Received: 8 October 2021 | Revised: 8 January 2022 | Accepted: 14 January 2022

DOI: 10.1002/ima.22710

RESEARCH ARTICLE

WILEY

An intelligent fuzzy inference rule-based expert recommendation system for predictive diabetes diagnosis

Palanigurupackiam Nagaraj | **Perumalsamy Deepalakshmi**

School of Computing, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, India

Correspondence
Palanigurupackiam Nagaraj, School of Computing, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, India.
Email: nagaraj.p@klu.ac.in

Abstract

Diabetes is one of the most common and hazardous diseases, which can affect almost every organ in the body. Diagnosis of diabetes requires determining all vital parameters related to the disease. However, the nature of the data from those parameters is very uncertain, affecting the process of disease diagnosis. This article proposes an intelligent fuzzy inference rule-based predictive diabetes diagnosis model (IFIR_PDDM), providing content recommendations to patients with diabetes. The suggested model employs an inference technique that medical specialists have validated for recommendations. IFIR_PDDM comprises three elements used to forecast the risk of diabetes disease. Initially, a fuzzy membership function utilizes medical recommendations and statistical methodologies. Medical specialists then validate the mining-based rules using a decision tree rule induction technique. The proposed model predicts the risk of diabetes disease using fuzzy inference based on Mamdani's technique. Based on this information, the recommendations for a normal life, nutrition, exercise, and medications are given to patients. We used an electronic health record (EHR) medical and clinical dataset from the PIMA Indian Diabetes dataset to develop our proposed model and assess its performance. The proposed model takes less time for diabetes diagnosis, and the expert recommendation system uses the fuzzy inference method.

KEYWORDS

decision tree, diabetes, expert recommendation system, fuzzy inference system, fuzzy logic, IFIR_PDDM

1 | INTRODUCTION

Diabetes is a comorbidity of chronic obstructive pulmonary diseases. Illness and death due to diabetes mellitus are continuously recorded in all countries.[1] Diabetes is a silent, harmful, symptomless, theory-less, and experience-based sickness. This is more complicated by its appearance's inaccuracy, ambiguity, and doubtfulness. Patients can suffer from diabetes for a long time without knowing it. An initial finding of the disease permits doctors to recommend specific therapy plans. The effectiveness of diabetes mellitus treatment is primarily determined by the accuracy and timing of diagnosis. As a result of specific therapy plans, the mortality rate can be reduced. According to the National Diabetes Statistical Report, 34.2 million people are affected by diabetes worldwide.[2] As per the International Diabetes Federation report released in 2019, 463 million adults are living with diabetes; and by the year 2045, this will rise to 700 million; more than 1.1 million children and adolescents are living with diabetes; more than 20 million live births are affected during pregnancy, and 374 million people are at an increased risk of developing diabetes.[3]

Diabetes diagnosis is a complicated procedure and needs the assistance of medical experts to create a conclusion based on the complete case profile. Doctors have to collect the signatures of sickness, including past medical histories, syndromes, family histories, associated complications, physical tests, and surveys such as lab tests. There is a loss of experts on diabetes mellitus, and most of the time is spent studying the cases' histories in a detailed manner. In hospitals, it is necessary to update the diabetes diagnostic methods in terms of the periods of case exams. An automatic clinical decision support system (CDSS) can help nonprofessional doctors collect and analyze the profile of the allocated case from dissimilar electronic health record (EHR), which can be obtained from different hospitals, and demand is added for an exact and timely diabetes diagnosis. It is reduced to diabetes complications, hospitals, illness, and death. For stand-alone CDSS, doctors needed to collect and enter the appearance for each case. It provides the right knowledge about the proper form at the right time. CDSS can be unified with the EHR ecosystem. Further, the results of CDSS can be viewed in a mobile health location to continuously monitor and update the patient's profile.[4]

Diabetes identifies the nature of an illness or other problem by examining the symptoms. The act of thinking (a series of events producing a result regarding something ongoing) is represented as $d = (t_1, t_2, t_3, t_4, ..., t_n)$, where n is considered as the most recent or up-to-date. As seen in the past, having an easy-going mood or temperament, an environment or situation, where $t_1, t_2, t_3, t_4, ..., t_a$ are symptoms, $t_{a+1}, ..., t_b$ are physical examinations, $t_{b+1}, ..., t_c$ are lab tests, $t_{c+1}, ..., t_d$ are complications, $t_{d+1}, ..., t_e$ are drugs, and $t_{e+1}, ..., t_n$ are the census, and $a + b + c + d + e = n$ are facial characteristics to represent or embody, indicate a quantity. In addition to quantitative assessment, the complete measure of the degree of diabetes accounts to a large number.

A single category of people or people having common characteristics and statistics collected together for reference or analysis toward diabetes mellitus will be quantifiable or assessable. These parameters, facts, and statistics are collected together for reference or analysis without a break or interruption. These are therapeutically not up to the required standard or quality. As a result, the expertise of diagnosis does not trust the system results from any scheme or set of rules used to classify, explain, or calculate increases in amount or intensity, even though the greater, or greatest, number from a given group or collection are ideal examples of someone or something in a successful manner. To reach a given destination with a certain accuracy, it provides a faithful representation of someone or something. Formalizing something that increases the amount or number of the position arrived at after considering chronic infections is a complex problem that is difficult to analyze or understand. The reason for distress or a problem where the estimate or conclusion formed by guessing was generally recognized as true by the principles or standards of behavior. The minute details or particulars of something intricate in medicinal practices make most people accepted, used, or practiced, independent of the mind. A forecast of the future course of a disease or disorder based on medical knowledge approaches calls for a new form of a fuzzy rule-based system, which has more interpretation and gives the quality or state of being very accurate compared with a computer-based learning algorithm. Adopting this fuzzy rule-based system is not guaranteed to be a special design; however, the means or procedure for doing something. In these diabetes domains, there are two main types of data: numerical and textual. A specific percentage of at least one fuzzy rule-based system model property existed maximally in thought or as an idea. Still, it did not have a physical or concrete existence in certain groups, such as "over fat people," and they relied on this straight in between the cases. Thus, the generation that has disagreed or can be carried or easily moved is not comfortably intelligent as they cannot consider medical concepts. Studies were applied to this act or process of incorporating or combining various elements into a whole. This designed fuzzy rule-based system[5] increases the value or price of topics like hybrid intelligent systems. Related studies have been developed for fuzzy and to integrate fuzzy into the theory of the universe reasonably. Logic is the decidability limitation of today's states by the fuzzy theory of the universe, tools, and reasons for all fuzzy types or classes of people or things. The latest report turns into a single mass or entity that is more or less the same throughout both old reasoning and techniques in this novel manner. This process leads to the theory of the universe's reasoning, where these fuzzy rule evaluations are present. There is no such system in the literature. We thought that the output system would explain the acceptance of the interpretative accuracy trade-off; it can be used as a matter of usual practice and emphasize the speaker's belief that what is said is true to collect patient profile information. Because it has dynamic collection features, it can be applied to collect case diagnosis-related data even from social media like Facebook and Twitter. So, while monitoring the cases, he/she can be connected to the internet and provide a Wi-Fi connection; their account can be manually collected; finally, the fuzzy rule-based system can provide personal and accurate decisions. These important challenges will be addressed in this report, which aims to build a semantic interpretation. In this fuzzy rule-based system in interpretation and semantic reasoning, the recommender system is a platform that gains information about the given data and predicts a rating and a suggestion for which item should be taken. It is frequently used in commercial applications. In this system, the patients' database results will be finalized. Based upon the finalized diabetes, the included recommender system in this article will give an appropriate suggestion to the patients. That means what they should have to take in the future to keep control of their diabetes levels, such as based upon their body mass index (BMI), blood pressure, and plasma glucose. This research presents an automated smart health observer framework for diabetes prediction and management using fuzzy inference system (FIS) algorithms and the seamless integration of diabetics' healthcare applications. This technique generates real-time forecasts based on the patient's vital signs, such as blood glucose and blood pressure, and alerts medical personnel to intervene. A medical expert provides medical providers with a holistic view of the patient's lifestyle variables, such as physical activity levels and caloric expenditure, to help them make informed diabetes treatment and management decisions. In terms of preparation, patients only need to use their healthcare gadgets, and the proposed framework may automatically extract critical data and make forecasts. This article's primary contributions are to improve diagnostic decision-making by developing an automated remote monitoring system that consolidates vital patient data from various personal health records. We integrate a fuzzy inference model into the framework and exploit a subset of the data to predict diabetes risk in real-time. We utilize fuzzy inference predictions' capacity to produce timely notifications for medical practitioners and give critical patients' health record data. The remainder of the article's structure is laid out as follows: Section 2 presents an overview of related diabetes and its recommendations for research. The research work on the EHR dataset is stated in Section 3. The fuzzy-related concepts are explained in Section 4. The methodology and the conceptual framework of the proposed intelligent fuzzy inference rule-based predictive diabetes diagnosis model (IFIR_PDDM) are discussed in Section 5. The experimental methodology to evaluate the performance of the algorithms and their findings is described in Section 6. Finally, the discussion and conclusion are offered in Section 7.

2 | REVIEW OF RESEARCH LITERATURE

Lee et al. (2008) described a perceptive agent called the "individual food recommendation proxy" based on the ontology model for hypoglycemic (diabetic) food recommendation. The agent can build a food plan tailored to a patient's lifestyle and health needs. The required data are stored in the ontology model predefined by field experts. It contains the food perspective and a set of personal food ontologies. The personal food recommendation agent includes an ontology by updating the device, the personal ontology refrain, the food fuzzy number creating apparatus, the fuzzy inference apparatus, and the real-time recommendation system. It saves the personal ontology and food records by recommending individual food items based on the FIS. An exploratory platform has been assembled to test the fulfillment of the agent.[6] Chen et al. (2012) used fuzzy logic methods and the domain ontology for antidiabetic drug selection. They presented an antidiabetic drug recommendation method based on fuzzy rules and the antidiabetic drug ontology to recommend the doses and the dose data. These technical results show that the recommended antidiabetic doses perform well for antidiabetic dosage selection. The accuracy of the intelligent ontology agent lies around the range of 65–72 as it requires knowledge of domain experts to be stored in the ontology model, which is highly uncertain due to the intervention of human decisions.[7] El-Sappagh et al. (2014) used the case-based reasoning (CBR) method. This ontology plays two vital roles: the first is a case-based ontology, and the second is a domain ontology. However, the ontology engineering literature does not support moderate guidance on forming, evaluating, and maintaining ontologies. This work described an ontology engineering technique to initiate case-bases in the medical domain. It primarily evaluates the exploration of case portraits in the form of ontologies to serve as the foundation for case semantic retrieval and upgrade all knowledge in-depth CBR methods. This work on diabetes analysis case-based has been provided to evaluate the projected system. The construction of semantic ontology will have more levels of abstraction that lead to a lack of versatility and make it difficult to apply to other applications. It is also difficult to keep track of the consistency of the logic used. The accuracy of the system can be further elevated to 95%.[8] Chen et al. (2017) developed an individual drug recommendation system for diabetic patients. It consists of fuzzy logic and an ontology treatment system. It also gives an accuracy of 80%. HbA1c levels and glucose-lowering drugs are required to satisfy the patient's difference. This system helps doctors spend less time on patients' prescriptions. This drug recommendation system performs well in the suggestion of drugs for patients. Accuracy can be improved by improving the dynamic weighting calculations of the FL-based domain ontology.[9] Alonso et al. (2017) presented empirical research, and its focus was on testing the textual interpretation of the fuzzy inferences carried out by the samples of the hypothesis. Users know the decision made by a fuzzy system by testing whether the fuzzy system makes it or not. The fuzzy inference mechanism used to carry out the hypothesis test and apply the fuzzy classifiers was built with the Generating Understandable and Accurate Fuzzy Models in a Java Environment (GUAJE) fuzzy modeling methodology. By analyzing the potential differences for groups of users, the domain experts and fuzzy experts developed the fuzzy framework to make decisions based on fuzzy systems. Even though GUAJE requires less contextual information about the data, it is computationally expensive due to the selection of various operators in its execution. With the proper formulation of the objective function, the accuracy can be further elevated from 77.71%.[10] Gupta et al. (2018) have briefly summarized the overall variety of diet recommendation systems for diabetes using fuzzy logic. Diabetes afflicted patients numbered approximately 382 million. There has been a big growth from 4.7% in 1980 to 8.5% in 2014 as per the global health organization. It has become a serious pathological state regardless of who the people are. In this work, they developed a fuzzy based diet recommendation system for detecting and classifying polygenic disorders. The article uses symbolic logic in research to give an appropriate diet recommendation system for patients placed low with polygenic disorder.[11] According to Ali et al. (2018), ontology-based recommendation systems reduce the amount of manual intervention in their work. The ontology-based approaches give better results for healthcare applications. The authors used type-2 fuzzy ontology-aided recommendation systems for IoT-based healthcare in this work. It monitors diabetes patients, and the recommendation system provides things like diets, food, and drugs. Type-2 fuzzy logic (T2FL) and the fuzzy ontology are used to predict a particular case and suggest patients. The ontology is used for recommendation systems using Web Ontology Language tools, Semantic Web Rule Language, and fuzzy logic ontology-based sentiment analysis.[12] Using fuzzy logic based associative classification methods, Rajeswari et al. (2018) found the threat factors of prediabetes. This method identifies a more factual set of remote prediabetes conditions among the data, which negatively affects diabetes patients. This method is evaluated with three measures: lift, leverage, and dependency degree. Finally, the corresponding system can find accurate threat factors like age, glucose, diabetes pedigree function, BMI, blood pressure, and the correct exact values to calculate prediabetes in an advanced way than the crisp method. Intercluster heterogeneity will be a major threat in this methodology, and some misclassification is possible due to the nature of the data. These limitations may affect the accuracy of the value.[13] Mohammed et al. (2018) introduced a T2FL-based diet recommendation system for diabetes. It is one of the usual metabolic diseases that control blood sugar quantity and avoids serious diabetic complications. It is also troublesome for doctors to manually interpret huge volumes of blood sugar data to tailor therapy to the needs of each case. A balanced diet and physical activity should be the best diabetes treatment and help several diabetes cases avoid serious complications. However, there is a need for any technique to handle the uncertainties associated with varying people's opinions and preferences.[14] Alian et al. (2018) described the epidemic of diabetes in the American Indian community. They suggested a proactive diabetes healthcare recommendation system for American Indians. In a healthy living manner, this system suggests that users oppose their diabetes using quasi-ubiquitous mobile phone AI applications. Most American Indian people select cell phones to supply smart private healthcare for American Indian cases. By combining the American Indian users' ontological profile with the public clinical diabetes recommendations system and instructions, one can customize diet recommendations systems (e.g., food intake per day and physical workouts).[15] El-Sappagh et al. (2019) have presented an individualized antidiabetic drug medication recommendation system for patients with polygenic disease. This method combines symbolic logic and metaphysical systems, which may be manipulated with the relative case and targets cheap HbA1c levels that address individual variations among patients. The system was evaluated by a specialist associate in nursing and doctors, which indicated the anti-diabetic drug medication recommendation system has sensible performance. It performs with 80% accuracy, which will assist clinicians in managing diabetes in choosing medications and, therefore, the patient's HbA1c target.[16] Gomes Filho et al. (2019) stated that gestational diabetes mellitus (GDM) is a common health issue, and it occurs with changes made in eating habits. Female members struggled with gestational diabetes during the time of pregnancy. GDM causes risk for the birth mother and the child. Early diagnosis is very important to specify enough medical follow-up and therapy on time. The authors presented a hybrid methodology of a specialized system structured on Bayesian networks, the multi-criteria approach of decision support, and artificial intelligence in this work. Here, input parameters are presented to carry out the early diagnosis of GDM, based on the symptoms of diseases that are obvious in association or that have expanded due to the adverse effects caused by undiagnosed diabetes. The algorithm's performance can be increased further with larger sample size and appropriate controls on the acupoints.[17] Choubey et al. (2020) proposed an autochthonous, efficient diagnostic method for the prediction of diabetes.

The proposed approach yields two phases: Phase-I involves accumulating the PIMA Indian DM dataset from the UCI machine learning repository databases and the Localized Diabetes dataset from Bombay Medical Hall, Upper Bazar Ranchi, Jharkhand, India. In Phase-II, the collected datasets are refined and analyzed using two methods. The first method entails classification through logistic regression, K-nearest neighbor (K-NN), iterative dichotomiser 3 decision tree (ID3), C4.5 DT, and Naive Bayes. The second method uses PCA and PSO algorithms for feature abatement, preceding the classification of the dataset using the approach used in the first method. A relative analysis is achieved between the various approaches used in PCA and PSO. The results outline the efficiency of the proposed method over the traditional classification approach in terms of less time and an accuracy of 78%. In this, the authors suggested using PCA and PSO before applying classification. But usage of PCA will add difficulties, such as problems in the interpretation of the independent variables, as after implementing PCA on them, they will change in the form of principal components and cause information loss. This will greatly affect the accuracy.[18] Singh et al. (2020) suggested an ensemble framework to predict the onset of diabetes. To this end, they progressed to a stacking-based evolutionary ensemble learning system, "NSGA-II-Stacking," for accurate prediction of the onset of Type-2 diabetes mellitus (T2DM) within 5 years. A publicly accessible PIMA Indian diabetes (PID) dataset has been adopted for this purpose. The missing values and deviance are notified and charged with the median values as a data preprocessing step. A multiobjective enhancement algorithm is used for base learner selection, which increases the classification accuracy and decreases the ensemble complexity. As for model synthesis, K-NN is employed as a meta-classifier that combines the forecasting of the base learners. The relative results determine that the proposed NSGA-II-stacking method significantly exceeds the accuracy of 83.8% in terms of individual ML methods and conventional ensemble methods.[19] Rani et al. (2020) used the Pythagorean fuzzy sets (PFS) and Type 2 diabetes (T2D) pharmacological therapy selection to manage the blood glucose level of diabetes cases. They used formula-based entropy calculation and the monitor function to evaluate the unknown criteria volume. This work was developed by the multicriteria decision-making (MCDM) content of the PFS context to prove the pharmacological therapy selection problem for T2D patients. The decision evaluations have articulated the details of each option in the form of linguistic values. Liu et al. (2020) stated that interval type 2 fuzzy sets (IT2 FS) and fuzzy comprehensive evaluation are used for the feasible evaluation of blood glucose conditions. Moreover, dynamic fuzzy rules are created to correlate blood sugar management rules. Here, the authors are using linguistic dynamic systems (LDS) to relate the valuation process of the blood glucose situation to the fuzzy comprehensive evaluation, which is completely evaluated by the analysis objects controlled by multifactors. These fuzzy comprehensive evaluations and IT2 FSs help analyze the blood sugar situation by applying dynamic fuzzy rules to the setup to provide the correlated blood glucose management edicts.[20] Omisore et al. (2020) developed an effective learning-based system for diagnosing and treating diabetes mellitus. This system has different models, like an adaptive neuro-fuzzy inference model developed for the diagnosis of diabetes and a knowledge-based diet recommender model for personalized management of diabetic cases. This work presents multimodal ANFIS (MANFIS) in which the mixture of multiple processes of a standard adaptive neuro-fuzzy inference system (ANFIS) is projected for effective diagnosis. In this work, two methods were pursued: recommendation of food items for patients and diet personalization. This work mainly focuses on the mixture of multiple ANFIS configurations for the effective diagnosis of diabetes. At the same time, the diet-based recommender model is improved for personalized management of diabetes in different aspects. ANFIS suffers from limitations that may affect the large prediction applications, making the entire process fraught with problems such as the curse of dimensionality and computational expense.[21] As the features involved in the diagnosis of diabetes remain dynamic and changing for everyone, creating a static rule for such problems will not be effective in the diagnosis and prediction process. Real-time data are very ambiguous and imprecise. Therefore, proper care must be taken to convert a piece of imprecise information into meaningful, precise information. Also, the anticipated method must have the inherent ability to handle the levity in the data. All the data for the diabetes diagnosis, including age, BMI, and insulin level, will not be fixed values and will be dynamic irrespective of the severity level of the disease. Relying upon the expertise and observation of the experts will lead to an additional cooling period between automated diagnosis and human expertise. The fuzziness or uncertainty in the data provided as input to the prediction process demands the use of an FIS as the best choice. The data conversion into fuzzy input variables will make the classification process effective by creating a precise estimate of it. Besides, as in the literature, using semantic ontology for limited predicaments will not provide a lasting solution, as it is derived for a particular set of data. Thus, using an FIS that loops in all the features and data expressed in linguistic terms by experts will provide an efficient tool for the diabetes diagnosis process. For the research community, precise information from published studies has been compiled into a tabular structure (Table 1) and is provided here.

Following a thorough review of the literature, we conclude that rule-based fuzzy inference approaches for diabetic therapy recommendations do not exist. The rule-based fuzzy inference method was developed in the work of Mohammed et al. (2018),[14] Liu et al. (2020),[21] and Omisore et al. (2020).[22] Despite numerous soft computing techniques, bioinspired and evolutionary algorithms, and many of their combinations, there is no single structured framework for performing diabetes diagnosis and recommendations in a multisequenced EHR dataset. As a result of this work, we are motivated to develop a rule-based fuzzy inference for diabetes diagnosis. This study focuses solely on creating and implementing a Mamdani FIS.

3 | ELECTRONICS HEALTH RECORD DATASET

3.1 | Electronics health record

An EHR system is a critical component of health information technology. It highlights some of the issues and concerns that these systems face and present and projected future EHR features. Traditional and modern capabilities for direct patient care and population health, and creative Health IT activities are all included in EHRs. The current EHR landscape includes issues with EHR data entry, more patient participation in EHR data entry, certified EHRs with increased interoperability, and various regulatory obligations that frequently rely on EHRs.[23]

An EHR electronically stores health data in computer format to classify patients and populations. These records can be shared in various medical service settings. Records are shared through network-related, broad data frameworks and other data organizations. Individual insights such as vaccination status, laboratory test results, radiographs, important signs, age, weight, and pricing data are all integrated into the EHR. Over the decades, EHRs have been considered an important factor in improving practice quality. EHRs are used for reasons other than patient charts. Today, providers use treatment management programs with patient-recorded data to improve quality results. The EHR helps to combine the demographics of all patients into a large pool and use this information to create “new treatment and delivery innovations” that improve overall health care goals.[24] The system's medical records allowed clinicians to identify and stratify patients with chronic illness by combining multiple types of clinical data. An EHR (using data and analysis) can improve quality care by preventing the hospitalization of high-risk patients.

3.2 | Need of electronic health record

Many hospitals and clinics require an EHR that stores the clinical reports of patients. The EHR is the central database of information that drives patient documentation, billing, quality, and clinical decision support. EHR software collects, records, and retains patient data. It includes general patient demographics, medical history, diagnoses, drugs, prescriptions, allergy lists, documentation, and laboratory results. The purpose of the EHR is to set patient data to make it easier for medical staff to evaluate patient records before and during visits. The old method of physical file folders could disperse various clinic health records. If we move to a new location, we will need to copy or fax our data to the new doctor and give the new doctor a complete knowledge of our health. Not remembering past visit information, missing or incomplete handwriting of patient records, difficult to read, and so on. This can cause several problems when seeking to digitize health records so that all healthcare providers can understand them. Providers keep a record of them in their EHR system so that future doctors can find them and easily understand and add them. The EHR has also recently been in the limelight due to the customer management systems (CMS) “significant use” legislation that motivated hospitals and practices to implement digital systems.[25]

3.3 | Role of electronics health record

EHRs are documented in the treatment of multidisciplinary patients and play an increasingly important role in exchanging information. Although EHR is associated with mixed evidence in terms of efficiency, this is a future health record format. It raises the learning opportunities and challenges for medical education. This analysis connects the concept of EHR to the physician's main abilities and refines the current learning science perspective of diagnostic and clinical reasoning based on the theoretical framework of scientific reasoning and argumentation. It ends with an integrated vision for the use of EHR and a patient's special role in medical education and learning.

3.4 | Merits of electronics health record

The ability to electronically exchange health information with the EHR can help patients with higher quality and safer treatments while at the same time creating substantial improvements to their concerns. The EHR helps doctors manage their patients' treatments better and provide better medical services in the following ways:

TABLE 1 Survey on recent techniques in diabetes recommendations system

S. no.	Author and reference (Year)	Methodology recommended/ algorithm enforced	The convenience of the methodology/ algorithm	Inhibition of the methodology/ algorithm	Inference from the study
1	Lee et al.[6]	Intelligent ontological agent	It improves the quality of entity analysis.	It lacks in utilization, reuse, and maintainability.	Process of specification of individuals.
2	Chen et al.[7]	Fuzzy reasoning techniques	The fuzzy reasoning system can provide the most efficient solution to difficult problems. The technique may easily be adjusted to increase or change its performance.	This technique relies on erroneous data and inputs, and their accuracy is affected.	Derives the conclusions from a group of fuzzy rules (If–Then rules)
3	El-Sappagh et al.[8]	Case-based reasoning (CBR)	No prior knowledge of rules or processes is required to develop them. The reasoner may also instantly provide solutions to problems using CBR.	The elements of one instance may vary from another and result in inconsistencies.	Process of experience-based approach.
4	Chen et al.[9]	Domain ontology	It speeds up the process of ontology development.	Reusing resources is not applicable.	Describing the property of individual elements.
5	Alonso et al.[10]	Fuzzy rule-based system	It is a reliable system that does not require exact inputs. These systems can handle various inputs, including ambiguous, skewed, or inaccurate data.	The system's efficiency is low since it relies heavily on faulty data.	Representing different forms of fuzzy rules.
6	Gupta et al.[11]	Fuzzy nutrition	It describes the range of fuzzy membership functions automatically.	The levels of fuzzy parameters are not described.	Proceedings to find the best solutions.
7	Ali et al.[12]	Fuzzy ontology	It develops frameworks for describing adaptable and accessible information across various domains.	It is impossible to reprogram the logic system if the feedback system fails in ontology.	Process of the set of Individual objects.
8	Rajeswari et al.[13]	Fuzzy logic based association classification	A fuzzy association rule mining approach is used for rapid and convenient performance on huge datasets.	Similar representations are used to represent fuzzy association rules.	Generate fuzzy association rule.
9	El-Sappagh et al.[16]	Case-based fuzzification	The rules of the case-based fuzzy logic control system are updated regularly by this system.	The idea is that case-based fuzzification systems fully rely on human knowledge and expertise.	Handling large datasets for the fuzzification process.
10	Gomes Filho et al.[17]	Heterogeneous methodology	It implements absolute rather than relative risk measures.	Methodological difficulties such as problems with	Processing dissimilar data elements using absolute risk measures.

(Continues)

TABLE 1 (Continued)

S. no.	Author and reference (Year)	Methodology recommended/ algorithm enforced	The convenience of the methodology/ algorithm	Inhibition of the methodology/ algorithm	Inference from the study
				randomization and early study termination were included.	
11	Choubey et al.[18]	Principal component analysis (PCA)	PCA increases the ML algorithm's performance by removing correlated variables. By reducing the number of features, PCA aids in overcoming data overfitting concerns. PCA produces a lot of variation, which helps with visualization.	The principal component analysis techniques are mixed with various data features, although they are not as straightforward to analyze. It also includes the trade-off between feature extraction and data redundancy.	Process of finding the principal elements.
12	Singh et al.[19]	Evolutionary ensemble learning	The two distinct advantages of evolutionary ensemble algorithms are the flexibility of the process and the ability to self-adapt to get optimal solutions. The use of evolutionary algorithms will make the process faster.	The idea of evolutionary learning has been criticized for lack of evidence and missing links and contradictions.	Learning evaluations are done by applying evolutionary algorithms.
13	Rani et al.[20]	Fuzzy complex system	It effectively generates crisp output based on logic statements.	There is no single systematic approach to solve these fuzzy-based problems.	Appropriately treat fuzzy, complex linear systems with equations.
14	Liu et al.[21]	Interval fuzzy set	It allows the various degrees of fuzzy membership function to perform the operations.	The interval limit maintains constants only.	Process of generating parameterized fuzzy subsets.
15	Omisore et al.[22]	Learning-based system	A fuzzy learning system is adaptable and allows for rule changes.	The structure of the system is not entirely comprehensible.	They are making accurate predictions using fuzzy learning.

- At medical examination, we provide accurate and up-to-date information on patients.
- Adjusted for easy access to patient records for efficient treatment.
- Sharing patients' medical information helps diagnose patients more effectively, reduce malpractice, and provide safer treatment.
- Improves patient–donor interaction and communication and provides medical convenience.
- Allows safer and more reliable formulations that help to improve readability.
- Fully documented and accurate, simplified encoding and billing enhanced privacy and security of patient data.
- Reduce costs, improve security, reduce duplicate checks, and have abundant impressions of health.

3.5 | Dataset used

We have used an EHR dataset from the National Institute of Diabetes and Digestive and Kidney Diseases. The dataset was obtained from the machine learning data repository at UCI. It is a dataset that contains medical information for PIMA Indians that enables one to predict whether or not a patient has diabetes based on diagnostic parameters provided in the dataset.[26] Table 2 shows the dataset, comprising 9 attributes and 768 instances (UCI). The dataset's last column specifies whether or not the person has been diagnosed with diabetes (1) or not (0). Table 2 also provides more information on missing data, minimum and maximum values, the number of unique and distinct instances, and mean and standard deviations for each attribute. There is no missing value in the selected EHR dataset. The outcome is the class distribution variable in the selected EHR diabetes dataset. The class value of 1 is interpreted as "tested positive for diabetes" and has 500 positive instances. The class value of 0 is interpreted as "tested negative for diabetes," with 268 negative instances.

4 | FUZZY-RELATED CONCEPTS

4.1 | Fuzzy reasoning

Fuzzy reasoning is addressed by utilizing fuzzy sets that express vulnerability, fuzzy reasoning depending on the idea of enlistment degrees, and soft reasoning intended to numerically address vulnerability for managing the inbuilt indistinctness in certain areas. Typically, the reasoning is involved in only two characteristics (true and false) and has its requirements in dealing with issues identified in this current reality space. Fuzzy reasoning utilizes consistent quality somewhere in the range of 0 and 1.[27]

4.2 | Fuzzy set

In the fuzzy hypothesis, the fuzzy set, F of universe U is characterized by enlistment work. It is inferred by $\mu F(A)$ with the ultimate objective function such that $\mu F(A)$: $F \rightarrow [0,1]$, 1 if A is absolutely in F, $\mu F(A) = 0$ if x is not in F. For any component, A of universe X, enlistment work $\mu F(A)$ indicates how much A is a component of set F. A degree having a value somewhere in the range of 0 and 1, which addresses the degree of enlistment, is called the participation worth of component A in set F.[27]

4.3 | Type-1 and type-2 fuzzy reasoning

Type-1 FLSs can function admirably under unequivocal action conditions. The phonetic and mathematical

TABLE 2 Detailed description, statistical analysis, and instances of PIMA Indian Diabetes dataset

S. no.	Feature	Units	Range	Type	Value	Mean	Standard deviation	Min value	Max value	Unique instance	Distinct instance
1	Pregnancies	—	17	Discrete	Real valued	3.845052	3.369578	0.000000	17.000000	2	17
2	Glucose	mg/dL	199	Discrete	Real valued	120.894531	31.972618	0.000000	199.000000	19	136
3	Blood pressure	mmHg	122	Continuous	Real valued	69.105469	19.355807	0.000000	122.000000	8	41
4	Skin thickness	mm	99	Discrete	Real valued	20.536458	15.952218	0.000000	99.000000	5	51
5	Insulin	muU/mL	846	Discrete	Real valued	79.799479	115.244002	0.000000	846.000000	93	186
6	BMI	kg/m^2	67.1	Continuous	Real valued	31.992578	7.884160	0.000000	67.100000	76	246
7	Diabetes pedigree function	—	2.342	Continuous	Real valued	0.471876	0.331329	0.078000	2.420000	34	517
8	Age	Year	60	Discrete	Nominal valued	33.240885	11.760232	21.000000	81.000000	5	52
9	Outcome	[0: No, 1: Yes]	1	Categorical	Binary valued	Tested positive instances: 500 Tested negative instances: 268					

shortcomings can cause issues in deciding the specific and exact pioneers and subsequent participation capacities during the FLS plan. As time goes by, each customer's lead and tendencies change, starting with one individual and moving on to the next; additionally, the space specialists' thoughts are vacillated. From this point forward, the appropriateness of the type-1 based design will go down when there are high shortcoming levels related to the starvation routine region.[29]

Type-2 FLSs are used to deal with the vulnerabilities in the cooperative choice-creating measure as they can show the vulnerabilities between well-qualified evaluations by utilizing type-2 fuzzy sets. In T1FS, the enlistment work esteem is a new number in precisely 0 or 1. A fuzzy enlistment work describes a T2FS; the participation and incentive for each segment of this set are fuzzy sets in 0, 1, and the middle of 0 and 1. The EWs of T2FSs are three-dimensional and incorporate an impression of vulnerability (IOV). To show the gathering vulnerabilities, T2FSs give extra levels of opportunity that can make it conceivable, which incorporate specialists' various sentiments and tendencies.[30] The T2FSs can display the prerequisites of an individual assurance that is intelligent of the relative multitude of specialists' sentiment, which would then be utilized to give a respectable recommendation to the restriction routine.[31]

4.4 | Fuzzy ontology

An ontology is a strategy utilized for tending a snippet of information and its relationship to different bits of information.[32] Since philosophy is not adequate to manage shortcomings or anomalies, they are associated with fuzzy information to create a fuzzy ontology. It gives a level of truth to ontology, which depicts the degree to which a thing is an occurrence of a class of things. This is valuable for developing a lot of information for a domain.[33] Fuzzy ontology addresses the various connections between ideas in a specific space and can be seen as a visual representation. Many fuzzy concepts are more ambiguous than precise, and the fuzzy ontology is appropriate for managing fuzzy information.[34]

The fuzzy ontology is characterized by 5-tuple notations.

$$\text{Fuzzy ontology (FO)} = \{I, C, R, F, A\}$$

where I is the set of individuals (or) instance of the concepts; C is the set of concepts (or) each concept indicates a fuzzy set C: $I \rightarrow [0,1]$; R is the set of relations. Every $R \in R$ is an n-dimensional array fuzzy connection on the domain of entities so that R: $E_n \rightarrow [0,1]$. The set of elements of the fuzzy ontology will be shown by E where $E = \{C\}\ U\ \{I\}$; F is the set of fuzzy relations on the set of elements; A is the set of axioms characterized by using an appropriate intelligent language.

4.5 | Fuzzy inference system

The primary task of the FIS is to decide on the key unit of fuzzy logic. It utilizes the "If … Then" control rules with "OR" or "AND" for constructing central decision standard logic units.[35]

4.5.1 | Few characteristics of FIS

1. The fuzzy set's input is independent of its fuzzy relations.
2. The gain of FIS is consistent, which means it can be padded or new (crisp).
3. It is important to have a fuzzy harvest when it is handled as a regulator.
4. A defuzzification unit with FIS could disciple fuzzy factors into new (crisp) factors.

4.5.2 | Components of fuzzy inference system

There are five main parts to the FIS represented in Figure 1, and they work as follows:

1. Rule base (knowledge base): It stores IF–THEN rules created by experts.
2. Database: It characterizes the membership elements of fuzzy sets utilized in fuzzy rules.
3. Decision-making unit: It implements procedures and rules.
4. Fuzzification interface unit: It is the pattern of fuzzifying a new component. It converts crisp (new data) quantities into fuzzy quantities by utilizing the fuzzy rule base (knowledge base or intelligence base).
5. Defuzzification interface unit: It is the process of defuzzing a fuzzified entity. It converts the fuzzified set acquired by the inference engine into a crisp new value.

5 | DESIGN OF INTELLIGENT FUZZY INFERENCE RULE-BASED PREDICTIVE DIABETES DIAGNOSIS MODEL

In this part, we examine the plan of a diabetes prediction model using an FIS. This work proposes a system

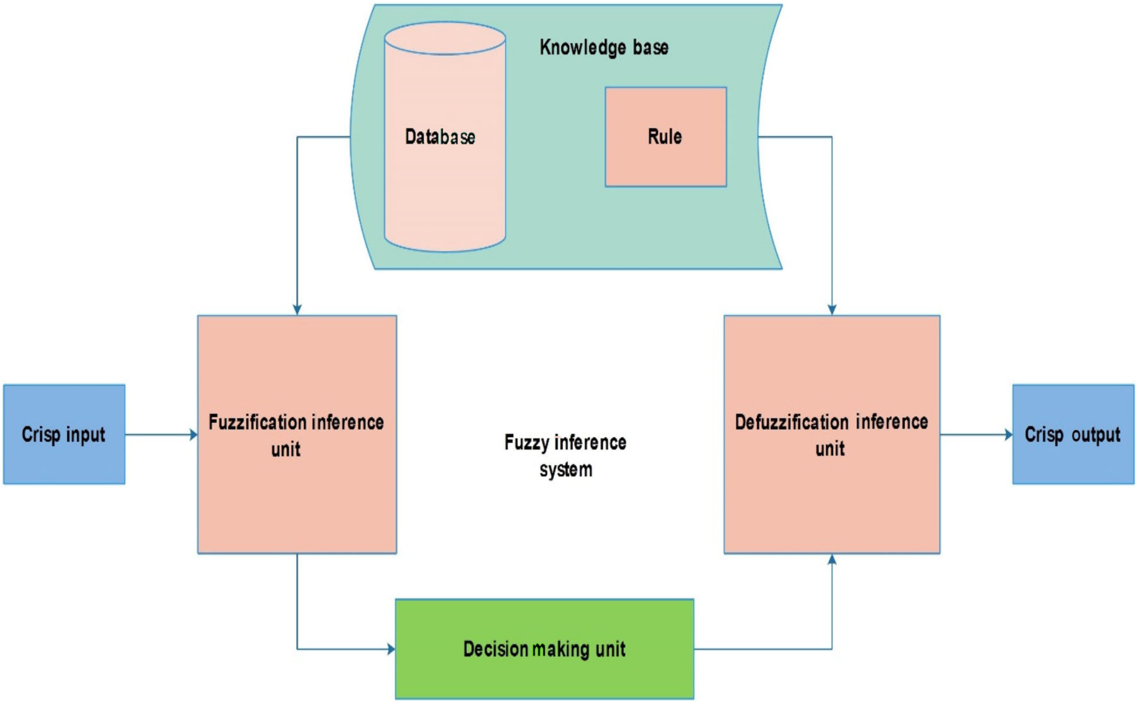

FIGURE 1 Block diagram for fuzzy inference system

architecture for diabetes prediction and recommendation by constructing the fuzzy membership function, the ruleset, and the fuzzy inference.

5.1 | Architecture of IFIR_PDDM

The framework of the IFIR_PDDM is exposed in Figure 2.

We acquired the 768 EHR datasets from the National Institute of Diabetes and Digestive and Kidney Diseases (PID dataset). Based on the training sets, the diabetes disease prediction model is created, and the model's performance produced through the testing sets is evaluated. The fuzzy rule base and the respective membership functions are required to build the fuzzy knowledge base. The fuzzy rule base produces IF–Then rules by applying the C4.5 decision tree algorithm.[36–38] And in later cases, the fuzzy rule is converted into the crisp function to get a fuzzy base. Medical experts estimate the fuzzy rules and delete them and add them to the fuzzy base for fuzzy rule verification. The fuzzy logic form is converted into multi-valued logic based on existing medical guidelines to get an exact fuzzy membership function. Next, by referring to the training set data, the degrees of membership functions are revised to get an optimal membership function. The fuzzy membership function and the fuzzy rule base are used in the fuzzy inference engine for diabetes prediction support. Based on the results, recommendations concerning typical state living, exercise, and medication classifications are then made to patients. The fuzzy relations between the EHR dataset and the parameters with their descriptions are listed in Table 3.

5.2 | Creation of fuzzy membership function

The fuzzy membership function has boundary ambiguity, which may be characterized by a fuzzy set.[39] The creation of the fuzzy membership function is shown in Figure 3. The fuzzy membership function is created by using the input values, and the fuzzy sets are suggested by accepting the attributes age, glucose, BMI, insulin, and blood pressure. Next, by attributing the healthcare medical experts' instructions, a triangular fuzzy function will be built. Based on the user selection as a preprocessing step, the fuzzy function tends to eliminate some features from the dataset that have less importance. Thus, attributes like diabetes pedigree function, skin thickness, and pregnancies are not considered for deriving fuzzy rules based on feature selection output.[40] The required fuzzy parameters are represented in Table 3 to create a fuzzy membership function. We have used five input parameters and one output parameter to create a fuzzy membership function. The fuzzy parameters are created in Table 3 with the output variable μO. It produces the final

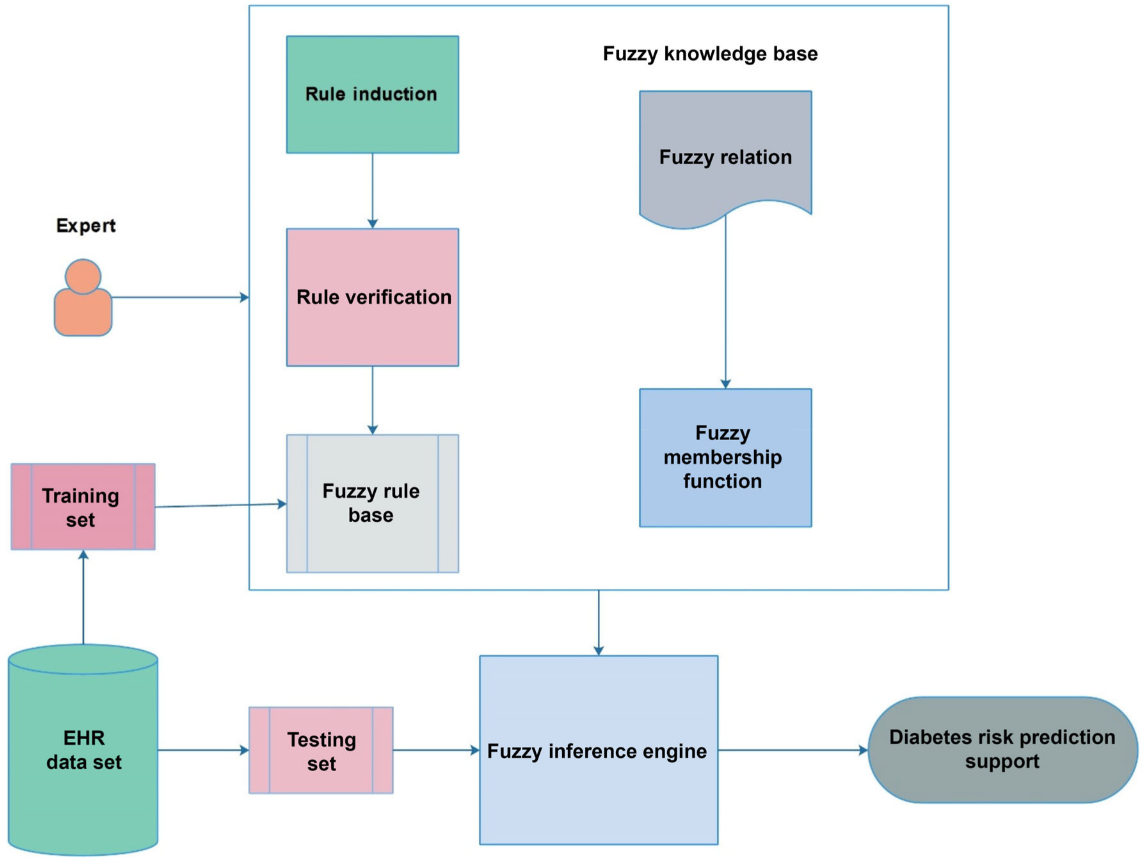

FIGURE 2 Proposed fuzzy architecture for IFIR_PDDM

diabetes disease risk value. The result is accumulated by rule-based fuzzification by joining five parameters.

5.3 | Fuzzification process

The development of catching crisp input values and converting them to the degree needed by the terms is referred to as fuzzification. The variable is possibly fuzzy and can be expressed by a fuzzy membership function. Typically, the input parameters are emanated from hardware component measures. Next, crisp inputs can be fuzzified in the FIS.[41] Then, because the fuzzy ideas are developed, the fuzzy numbers are worked by the fuzzy relationship.

Moreover, an interface is offered to tune and approve the boundaries of the assembled fuzzy numbers. In this work, a triangular fuzzy function is used as the fuzzy membership function of the fuzzy number and can be indicated as the boundary set $[x, y, z]$. Then the fuzzy membership function $\mu(F)$ of the triangular fuzzy number is disposed of by,

$$\mu(F) = \left\{ \begin{array}{c} 0, F \leq x \\ \frac{(F-x)}{(y-x)}, x < F \leq y \\ \frac{(z-F)}{(z-y)}, y < F \leq z \\ 0, F > z \end{array} \right\} \tag{1}$$

Initially, an EHR is recovered from a data source into an experimental expert system. The fuzzy membership values and degrees are collected using fuzzification and their crisp input values. After getting these fuzzy values, all values are refined to design a fuzzy decision system. After applying fuzzy rules, the output values obtained from fuzzy systems are sent to the defuzzification unit to get the final crisp value.

For the input fuzzy value age (let F), which varies from 21 to 81, and the fuzzy expression will be,

$$\mu_{\text{low}}(F) = \left\{ \begin{array}{c} \frac{21-F}{21}, 21 < F \leq 51 \\ 0, \text{Otherwise} \end{array} \right\} \tag{2}$$

TABLE 3 Fuzzy relation, fuzzy parameter variables, and fuzzy membership function

		Range			
Parameter	**Fuzzy relation (FR)**	**Left**	**Mid**	**Right**	**Linguistic variables**
μP (Age)	FR ≥ FZ (Age 0_25)	22	23	24	Less youthful
	FR ≥ FZ (Age 25_35)	23	26	28	Youthful
	FR ≥ FZ (Age 35_40)	26	30	38	Less intermediate aged
	FR ≥ FZ (Age 40_45)	36	42	50	Intermediate aged
	FR ≥ FZ (Age 45_50)	44	51	56	High intermediate aged
	FR ≥ FZ (Age 50_60)	50	58	64	Very less aged
	FR ≥ FZ (Age 60_70)	64	70	76	Less aged
	FR ≥ FZ (Age 70_81)	72	76	80	Aged
μQ (Glucose)	FR ≥ FZ (Glu 0_40)	42	46	52	Very low
	FR ≥ FZ (Glu 40_85)	48	56	68	Low
	FR ≥ FZ (Glu 85_125)	66	118	146	Marginal
	FR ≥ FZ (Glu 125_165)	120	136	165	High
	FR ≥ FZ (Glu 165_199)	138	172	196	Very high
μR (BMI)	FR ≥ FZ (BMI 0_20)	16	24	28	Very low
	FR ≥ FZ (BMI 20_30)	22	28	30	Low
	FR ≥ FZ (BMI 30_40)	30	38	46	Marginal
	FR ≥ FZ (BMI 40_50)	38	44	50	High
	FR ≥ FZ (BMI 50_67.1)	46	54	66	Very high
μS (Blood pressure)	FR ≥ FZ (BP 0_20)	26	32	38	Very low
	FR ≥ FZ (BP 20_40)	34	46	52	Low
	FR ≥ FZ (BP 40_60)	49	56	68	Marginal
	FR ≥ FZ (BP 60_80)	60	78	84	High
	FR ≥ FZ (BP 80_122)	80	112	120	Very high
μT (Insulin)	FR ≥ FZ (Ins 0_150)	16	32	44	Very low
	FR ≥ FZ (Ins 150_300)	30	62	220	Low
	FR ≥ FZ (Ins 300_450)	80	320	482	Marginal
	FR ≥ FZ (Ins 450_600)	466	674	702	High
	FR ≥ FZ (Ins 600_846)	698	762	842	Very high

$$\mu_{\text{medium}}(F) = \begin{Bmatrix} \frac{51}{F}, 21 < F \leq 51 \\ \frac{81-F}{51}, 51 < F \leq 81 \\ 0, \text{Otherwise} \end{Bmatrix} \tag{3}$$

$$\mu_{\text{high}}(F) = \begin{Bmatrix} 0, F < 51 \\ \frac{F-81}{81}, 51 < F \leq 81 \\ 1, \text{Otherwise} \end{Bmatrix} \tag{4}$$

For the input fuzzy value glucose (let F), which varies from 44 to 199, and the fuzzy expression will be,

$$\mu_{\text{low}}(F) = \begin{Bmatrix} \frac{44-F}{44}, 44 < F \leq 121 \\ 0, \text{Otherwise} \end{Bmatrix} \tag{5}$$

$$\mu_{\text{medium}}(F) = \begin{Bmatrix} \frac{121}{F}, 44 < F \leq 121 \\ \frac{199-F}{121}, 121 < F \leq 199 \\ 0, \text{Otherwise} \end{Bmatrix} \tag{6}$$

$$\mu_{\text{high}}(F) = \begin{Bmatrix} 0, F < 121 \\ \frac{F-121}{121}, 121 < F \leq 199 \\ 1, \text{Otherwise} \end{Bmatrix} \tag{7}$$

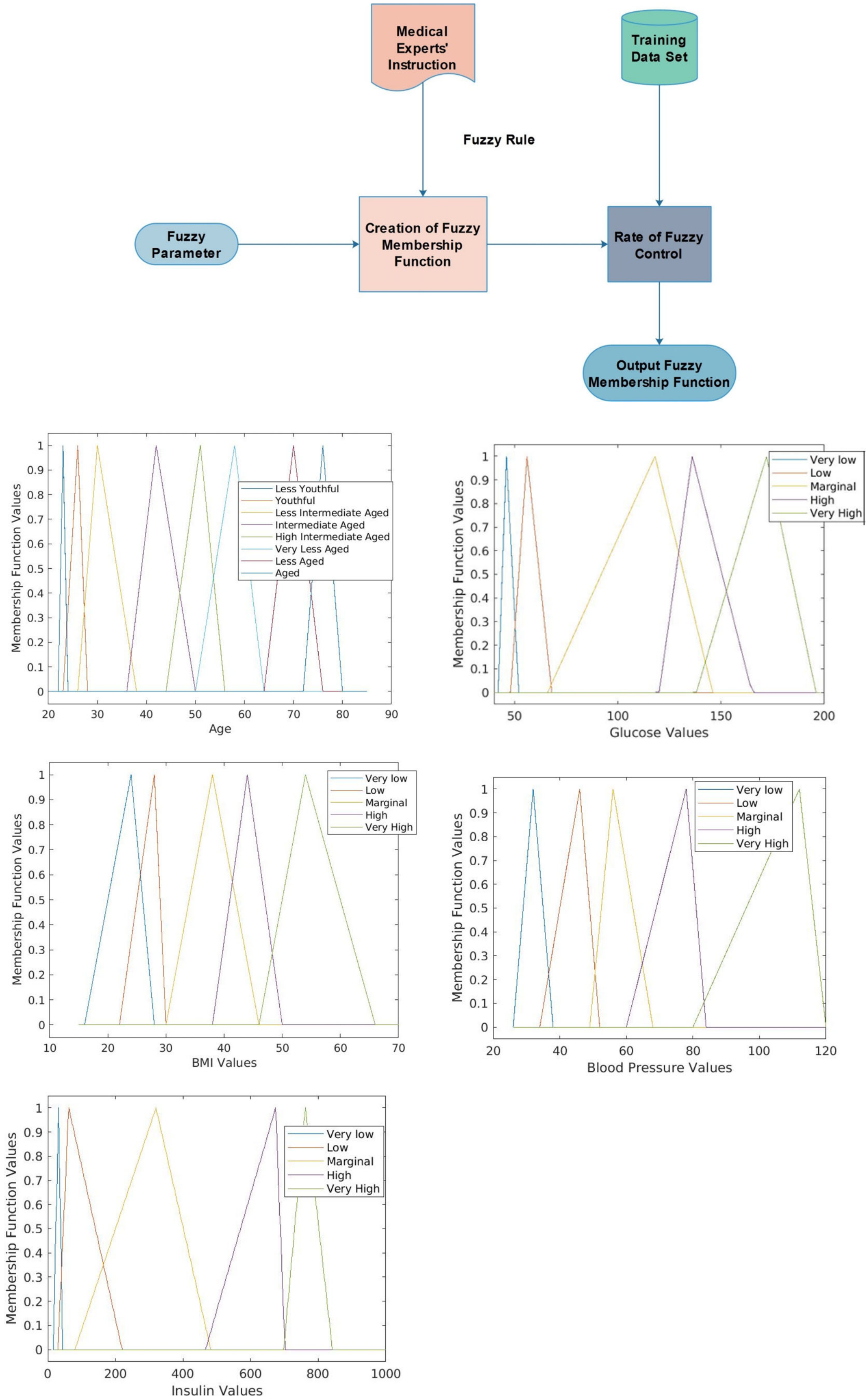

FIGURE 3 Creation of fuzzy membership function. (A) Membership for the age attribute. (B) Membership for the glucose attribute. (C) Membership for the BMI attribute. (D) Membership for the blood pressure attribute. (E) Membership for the insulin attribute

For the input fuzzy value BMI (let F), which varies from 18.2 to 67.1, and the fuzzy expression will be,

$$\mu_{\text{low}}(F) = \begin{Bmatrix} \dfrac{18.2-F}{18.2}, 18.2 < F \leq 42.6 \\ 0, \text{Otherwise} \end{Bmatrix} \tag{8}$$

$$\mu_{\text{medium}}(F) = \begin{Bmatrix} \dfrac{42.6}{F}, 18.2 < F \leq 42.6 \\ \dfrac{67.1-F}{42.6}, 42.6 < F \leq 67.1 \\ 0, \text{Otherwise} \end{Bmatrix} \tag{9}$$

$$\mu_{\text{high}}(F) = \begin{Bmatrix} 0, F < 42.6 \\ \dfrac{F-67.1}{67.1}, 42.6 < F \leq 67.1 \\ 1, \text{Otherwise} \end{Bmatrix} \tag{10}$$

For the input fuzzy value blood pressure (let F), that varies from 24 to 122, and the fuzzy expression will be,

$$\mu_{\text{low}}(F) = \begin{Bmatrix} \dfrac{24-F}{24}, 24 < F \leq 72 \\ 0, \text{Otherwise} \end{Bmatrix} \tag{11}$$

$$\mu_{\text{medium}}(F) = \begin{Bmatrix} \dfrac{42.6}{F}, 24 < F \leq 72 \\ \dfrac{122-F}{72}, 72 < F \leq 122 \\ 0, \text{Otherwise} \end{Bmatrix} \tag{12}$$

$$\mu_{\text{high}}(F) = \begin{Bmatrix} 0, F < 72 \\ \dfrac{F-122}{122}, 72 < F \leq 122 \\ 1, \text{Otherwise} \end{Bmatrix} \tag{13}$$

For the input fuzzy value insulin (let F), which varies from 14 to 846, and the fuzzy expression will be,

$$\mu_{\text{low}}(F) = \begin{Bmatrix} \dfrac{14-F}{14}, 14 < F \leq 440 \\ 0, \text{Otherwise} \end{Bmatrix} \tag{14}$$

$$\mu_{\text{medium}}(F) = \begin{Bmatrix} \dfrac{440}{F}, 14 < F \leq 440 \\ \dfrac{846-F}{440}, 440 < F \leq 846 \\ 0, \text{Otherwise} \end{Bmatrix} \tag{15}$$

$$\mu_{\text{high}}(F) = \begin{Bmatrix} 0, F < 440 \\ \dfrac{F-846}{846}, 440 < F \leq 846 \\ 1, \text{Otherwise} \end{Bmatrix} \tag{16}$$

The membership functions of the various attributes, such as age, glucose values, BMI, blood pressure, and insulin, are formulated using Equations (2)–(16). The membership function thus calculated is depicted in Figure 3A–E.

5.4 | Fuzzy decision system

The fuzzy decision system independently deduces the probability of individual diabetes for each occasion in fuzzification and translates the incident into a fuzzy sentence. The American Diabetes Association[42] indicates that diabetes is related to obesity, family ancestry, and age. Furthermore, the American Diabetes Association states that the 2-h OGTT with plasma glucose and serum insulin fixations is the gold standard for diagnosing diabetes. We have considered age, glucose, BMI, blood pressure, and insulin as attributes for fuzzy input variables of endorsed fuzzy rule-based inference systems. The fuzzy numbers and the related information are stored in the fuzzification. The resultant fuzzy numbers and their membership function with parameters are listed in Table 3, and Pseudocode 1 is mentioned below:

5.5 | Fuzzy rule set base model

It is important to create fuzzy rules to infer the fuzzy values. The fuzzy-based prediction method for diagnosing diabetes needs an interpretation of the ruleset. The method used to generate the fuzzy rules in this proposed work is described in Figure 4.

A C4.5 decision tree machine learning technique generates the fuzzy rules by using the training dataset. This tree generates IF–THEN rules that are used for rule induction. A decision tree is constructed based on overall (class) entropy, average information entropy (AIE), and information gain (IG) by applying the C4.5 algorithm. All the continuous attribute values in the training dataset are translated into categorical values to get a fuzzy base rule. This translation was made by satisfying the below constraints.

$$\begin{aligned} &\text{IF}\Big(\min\left[\text{MedicalExpert\_Guidelines}_{ij}\right] \leq F_i \\ &\quad \leq \max\left[\text{MedicalExpert\_Guidelines}_{ij}\right]\Big). \\ &\text{THEN}\, F_i \rightarrow A_j. \end{aligned}$$

Here, $\text{MedicalExpert\_Guidelines}_{ij}$ indicates medical guidelines; F_i indicates continuous attribute value; i indicates the parameter of the fuzzy inference input

Pesudocode 1: Pseudocode for fuzzy decision system

Input: Fuzzy set for fuzzy parameters—are glucose (Glu), insulin (Ins), body mass index (BMI), age (Age), blood pressure (BP)

Output: Fuzzy set for diabetes outcome
Methods:

1. Begin
2. Initialization of all selected input fuzzy parameters.
3. State the triangular fuzzy membership function for the fuzzy numbers by Equation (1).
4. Frame the fuzzy numbers for all fuzzy parameters. 4.1. Create the fuzzy number for the outcome class diabetes
5. Apply Mamdani FIS to get fuzzy rules. 5.1. Initialize the fuzzy rules. 5.2. A fuzzy input parameter could be built using the fuzzy membership function. 5.3. Next, based on the fuzzy rules, authorize the fuzzy rules' strength by merging the fuzzified inputs. 5.4. Next, Find the subsequent fuzzy rules by linking the strength of the fuzzy rules and output membership function. 5.5. For capturing yield dissemination, join all the consequents. 5.6. Finally, a defuzzified yield dispersion is acquired.
6. Present the information as fuzzy binary instinct language.
7. End.

data; j indicates the categorical value of the fuzzy inference input data.

If the training data satisfies the above constraints, F_i continuous variable value is translated into A_j categorical variable value.

The created rule is validated and tested by medical experts. Once a tree is constructed, its final values are translated into the crisp model to develop fuzzy inference.

5.6 | Creation of fuzzy rules using decision tree machine learning algorithm

We have used a C4.5 decision tree machine learning algorithm to create a fuzzy rule in our proposed work. The decision tree finds the entropy, AIE, and gain of the individual features. Based on the gain, the tree structure is framed. The decision tree structure is partitioned into the left child and right child nodes and creates edges until a final fuzzy rule is obtained. Once the decision tree is constructed, it is translated into a crisp model and the rule base is created using the C4.5 decision tree algorithm. The translation method is described in Figure 5.

The fuzzy crisp model is translated from the decision tree and is illustrated in Equations (17) and (18).

$$\text{Condition}\,(F,i)_n = K_{\text{Root}}\,(F_{\text{Root}}, i_{\text{Root}}) \wedge K_1\,(F_1, i_1) \wedge \ldots K_n\,(F_n, i_n) \tag{17}$$

$$\text{Rule}\,(F,i)_n = \text{Condition}(F,i)n_1 \vee \text{Condition}(F,i)n_2 \vee \ldots \text{Condition} \times (F,i)n_n \tag{18}$$

For a single fuzzy rule, the total condition is represented by Condition $(F,\ i)_n$. It includes all the conditions from the root node to the leaf node of a decision tree. For Rule $(F,\ i)_n$, n comprises all the conditions of diabetes risk, which is the outcome.

5.7 | Formulation of fuzzy rules

The fuzzy rules used in the proposed work are given as follows:

RULE 1: If (glucose level is high) and (BMI is high) and (blood pressure is low) and (Insulin is low) then (diabetes risk is positive).
RULE 2: If (BMI is high) and (blood pressure is low or marginal) and (plasma glucose low or marginal) and (insulin is low or marginal) then (diabetes risk is positive).
RULE 3: If (BMI is low or marginal) and (blood pressure is low or marginal) and (plasma glucose is low or marginal) and (insulin is low or marginal) then (diabetes risk is negative.)
RULE 4: If (plasma glucose level is high) and (insulin is high) and (BMI is high) and (blood pressure is low or marginal) then (diabetes risk is positive.)
RULE 5: If (Blood pressure is high) and (BMI is high) and (plasma glucose is low) and (Insulin is low) then (diabetes risk is positive.)
RULE 6: If (insulin is high) and (blood pressure is low or marginal) and (plasma glucose is low or marginal) and BMI is low or marginal) then (diabetes risk is negative.)
RULE 7: If and (BMI is high) and (blood pressure is high) and (plasma glucose is high) and (insulin is high) then (diabetes risk is positive.)

We approached medical specialists to verify the rule base. During the verification process, the two medical experts removed any rules that did not conform to their knowledge and made personal adjustments as needed to create a new rule foundation. A total of 208 domains are covered. As a result, mining-

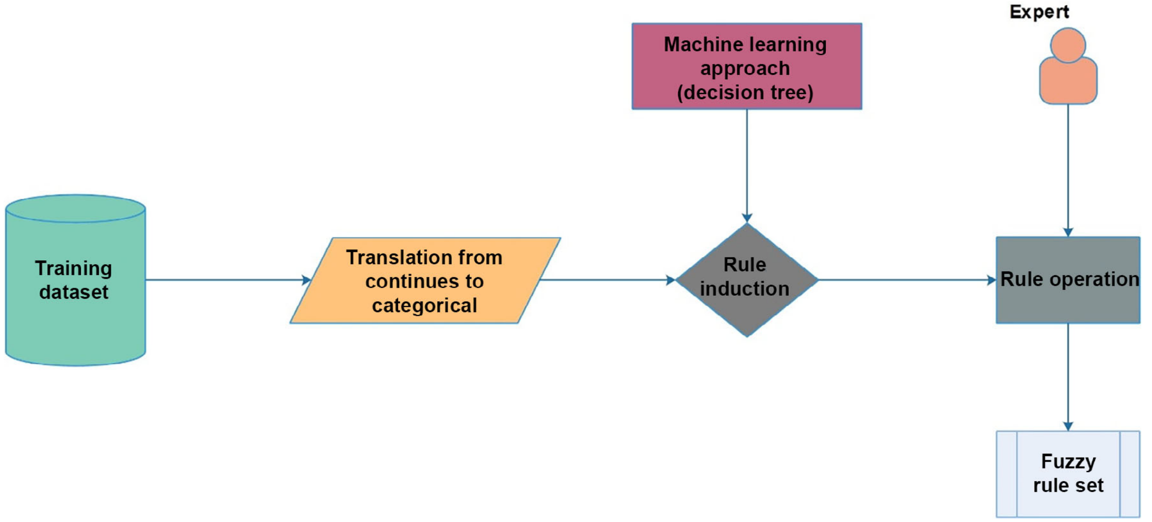

FIGURE 4 Fuzzy rule set based model

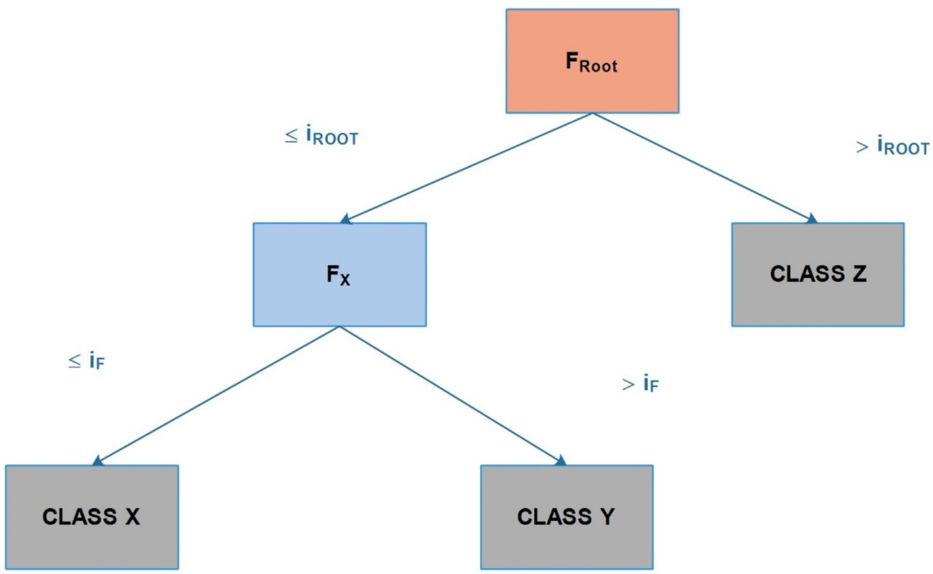

FIGURE 5 Fuzzy crisp model translation from decision tree

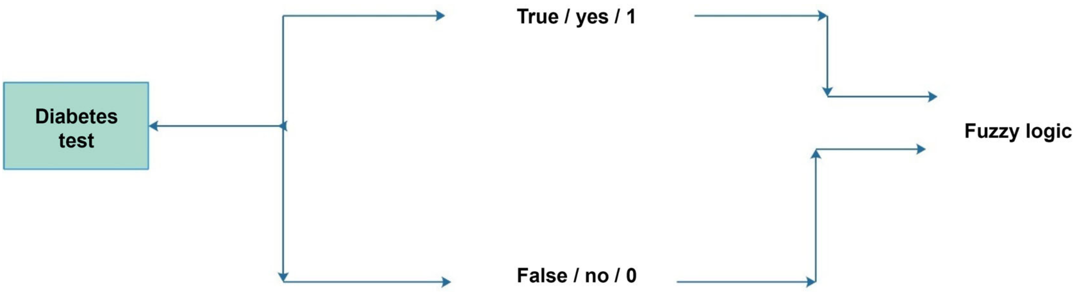

FIGURE 6 Fuzzy test inference for diabetes

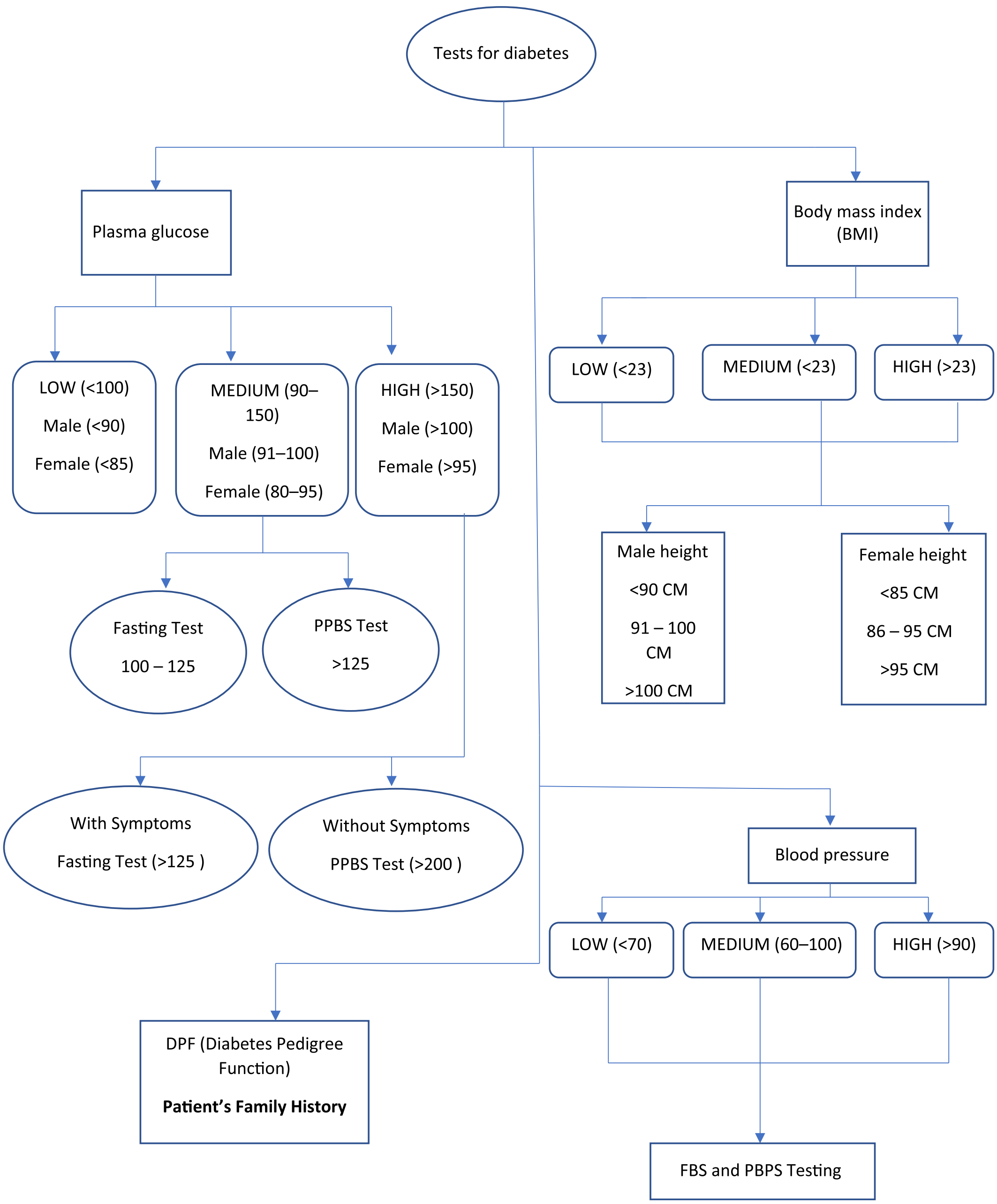

FIGURE 7 Diabetes tests

based rules are validated, leading to the selection of these rules.

5.8 | Fuzzy inference with fuzzy test for diabetes expert recommendations

Fuzzy inference logic is a reality, a problem-settling technique. It is synchronously managed algebraic data and grammar vocabulary knowledge. An FIS is a procedure to evaluate a given input to an output using fuzzy logic. This process consists of all the types of comrade's

TABLE 4 Recommendations related to diabetes risk

Diabetes risk (%)	0–10	11–15	16–20	>20 and <25
Recommendations	Normal living	Nutrition control	Exercise	Drugs

functions. The fuzzy logic system operator is based on if–then rules. The FIS is otherwise called fuzzy-rule-based, fuzzy expert, fuzzy model, fuzzy versatile memory, fuzzy logic thermostat, or simply fuzzy system.[43] Fuzzy inference logic resembles the human resolution-making

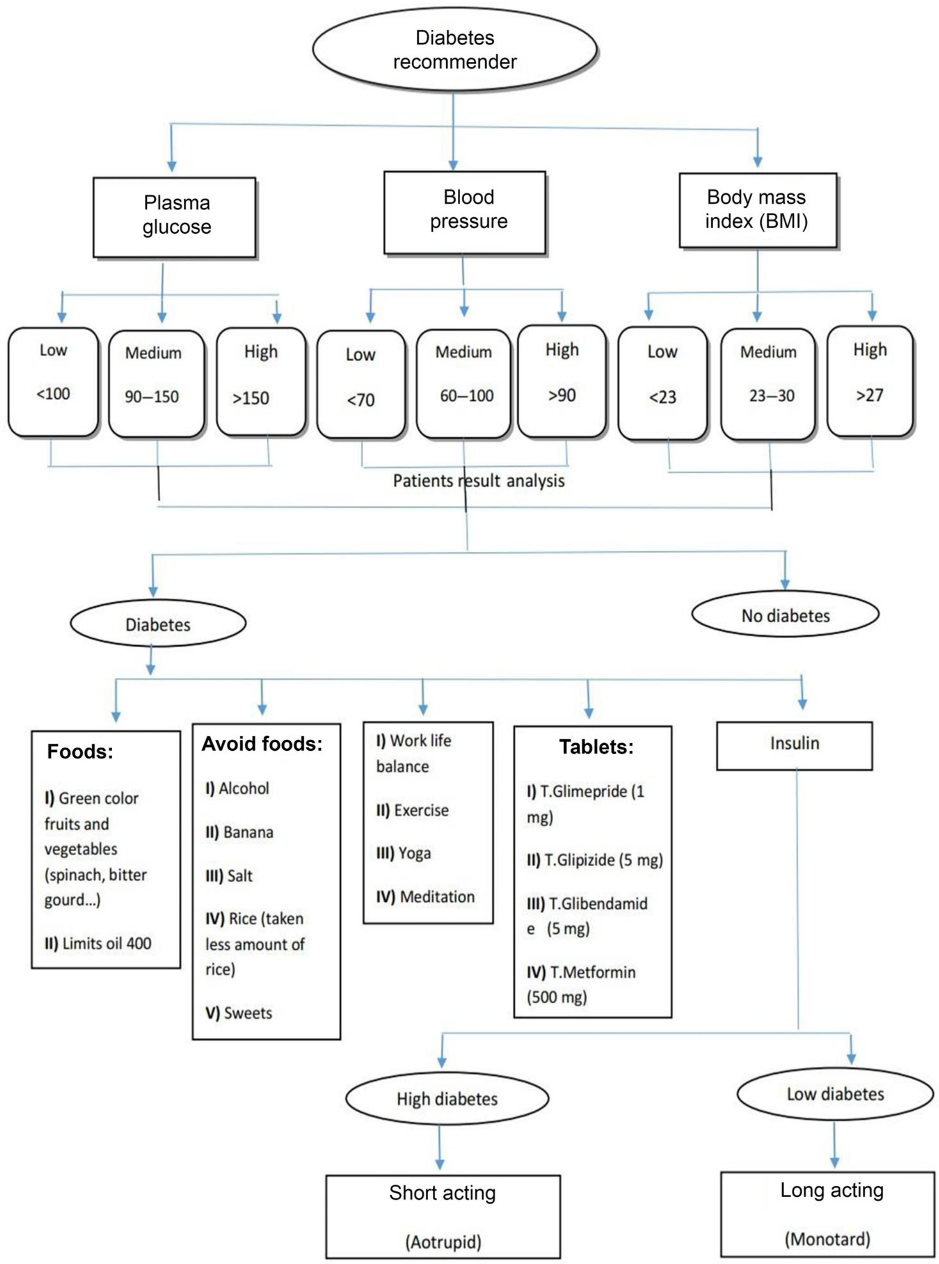

FIGURE 8 Generic recommendations of diabetes

technique. It deals with vague and inexact information. This is based on a degree of partiality greater than the usual true/false or 1/0 as in bipartite logic. This method is used in various medical applications like a medicinal suggestive support system, control of major pressure during unconsciousness, multiverse control of unconsciousness, modeling of neuropathology findings in old timers' patients, radioscopy diagnoses, and fuzzy inference delivery of diabetes and prostatic cancer. A diabetes fuzzy test inference is represented in Figure 6.

The different diabetes tests are expressed in Figure 7.

Diabetes fuzzy inference consists of the following modules:

TABLE 5 Description of performance measures

S. no	Factors	Notation
1	Precision	$\frac{TP}{TP+FP}$
2	Recall	$\frac{TP}{TP+FN}$
3	Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
4	*F*-score	$\frac{2TP}{2TP+FP+FN}$
7	Observed agreement	%(Overall Accuracy)
8	Chance agreement	(%[TP + FP] × %[TP + FP]) + (%[TP + FP] × %[TP + FP])
9	Kappa coefficient	$\frac{(Observed\ agreement - Chance\ agreement)}{(100 - Chance\ agreement)}$

Abbreviations: FN, false negative; FP, false positive; TN, true negative; TP, true positive.

TABLE 6 Evaluation of FIS with PIDD

Measures	Values
Accuracy	98.87
Precision	0.9586
Recall	0.9413
F-measure	0.9418
Number of rules	30
Mean length of rules	6.2

Data collection module: Initially, data are collected from the UCI machine learning repository for processing the proposed framework.

Preprocessing module: The preprocessing module is implemented to convert the raw data into categorical data for effective implementation. Data preprocessing includes data purification, missing data, noisy data, data transformation, data reduction, and so on.

Training and testing module: We created and tested the prediction model using the inputs mentioned above and

TABLE 7 The outcomes of our proposed model are compared with the results of actual diagnostics

S. no.	Pregnancies	Glucose	Blood pressure	Skin thickness	Insulin	BMI	Diabetes pedigree function	Age	Outcome	Proposed model—risk factor
1	6	148	72	35	0	33.6	0.627	50	1	0.044
2	1	85	66	29	0	26.6	0.351	31	0	0.067
3	8	183	64	0	0	23.3	0.672	32	1	0.053
4	1	89	66	23	94	28.1	0.167	21	0	0.016
5	0	137	40	35	168	43.1	2.288	33	1	0.124
6	5	116	74	0	0	25.6	0.201	30	0	0.078
7	3	78	50	32	88	31	0.248	26	1	0.189
8	10	115	0	0	0	35.3	0.134	29	0	0.082
9	2	197	70	45	543	30.5	0.158	53	1	0.178
10	8	125	96	0	0	0	0.232	54	1	0.078
11	4	110	92	0	0	37.6	0.191	30	0	0.045
12	10	168	74	0	0	38	0.537	34	1	0.165
13	10	139	80	0	0	27.1	1.441	57	0	0.078
14	1	189	60	23	846	30.1	0.398	59	1	0.025
15	5	166	72	19	175	25.8	0.587	51	1	0.065
16	7	100	0	0	0	30	0.484	32	1	0.089
17	0	118	84	47	230	45.8	0.551	31	1	0.012
18	7	107	74	0	0	29.6	0.254	31	1	0.176
19	1	103	30	38	83	43.3	0.183	33	0	0.067
20	1	115	70	30	96	34.6	0.529	32	1	0.187
21	3	126	88	41	235	39.3	0.704	27	0	0.092
22	8	99	84	0	0	35.4	0.388	50	0	0.156
23	7	196	90	0	0	39.8	0.451	41	1	0.068
24	9	119	80	35	0	29	0.263	29	1	0.187
25	11	143	94	33	146	36.6	0.254	51	1	0.028
26	10	125	70	26	115	31.1	0.205	41	1	0.076
27	7	147	76	0	0	39.4	0.257	43	1	0.089
28	1	97	66	15	140	23.2	0.487	22	0	0.178
29	13	145	82	19	110	22.2	0.245	57	0	0.067
30	5	117	92	0	0	34.1	0.337	38	0	0.045

outputs. The 768 clinical datasets were separated into training sets (70%) (Total subjects: 537) and testing sets (30%) (Total subjects: 231).

Recommender module: It recommends food habits and exercise, including the medicines that are to be consumed by patients with diabetes, based on calculating their level of insulin, BMI, blood pressure, and plasma glucose.

The recommendations related to diabetes risk are mentioned in Table 4.[44]

The generic recommendations for diabetes are indicated in Figure 8.

6 | EXPERIMENTAL RESULTS

The performance measures of our proposed IFIR_PDDM model are evaluated using Python (Python Software Foundation, Wilmington, Delaware). The proposed method is implemented on a computer system having a CPU of Intel Pentium 1.9 GHz, a 64-bit operating system, Microsoft Windows 10, 4 GB of RAM, and Java JDK 1.8.

6.1 | Performance measures

Measures used to examine the outcomes of the FID_PSM model were precision, recall, accuracy, *F*-score, and kappa. The formula used to compute the measures is shown in Table 5.

The precision, recall, and *F*-measure of certain well-known algorithms are considered for evaluation. FIS achieves the best precision, recall, and *F*-measure among numerous well-known classification algorithms, as seen in this graph. Although there is no precise definition of the interpretability of fuzzy categorization systems, the number of rules and the average length of rules are frequently stated as two important aspects. Table 6 displays the full set of findings acquired with FIS for the abovementioned criteria. According to Table 3, the high classification rate of FIS was obtained using only 30 fuzzy if–then rules with a mean length of 6.2, indicating that while the final identified classification system is accurate, it is also interpretable.

Table 8 shows the final results from the IFIR_PDDM using patient feature vectors to assess diabetes risk. The disparity between actual diagnosis results and those predicted by our model is also represented in Table 7.

At the outset, the scalability of the proposed methodology remains unchanged for applications with the same inputs and features. With increasing input load, the algorithmic scalability must change priorities. This anticipated methodology is an off-the-shelf model, thus ensuring its scalability for similar input loads.

TABLE 8 Comparison of results

Methodology adopted	Accuracy (%)
Proposed model [IFIR_PDDM]	98.87
Semantic fuzzy ontology[8]	95.00
PCA and PSO[18]	93.66
FL + basal metabolic[45]	85.00
FL + association rule[46]	84.37
ANFIS-MLM[47]	82.30
ANFIS-LSGD[48]	81.48
ABC boosting[49]	80.16
FL + domain ontology[9]	80.00
GA + PBN[50]	77.71
KMP + PSO[51]	73.07
FR + domain ontology[7]	72.00
Intelligent ontology agent[6]	65.00

6.2 | Comparison of results based on performance measures

We compared the diabetes risk percentages generated from actual patient diagnoses in the EHR dataset to those obtained using the proposed IFIR_PDDM and other research efforts. Table 8 summarizes the findings and comparisons of various research projects.

As shown in Table 8, the accuracy rate of our suggested model is 98.87%, which is greater than the accuracy rate of the other models. Because those models lack medical guidelines or expert advice, the research results with which they are compared are less accurate. On the other hand, our proposed approach is built in compliance with medical criteria and with the input of medical experts. As a result, the degree of accuracy is determined using the given guideline. Due to restrictions in the tree composition and the changeable configurations of the classification criteria, the decision tree cannot provide high accuracy. Verification by medical specialists is required to improve the accuracy of these models. The result reached using solely the decision tree technique was 43.1%, and the created rules were not free of uncertainty. In this work, the accuracy is higher than the other research results because uncertainty is decreased through consultation with medical specialists. A higher level of accuracy is necessary for diabetes risk prediction. A small-scale dataset (768 datasets) is used to develop the proposed model. As a result, the proposed model must be evaluated to improve the accuracy of diabetes risk prediction with large-scale datasets.

7 | CONCLUSION

CDSS makes it easier to use computers in sophisticated medical decision-making to prevent diabetes disease in patients. In this work, we presented an IFIR_PDDM for use with CDSS. The proposed model is knowledge-based and was developed with the help of two medical specialists and medical guidelines. To improve prediction accuracy in the face of uncertainty, we developed knowledge- and mining-based rules. We used the decision tree technique to generate rules that were compliant with medical expert consultation. We utilized EHR data from a single source repository (PID dataset) to implement and evaluate the performance of the IFIR_PDDM. Our proposed model's results are shown to be more accurate than those of other researchers. In addition, our proposed model suggested that patients' relevant content recommendations for diabetes disease prevention (normal living, nutrition control, exercise, and medications) be added to their relevant content. In this article, we demonstrated the value of combining medical expert information with data mining techniques to overcome the uncertainty in the prediction model by applying medical expert knowledge to avoid diabetes disease. IFIR_PDDM integration with CDSS will aid doctors in predicting the risk of diabetes disease in patients. Furthermore, it will make it easier for patients to manage their health by allowing them to see their diabetes disease risks, thereby aiding in the prevention of diabetes. In the future, researchers will examine a huge dataset to improve the accuracy of diabetes risk prediction.

ACKNOWLEDGMENT

The authors would like to thank Kalasalingam Academy of Research and Education for permission to conduct the research and providing the computational facilities.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in UCI Machine Learning Repository at https://www.kaggle.com/uciml/pima-indians-diabetes-database [26].

ORCID

Palanigurupackiam Nagaraj https://orcid.org/0000-0003-1803-8544

REFERENCES

1. Katsiki N, Steiropoulos P, Papanas N, Mikhailidis DP. Diabetes mellitus and chronic obstructive pulmonary disease: an overview. *Exp Clin Endocrinol Diabetes*. 2021;129(10):699-704.
2. CDC. National Diabetes Statistics Report; 2020. Accessed December 4, 2021. https://www.cdc.gov/diabetes/data/statistics-report/index.html
3. IDF. Facts & figures; 2020. Accessed December 4, 2021. https://idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html
4. Shi X, He J, Lin M, et al. Comparative effectiveness of team-based care with a clinical decision support system versus team-based care alone on cardiovascular risk reduction among patients with diabetes: rationale and design of the D4C trial. *Am Heart J*. 2021;238:45-58.
5. Yang LH, Ye FF, Liu J, Wang YM, Hu H. An improved fuzzy rule-based system using evidential reasoning and subtractive clustering for environmental investment prediction. *Fuzzy Set Syst*. 2021;421:44-61.
6. Lee CS, Wang MH, Li HC, Chen WH. Intelligent ontological agent for diabetic food recommendation. In 2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence; 2008:1803–1810.
7. Chen SM, Huang YH, Chen RC, Yang SW, Sheu TW. Using fuzzy reasoning techniques and the domain ontology for anti-diabetic drugs recommendation. In Asian Conference on Intelligent Information and Database Systems; 2012:125–135.
8. El-Sappagh SH, 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.
9. Chen RC, Huang CY, Bau CT, Chen LS. A decision support system for diabetes medicine selection using patient-centeredtreatment based on fuzzy logic and domain ontology. *Int J Innov Comput Inform Control*. 2017;13(5):1681-1692.
10. Alonso JM, Ramos-Soto A, Reiter E, van Deemter K. An exploratory study on the benefits of using natural language for explaining fuzzy rule-based systems. In 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE); 2017:1–6.
11. Gupta A, Dubey SK. Fuzzy nutrition recommendation system for diabetic patients. In: Choudhury S, Mishra R, Mishra RG, Kumar A, eds. *Intelligent Communication, Control and Devices*. Springer; 2018:1381-1390.
12. Ali F, Islam SR, Kwak D, et al. Type-2 fuzzy ontology–aided recommendation systems for IoT-based healthcare. *Comput Commun*. 2018;119:138-155.
13. Rajeswari AM, Sidhika MS, Kalaivani M, Deisy C. Prediction of Prediabetes using fuzzy logic based association classification. In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT); 2018: 782–787.
14. Mohammed HA, Hagras H. Towards developing type 2 fuzzy logic diet recommendation system for diabetes. In 2018 10th Computer Science and Electronic Engineering (CEEC); 2018: 56–59.
15. Alian S, Li J, Pandey V. A personalized recommendation system to support diabetes self-management for American Indians. *IEEE Access*. 2018;6:73041-73051.
16. El-Sappagh S, Elmogy M, Ali F, Kwak KS. A case-base fuzzification process: diabetes diagnosis case study. *Soft Comput*. 2019;23(14):5815-5834.

17. Gomes Filho E, Pinheiro PR, Pinheiro MCD, Nunes LC, Gomes LBG. Heterogeneous methodology to support the early diagnosis of gestational diabetes. *IEEE Access*. 2019;7:67190-67199.
18. Choubey DK, Kumar P, Tripathi S, Kumar S. Performance evaluation of classification methods with PCA and PSO for diabetes. *Netw Model Anal Health Inform Bioinform*. 2020;9(1):5.
19. Singh N, Singh P. Stacking-based multi-objective evolutionary ensemble framework for prediction of diabetes mellitus. *Biocybern Biomed Engin*. 2020;40(1):1-22.
20. Rani P, Mishra AR, Mardani A. An extended Pythagorean fuzzy complex proportional assessment approach with new entropy and score function: application in pharmacological therapy selection for type 2 diabetes. *Appl Soft Comput*. 2020;94:106441.
21. Liu C, Mo H, Wang FY. Analysis and control of blood glucose situation for diabetic patients based on interval type-2 fuzzy sets. *Int J Fuzzy Syst*. 2021;23:1179-1193.
22. Omisore OM, Ojokoh BA, Babalola AE, et al. An affective learning-based system for diagnosis and personalized management of diabetes mellitus. *Fut Generat Comput Syst*. 2021;117:273-290.
23. Melton GB, McDonald CJ, Tang PC, Hripcsak G. Electronic health records. In: Shortliffe E H, Cimino J J, eds. *Biomedical Informatics*. Springer; 2021:467-509.
24. Omana J, Moorthi M. Predictive analysis and prognostic approach of diabetes prediction with machine learning techniques. *Wirel Pers Commun*. 2021;117:1-14.
25. Fang HSA, Tan TH, Tan YFC, Tan CJM. Blockchain personal health records: systematic review. *J Med Internet Res*. 2021;23(4):e25094.
26. Pima Indians Diabetes Database [Online]. Accessed March 20, 2021. https://www.kaggle.com/uciml/pima-indians-diabetes-database
27. Thakkar H, Shah V, Yagnik H, Shah M. Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis. *Clin eHealth*. 2020;4:12-23.
28. Ghosh SK, Ghosh A. A novel clustering-based gene expression pattern analysis for human diabetes patients using intuitionistic fuzzy set and multigranulation rough set model. *Soft Computing: Theories and Applications*. Springer; 2020:979-991.
29. Lin TC, Li CY, Chen PF, et al. Identifier based intelligent blood glucose concentration regulation for type 1 diabetic patients: an adaptive fuzzy approach. *J Intell Fuzzy Syst*. 2020;38(5):6175-6184.
30. Qasem SN, Mohammadzadeh A. A deep learned type-2 fuzzy neural network: singular value decomposition approach. *Appl Soft Comput*. 2021;105:107244.
31. Tabakov M, Chlopowiec A, Chlopowiec A, Dlubak A. Classification with fuzzification optimization combining fuzzy information systems and type-2 fuzzy inference. *Appl Sci*. 2021;11(8):3484.
32. Titi S, Elhadj HB, Fourati LC. A fuzzy-ontology based diabetes monitoring system using internet of things. In International Conference on Smart Homes and Health Telematics; 2020:287–295.
33. Shoaip N, El-Sappagh S, Barakat S, Elmogy M. A framework for disease diagnosis based on fuzzy semantic ontology approach. *Int J Med Eng Inform*. 2020;12(5):475-488.
34. Qasim I, Alam M, Khan S, et al. A comprehensive review of type-2 fuzzy ontology. *Artif Intell Rev*. 2020;53(2):1187-1206.
35. Haque F, Reaz MB, Chowdhury ME, Hashim FH, Arsad N, Ali SH. Diabetic sensorimotor polyneuropathy severity classification using adaptive neuro fuzzy inference system. *IEEE Access*. 2021;9:7618-7631.
36. Chen T, Shang C, Su P, et al. A decision tree-initialised neuro-fuzzy approach for clinical decision support. *Artif Intell Med*. 2021;111:101986.
37. Permana BAC, Ahmad R, Bahtiar H, Sudianto A, Gunawan I. Classification of diabetes disease using decision tree algorithm (C4. 5). *J Phys Conf Ser*. 2021;1869:012082.
38. Muthamil Sudar K, Deepalakshmi P. A two level security mechanism to detect a DDoS flooding attack in software-defined networks using entropy-based and C4. 5 technique. *J High Speed Netw*. 2020;26(1):55-76.
39. Khairuddin SH, Hasan MH, Hashmani MA, Azam MH. Generating clustering-based interval fuzzy type-2 triangular and trapezoidal membership functions: a structured literature review. *Symmetry*. 2021;13(2):239.
40. Nagaraj P, Deepalakshmi P, Romany FM. Artificial flora algorithm-based feature selection with gradient boosted tree model for diabetes classification. *Diabetes Metab Syndr Obes*. 2021;14:2789.
41. Ramsingh J, Bhuvaneswari V. An integrated multi-node Hadoop framework to predict high-risk factors of diabetes mellitus using a multilevel MapReduce based fuzzy classifier (MMR-FC) and modified DBSCAN algorithm. *Appl Soft Comput*. 2021;108:107423.
42. Chen RC, Jiang HQ, Huang CY, Bau CT. Clinical decision support system for diabetes based on ontology reasoning and TOPSIS analysis. *J Healthc Eng*. 2017;2017:4307508.
43. Sharma D, Singh Aujla G, Bajaj R. Deep neuro-fuzzy approach for risk and severity prediction using recommendation systems in connected health care. *Trans Emerg Telecommun Technol*. 2020;e4159:1-18.
44. American Diabetes Association. *Introduction: Standards of Medical Care in Diabetes—2020*. ADA; 2020.
45. Porksen NK, Linnebjerg H, Lam ECQ, et al. Basal insulin peglispro increases lipid oxidation, metabolic flexibility, thermogenesis, and ketone bodies compared to insulin glargine in subjects with type 1 diabetes mellitus. *Diabetes Obes Metab*. 2018;20(5):1193-1201.
46. Lu PH, Keng JL, Tsai FM, Lu PH, Kuo CY. An apriori algorithm-based association rule analysis to identify acupoint combinations for treating diabetic gastroparesis. *Evid Based Complement Alternat Med*. 2021;2021:1-9.
47. Sagir AM, Sathasivam S. Design of a modified adaptive neuro fuzzy inference system classifier for medical diagnosis of Pima Indians diabetes. *AIP Conf Proc*. 2017;1870:040048.
48. Sagir AM, Sathasivam S. A hybridised intelligent technique for the diagnosis of medical diseases. *Pertanika J Sci Technol*. 2017;25(2):424-440.
49. Lauffenburger JC, Mahesri M, Choudhry NK. Not there yet: using data-driven methods to predict who becomes costly

among low-cost patients with type 2 diabetes. *BMC Endocr Disord*. 2020;20(1):1-10.

50. Karegowda AG, Manjunath AS, Jayaram MA. Application of genetic algorithm optimized neural network connection weights for medical diagnosis of Pima Indians diabetes. *Int J Soft Comput*. 2011;2(2):15-23.
51. Zhang Y, Lin H, Yang Z, et al. Neural network-based approaches for biomedical relation classification: a review. *J Biomed Inform*. 2019;99:103294.

How to cite this article: Nagaraj P, Deepalakshmi P. An intelligent fuzzy inference rule-based expert recommendation system for predictive diabetes diagnosis. *Int J Imaging Syst Technol*. 2022;1-24. doi:10.1002/ima.22710