J Biomed Inf* 2012;45:120–8.
- [81] Heras S, Botti V, Julian V. ArgCBROnto: a knowledge representation formalism for case-based argumentation. *Agreement Technol* 2013;8068:105–19.
- [82] Guo Y, Hu J, Peng Y. A CBR system for injection mould design based on ontology: a case study. *Comput -Aided Des* 2012;44(6):496–508.
- [83] Zhukova I, Kultsova M, Navrotsky M, Dvoryankin A. Intelligent support of decision making in human resource management using case-based reasoning and ontology. *Knowledge-Based Softw Eng* 2014;466:172–84.
- [84] Djedidi R, Aufaure M. ONTO-EVOAL an ontology evolution approach guided by pattern modeling and quality evaluation. In: Foundations of information and knowledge systems, Sebastian Link and Henri Prade. Berlin, Heidelberg, Germany: Springer; 2010. p. 286–305.
- [85] Zhang D, Ye C, Yang Z. An evaluation method for ontology complexity analysis in ontology evolution, managing knowledge in a world of networks, Steffen Staab and Vojtěch Svátek, 4248. Berlin, Heidelberg, Germany: Springer-Verlag; 2006. p. 214–21.
- [86] Fernández M, Overbeeke C, Sabou M, Motta E. What makes a good ontology? A case study in Fine-Grained knowledge reuse. *Semant Web* 2009;5926:61–75.
- [87] Satter R, Cohen T, Ortiz P, Kahol K, Mackenzie J, Olson C, et al. Avatar-based simulation in the evaluation of diagnosis and management of mental health disorders in primary care. *J Biomed Inform* 2012;45:1137–50.
- [88] Petrovic S, Mishra N, Sundar S. A novel case based reasoning approach to radiotherapy planning. *Expert Syst Appl* 2011;38:10759–69.
- [89] Marling C, Wiley M, Cooper T, Bunescu R, Shubrook J, Schwartz F. The 4 diabetes support system: a case study in CBR research and development. In: Ram Ashwin, Wiratunga Nirmalie, editors. The proceeding of the 19th international conference on case-based Reasoning (ICCBR). Berlin, Heidelberg, Germany: Springer; 2011. p. 137–50.
- [90] Armengol E, Esteva F, Godo L, Torra V. On learning similarity relations in fuzzy case-based reasoning. *Trans Rough Sets II* 2005;3135:14–32.
- [91] Fernandes R, Grosse I, Krishnamurti S, Witherell P, Wileden J. Semantic methods supporting engineering design innovation. *Adv Eng Inf* 2011;25:185–92.
- [92] Anoucia S, Madonna L, Jeevitha P, Nandhini R. Design of a diabetic diagnosis system using rough sets. *Cybern Inf Technol* 2013;13(3):124–39.
- [93] Montani S, Bellazzi R, Portinale L, d'Annunzio G, Fiocchi S, Stefanelli M. Diabetic patients management exploiting case-based reasoning techniques. *Comput Methods Programs Biomed* 2000;62:205–18.
- [94] Mobyen A, Begum B, Funk P, Xiong N, Schéele B. Case-based reasoning for diagnosis of stress using enhanced cosine and fuzzy similarity. *Trans Case-Based Reasoning Multimedia Data* 2008;1(1):3–19.
- [95] Rodriguez Y, Garcia M, Baets B, Morell C, Bello R. A connectionist fuzzy case-based reasoning model, MICAI: advances in artificial intelligence, Alexander Gelbukh and Carlos Reyes-Garcia. Berlin, Heidelberg, Germany: Springer; 2006. p. 176–85.
- [96] Fan C, Chang P, Lin J, Hsieh J. A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. *Appl Soft Comput* 2011;11:632–44.
- [97] Begum S, Ahmed M, Funk P, Xiong N, Schéele V. A case-based decision support system for individual stress diagnosis using fuzzy similarity matching. *Int J Comput Intell Appl* 2009;25(3):180–95.
- [98] Li S, Ho H. Predicting financial activity with evolutionary fuzzy case-based reasoning. *Expert Syst Appl* 2009;36:411–22.
- [99] Arias-Aranda D, Castro J, Navarro M, Zurita J. A CBR system for knowing the relationship between flexibility and operations strategy. *Found Intell Syst* 2009;5722:463–72.
- [100] Han M, Cao Z. An improved case-based reasoning method and its application in end point prediction of basic oxygen furnace. *Neurocomputing* 2015;149(Part C):1245–52.
- [101] Sushmita S, Chaudhury S. Hierarchical fuzzy case based reasoning with multi-criteria decision making for financial applications. *Pattern Recognit Mach Intell* 2007;4815:226–34.
- [102] Xiong N. Learning fuzzy rules for similarity assessment in case-based reasoning. *Expert Syst Appl* 2011;38:10780–6.
- [103] Godo L, Sandri S, Dutra L, Freitas C, Carvalho O, Guimarães R, et al. Classification of schistosomiasis prevalence using fuzzy case-based reasoning. *Bio-Inspired Syst: Comput Ambient Intell* 2009;5517:1053–60.
- [104] Jin Q, Jie H, Ying-hong P, Wei-ming W, Zhen-fei Z. New weighted fuzzy case retrieval method for customer-driven product design. *J Shanghai Jiaotong Univ (Sci)* 2010;15(6):641–50.
- [105] Marling C, Shubrook J, Schwartz F. Case-based decision support for patients with type 1 diabetes on insulin pump therapy. Advances in case-based reasoning: ninth European conference (ECCB), Klaus-Dieter Althoff, Ralph Bergmann, Mirjam Minor, and Alexandre Hanft, vol. 5239. Berlin, Heidelberg, Germany: Springer; 2008. p. 325–39.
- [106] Balakrishnan V, Shakouri M, Hoodeh H, Loo H. Predictions using data mining and case-based reasoning: a case study for retinopathy. *World Acad Sci Eng Technol* 2012;63:573–6.
- [107] Bellazzi R, Montani S, Portinale L. Retrieval in a prototype-based case library: a case study in diabetes therapy revision. In: Smyth Barry, Cunningham Pádraig, editors. *Advances in Case-Based Reasoning*, vol. 1488. Berlin, Heidelberg, Germany: Springer; 1998. p. 64–75.
- [108] Shankaracharya D, Odedra S, Vidhyarthi Samanta A. Computational intelligence in early diabetes diagnosis: a review. *Rev Diabetes Stud* 2010;7:252–62.