INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Check for updates

OPEN ACCESS

Received: 16-10-2024

Accepted: 28-10-2024

Published: 30-12-2024

Citation: Kere H, Meshesha M, Rashid F, Singh G, Abate J, Shiferaw T, Jifara W, Ketema D (2024) An Intelligent Knowledge Based System for Diagnosis and Treatment of Diabetes. Indian Journal of Science and Technology 17(48): 5033-5045. https://doi.org/10.17485/IJST/v17i48.3412

* **Corresponding author**.

abatejemal@gmail.com

Funding: This study is funded by Haramaya University

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

An Intelligent Knowledge Based System for Diagnosis and Treatment of Diabetes

Hasan Kere[1], Million Meshesha[2], Faizur Rashid[3], Gavendra Singh[4], Jemal Abate[1]*, Tilahun Shiferaw[1], Wabi Jifara[1], Demeshi Ketema[1]

1 Department of Information Science, College of Computing and Informatics, Haramaya University, Dire Dawa, Ethiopia
2 School of Information Science, Addis Ababa University, Addis Ababa, Ethiopia
3 School of Computational sciences, JSPM University, Pune, Maharashtra, India
4 Department of Computer science and Engineering, SDEC , SDGI GLOBAL UNIVERSITY, NH9(24) Delhi Hapur Road, Ghaziabad, 201015, U.P, India

Abstract

Background/Objectives: Diabetes poses a significant health challenge in Ethiopia due to factors such as lifestyle choices, a lack of medical professionals, and limited laboratory resources. Knowledge-based system (KBS) are increasingly being explored to assist healthcare providers in diagnosing and managing diseases, including diabetes. Rule-based and case-based reasoning are two primary approaches to developing these knowledge-based systems. The major goal of this research is to create an intelligent KBS for diabetes diagnosis and therapy utilizing data mining approaches. **Methods:** This study uses a design science research approach to develop a smart, knowledge-based system for diagnosing and managing diabetes, combining rule-based and case-based reasoning. Data were collected from Hiwot Fana Specialized Hospital at Haramaya University and Harari Regular Hospital, with the dataset undergoing preprocessing to address errors, noise, outliers, duplicates, missing values, and normalization. Rule-based reasoning was implemented using the Prolog programming language, while the case-based reasoning system was built with JCOLLIBIR studio using its CBR main cycle. For predictive and descriptive modeling, the J48 decision tree, PART, and JRip algorithms were evaluated for rule-based reasoning, while K-means and Farthest First were evaluated for case-based reasoning. **Findings:** The JRip rule induction algorithm produced superior results for the rule-based system, while the K-means clustering algorithm performed best for case-based reasoning. These outputs were used to construct the rule base and automatically generate the case base, respectively. The integrated system achieved an overall accuracy of 95.23%, with a precision of 100% and a recall of 89.7%. Additionally, 93.2% of users passed the user acceptance test, demonstrating the system's practical utility. **Novelty and Applications:** This study presents a novel integration of rule-based and case-based reasoning using a rule-dominant approach to create an intelligent knowledge-based system for diabetes diagnosis and management. The system has shown high accuracy and user satisfaction, making it a viable tool for aiding healthcare professionals in Ethiopia, especially in resource-limited settings.

Keywords: Rule Based Reasoning; Case Based Reasoning; Diagnosis and Treatment of Diabetes; Intelligent Knowledge-Based System

1 Introduction

According to[1], non-communicable diseases (NCDs) like diabetes, cardiovascular disease, cancer, and chronic respiratory illnesses are presently the leading causes of death in much of the world, including countries in sub-Saharan Africa. Diabetes is a substantial public health burden, with increasing incidence, morbidity, death, and economic repercussions. It is one of the most common non-communicable illnesses worldwide[2]. The American Diabetes Association (ADA) asserts that diabetes creates a substantial financial burden on healthcare overheads. This is because diabetes accounted for 11% (465 billion) of global healthcare expenses in 2011. By 2030, this figure is expected to surpass 595 billion[3].

It has been widely acknowledged for many years that there are three main types of diabetes: type I diabetes, type II diabetes, and gestational diabetes mellitus (GDM). The classification, diagnosis, and treatment of diabetes are complicated and have been the focus of significant consultation, discussion, and modification over many decades. Because the immune system, which normally protects you from infection by clearing your body of bacteria, viruses, and other harmful substances, has attacked and destroyed the cells that manufacture insulin, people with type I diabetes are unable to produce enough insulin in their bodies.

The root cause of type I diabetes is an autoimmune response where the body's immune system attacks the insulin-producing beta cells in the islets of the pancreas gland[4]. Type II diabetes is the most common type of diabetes accounting for around 90% of all cases of diabetes and in this type of diabetes, the pancreas produces insulin, but it doesn't yield a sufficient amount, or the body may have difficulty in using it properly[1]. Intelligent knowledge based system (IKBS) is a computer program and division of Artificial Intelligence (AI) that imitators the behavior of a human being and that has expert knowledge and experience in a particular domain towards deciding as the human expert[5]. It is the type of application program that makes decisions or solves problems in a specific field by using stored knowledge and analytical expertise defined by domain experts in that field. IKBS is meant to solve a real problem, which normally would require a specialized human expert such as a doctor in the medical field. Such a system also enables to deliver an intelligent decision making where there are insufficient medical experts, especially in developing countries like Ethiopia.

Using the IKBS technique the current problem is solved by examining and reusing preceding knowledge to turn that knowledge into useful knowledge-based system for the diagnosis and treatment of diabetes[6]. In general, data mining or knowledge discovery from historical data became the most preferable automatic knowledge acquisition technique used in recent years[7]. Knowledge acquisition using data mining technique reduces the difficulty caused by the knowledge acquisition bottleneck of IKBS and automates knowledge acquisition by obtaining a low-cost and high-quality knowledge base. Thus, this study was instigated with the main objective of developing an intelligent knowledge-based system for the diagnosis and treatment of diabetes.

There has been a lot of research conducted on diabetes diagnosis and treatment by integrating knowledge-based systems with data mining and machine-learning techniques. The study done by[2] used an experimental research design approach to conduct studies on knowledge based system for diabetes prediction and recommendation using data mining techniques. The data sources used for this study were collected from

Jimma Medical Center, Saint Paul Millennium Medical Hospital, and Adama Medical College Hospital and applied the data mining preprocessing step for finding quality of information to acquire quality of knowledge resource for building an integrated data mining and knowledge-based system. He used Clisp.net for integrating data mining results that were imbedded in Java to prolog programming language and developing prototype knowledge-based systems that diagnose diabetes disease.

For this study, two categories of classifiers rules-based classifiers JRIP, PART, and tree-based classifiers J48, LMT, and REPTree were chosen from data mining classifier algorithms. The system's precision and recall averages were 71% and 76%, respectively, according to its performance.[1] performed a study on the integration of prediction models with the knowledge-based system for diagnosis and treatment of diabetes using an experimental research design approach.

To design this prototype the researcher used an arrangement of conditions for comparison and analysis of data in a manner that aims to address the research problem. To acquire the necessary knowledge the researcher selected classification algorithms such as decision tree (J48, REPTree and Random Tree) and rule classifiers (PART and JRiP) for experiment. In this study, the researcher undertakes a total of five experiments aimed at building predictive models. The processed dataset contains 3112 instances with 16 attributes and five class labels, and all the datasets engage in all experiments. The researcher conducts all experiments under a 10-fold cross-validation test option for training and testing.

A thorough review of the literature revealed that diabetes is one of the chronic diseases in the world that is affecting the economic development of countries. Healthcare institutions are taking diverse diagnosis and treatment measures to tackle the problems arising from diabetes. As for the review, different techniques and tools must be considered for diabetes disease management. However, such a process requires enormous resources to invest in, which is difficult for developing countries where there is an insufficient health expert and laboratory facilities like Ethiopia. Computer-based diagnostic and treatment systems such as KBS can support physicians in diagnostic and treatment processes for healthcare systems. Different studies have been applying for the implementation of a KBS in healthcare. Most of the researchers,[1,2,8], are limited only on rule based reasoning approach to design and develop the knowledge based system for the diagnosis and treatment of diabetes.

An intelligent knowledge-based system applies the power of rule-based reasoning and case-based reasoning on a single problem towards increasing the reasoning capacity of KBS system by incorporating both general and specific knowledge of the domain. The data collection method of many researchers is making interviews with domain experts and analysis of relevant documents for acquiring the required knowledge without using automatic knowledge acquisition. Keeping all the above-related and conceptual literature review in mind, this study therefore emphasizes exploring a way to design and develop an intelligent knowledge-based system that takes advantage of both rules based and case-based reasoning for diagnosis and treatment of diabetes using automatically acquired knowledge.

2 Material and methods

2.1 Research Design

Method is the way that deals with data collection, data analysis, algorithm selection techniques and interpretations in order to achieve the objectives and answer the research question[9]. It is a discipline of studying how investigation is conducted scientifically. In the development of intelligent KBS for diagnosis and treatment of diabetes, design science research methodology has been used. Design science research includes processes of identifying a practical research problem, creating artefacts to address the problem, evaluate the result and communicates the findings to appropriate audiences effectively[10]. According to[11], design science research includes six steps: Problem identification and motivation, objective of the solution, design and development, demonstration, evaluation and communication.

In the field of informatics design science research method contributes to the development of practical and innovative solutions, evaluation, theory development, and practical relevance[12]. It enables researchers to create artifacts that address real-world problems, leading to advancements in the field and improvement of technologies. To conduct research using design science research starting from problem-centered point of entry and diabetes dataset using data mining algorithms the researcher follows these steps: problem identification, define research objectives, design artifact, prototype development, evaluation, documentation and reporting and dissemination of results. The main reason the researcher selects the DSR is particularly well-suited for solving problems or addressing challenges in a specific domain.

2.2 Study Site and Area Coverage

The study focuses on integrating rule-based and case-based reasoning to give a medical diagnosis and treatment for diabetes patients and to support clinical decision-making. For automatic knowledge discovery, data mining is the best and the most important area for knowledge acquisition in order to design IKBS[5]. The hidden knowledge is mined from a large dataset which is collected from Haramaya University Hiwot Fana specialized hospital, Haramaya general hospital and Harari general hospital. The result or the discovered knowledge is used for building the IKBS. To this end, for this study data mining predictive technique is performed on experimental data set to acquire rules whereas descriptive technique of data mining is performed on experimental data set to acquire cases via the KDD process. Once rules and cases are automatically acquired from the data set, the next step is to design an intelligent knowledge-based system using these acquired rules and cases from dataset. Rules in this case are used for building a rule-based whereas cases in this case are used for case-base construction.

2.3 Data Mining Techniques

The data mining process inputs cleaned and transformed data, searches the data by using different techniques and algorithms, and then outputs patterns and relationships to the interpretation/evaluation step of the knowledge discovery from the data [13]. To create effective data mining models in terms of quality and performance, the raw dataset needs to undergo preprocessing steps in the form of data cleaning. Because real world data are mostly dirty and unclean, which need to correct bad data that is encountered from data redundancy, incompleteness or missing attributes value, noise, and inconsistency to make knowledge searching paths ease for mining algorithms.

2.4 Classification Methods

Classification is a widely known data mining technique that finds application in various domains for a variety of purposes. To be able to regularly use the model to forecast the class whose label is unknown, one must first identify a set of models that describe and differentiate various concepts and data classes [14]. Different classification techniques exist, including rule-based, decision tree, Bayesian network, neural network. A rule-based classification extracts a set of rules that show relationships between attributes of the dataset and the class label. For this study, three algorithms; namely J48, PART, and JRip were used. J48 decision trees are used in the classification and prediction. It is a simple and powerful way of representing knowledge. PART is a rule-based classifier that uses a set of IF-THEN rules for classification. An IF-THEN rule is an expression of the form IF condition THEN conclusion. JRip is also a rule-based classifier that uses a set of IF-THEN rules for classification; this experiment was conducted with default parameters of WEKA, and the algorithm generates a model with rules.

2.5 Clustering Methods

According to [1], clustering is a description model that entails classifying a collection of physical objects into interrelated groups. This dataset consists of items that are distinct from those in other clusters but related to each other within the same cluster. Finding homogenous group objects based on the values of their attributes is the goal of this descriptive task. Items within a cluster are arranged to be highly like each other but differ from items in other clusters. This is how item clusters are created. To look for distinct groupings within a data set, clustering is frequently used. In the process of cluster analysis used in data mining, k-means clustering and farthest-first algorithms are used. The k-means clustering algorithm seeks to divide n observations into k clusters, each of which is allocated to the cluster with the closest mean [7]. A splitting algorithm partitions the dataset D, which contains n objects, into k partitions ($k \leq n$), each of which represents a cluster. It is the intention of the clusters to improve an objective splitting principle, like a distance-based dissimilarity function. The distance between the selected centroid and the remaining data elements is then calculated using the Euclidean distance formula, and the process is repeated repeatedly. A modified k-means clustering called ”farthest first” places each cluster center successively at the point farthest from the current cluster center. This algorithm chooses k points as the cluster centers. The initial center is taken at random. As the point that is farthest from the first, the second center is randomly chosen. The cluster with the closest center is the one to which the outstanding points are added.

2.6 Knowledge Acquisition from Domain Experts

Various strategies have been employed to comprehend the primary issue of the study. To determine the study's problem, a domain expert interview, phenomenon observation, a review of related literature, document analysis, and a review of past researchers' recommendations are conducted. The researcher interviews domain experts to grasp the type of diabetes disease that is frequently occurring in Ethiopia and the relevant qualities utilized to diagnose the types of diabetes. Interviewees were asked to detail any factors (lab results and patient symptoms) that were employed in the diagnosis and management of various forms of diabetes during the interviews. On the other hand, the data mining approach is particularly fruitful in automating the knowledge acquisition task of rule and case-based system. The development of an efficient and effective intelligent knowledge-based system involves the development of an efficient knowledge base that has to be complete, coherent, and non-redundant [15].

Knowledge acquisition is the most difficult and error-prone task in the development of KBS since knowledge acquisition involves communication between people with completely diverse backgrounds, human experts, and knowledge engineers, who must formulate the concepts, relations and control mechanisms needed for the knowledge-based system. Knowledge acquisition using data mining techniques eliminates or reduces the difficulty caused by the knowledge acquisition bottleneck of rule-based systems and automates knowledge acquisition by obtaining low cost, adding better values to traditional methods, and producing a high-quality knowledge base.

2.7 System Architecture

Architecture functions as a blueprint throughout the creation and justification of knowledge bases, showing how the components of prototype knowledge bases interact and relate to one another [16]. It gives instructions for someone (the audience) to understand the beginnings and ends of the system. It also explains the relationships and interactions between the subsystems, how the system works, and what parts it should have. Just by looking at the building, one may comprehend how the researcher finished their task. Thus, the accompanying Figure 1 depicts the general architecture of an intelligent knowledge-based system for diagnosis and therapy.

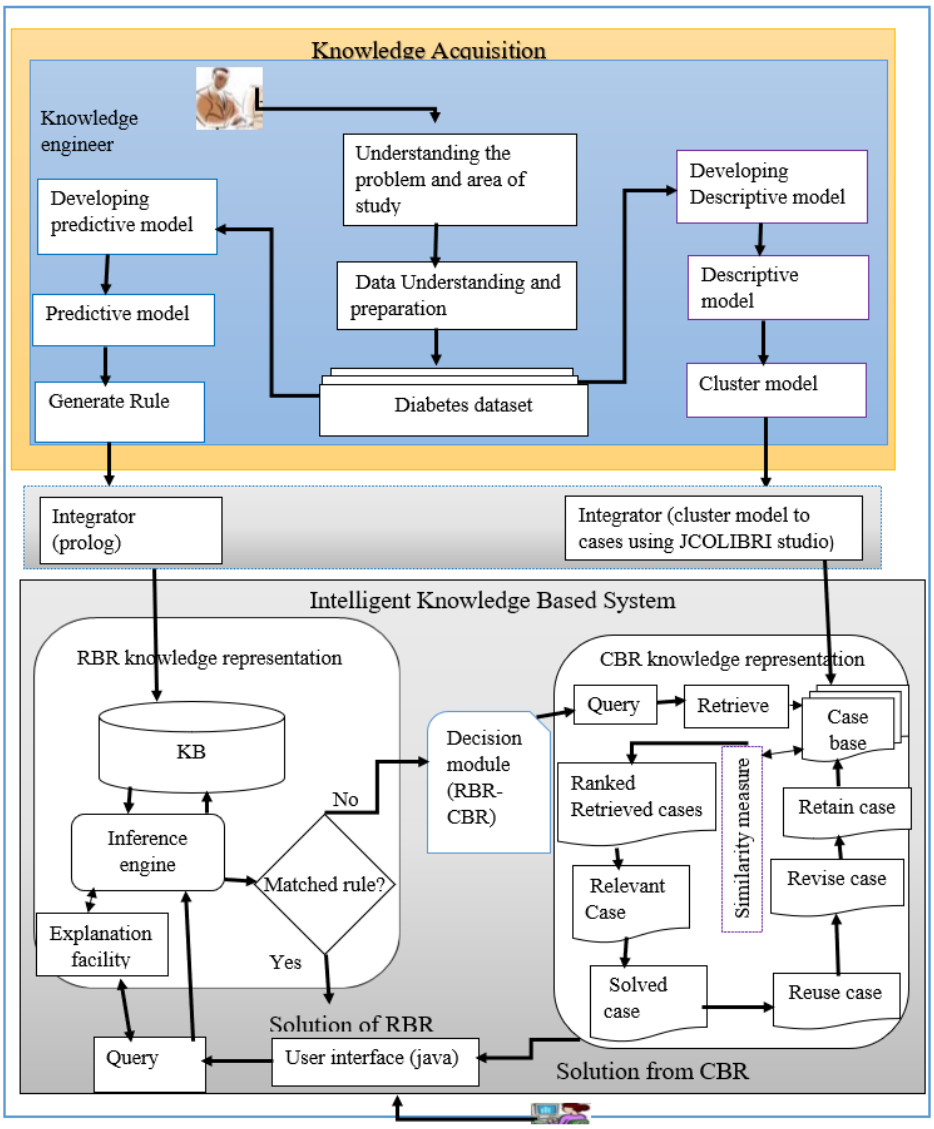

Fig 1. Architecture of the proposed system

2.8 KBS Development Tools

This work aims to construct an intelligent knowledge-based system for diabetes diagnosis and treatment by merging CBR and RBR. After the RBR and CBR systems were developed independently, a test was run to see which system should be utilized first for diabetes diagnosis and treatment. The researcher used SWI prolog 9.04 to create the automatic integration of the prediction model to the RBR module of the intelligent KBS. This was accomplished by translating the JRip rules' usual format into prolog formalism. Based on its objective evaluation criteria, JRip was chosen as the best algorithm among the tested classification algorithms for creating a prediction model. IF-Then rules are generated by the JRip algorithm. The JRip rules' conclusion is found in the THEN component, whereas the IF part contains the rules' condition. As a domain expert evaluation, JRip extracted 19 rules, all of which can diagnose the type of diabetes. One of the study's other specific goals is to automatically combine an intelligent KBS system's CBR module with a descriptive data mining model. To automatically incorporate examples retrieved from the clustering technique to CBR, is being attempted. Every extracted case was automatically saved as a plaintext file to represent the cases.

Because the dataset cases are saved in plaintext file format in the COLIBRI Studio application, the researcher employed a plaintext connector to implement case-based diabetes diagnosis. In this connector, the researcher has provided the path to the case structure as well as the route to a plain text file. Using rule-based and case-based knowledge representation techniques, respectively, the researcher saves all the knowledge gathered via data mining by verifying domain experts in the knowledge base as a set of rules and cases.

2.9 Knowledge Representation

Production rules, frame, and network representations are the most often utilized techniques for knowledge representation. Knowledge gathered from specialists and other sources needs to be arranged such that conclusions can be drawn from it whenever needed by a computer inference program. An example rule produced from the diabetes dataset using JRip is:

Rule 1: Type I Diabetes (814.0/0.0) is present if the following conditions are met: FBS = impaired, polydipsia = yes, pregnant time (week) = no, age = young, BMI = underweight, and pedigree = no.

Rule 2: Type I Diabetes (16.0/2.0) if FBS = impaired, polydipsia = no, pregnancy time (week) = yes, age = young, BMI = underweight, pedigree = no, sex = f, blood pressure = prestage, and kidney disease = yes.

Rule 3: If the following conditions are met: age = child; blood pressure = prestage; kidney disease = no; polyuria = yes; visual complaints = yes; physical activity = no; FBS = impaired; polydipsia = no; pregnant time (week) = no; and BMI = underweight.

Rule 4: If the following conditions are met: blood pressure is prestage, kidney disease is present, polyuria is present, visual complaints are present, physical activity is not present, history of alcohol use is present, history of smoking is present, pregnant time (week) is not present, age is young, BMI is underweight, pedigree is not present, and sex is not present, then type I diabetes (1444.0/0.0).

Rule 5: If the following conditions are met: blood pressure is prestage, kidney disease is present, polyuria is present, visual complaints are present, physical activity is not present, history of alcohol use is not present, history of smoking is not present, polyphagia is present, pregnant time (week) is present, age is young, BMI is underweight, pedigree is not present, and sex is not present, then type I diabetes (612.0/2.0).

Rule 6: If the following criteria are met: blood pressure = stage 1, kidney disease = no, polyphagia = yes, pregnancy time (week) = yes, BMI = overweight, pedigree = yes, sex = f, and FBS = high, then Type II Diabetes (1231.0/2.0).

2.10 Integration of RBR and CBR

Java Eclipse was used to bring CBR and RBR together as a combined reasoning process. An intelligent KBS is used to diagnose and cure diabetes by asking a series of questions to its users in order to establish communication. Users are prompted with queries by the system that has useful characteristics. Questions that users respond to are likely to be the source from which the RBR system gets its early insights. Therefore, in cases where a remedy is available, the RBR diagnoses the various forms of diabetes and makes recommendations for its diagnosis. GUI development increases the usability of the system by end users, which is one of the evaluation metrics. Developing a graphical interface is crucial to display the results of a knowledge-based system without overloading users. Developing a suitable user interface has a significant impact on increasing the doctor's well-being in entering and accessing data during the medical diagnosis process. The users are expected to answer the questions, based on which the RBR system starts with the first reasoning. The query is sent to the CBR system if the RBR is unable to diagnose the types of diabetes that users inquire about throughout the diagnosis procedure.

If users are unfamiliar with any of the attributes, an explanation facility is utilized to identify the attribute and explain why it is required for the diagnosis of diabetes. After reading a description of each attribute by clicking the "what" button, users can click the ”why” button to learn why that attribute is necessary. Figure 2 shows the situation where the users cannot know what the attribute of body mass index (BMI) means. So, when he/she clicks on “what” button, he/she gets a description of that attribute and clicks on “why” button then the system displays why this attribute was required for the diagnosis of diabetes, as shown in Figure 2.

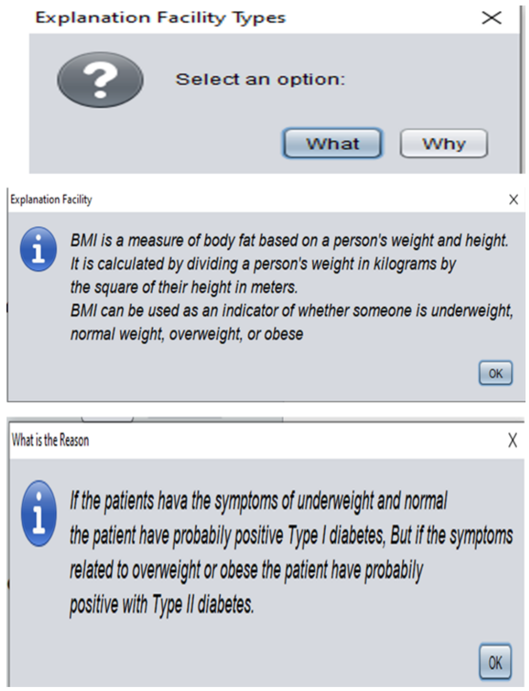

Fig 2. Explanation Facilities Description

To obtain the diagnostic result, users can select their response from the answer choices in the combo box or respond to each characteristic (question description) with a simple ”yes” or ”no.” The precise outcome of the user’s interest-based Type I diabetes diagnosis is displayed in Figure 3 below by the RBR user interface. If the users want to select other symptoms to diagnose other types of diabetes, they can click the Clear button to start again. To quiet the system, the users can click on the “Exit” button. After the user’s diagnoses the types of diabetes and displays the result of diabetes as “Type I diabetes” of the below diagnostic result, the diagnosis of the RBR user interface automatically accepts the diagnosis result to retrieve the recommendation as treatment.

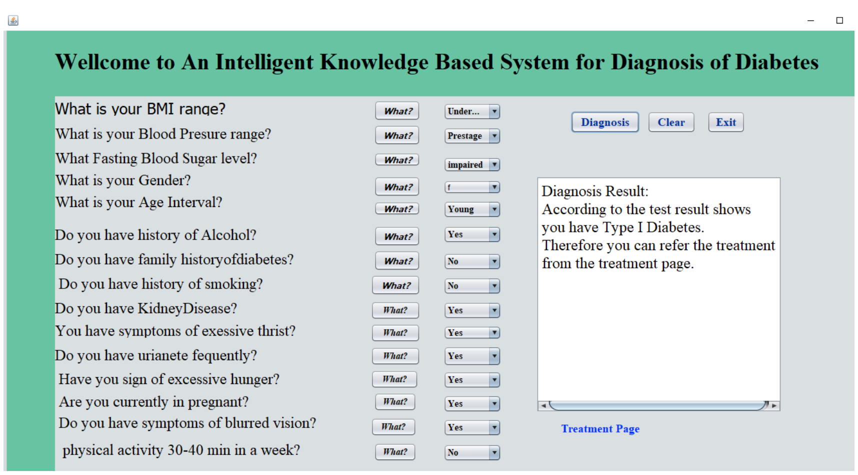

Fig 3. RBR diagnosis result user interface

To obtain the desired outcome, choose the attribute values from the combo box and click the diagnosis button. The user's question will be sent to CBR if it is not answered by the RBR system. Cases can be retrieved, reused/adapted, revised, and retained by the CBR system. The symptoms of a specific kind of diabetes cannot be accurately detected by the RBR system, as illustrated in Figure 4.

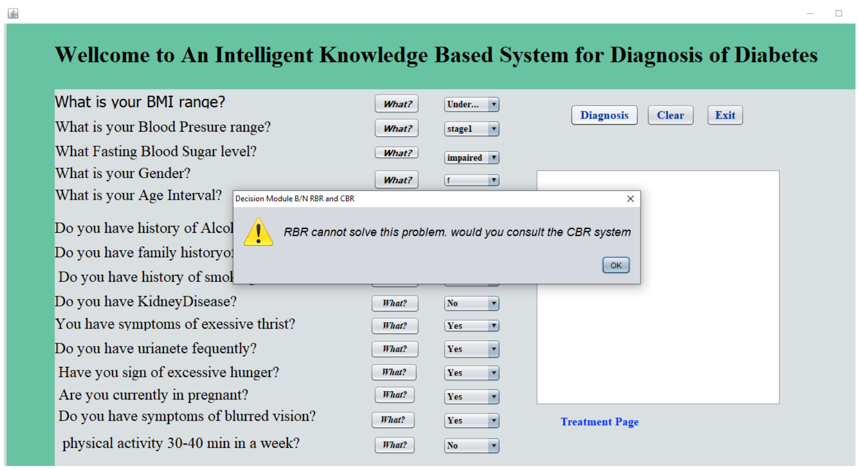

Fig 4. RBR is unable to solve problems and invites CBR to solve problems

As a result, the notice "RBR cannot solve this problem" was shown on the RBR system. Would you check with the CBR system first? In this instance, users can configure the query in the system to diagnose and treat diverse types of diabetes. The system transmits the query from RBR to CBR and automatically opens the CBR query dialog interface when users click on the controller that indicates "OK."

2.11 Model Evaluation

Understanding the findings, determining whether the knowledge that was extracted is novel and fascinating, having domain experts analyze the findings, and assessing the usefulness of the knowledge that was found in the field (problem area) are all important components of the evaluation of knowledge that has been discovered. To assess the prediction model as well as the descriptive model and determine whether the new information should be applied to address the identified issue, the researcher employed both objective and subjective evaluation techniques. Subjective evaluation gauges the qualitative elements and practicality, whereas objective evaluation gauges the quantitative result. In the end, the algorithms that met both evaluation criteria were chosen to develop a clever, knowledge-based diabetes diagnosis and treatment system. Therefore, the JRip and k-means algorithms were chosen since they registered the best accuracy compared to other algorithms.

3 Results

To create prediction models, three trials are conducted in total. The prepared dataset, which is used in all tests, consists of 6086 instances with 15 attributes and one class label. Additionally, after conducting several tests, each classifier algorithm's default parameter value is considered because it allows for higher accuracy than changing the default parameter values. To extract interesting rules from the diabetes dataset, three predictive models involving the J48, PART, and JRip classification algorithms are purposefully chosen at this point. PART and JRip are rule-based classifiers, whereas J48 is a tree-based classifier.

Table 1. Comparison of classification algorithms

Classifier	Correctly classified instances		Incorrectly classified instances		Time takes to build the model
	In No.	percentage	In No.	percentage	
J48	5255	99.75%	13	0.25%	0.14 seconds
PART	5258	99.80%	10	0.20%	0.03 seconds
JRip	5263	99.90%	5	0.10%	0.41 seconds

Table 1 illustrates how the model created with the J48 classifier has a tree with a size of 28 and 19 leaves.After constructing the model in 0.14 seconds, the J48 algorithm properly identified 5255 examples and wrongly classified 13 instances, yielding accuracy rates of 99.75% and 0.25%, respectively. The PART rule induction approach took 0.03 seconds to generate the model, and it successfully classified 5258 cases and wrongly identified 10 instances, yielding an overall accuracy of 99.80% and 0.20%, respectively. In contrast, the JRip algorithm took 0.41 seconds to build the model, but it properly recognized 5263 occurrences and mistakenly classified only 5 instances, yielding accuracy rates of 99.90% and 0.10%. Among all classification algorithms, the JRip algorithm has the highest accuracy record. This classifier uses the 10-fold cross-validation test option and applies propositional IF-THEN rules from the dataset with default values of the parameter. The JRip method correctly classified 5263, or 99.90%, of the 5268 occurrences; the other 5 instances, or 0.10%, were wrongly classified, resulting in a 0.41-second model building time. Therefore, the JRIP algorithm is selected because of registering highest accuracy than other algorithms. The chosen models are automatically connected to the Colibri studio via the rule base from the predictive model (programmed with SWI Prolog).

To better illustrate the study's findings, three key visualizations have been included to compare the performance of the algorithms and enhance the clarity of the results. Figure 5 presents a bar chart comparing the accuracy, precision, and recall of the three classification algorithms: J48, PART, and JRip.

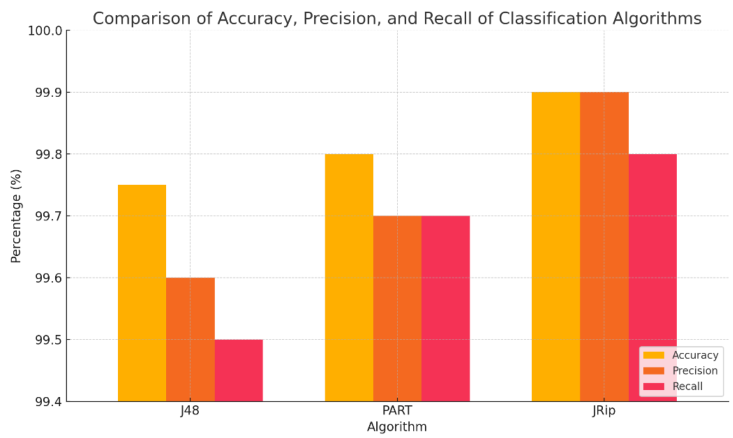

Fig 5. Bar chart comparing the three classification algorithms

The bar chart shows distinctly that the JRip algorithm achieves the highest performance across all metrics, with an accuracy of 99.9%, precision of 99.9%, and recall of 99.8%, making it the most effective choice among the evaluated models. This graphical comparison highlights the superior ability of the JRip classifier in correctly diagnosing diabetes cases while maintaining minimal misclassification errors.

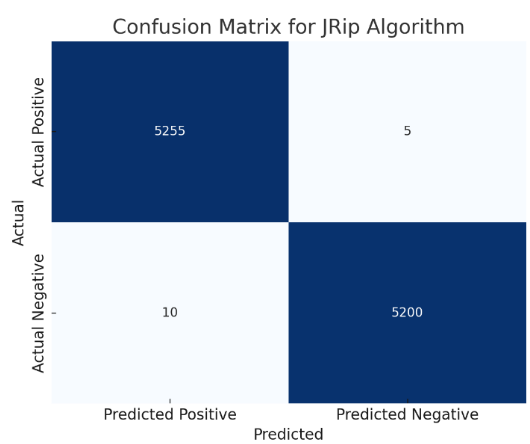

Fig 6. Confusion matrix heatmap

In Figure 6, a confusion matrix heatmap is included to visualize the classification performance of the JRip algorithm in more detail. The confusion matrix shows the distribution of true positives, false positives, true negatives, and false negatives. It reveals that the JRip model correctly classified most instances with a negligible number of misclassifications, providing a clearer understanding of the model's accuracy in distinguishing between diabetes types.

As previously stated, the two clustering methods that are specifically chosen to construct the descriptive model are the farthest first and k-means. The cluster size, or the number of clusters, is two by default. This is achieved by using all other default parameters for each method and by not changing the number of clusters. Apart from the class-to-cluster test mode, two algorithms are evaluated to determine their accuracy in matching the assigned class with the instance.

Table 2. Cluster algorithm with each performance accuracy

Cluster algorithm	Correctly clustered instances		Incorrectly clustered instances		Time takes to build the model
	In No.	Percentage	In No.	Percentage	
K-means	4545	86.3%	723	13.7%	0.13 Seconds
Farthest first	4038	76.7%	1230	23.3%	0.02 Seconds

Table 2 above illustrates how the K-means clustering approach, which takes 0.13 seconds to create the model, correctly clusters 4545 out of 5268 trained instances to their class and mistakenly clusters 723 instances to the other class with an accuracy of 86.3% and 13.7%, respectively. In contrast, the Farthest First technique took 0.02 seconds to create the model and properly classified 4038 out of the 5268 trained instances, while wrongly clustering 1230 instances to another class with anaccuracy of 76.7% and 23.3%, respectively. The outcomes of the two cluster algorithms were different. The K-means algorithms' output was chosen because it showed superior clustering accuracy than the farthest first algorithm. The chosen models are automatically connected to the Colibri studio to construct case base reasoning.

Additionally, Figure 7 compares the clustering accuracy of the K-means and Farthest First algorithms. The bar chart shows the percentage of correctly and incorrectly clustered instances for each algorithm. The visualization demonstrates that K-means achieves an accuracy of 86.3%, significantly outperforming Farthest First, which achieved only 76.7%. This indicates that K-means is more dependable in grouping the dataset into relevant clusters based on the attributes of diabetes cases.

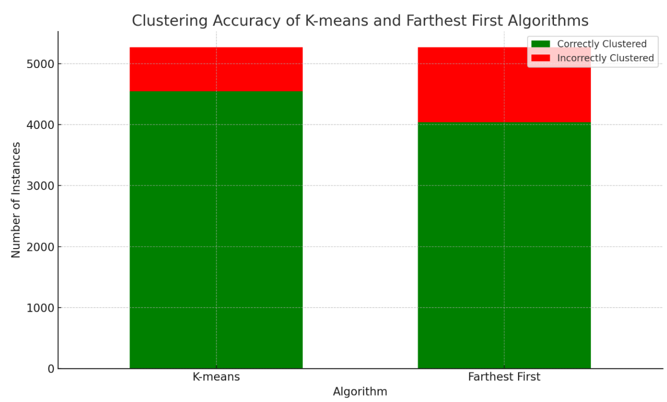

Fig 7. Clustering accuracy of the K-means and Farthest First algorithms

4 Discussion

The datasets used in the tests were created with WEKA software version 3.8.6. To achieve this goal of creating an intelligent knowledge-based system for diabetes diagnosis and treatment, classification algorithms such as cluster, rule, and tree-based classifiers are selected. Furthermore, after running a series of trials, the default value of parameters is taken into consideration for each classifier algorithm, since it allows them to obtain higher accuracy compared to adjusting the default parameter values.

The three classification algorithms (J48, PART, and JRiP) were chosen by the researcher because they are simple to apply and allow for easy interpretation of the results. Among all classification algorithms, the JRip algorithm has the highest accuracy record. The JRip method correctly classified 5263, or 99.90%, of the 5268 occurrences; the other 5 instances, or 0.10%, were wrongly classified, resulting in a 0.41-second model building time. To increase the prediction model's effectiveness, the researcher assesses the model's performance by considering its precision, recall, F-measure, and true positive rate value. With values of 99.9%, 99.7%, 99.8%, and 99.9%, respectively, the JRip algorithm therefore continuously achieves high values for true positive rate, precision, f-measure, and recall, indicating that the model has a good ability to correctly detect positive cases.

To construct the descriptive model, the researcher has also deliberately selected two clustering algorithms: k-means and farthest first. All other default settings for each algorithm are used, and the cluster size (number of a cluster) has two by default. This number of clusters cannot be modified. The algorithms' output is assessed according to how well the dataset's examples are clustered into the Type I and Type II diabetes classes. The K-means clustering algorithm correctly clustered 4545 out of 5268 trained instances to their class, and 723 instances incorrectly clustered to the other class with an accuracy of 86.3% and 13.7%, respectively. In contrast, the farthest first algorithm correctly clustered 4038 instances out of the total number of 5268 trained instances, and incorrectly clustered 1230 instances to the other class with an accuracy of 76.7% and 23.3%, respectively. The researcher chooses the K-means approach to construct a model since it produces results with better clustering accuracy than the farthest first algorithm.

The chosen models are automatically connected to the Colibri studio via the rule base from the predictive model (programmed with SWI Prolog) and the case base from the descriptive model (using the descriptive model). The researcher built independent prototypes for both the RBR and CBR, and conducted experiments to verify that each module can independently diagnose different forms of diabetes and offer treatment recommendations. We also conduct experiments to determine the optimal method for combining RBR and CBR to create a sophisticated knowledge-based system for diabetes type identification and treatment.

The rule-based approach—that is, RBR followed by CBR—was chosen by the researcher based on the trial results to incorporate both thinking mechanisms. The two reasoning mechanisms, CBR and RBR, were integrated using Java Eclipse. An intelligent KBS is used to diagnose and cure diabetes by asking a series of questions to its users to establish communication. Users are prompted with queries by the system that have useful characteristics. 95.23% system performance and 93.2% user approval are attained by the suggested system when compared with previous works such as[1] conducted An Integration of Prediction Model with the KBS for Diagnosis and Treatment of Diabetes which registered 92% user acceptance testing and[2] investigated KBS Using Data Mining Techniques for Diagnosis and Treatment of Diabetes and registered 91.43% user acceptance testing because of using only RBR reasoning system. The dataset used by[1] is only about 3112 instances with 16 attributes and five class labels and all the dataset are involved in all experiments when compared with us as used 6086 instances with two classes. As datasets increase the quality of experimental results also increases, so this is the primary factor for registering a favorable outcome of our investigation. We conducted an intelligent knowledge-based system for diagnosis and treatment of diabetes which incorporates both reasoning mechanisms which are RBR and CBR. Because of using automatic knowledge acquisition techniques with more dataset and using both reasoning mechanism, we get a promising result when compared with other researchers such as[1][17] who used only RBR reasoning approach. Usually, the system automatically integrates into its knowledge base utilizing data mining approaches, in addition to merging two reasoning mechanisms, which is why a beneficial outcome is realized.

According to the testing results, the intelligent knowledge-based system that has been built is beneficial for aiding medical professionals in the diagnosis and treatment of diabetes. Nevertheless, this study has a restriction in that it compares the results and chooses the most pertinent solution to address the specific difficulties by receiving and sending the question simultaneously for both justifications. Only common forms of diabetes, such as type I and type II diabetes, were covered in the conducted study. To develop a generic intelligent knowledge base system, it will be necessary to incorporate the other forms of diabetes into the system that is now being suggested.

This study presents a novel intelligent knowledge-based system (IKBS) specifically designed for the diagnosis and treatment of diabetes. What sets this system apart is its unique integration of both rule-based reasoning (RBR) and case-based reasoning (CBR), creating a dual reasoning framework that enhances the adaptability and effectiveness of the system. Unlike previous approaches that rely on traditional rule-based systems, which can often be rigid and limited in scope, this innovative framework allows the IKBS to effectively manage a broader range of cases. This includes those complex scenarios that do not conform to predefined rules, thereby providing a more comprehensive solution for diabetes management.

In addition to its dual reasoning capabilities, this research employs advanced data mining techniques for automatic knowledge acquisition. This approach effectively addresses a common challenge faced by many healthcare systems: reliance on manual expert input, which can be time-consuming and prone to human error. By automating the knowledge acquisition process, the system not only streamlines the development of its knowledge base but also ensures that it remains current and relevant in the face of evolving medical knowledge.

To achieve its impressive performance metrics, the system combines the JRip algorithm for rule-based reasoning with K-means clustering for case-based reasoning. This strategic combination results in an accuracy rate of 95.23%, demonstrating the system's ability to make precise and reliable diagnoses. Furthermore, the user acceptance rate stands at an impressive 93.2%, indicating that healthcare professionals are not only able to utilize the system effectively but also find it beneficial in their practice. These performance metrics surpass those of existing models, highlighting the superiority of this approach in the field of diabetes management.

The integration of dual reasoning, automated knowledge acquisition, and a contextual design specifically tailored for Ethiopia's resource-limited healthcare environment collectively underscores the system's novelty. This thoughtful design consideration is crucial, as it ensures that the IKBS is not only technically advanced but also practical and applicable in real-world settings where resources may be limited. The potential for significant practical impact is evident, as this system could enhance the quality of diabetes care in such environments, leading to better health outcomes for patients. By bridging the gap between advanced technology and the realities of healthcare delivery in resource-limited settings, this research paves the way for future innovations in intelligent healthcare systems.

5 Conclusion

In this work, rule-based and case-based reasoning have been used to build and construct an intelligent knowledge-based system for diabetes diagnosis and management. Data mining techniques and an algorithm for building descriptive and predictive models were used to automate the knowledge gathering operation. A sample of the diabetes dataset was obtained for experimental purposes from the following hospitals: Haramaya University Hiwot Fana Specialized Hospital, Haramaya General Hospital, and Harari General Hospital. Then, to prepare the dataset for data mining, preprocessing is applied. To create predictive models, data mining algorithms like PART, J48, and JRip are being evaluated; to create descriptive models, K-means clustering and farthest-first algorithms have been evaluated. Ultimately, based on their respective best accuracy of 99.91% and 86.3%, the researcher decided to incorporate the results of the JRip classification algorithm and the K-means clustering algorithm with the intelligent knowledge-based system.

Rule-based and case-based reasoning systems are automatically constructed using the best predictive and descriptive models. The RBR system initially helps the user's query to identify an exact match between the query and the rule in the rule base based on the best result found in the produced IKBS. If a precise match is detected, the algorithm makes conclusions and provides diagnostic results to consumers. The treatment page automatically determines the type of diabetes and displays recommendations based on the patient's condition if the RBR system accurately diagnoses the condition.

On the other hand, the user inquiry is immediately forwarded to the CBR query interface if the rule-based logic for the query returns no results. Because the CBR system utilizes a similarity metric by default to provide potential diagnosis outcomes. The CBR system then uses similarity measures to find relevant cases, reuses them so domain experts can apply the case solution to the current case, validates the solution, and keeps them until the next step to determine whether the knowledge gained from solving a new case is significant enough to be added to the system for use in solving other problems in the future.

The prototype system's total performance shows 95.23% accuracy with average precision and recall of 97.7% and 97.6%, respectively. Additionally, the user acceptance testing reveals an average of 93%. The outcome of the trial indicates that the intelligent knowledge-based system that was built is beneficial for helping medical professionals diagnose and treat diabetes. Due to the availability of their dataset and necessary attributes in the hospitals, the study only examined type I and type II diabetes. Furthermore, Ethiopians aged 20 to 79 are more likely to develop those forms of diabetes.

It was determined that the study's positive findings may be applied to address issues with diabetes diagnosis and treatment. Nevertheless, this study has a flaw in that it compares the results and chooses the most pertinent answer to address the problem after receiving and sending the question simultaneously for both arguments. As a result, the following suggestions are for additional research:

- The research done was restricted to prevalent forms of diabetes, specifically type I and type II diabetes. Therefore, it will be necessary to incorporate the remaining forms of diabetes into the suggested system in the future to create a generic intelligent knowledge base system.
- The next challenge is creating a knowledge-based system that processes queries using both Case-Based Reasoning and Rule-Based Reasoning at the same time, which improves the system's adaptability and flexibility.
- A concern raised by subject matter specialists is the ignorance of diabetic patients. Therefore, future work will also focus on creating an intelligent knowledge base system for diabetes awareness and self-care assistance utilizing machine learning.
- The next challenge is to create a mobile-based knowledge base system that uses machine learning to diagnose and cure diabetes. Such a system can help people manage their diabetes by offering useful information, tailored suggestions, and

support.

References

1) Tilahun B. An Integration of Prediction Model with the Knowledge Based System for Diagnosis and Treatment of Diabetes: In Case of Debre Berhan Referral Hospital. 2018.
2) Kedir E, Worku J, Takele T. Developing a Prototype Knowledge-Based System for Diagnosis and Treatment of Diabetes Using Data Mining Techniques. *Ethiopian Journal of Health Sciences*. 2020;30(1). Available from: https://doi.org/10.4314/ejhs.v30i1.15.
3) Eyasu K. Developing Knowledge Based System Using Data Mining Techniques for Diagnosis and Treatment of Diabetes. 2018. Available from: https://repository.ju.edu.et/bitstream/handle/123456789/4590/Developing%20KBS%20Using%20Data%20Mining%20Techniques%20for%20Diagnosis%20and%20Treatment.pdf?sequence=1&isAllowed=y.
4) Salih NK, Elbashier H, Zang T, Eshtiag AAE. Self-Diagnosis of Diabetes Using CBR Algorithm. *J Comput Sci Syst Biol*. 2021;11(4):235–239. Available from: https://www.hilarispublisher.com/open-access/selfdiagnosis-of-diabetes-using-cbr-algorithm-jcsb-1000279.pdf.
5) Wendimu D, Biredagn K. Developing a knowledge-based system for diagnosis and treatment recommendation of neonatal diseases. *Cogent Engineering*. 2022;10(1). Available from: https://doi.org/10.1080/23311916.2022.2153567.
6) Birhane T, Teferi D, Mohammed T. An integrated knowledge-based system for early detection of eye refractive error using data mining. *SINET: Ethiopian Journal of Science*. 2021;44(2):223–233. Available from: https://www.ajol.info/index.php/sinet/article/view/219306.
7) Indebu I. Developing a hybrid knowledge based system for tuberculosis disease diagnosis. Haramaya, Ethiopia. 2021.
8) Koca G. An Intelligent System for Type II Diabetes Mellitus Diagnostic. Regina, Saskatchewan. 2020. Available from: https://www.proquest.com/openview/f94a6161ba73c10757e64576d375dd9b/1?pq-origsite=gscholar&cbl=18750&diss=y.
9) Kiragu MK, Waiganjo PW. Case based Reasoning for Treatment and Management of Diabetes. *International Journal of Computer Applications*. 2016;145(4):20–29. Available from: https://www.ijcaonline.org/archives/volume145/number4/25266-2016910552/.
10) Nguyen A, Gardner L, Sheridan D. Towards Ontology-Based Design Science Research for Knowledge Accumulation and Evolution. In: and others, editor. Proceedings of the 52nd Hawaii International Conference on System Sciences;vol. 6. 2019;p. 5755–5764. Available from: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/c140f10d-e84a-4efc-889f-0f7d7b9b2e65/content.
11) Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*. 2007;24(3):45–77. Available from: https://doi.org/10.2753/MIS0742-1222240302.
12) Mekonnen Y. Information Security Management Framework for Effective Implementation of Integrated Financial Management Information System (IFMIS): The Case of MoF. Addis Ababa, Ethiopia. 2020. Available from: https://etd.aau.edu.et/server/api/core/bitstreams/aa591c9d-b451-4afc-ba1a-f7d9454581ac/content.
13) Adugnaw B. Knowledge based system for diagnosis of under nutrition status of children under five. Bahir Dar, Ethiopia. 2019. Available from: http://ir.bdu.edu.et/handle/123456789/10344.
14) Nadjette D, Rayene A. Rules-based decision support system and domain ontology for diabetes diagnosis. *30es Journées Francophones d'Ingénierie des Connaissances, IC 2019, AFIA*. 2019;p. 224–230. Available from: https://hal.science/hal-02329609/document.
15) Embabo A, Jimma W, Diriba C. A Self-Learning Knowledge Based System to Mitigate the Impact of Drug Interactions in Type Two Diabetes Mellitus. 2022;10(3):241–250. Available from: https://www.hrpub.org/journals/article_info.php?aid=12153.
16) Hailu B, Meshesha M, Mohamed T. Developing A Bilingual Self-Learning Medical Knowledge Based System For Diagonoses, And Treatment Of Malnutrition Caused Diseases For Children. 2022. Available from: http://ir.haramaya.edu.et//hru/handle/123456789/4771.
17) Eyasu K, Jimma W, Tadesse T. Developing a Prototype Knowledge-Based System for Diagnosis and Treatment of Diabetes Using Data Mining Techniques. *Ethiop J Health Sci*. 2020;30(1):115–124. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC7036459/pdf/EJHS3001-0115.pdf.