Clinical eHealth 4 (2021) 12–23

Contents lists available at ScienceDirect

Clinical eHealth

journal homepage: ww.keaipublishing.com/CEH

Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis

Harshil Thakkar [a], Vaishnavi Shah [a], Hiteshri Yagnik [b], Manan Shah [c,*]

[a] Department of Computer Engineering, LDRP Institute of Technology and Research, Gandhinagar, Gujarat, India
[b] Gujarat Info Petro Limited (GIPL), Gandhinagar, Gujarat, India
[c] Department of Chemical Engineering, School of Technology, Pandit Deendayal Petroleum University, Gandhinagar, Gujarat, India

ARTICLE INFO

Article history:
Received 5 October 2020
Revised 30 October 2020
Available online 17 November 2020

Keywords:
Data mining
Fuzzy logic
Diabetes
Health care

ABSTRACT

Diabetes is an ailment in which glucose level increase in at high rates in blood due to body's inability to metabolize it. This happens when body does not produce sufficient amount of insulin or it does not respond to it properly. Critical and long-term health issues arise if diabetes is not handled or properly treated which includes: heart problems, disorders of the lungs, skin and liver complications, nerve damage, etc. With increasing number of diabetic patients, its early detection becomes essential. In this paper, our major focus areas are data mining and fuzzy logic techniques used in diabetes diagnosis. Data mining is used for locating patterns in huge datasets using a composition of different methods of machine learning, database manipulations and statistics. Data mining offers a lot of methods to inspect large data considering the expected outcome to find the hidden knowledge. Fuzzy logic is similar to human reasoning system and hence it can handle the uncertainties found in the data of medical diagnosis. These systems are called expert systems. The fuzzy expert systems (FES) analyze the knowledge from the available data which might be vague and suggests linguistic concept with huge approximation as its core to medical texts. In this paper, the methodology section delivers the pipeline of various tasks such as selecting the dataset, preprocessing the data by applying numerous methods such as standardization, normalization etc. After that, feature extraction technique is implemented on the dataset for improving the accuracy and finally dataset worked on data mining and fuzzy logic various classification algorithms. While analyzing different data mining methods, the accuracy computed through random forest classifiers as high as 99.7% and in case of numerous fuzzy logic approaches, high precision and low complexity was found to contribute a fairly high accuracy of 96%.

Contents

1. Introduction . . . 13
2. General implementation of technologies in diabetes detection . . . 14
3. Methodology . . . 14
   3.1. Dataset description . . . 14
   3.2. Data preprocessing . . . 14
   3.3. Classification algorithm . . . 15
4. Elaborate evaluation of data mining techniques in diabetes prognosis . . . 15
5. Elaborate Evaluation of Fuzzy Logic Technologies in Diabetes Prognosis . . . 19
6. Future scope and limitations . . . 21
7. Conclusion . . . 21
   Author contributions . . . 21
   Declaration of Competing Interest . . . 21
   Acknowledgements . . . 21

* Corresponding author.
E-mail address: manan.shah@spt.pdpu.ac.in (M. Shah).

https://doi.org/10.1016/j.ceh.2020.11.001
2588-9141/

Availability of data and material . . . 21
Funding . . . 22
Consent for publication. . . . 22
Ethics approval and consent to participate . . . 22
References . . . 22

1. Introduction

Health care is a profound sector with a wide scope of indeterminate improvement. Health not only depicts the presence or absence of a disease, but it is also a measure of a person's overall well-being. Health delineates the maladies as well as proper functioning of a human across multiple domains making health care a crucial domain for computer science. Healthy living is clearly one of human beings' most important goals and health is among the basic abilities that make human life worthwhile. One of the major areas of concern in health care is diabetes and its detection. With more and more people leading a sedentary lifestyle and poor dietary habits, diabetes has become a common dread. However, it can also be caused due to genetics factor. Diabetes is an ailment in which glucose levels increase in at high rates in blood due to body's inability to metabolize it. This happens when body does not produce sufficient amount of insulin or it does not respond to it properly. This condition is known as hyperglycemia. It is incurable and the only way to deal with this disease is by controlling it.[1] With increasing number of diabetic patients, its early detection becomes essential. Researchers believe that the data related to diabetes can help in its early detection and it can also convey that how likely someone is to get diabetes in future (see Fig. 1).[2,3]

There have been major breakthroughs in the field of health care owing to newer technologies used in this sector. One of these technologies is Machine Learning. Machine Learning is one of the most propitious tool to refine the existing technologies in this era which is evidently moving towards a fourth industrial revolution, aiding expertise like IOT, automation, artificial intelligence, data mining, neural networks and many more.[4] Data mining is an important branch of machine learning and exists as an integral part under its umbrella. Data mining is used for locating patterns in huge datasets using a composition of different methods of machine learning, database manipulations and statistics. It involves computer and statistics discipline into one field that is used for extracting data patterns from dataset and it uses data for further use. These patterns can be deterministic factors for major predictions and can also act as an observational likelihood measurement for the data of the dataset. Data mining is being utilized in many applications, but it is progressively getting famous in medical services. Abundant data without correct processing and distinguishment can prove to be useless. Data mining gives innovation to actuate and transform data into helpful information for decision making that is an essential piece of medical service. It enables doctors to identify and give powerful medications and patients to acquire better and more affordable medical services.[5] From data mining it has a high expectation for predicting illnesses in therapeutic administrations industry.[6] The aim of data mining is to induce sense in the unorganized and unsupervised clusters in any domain. Data mining is the process of finding useful data patterns and knowledge that requires knowledge discovery in databases (KDD).[7]

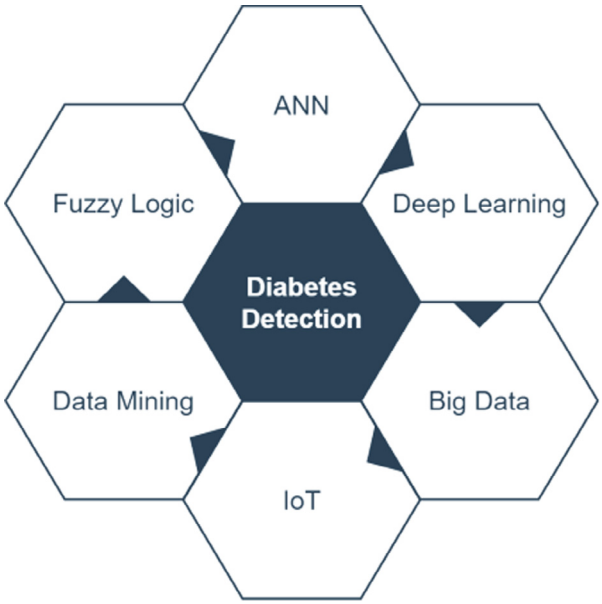

Fig. 1. Various technologies used in diabetes detection.

Artificial intelligence is another technology that is widely used in the field of health care and medical diagnosis. Some techniques in soft computing which is an interdisciplinary branch of computer science and artificial intelligence include neural networks like deep learning and artificial neural networks and fuzzy logic. These approaches have the supremacy of solving ambiguous problems that contain vagueness included in this field.[8] Fuzzy logic is similar to human reasoning system and hence it can handle the uncertainties found in the data of medical diagnosis. It is an extension of boolean logic based of the degrees of truth lying between 0 and 1 or true/false. Fuzzy logic is based on the theme of degree, inaccuracy, linguistic and observation. An expert system undertakes the task to personify the expertise of a human being that functions and makes decisions without any personal propensity, that is, complete unbiasedness.[8–10] The fuzzy expert system (FES) analyzes the knowledge from the available data which might be vague and suggests linguistic concept with huge approximation as its core to medical texts. Fuzzy logic makes the imprecise data precise and solid. FES shows a vital role in medicine for symptomatic diagnostic cures.[11] Hence, fuzzy logic is a computation method that is based on linguistic language, that is, natural language that is used by humans for logical reasoning.[12] It is quite evident that data mining and fuzzy logic prove to be promising approaches for medical diagnosis.

Considering the size of healthcare databases, it makes the task of finding or discovering the hidden knowledge a priority in order to establish diagnostic effectiveness. By early detection of diabetes one can not only be saved from that ailment but can also be saved from other major side effects that follow. The existing solutions have definitely helped the cause. However, in order to reach an ideal result, we must traverse through the different ways that data mining and fuzzy logic techniques can be used for the given purpose along with drawing a comparative analysis of both. By doing so we can map the existing solutions to the real world and comprehend what is best for the prognosis along with keeping in track the chronological development in the approaches used. This side by side analysis would not only help the researchers get a complete

overview of the different propositions but also help doctors and physicians put in action a real time system for diabetes detection with maximum accuracy. Hence, in this paper we will discuss in detail both the approaches by carrying out a comparative study.

2. General implementation of technologies in diabetes detection

Many approaches and algorithms governing a lot of different technologies have been used to detect diabetes. In this section we will be discussing some of these approaches. Staring with artificial neural networks (ANN), which is a human brain based computational model that simulate the concept of neurons present in the brain. It has decision making capabilities similar to a human brain and hence its major application is diagnosis and prediction. ANNs help the doctors in diagnosis and influences their decisions with high accuracy which in turn increases the confidence. The parallel processing capabilities of ANN makes it efficient in identifying complex patterns.[13] Using ANN in diabetes detection is sometimes tedious as a single layer network won't be able to draw accurate predictions and hence non-linear part must be integrated which can be done by forming a multi-layer complex structure. This causes time delay due to high processing.

Another technology that is majorly based on ANN is deep learning. Deep learning aims on simulating the human brain in the hope that one day it will act as a virtual replication of the brain itself and transpose it into systems.[14] DNNs (deep neural network) have been used comprehensively in problems of classification owing to its exceptional results in classification showing prominent growth in artificial intelligence.[15] Some of the deep learning algorithms like support vector machine (SVM) requires high speed and large training as well as test sets.[16]

Hardware based devices are also built for this purpose which deploy internet of things (IOT) as its core technology. IOT along with deep learning neural networks for data processing is used to build a diabetes risk analysis device that may indicate a person's potential to have diabetes according to various parameters that are acquired using sensors like glucometer sensor; feet pressure sensors, blood pressure sensors, etc.[17] However, the integration part becomes strenuous. One can also find it difficult to incorporate all the essential parameters that govern this decision and may result in improper predictions due to the lack of data.

Predictive systems often rely on large amount of data and so do the machine learning or artificial intelligence algorithms that are used in such systems. Therefore, here comes in the picture big data and its analysis. As the term speaks for itself, big data is a technology that deals with excessive amount of data efficiently and effectively. The incomplete diabetes datasets were found to be major shortcoming for ANNs and DL.[18] This is overcome by big data analysis. Big data organizes the haphazard data which makes processing of this data comparatively easy.[19] Some of the challenges faced by big data are large memory requirements and privacy issues regarding patient's personal information.[20]

If we talk about Data Mining; it offers a lot of methods to inspect large data considering the expected outcome to find the hidden knowledge. The knowledge is vividly based on the relationships between the variables and the amount of dependency or effect they have on the outcome. This knowledge base is a repository for the decision-making process. However, the disparity among different cases and in datasets can prove disadvantageous while working with data mining as data mining does not leave a scope for vagueness. Fuzzy logic, on the other hand, offers an advantage of being less expensive in terms of development in comparison to other technologies that we discussed earlier. Drawback encountered while using fuzzy logic is that the transformation of fuzzy values to crisp real-world values may cause in a deterioration of precision.

3. Methodology

3.1. Dataset description

The below table [Table 1] refers to the list of publicly available diabetes datasets that contain independent features such as various symptoms of diabetes, and a dependent feature that indicate whether the individual is diabetic or not. In the table above, there are numerous fields such as dataset titles, papers that used this data collection and a concise description regarding the datasets are listed. In most of the research, Pima Indian Dataset is cited by the majority of authors from the above table.

3.2. Data preprocessing

Generally, a diabetes database includes noises, missing values, and perhaps in an unusable format which cannot be used directly for machine learning models. Feature pre-processing is necessary to clean up the data and make it suitable for the prediction model, which also makes a prediction model more accurate and more efficient. Pre-processing method includes details that involve zero values management, standardization, categorical variables management, one-hot coding and multiple-linearity.

Feature pre-processing tends to demonstrate that certain unrelated samples do not support or even decrease the detection precision. It is a widely used method that removes irrelevant data characteristics and reduces predictive pattern time complexity.[26] The consistency of the model depends primarily on the data input into the model. If data is obtained from processes of data mining and fuzzy logic, some of the details (which we term missing values) are incomplete. It is also particularly prone to noise control.

Feature extraction is the function pre-processing technique, which extracts and/or blends variables into features and decreases efficiently the volume of data that needs to be descriptive and

Table 1
Publicly available diabetes dataset.

Sr. No.	Database name	Database used by authors	Database description
1	Pima Indians Diabetes database[21]	5, 7, 33, 36, 40, 42, 46, 52, 54, 55, 59, 60, 62, 66	This data collection comes from the Diabetes and Digestive and Kidney Institute. The aim is to determine when a patient has diabetes based on diagnostic tests.
2	LARS Diabetes dataset[22]	57	The response component, a disease progression indicator one year after baseline, was requested for prediction model for442 patients with diabetes dependent on 10 baseline variables.
3	Diabetes Dataset[23]		The dataset containing 615 samples with 8 independent variables namely pregnant, glucose, pressure, triceps, insulin, mass, pedigree, age to discover a patient is diabetic or not.
4	Abel Vikas's Diabetes Dataset[24]		The dataset containing 615 samples with 7 independent variables namely age, bs_fast, bs_pp, plasma_rr, plasma_f, hb1ac, type, class to discover a patient isdiabetic or not.
5	Early stage diabetes risk prediction dataset[25]		This dataset of 16 independent variables includes newly diabetic sign and symptom data or is diabetic.

non-redundant, defining the initial data set in its entirety. Many methods of extraction involve linear transformations of the new vectors with less dimensionality are the original pattern vectors. The extraction technique translates or projects the original function Vectors in a reduced vector field improving the power of the predictive model where class differentiation is maximized. This process, however, has a few drawbacks. First, the features are not informed about the extraction process and second, the excessive area of inequalities in datasets is also a restriction. The findings can thus be inaccurate, contributing to a reduction in exactness.[27]

Neural networks are often susceptible to under-fitting or overfitting of data, which is substantially due to lack of selection of relevant features. An optimal subset of the existing features can help mitigate this problem.[28] This selection of relevant features is termed as feature selection. Moreover, the imbalance nature of the data can adversely affect the constructed machine learning models. These models were found to have higher bias and an increased rate of misclassification of data points in test data which can be alleviated using methods like resampling. However, to further increase the classification accuracy subset of apposite attributes feature selection can be used.[29]

Data-mining process for pre-processing features removes noisy, non-relevant features through the use of data input and improves classification accuracy, and the time complexity of learning models reduces if a predictive model trained on features less, the method of choosing a sub-set of specific features for use in the model building is often called feature selection. The extraction of features in feature pre-processing is a method of dimensional reductions that reduces the original set of raw features to more manageable application classes in data mining approach. There are two major styles of linear and non-linear characteristic extraction algorithms in data mining approach.[30]

It is clear that the key difference between feature selection and feature extraction is that, in feature selection the labeled features are not mutated, instead few of the essential ones are selected for the model, while in feature extraction new features are created stimulated by the original features. Neuro-fuzzy system (NFS) is a fuzzy based approach that can be used for feature selection.[31] Another algorithm that aims at dimensionality reduction is based on fuzzy rough sets which use the notion of feature significance, selecting simultaneously those features and extracting them. It focuses on relevance as well as significance of the features.[32]

3.3. Classification algorithm

After successfully pre-processing the data, the next step involves the core algorithms that are used to classify the data, which in this case is classification of diabetic and non-diabetic patients. The database is divided into two parts: the training set and the test set. This database is used on the various proposed Data mining techniques to diagnose diabetes after the operation is carried out. Data mining techniques consists of various algorithms such as SVM algorithms, decision tree algorithms, KNN algorithm, clustering algorithm etc. It consists of association rule mining (ARM) association rules, as the name implies, are simple if/then declarations that contribute to discovering connections between seemingly independent relation databases or other data repositories. It is divided into two parts: the previous and therefore Figs. 2 and 3.

If we talk about fuzzy logic approach a generalized after the fore step of pre-processing the data that helps clean the data set, the values of the dataset are converted into fuzzy values which then become fuzzy model appropriate and manipulations can be carried out on these values. Feature reduction which is optional can be carried out to decrease the complexity of the model and hence increase the results. The knowledge is acquired which helps build fuzzy rules for the model that is the most crucial processing unit of the model. An alternative approach for building the fuzzy rules includes ontology construction that well establishes the relationships between the features or attributes of the dataset. Finally, the fuzzy logic algorithm is applied and prospect is classified as diabetic or non-diabetic.

After applying various techniques, the result is carried out in form of Accuracy which is forwarded for performance validation stage. In this stage, the performance matrix and various conclusion matrix are generated and verified for building an optimized model. The optimized model is computed by calculating accuracy. This flow is iterative until it gets most accurate results. The results are represented in the form of Accuracy, which is sent for the performance validation stage, once applied for various techniques. At that stage, optimized models are developed and checked for the performance matrix and the different conclusion matrix. Calculating precision is used to measure the optimized model. The flow is iterative until accurate results are obtained.

4. Elaborate evaluation of data mining techniques in diabetes prognosis

Butwall and Kumar (2015) [Table 2], [Fig. 4] created a data mining approach for detection of diabetes milletus using random forest classifier. The system predicting attributes are sex, age, blood pressure and blood sugar and for one attribute named the chance of diabetic patient suffering heart attack is obtaining by Naïve Bayes classifier. Here the data we had collected are processed in data mining method using WEKA (Waikato Environment for Knowledge Analysis) Tool consisting of 10 cross fold validation. In this approach the classifier is used is an ensemble of individual large tree-shaped classifiers. Here the data passes to every decision tree as an input array and in each tree the voting is done for each elements of vector and on base of that this classifier predicts the result. By implementing confusion matrix for this system, it found out 99.7% accuracy with better outperformance. This approach's accuracy can be intensified by implementing hybrid classifications algorithm.[5]

A similar approach to the one used in the above-mentioned model was used by Kumari and Singh (2012) in their method for early detection of diabetes by using the Backpropagation neural network with the aim of building a quick and beneficial method for an automated diabetes detection. Plasma glucose concentration, blood pressure, triceps skin fold, serum insulin, body mass index (BMI), diabetes pedigree function number of times a person was pregnant and age are considered as inputs of this method. Neural network is split into 28 nodes, 13 are secret nodes, one visible node and 13 of which are data nodes in which the inputs are provided with some value of weights and with a bias value. This approach contains 3 stages are network designing and training, Back propagation error and hidden layer error. This system considered the parameters as symptoms of diabetes diagnosis and it found 92.8% accurate as 18 out of 20 cases were detected correctly.[33]

A predictive approach was proposed by Sankaranarayanan and Perumal, (2014) which uses data mining techniques such as rule classification and decision trees for diabetes detection. This technique uses attributes as inputs are age, gender, level of glycated haemoglobin (HbA1C). The rule classification technique uses rule set classifier such as IF-THEN prediction rules to discover high preoccupation level knowledge which consisting of 2 parts: rule predecessor and rule successive. In this technique the IF part containing one or more conditions with prognosticator value and the THEN part containing prognostication value of goal attribute. The decision tree technique is also known as classification tech-

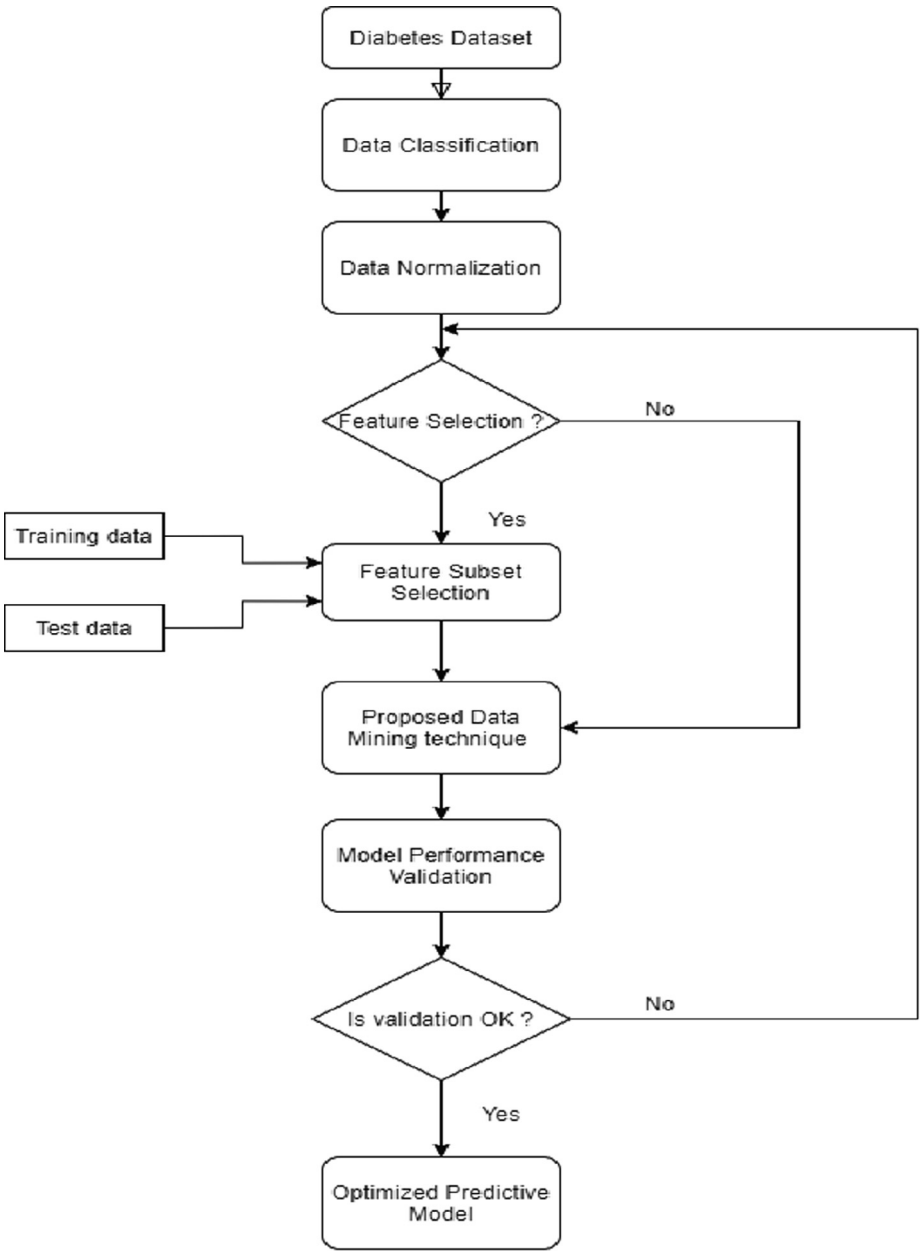

Fig. 2. Generalized structural schema for data mining processes.

nique that is implemented on data for decision making and knowledge representation. The result is carried on 768 instances along with 3 physical examination attributes and 2 classes. The patient occurs to be diabetic when micro albuminuria is appearing in them whose presence is traced by decision attribute.[34]

A similar concept to the one used in the above mentioned model was used by Iyer et al., (2015) in their approach describing diagnosis of diabetes in pregnant women by applying classification technique such as Naïve Bayes algorithm (probabilistic algorithm following the sequential steps of classification, valuation and projection) and J48 decision tree (classification method used for forecasting result in tree structure consisting of nodes) on existing dataset and forecast the result. The techniques used in this approach was applied on PIMA dataset which consisting of 12 features and 768 specimens. Initially, pre-processing and transformation method was applied on dataset followed by feature selection process finalizing the dataset for input. Here for following methods, the dataset is divided in 2 groups: tested positive and tested negative. In J48 decision tree, the dataset is split by two types: cross validation and percentage split technique and in Naïve Bayes algorithm, the dataset is split by percentage split technique. When results acquired, the small variance is seen in both method's error rate, the Naïve Bayes algorithm with 70:30 percentage split algorithm seems to be least error rate.[35]

Bai et al., (2019) presented a method proposing survey and diagnosis of diabetes using data mining techniques. The aim of the method was to survey the dataset and forecast the diabetes detection among the patients. The above-mentioned aim was achieved by data mining techniques such as gaussian Naïve Bayes, OPTICS (Ordering points to identify the clustering structure), BIRCH (balanced iterative reducing and clustering using hierarchies). Gaussian Naïve Bayes is described as they survey the dataset and it fragments the constant values into classes and predicts the diabetic status of patient. The OPTICS technique is known as

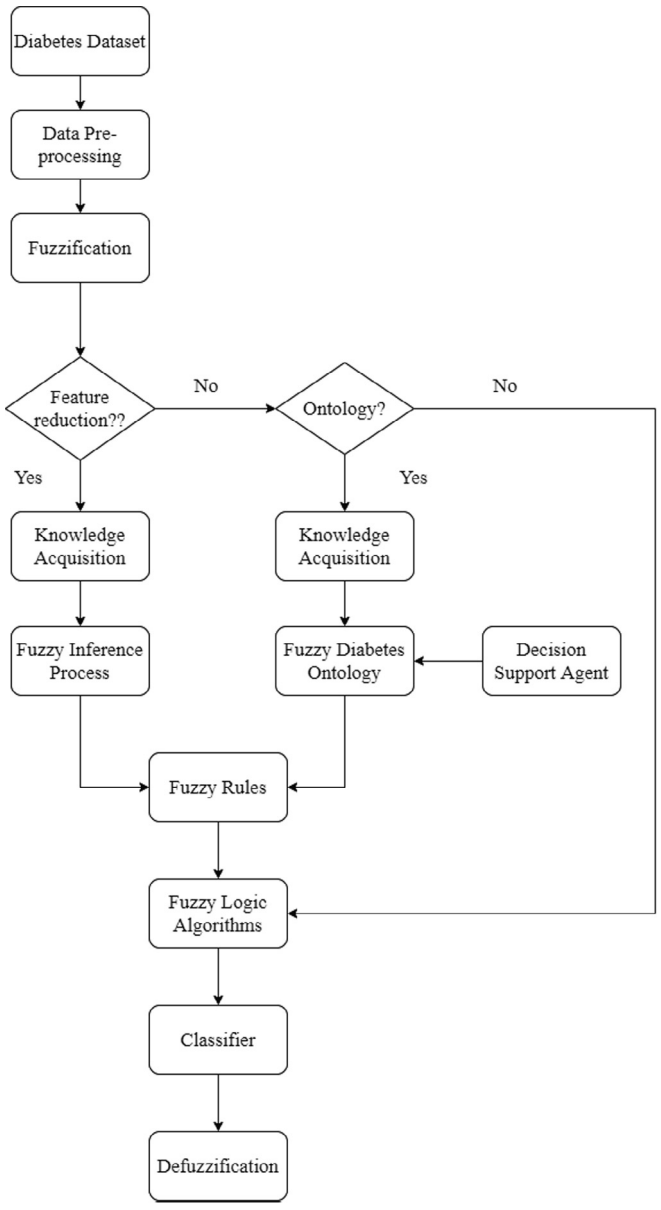

Fig. 3. Generalized structural schema for fuzzy logic processes.

clustering technique which was applied on database to group the datasets into 3 clusters named as standard, prediabetes and diabetes. BIRCH technique similar to OPTICS technique classified as hierarchical clustering which utilized aggregate characteristics and built a CF tree. It basically works on large datasets and the working was carried in 4 stages. The results were carried by different methods like silhouette method and performance metrics for accurate diabetes detection between OTICS and BIRCH technique and from both the method it is observed that OPTICS was working more efficiently and suitable technique.[36]

Padmaja et al. (2008) analyzed the features of diabetes in maximum number of women through data mining process by using clustering techniques. This analysis is referred in forecasting diabetes stages i.e. initial stage or advance stage. The clustering refers to partitioning the datasets in groups and in this technique; initially the data was partitioned into clusters. For determining the features of diabetes data, the technique consist of procedure where initially the data is provided as input to normalization technique after that, the normalized file works as input for clustering technique for generating set of clusters and the quality of clusters were surveyed with the help of attribute oriented induction technique. The clusters are also generated by other 4 techniques are K-Means, partitioning around medoids (PAM), minimum spanning tree (MST) and Nearest Neighbor algorithms. The data on which the algorithms are applied consisting of 268 women in target classes and when results is carried it seems that maximum (25%) women lies in cluster-4 (KNN-Neighbor) suffering from diabetes.[37]

Another data mining approach was discovered by Thangaraju and NancyBharathi (2014) for forecasting diabetes affection to different age group of people by utilizing feature selection technique. The knowledge is mined from dataset through data mining method through classification and clustering techniques which helps in classifying and allocating data. The feature selection is the choosing of a subset of the attributes for use in system construction. This technique offers three big advantages in the development of predictive models: enhanced sample interpretability, shortened exercise times, and increased overfitting generalization also it is expanding the classification accuracy. This approach uses look based technique such as best first search and greedy forward selection which assist in collecting data from database and updating data in database. Instead of using complex classifiers (using all features), the feature selection method is more advantageous to use as it is using subset of relevant features. The representation technique extracted the result but for more accuracy, more information should be provided to representation.[38]

Abed and Nasr (2019) proposed a model functioning as diabetes detection using data mining technique consisting of k-means clustering algorithm. The motive of this model is to exploring a knowledge which splits the data into clusters of diabetic and non-diabetic data. The model's dataset consisting of 2733 cases with 12 attributes. The proposed model comprises of 3 steps are: describing the dataset and its features, pre-processing the data, and implementing k-means clustering algorithm. As k-means algorithm is one of unsupervised learning and for this reason it is best to choose for clustering algorithm and it is repeated algorithm which divide in clusters repetitively. Pre-processing of dataset step includes number of actions: data cleaning, improving, normalization, transformation, and feature selection which assures data quality for prognosis of diabetes. After pre-processing step, the model's final training set is prepared which execute as input to k-means clustering algorithm. This model is built by partitioning into clusters with 4 repetitions in 0.05 s. This model is designed very efficiently for discovering the data intelligence about diabetes dataset.[39]

Sanakal and Jayakumari (2014) proposed a comparative analysis of discovering diabetes through data mining techniques between support vector machine (SVM) and fuzzy C-measns (FCM) clustering algorithm. The analysis consists of database have 10 attributes which comprises 9 input and 1 output attributes and it have 768 samples. FCM and K-Means algorithm are having same principles but only the difference is in FCM, each datapoint is split in clusters on basis of degree of membership grade. For splitting the dataset, the FCM uses minimized cost function and it is an iterative algorithm for computing new fuzzy centroids and distance measures. SVM consists of 2 types: linear and non-linear types. The linear SVM is a hyperplane / linear line splitting datapoints in 2 clusters with maximum margin. The classification algorithm separates dataset in 576 data-points in training dataset and 192 data-points in testing dataset. By implementing confusion matrix for this analysis, it found out that FCM implemented with 1.25 fuzziness co-efficient(m) with 94.3% accuracy and SVM is implemented with SMO training algorithm with accuracy of 59.5% which is quite low.[40]

With the objective of classifying the type of diabetes and its severity with respect to each patient, Kumar and Umatejaswi (2017) proposed a model that used data mining techniques for

Table 2
Accuracies of different data mining methods for comparative anatomization.

Sr. No	Method	Accuracy	Author
1	ADAP Learning algorithm	76%	Smith et al. (1988)[42]
2	flexible neural-fuzzy inference system (FLEXNFIS)	78.6%	Rutkowski and Cpalka (2003)[43]
3	Relational Fuzzy Neural Networks for Enhancing Pattern Classification	81.8%	Davis and kohout (2006)[44]
4	GDA-LSVM method	82.05%	Polat et al. (2008)[45]
5	HPM includes Simple k-Means Clustering Algorithm and C4.5 classification algorithm	92.38%	Patil et al. (2010)[46]
6	LDA-ANIFS	84.61%	Dogantekin et al. (2010)[47]
7	Hybrid-system containing genetic model and Back propagation Network	76%	Karegowda et al. (2011)[48]
8	Linear Discriminant Analysis Morlet Wavelet Support Vector Machine (LDA-MWSVM)	89.74%	Çalişir and Dogantekin (2011)[49]
9	J48 Decision Tree	76.68%	Asma A. Al Jarullah (2011)[50]
10	Radial Basis Function Hybrid model (RBF-SVM)	83.71%	Karatsiolis and Schizas (2012)[51]
11	(EM), h-means + clustering and -Genetic Algorithm	–	Velu and Kashwan (2012)[52]
12	CART Method	96.37%	Kavitha and Sarojamma (2012)[53]
13	Data Mining technique with C4.5 Algorithm	91%	K Rajesh and V Sangeetha (2012)[7]
14	KNN and K-Means algorithm using GA_CFS Feature selection algorithm	96.67%	Karegowda et al. (2012)[54]
15	Neural Network	92.8%	Kumari and Singh (2013)[33]
16	PLS-DA algorithm	74%	Lakshmi and Kumar (2013)[55]
17	Amalgam KNN and ANFIS	80%	Vijayan et al. (2014)[56]
18	FCM and SVM algorithm	94.3%	Sanakal and Jayakumari (2014)[40]
19	K-nearest Neighbor at k = 3	70%	Saxena et al. (2014)[57]
20	Support Vector Regression (SVR) and the Non-dominated Sorting Genetic Algorithm-II (NSGA-II)	86.13%	Zangooei et al. (2014)[58]
21	ROSETTA software by applying rough sets	82.6%	Joseph.L and Breault. (2014) [59]
22	SVM Algorithm	81.77%	Farahmandian et al. (2015).[60]
23	Random Forest Classifier	99.7%	Butwall and Kumar (2015)[5]
24	ARM technique with Aprori Algorithm	65%	Azra et al. (2015)[61]
25	Genetic Algorithm (GA) and J48Graft Decision tree	74.826%	ChoubeyandPaul (2015)[62]
26	Bayesian Network	90.4%	Mohammadi et al. (2015)[63]
27	Duo-Mining Tool with SVM algorithm	96.39%	Jaya Rama Krishnaiah et al. (2017)[64]
28	K-Means Clustering Algorithm and Naïve Bayes Classification Algorithm	90.9%	Kumarand Umatejaswi* (2017)[41]
29	Bayesian Network	99.3502%	Priyadarshini and Lakshmi (2018)[65]
30	SVM and DNN	77.86%	Wei et al. (2019)[66]

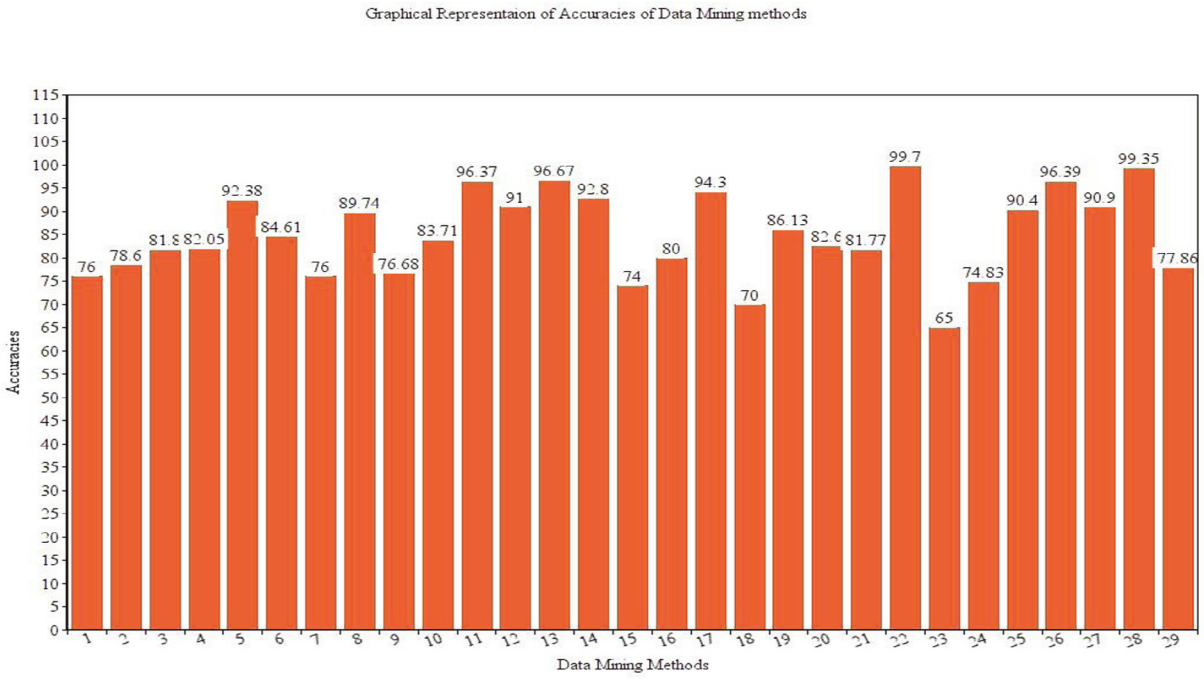

Fig. 4. Graphical representation of accuracy of data mining methods.

the same. Clustering and classification were done using data mining techniques like k-means algorithm that clustered the entire dataset into three clusters for gestational diabetes, type-1 or juvenile diabetes and type-2 diabetes respectively. Pre- processing of the dataset is done before clustering. The last step involves the classification that depicts the risk of a patient to get diabetes. The dataset that is used contains 650 entries of different patients. Attributed like age, gender, insulin dependence, bg, pedigree, job type, food habit etc. were considered. For classification in the last step 4 different techniques namely Naive Bayes, random tree, C4.5 and simple logistics are used and their accuracies are recorded, that is, 90.9, 96.3, 100 and 99 respectively.[41]

The above bar graph represents the accuracies of the data mining techniques that are mentioned in Table 2 where the X-axis of

the graph represents the serial number of the methods in table and Y-axis represents the accuracies of the mentioned data mining method respectively.

5. Elaborate Evaluation of Fuzzy Logic Technologies in Diabetes Prognosis

Expert systems are made to work like humans and reason like human brain does. One such expert system was developed by Polat and Günes (2006) [Table 3], [Fig. 5] to detect diabetes. Their objective to improve the diagnostic capabilities and accuracies was fulfilled by combining principal component analysis (PCA) and adaptive neuro-fuzzy inference system (ANFIS) methods. The entire model works in 2 levels. The first stage consists of the PCA module whose main purpose is feature reduction without affecting the authenticity and impact of those features on the end results (here 8 features are reduced to 4). The next step involves the ANFIS module that is proposed to analyze and prompt decision making process and classifies the patients. The ANFIS structure is well established in the adaptive system framework to make its modeling methodical. The accuracy of this model is found to be 89.47% (10-fold cross validation used).[67]

Lee and wang (2011) introduced a fuzzy expert system that supported the diabetes decision application. It is a 5-layerontology-based model. Ontology is a concept of data management which is mostly used to well establish the relationships among the variables. To model the diabetes knowledge acquired from the dataset fuzzy Diabetes ontology (FDO) is structured. The 5 layers of the ontology structure include fuzzy knowledge layer, fuzzy group relation layer, fuzzy group domain layer, fuzzy personal relation layer, and fuzzy personal domain layer. The concept construction mechanism constructs the concepts over the knowledge domain. The relations between these concepts are established by the relation construction mechanism. A semantic decision support agent (SDSA) contains the utilities like fuzzy group ontology generating mechanism, fuzzy personal ontology generating mechanism and semantic fuzzy decision-making mechanism. The accuracy of this model depending on the age attribute like slightly old, slightly young, more or less young, very young and very very young are 91.2%, 90.3%, 85.9%, 81.7% and 77.3% respectively.[68]

A fuzzy based model for diabetes diagnosis decision support was developed by Rajeswari and Vaithiyanathan (2011) using artificial neural network. The aim was to implicate a physician's decisive power into a decision support system based on fuzzy logic and ANN. The parameters used in this are processed using a fuzzy approach. The data is converted to fuzzy linguistic set that is data to symbol unit for knowledge representation by generating and using fuzzy membership functions. The concept of artificial neural networks is used for classification of the patients into the two categories: type 2 diabetic or non-diabetic. Classification is done using backpropagation algorithm. This network had 10 features like polyuria (increased urination), polydipsia (excessive thirst), polyphagia (increased appetite), nocturia (urination at night), tiredness, giddiness, sleeplessness, non-healing ulcer, and itching and shoulder pain. The relations among the symptoms are mapped using fuzzy logic. The classification accuracy of the training set is 82.9% and test set is 83.3%.[69]

A similar concept to the one used in the above mentioned model was used by Lukmanto and Irwansyah (2015) in their system for early detection of diabetes by using the fuzzy Hierarchical Model with the aim to predict someone's potential against diabetes mellitus since early depending on the symptoms like polydipsia, polyuria, polyphagia and other blood glucose level tests. Age of the patient is also considered in the model. The knowledge acquisition is based on the survey they conducted by interviewing medical doctors. From the knowledge they determined if the person 'can' have DM or not. After the knowledge acquisition they introduced the fuzzy hierarchical model with 3 variables namely input variable containing values of the above-mentioned parameters, temporary variable obtained from fuzzy inference process and out-

Table 3
Accuracies of different fuzzy logic methods for comparative anatomization.

Sr no	Method	Accuracy	Author
1	kNN	71.9%	Ster and Dobnikar (1996)[75]
2	BP	76.4%	Ster and Dobnikar (1996)[75]
3	ANFIS and PCA	89.47%	Polat and Günes (2007)[67]
4	GDA LS and SVM	82.05%	Polat et al. (2008)[45]
5	FNN and ANN	84.2%	Kahramanli and Allahverdi (2008)[76]
6	Fuzzy logic and ANN	83.3%	Rajeswari and Vaithiyanathan (2011)[69]
7	FDO and SDSA (slightly old)	91.2%	Lee and Wang (2011)[68]
8	FDO and SDSA (slightly young)	90.3%	Lee and Wang (2011)[68]
9	FDO and SDSA (very young)	81.7%	Lee and Wang (2011)[68]
10	FCS-ANTMINER	84.24%	Ganji and Abadeh (2011)[77]
11	FES using Correlation Fuzzy Determination Mechanism	89.32%	Kalpana and Kumar (2011)[78]
12	Intensified Fuzzy Verdict Mechanism	88.35%	Kalpana and Kumar (2011)[78]
13	Sim (similarity classifier)	75.29%	Pasi Luukka (2011)[79]
14	Sim + F1 (similarity classifier with feature extraction method 1)	75.84%	Pasi Luukka (2011)[79]
15	Sim + F2 (similarity classifier with feature extraction method 2)	75.97%	Pasi Luukka (2011)[79]
16	Fuzzy Logic without Genetics Algorithm	69%	E.P. Ephzibah (2011)[80]
17	Fuzzy Logic with Genetics Algorithm	87%	E.P. Ephzibah (2011)[80]
18	Fuzzy kNN algorithm	89.10%	Pradhan et al. (2012)[81]
19	FNC		Thirugnanam et al. (2012)[82]
20	Fuzzy Artificial Bee colony with blended crossover operator	84.21%	Beloufa and Chikh (2013)[83]
21	FLR	51.43%	Rahman and Afroz (2013)[84]
22	FIS	75.51%	Rahman and Afroz (2013)[84]
23	ANFIS	78.79%	Rahman and Afroz (2013)[84]
24	Particle Swarm Optimization algorithm with fuzzy classifier	85.15%	Sahebi and Ebrahimi (2015)[85]
25	FIS		Chandgude and Pawar (2016)[86]
26	MOE NSGA II fuzzy after GA FL	83.0435%	Vaishali et al. (2017)[87]
27	RBFNN with bat based fuzzy miner clustering algorithm	73.91%	Edlaa and Cheruku (2017)[88]
28	Enhanced Fuzzy Expert System	91.61%	Prajapati et al. (2017)[89]
29	FDT4CR	81%	Benamina et al. (2018)[74]
30	FIS	96%	Niswati et al. (2018)[8]

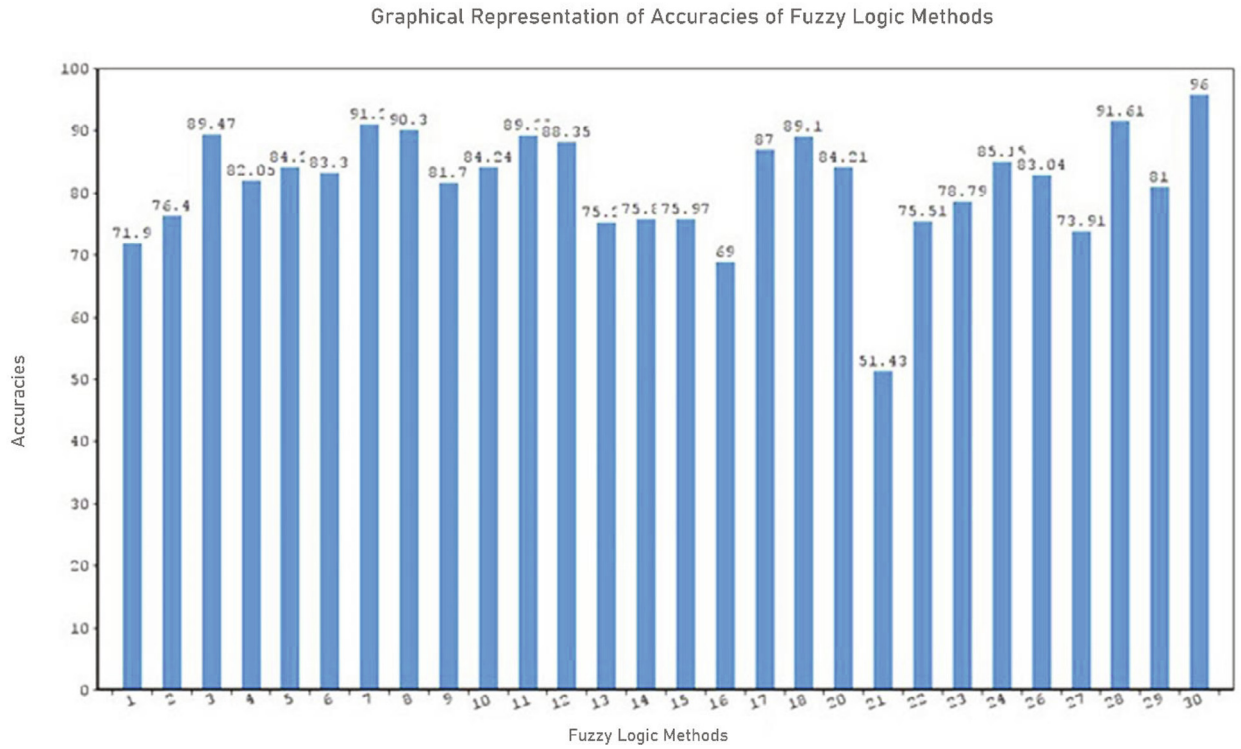

Fig. 5. Graphical representation of accuracy of fuzzy logic methods.

put variable containing the end result predications. There are 5 inference processes each containing 9 rules that help determine the values of temporary variable and output variable. To recapitalize, the model first accepts the fuzzy values which are converted from crisp values using fuzzification, then the 5-step inference process is conducted ending with defuzzification.[2]

A diabetes prediction system was proposed by Gomathi et al. (2015) which used fuzzy ontology approach to get through with the solution. Ontology is a concept of data management which is mostly used to well establish the relationships among the variables. In this case ontology represented the relationship among symptoms and the amount of weighted effect on the result. The proposed method includes the foremost step of preprocessing the Pima India diabetes dataset and removing the inconsistent data using min–max normalization. Next step is the knowledge acquisition which is done using fuzzy logic algorithm where the trapezoidal waveform is used as the derived membership function. The ontology is then constructed which represents the extracted knowledge in the form of ontology graph and enlightens the abstract relations between the symptoms. As a result, this system provides the accuracy of diabetes detection and with low complexity. The ontology makes the representation of knowledge more specific.[70]

Nnamoko et al. (2013) proposed a fuzzy expert system whose main purpose was type 2 Diabetes mellitus management that uses dual interface mechanism creating a solid framework with an effective combination of rule-base and case-base ratiocination. According to them, the outcome of factors like a person's age, gender, glucose levels, BMI, etc. that are somehow related to this ailment, directly or indirectly affect each other as well. To make diabetes management more effective a personalization must be used for better outcomes. But this can be a difficult as the management guidelines are generalized and reasoning/inference becomes difficult for the expert systems as well. Hence, they came up with the fuzzy approach to overcome these dynamic uncertainties of the above-mentioned factors. The system provides general advices and guidelines to the patients. The system is based on 9 input variables for the ratiocination and outputs are in the form of recommendations. The classification is aided by Membership functions (triangular and trapezoidal used here) which establishes the standards for rule generation. The data samples collected as the result of research are analyzed and the best is chosen after undergoing 3 selection techniques. Clustering of data is done and fuzzy logic is imposed on the search space of each cluster making the range optimum.[71]

Reddy and Khare (2016) came up with a more optimized algorithm by applying optimization techniques to the fuzzy logic to detect or classify whether the patient is diabetic or not. They developed a rule based fuzzy logic (RBFL) predictive system which was optimized by FireflyBAT (FFBAT). To begin with, the algorithm firsts undergoes feature reduction which makes the search space easy to traverse without losing its precision. Locality Preserving Projections (LPP) algorithm is used for this purpose. Next, the fuzzy logic is applied where the data undergoes fuzzification, rule generation by inference engine and defuzzification, carried out by the RBFL for classification purpose. The last step which includes optimization makes it a better system. The firefly algorithm and the BAT algorithm is used to generate the best results. The entire dataset is divided into training set which is used in fuzzy engine to generate rules and test set is used to test the accuracy of the entire system.[72]

Niswati et al. (2018) created an expert decision support system for early diagnosis of diabetes implemented using fuzzy logic. The variables used are blood pressure, plasma glucose concentration, body mass index (BMI), diabetes pedigree function and whether the person is pregnant. This data is processed using Mamdani's fuzzy inference system (FIS). The FIS consists of 5 steps: domain problem, fuzzification, creation of fuzzy rules, defuzzification and evaluation. The fuzzy rules are designed such that it aids the diagnostic process with keeping in mind all the input variables and

their effect on the result. This system was found to be 96% accurate; as 48 out of 50 cases were diagnosed correctly, they matched the doctor's diagnosis. The testing process also included validation testing which tested all the functionalities of the system features, questionnaire testing indicating its user experience and application testing.[8]

Prediction of prediabetes also becomes essential to dodge the greater harm which can be caused due to diabetes due to negligence. A fuzzy based association classification system was proposed by Rajeswari et al. (2018) for this objective. Pima Indian Diabetes dataset is used and some of the prediabetic symptoms group is taken as attribute with few deviated characteristics mentioned as outliers. The system handles the hazard risk boundary values whose vagueness can be overcome by fuzzy logic. The patterns are recognized among the attributes based on the association rules. The correspondence among the set of attributes is measured by S% support and C% confidence. The fuzzy association rules (FAR) are prepared by pairing the patterns with proper class labels by matching them. The proposed method is executed either as 3 linguistic terms or 5 linguistic terms with the latter option being more optimal and accurate. Hence, the system uses association classification to identify those risk factors, from the sets having negative result for diabetes, which depict the outlying prediabetic conditions.[73]

Case-based reasoning is a reasoning approach for problem solving purpose that analyzes the past experience acquired from the case retrieval process. A diabetes diagnosis using the concept of case-based reasoning (CBR) and fuzzy logic was proposed by Benamina et al. (2018). CBR involves re utilization of the similar cases and evaluating the new similar case with the approach used in past experience which requires knowledge acquiring. The similarity measurement among the cases is done using k nearest neighbors' algorithm. This case retrieval step is the foremost and essential step of CBR. Fuzzy interface system is realized by Fispro which acts as the modeling part of fuzzy system. The reasoning part which is complementary to the first step is done using JColibri. Fuzzy decision tree is integrated to make the case retrieval better in CBR.[74]

The above bar graph represents the accuracies of the fuzzy logic techniques that are mentioned in the Table 3 where the X-axis of the graph represents the serial number of the methods in the table and Y-axis represents accuracies of the mentionedfuzzy logic method respectively.

6. Future scope and limitations

Healthcare is a field where accuracy of models is of utmost importance and one cannot afford misclassification or incorrect diagnosis. The main challenge for all models surveyed is their ability to completely pose as a diagnostic expert like a real human physician. Accuracies of these models are debatable as they depend on various other factors like the dataset used and number of attributes considered while classifying the patient. While mapping the algorithms to their real time applications, accuracy cannot be the only parameter on which the best algorithm can be selected. It is important to understand the inference part of the solutions, that is, which predictors are associated with the response. Further investigation must be carried out and more training data can be used to make the model robust. To increase the overall efficiency of the data mining and fuzzy logic methods by utilizing the hybrid classification algorithms. In effort to align diabetes and other diseases, the methods of data mining should be performed implicitly to enable the physicians and patients to evaluate diabetes far faster and quicker. Using sample pre-processing methods will positively affect the rate of accuracy by improving. In future, smart and more powerful solutions which can accumulate and manage the diverse activities of diabetic patients are required automatically. Also, gene examination and past diabetes records can be referred as future research. The data mining method includes C4.5 classification algorithm that should be improved for getting a better prediction accuracy rate. In future, for discovery of new victims and for increasing decision-making potential, the data mining methods will separate rules from mixture neural network model. For improving the precision rates, the hybrid model should be developed by using KNN with various techniques of AI. Fuzzy logic techniques can be further extended with better rule generation capabilities by using similarity classification and association among the features can be scaled to reduce the complexity. Fuzzy logic is well known for handling the vagueness as well as the uncertainties and hence fuzzy logic could be the solution to make the diabetes diagnosis model personalized for different patients with different medical history which in turn will provide a more accurate classification of these patients.

7. Conclusion

As we went through different techniques that have been so far used and different models that are developed so far for diabetes diagnosis, we found that they have proved to provide fairly high accuracies. With our major areas of research that includes two techniques namely data mining and fuzzy logic we draw the following conclusions. Fuzzy logic techniques have successfully shown accuracies as high as 96%. Through this comparative study we found that Fuzzy Inference Systems are efficient with high accuracies and less complexity. However, swarm-based optimization techniques can be used to further optimize the model and increase its efficiency. Feature reduction has proved to decrease the complexity of the fuzzy classifier. Data mining plays a major role in the study on diabetes as it is able to discover and reveal secret facts from a vast volume of information on diabetes. The accuracy shown by data mining methods is as high as 99.7% computed through random forest classifier. This study shows how different data mining methods are efficient with respect to their accuracies. Among all data mining methods, the random forest classifier method gives the most consistent value and after executing on 2 or more different test features, it gives the effective result to the patient.

Author contributions

All the authors make a substantial contribution to this manuscript. HT, VS, HY, and MS participated in drafting the manuscript. HT and VS wrote the main manuscript. All the authors discussed the results and implication on the manuscript at all stages.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful to Department of Computer Engineering, LDRP Institute of Technology and Research, Gujarat Info Petro Limited (GIPL), Department of Chemical Engineering School of Technology for the permission to publish this research.

Availability of data and material

All relevant data and material are presented in the main paper.

Funding

Not applicable.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

References

[1]. Swapna G, Vinayakumar R, Soman KP. Diabetes detection using deep learning algorithms. *ICT Express*. 2018. https://doi.org/10.1016/j.icte.2018.10.005.
[2]. Lukmanto RB, Irwansyah E. The Early Detection of Diabetes Mellitus (DM) Using Fuzzy Hierarchical Model. *Procedia Comput Sci*. 2015;59:312–319.
[3]. Zhu J, Xie Q, Zheng K. An improved early detection method of type-2 diabetes mellitus using multiple classifier system. *Inf Sci*. 2015;292:1–14.
[4]. Jha K, Doshi A, Patel P, Shah M. A comprehensive review on automation in agriculture using artificial intelligence. *Artificial Intelligence in Agriculture*. 2019;2:1–12.
[5]. Butwall M, Kumar S. A data mining approach for the diagnosis of diabetes mellitus using random forest classifier. *Int J Comput Appl*. 2015;120(8):36–39.
[6]. Sorte RR, Bante CB, Thakare AP, Saha ST, Meshram MB, Gaikwad VP, 2017. An analytical review on prediction of diabetes using data mining technique. Int J Scientific Res Sci Eng Technol. 3(8), 962-966.
[7]. Rajesh K, Sangeetha V. Application of data mining methods and techniques for diabetes diagnosis. *Int J Eng Innov Technol*. 2012;2(3):224–229.
[8]. Niswati Z, Mustika FA, Paramita A. 2018. Fuzzy logic implementation for diagnosis of Diabetes Mellitus disease at Puskesmas in East Jakarta. IOP Conf. Series: Journal of Physics: Conf. Series 1114.
[9]. Siregar D, Arisandi D, Usman A, Irwan D, Rahim R. Research of simple multi-attribute rating technique for decision support. *J Phys Conf Ser*. 2017;930(1) 012015.
[10]. Suryanto T, Rahim R, Ahmar AS. Employee recruitment fraud prevention with the implementation of decision support system. *J Phys Conf Ser*. 2018;1028(1) 012055.
[11]. Greeda J, Mageswari A, Nithya R. A study on fuzzy logic and its applications in medicine. *Int J Pure Appl Math*. 2018;119(16):1515–1525.
[12]. Alesyanti A, Ramlan R, Hartono H, Rahim R. Ethical decision support system based on hermeneutic view focus on social justice. *Int J Eng Technol*. 2018;7 (2.9):74–77.
[13]. Soltani Z, Jafarian A. A new artificial neural networks approach for diagnosing diabetes disease type II. *Int J Adv Comput Sci Appl*. 2016;7(6):89–94.
[14]. Kim EK, Kwak SH, Jung HS, et al. The effect of a smartphone-based, patient-centered diabetes care system in patients with type 2 diabetes: a randomized, controlled trial for 24 weeks. *Diabetes Care*. 2018;dc172197. https://doi.org/10.2337/dc17-2197.
[15]. Oumaima G, Lotfi E, Fatiha E, Mohammed B. (2019). Deep Learning Approach as New Tool for Type 2 Diabetes Detection. Proceedings of the 2nd International Conference on Networking, Information Systems & Security – NISS19. doi: 10.1145/3320326.3320359.
[16]. Kamalakkannan S, Sharmila K, Muthuselvam M, Jebathangam J. A survey on big data analytics in diabetes disease using classification algorithm. *J Appl Sci Comput*. 2018;5(12):1746–1752.
[17]. Sujaritha M, Murugesan M, Bhuvana MK, Saleekha. Risk analysis of diabetes using IoT and deep learning. *Int J Innov Technol Exploring Eng*. 2019;8 (7s):222–227.
[18]. Shankaracharya, Odedra D, Samanta S, Vidyarthi AS, 2011. Computational intelligence in early diabetes diagnosis: a review. Rev Diabet Stud, 2010, 7 (4):252-262.
[19]. Suvarnamukhi B, Seshashayee M. Big data processing system for diabetes prediction using machine learning technique. *Int J Innov Technol Exploring Eng*. 2019;8(12):4478–4483.
[20]. Alexandru AG, Radu IM, Bizon ML. Big data in healthcare – opportunities and challenges. *Informatica Economica*. 2018;22(2):43–54.
[21]. https://www.kaggle.com/uciml/pima-indians-diabetes-database (Last Access on 13th August, 2020).
[22]. www.stanford.edu/~hastie/Papers/LARS/diabetes.data (Last Access on 13th August, 2020).
[23]. https://data.world/akshayginodia/diabetes-dataset (Last Access on 13th August, 2020).
[24]. https://data.world/abelvikas/diabetes-type-dataset (Last Access on 13th August, 2020).
[25]. Islam, MM Faniqul, et al. Likelihood prediction of diabetes at early stage using data mining techniques. Computer Vision and Machine Intelligence in Medical Image Analysis. Springer, Singapore, 2020. 113-125.
[26]. Ali L, Zhu C, Zhou M, Liu Y. Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection. *Expert Syst Appl*. 2019. https://doi.org/10.1016/j.eswa.2019.06.052.
[27]. Ali L, Zhu C, Zhang Z, Liu Y. Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network. *IEEE J Transl Eng Health Med*. 2019;1–1. https://doi.org/10.1109/jtehm.2019.2940900.
[28]. Ali L, Bukhari SAC. An Approach Based on Mutually Informed Neural Networks to Optimize the Generalization capabilities of Decision Support Sysytems Developed for Heart Failure Prediction. IRBM(2020), doi: 10.1016/j.irbm.2020.04.003.
[29]. Ali L, Bukhari SAC. An Approach Based on Mutually Informed Neural Networks Optimize the Generalization capabilities of Decision Support Sysytems Developed for Heart Failure Prediction. IRBM(2020), doi: 10.1016/j.irbm.2020.04.003.
[30]. Aziz R, Verma CK, Srivastava N. Dimension reduction methods for microarray data: a review. *AIMS Bioeng*. 2017;4(2):179–197.
[31]. Biswas S, Bordoloi M, Purkayastha B. Review on feature selection and classification using neuro-fuzzy approaches. *Int J Appl Evolutionary Computation*. 2016;7(4):28–44.
[32]. Maji P, Garai P. Fuzzy-rough simultaneous attribute selection and feature extraction algorithm. *IEEE Trans Cybern*. 2013;43(4):1166–1177.
[33]. Kumari S, Singh A. (2013). A data mining approach for the diagnosis of diabetes mellitus. 2013 7th International Conference on Intelligent Systems and Control (ISCO). doi: 10.1109/isco.2013.6481182.
[34]. Sankaranarayanan S, Pramananda PT. A predictive approach for diabetes mellitus disease through data mining technologies. *World Congress on Computing and Communication Technologies*. 2014;231–233.
[35]. Iyer A, Sumbaly R. Diagnosis of diabetes using classification mining techniques. *Int J Data Mining Knowledge Manage Process*. 2015;5(1):01–14.
[36]. Mamatha Bai BG, Nalini BM, Majumdar J. (2019). Analysis and Detection of Diabetes Using Data Mining Techniques—A Big Data Application in Health Care. In: Shetty N, Patnaik L, Nagaraj H, Hamsavath P, Nalini N. (eds) Emerging Research in Computing, Information, Communication and Applications. Advances in Intelligent Systems and Computing, vol 882. Springer, Singapore.
[37]. Padmaja P, Vikkurty S, Siddiqui NI, et al. Characteristic evaluation of diabetes data using clustering techniques. *Int J Comput Sci Network Security*. 2011;8 (11):244–251.
[38]. Thangaraju P, Bharathi NG. Data mining approaches for diabetes using feature selection. *Int J Comput Scie Information Technol*. 2014;5(4):5939–5943.
[39]. Abed AH, Nasr M. Diabetes disease detection through data mining techniques. *Int J Adv Netw Applications*. 2019;11(1):4142–4149.
[40]. Sanakal R, Jayakumari T. Prognosis of diabetes using data mining approach-fuzzy C means clustering and support vector machine. *Int J Comput Trends Technol*. 2014;11(2):94–98.
[41]. Kumar PS, Umatejswi V. Diagnosing diabetes using data mining techniques. *Int J Scientific Res Publications*. 2017;7(6):705–709.
[42]. Smith J, Everhart J, Dickson W, Knowler W, Johannes R. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In Proceedings of the symposium on computer applications and medical care 261–265. IEEE Computer Society Press.
[43]. Rutkowski L, Cpalka K. Flexible neuro-fuzzy systems. *IEEE Trans Neural Networks*. 2003;14(3):554–574.
[44]. Davis WL, Kohout L, 2006. Enhancing performance of relational fuzzy neural networks with square BK-products. Proceedings of the 9th joint international conference on information sciences (JCIS-06). 1-4.
[45]. Polat K, Günes S, Arslan A. A cascade learning system for classification of diabetes disease: generalized discriminant analysis and least square support vector machine. *Expert Syst Appl*. 2008;34(1):482–487.
[46]. Patil BM, Joshi RC, Toshniwal D. Hybrid prediction model for Type-2 diabetic patients. *Expert Syst Appl*. 2010;37(12):8102–8108.
[47]. Dogantekin E, Dogantekin A, Avci D, Avci L. An intelligent diagnosis system for diabetes on Linear Discriminant Analysis and Adaptive Network Based Fuzzy Inference System: LDA-ANFIS. *Digital Signal Process*. 2010;20(4):1248–1255.
[48]. Karegowda AG, Manjunath AS, Jayaram MA. Application of genetic algorithm optimized neural network connection weights for medical diagnosis of Pima Indians diabetes. *International Journal on Soft Computing*.. 2011;2(2):15–23.
[49]. Çalişir D, Doğantekin E. An automatic diabetes diagnosis system based on LDA-Wavelet Support Vector Machine Classifier. *Expert Syst Appl*. 2011;38 (7):8311–8315.
[50]. Al Jarullah AA (2011). Decision tree discovery for the diagnosis of type II diabetes. 2011 International Conference on Innovations in Information Technology. doi:10.1109/innovations.2011.5893838.
[51]. Karatsiolis S, Schizas CN (2012). Region based Support Vector Machine algorithm for medical diagnosis on Pima Indian Diabetes dataset. 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE). doi: 10.1109/bibe.2012.6399663.
[52]. Velu CM, Kashwan KR. (2013). Visual data mining techniques for classification of diabetic patients. 2013 3rd IEEE International Advance Computing Conference (IACC). doi: 10.1109/iadcc.2013.6514375.

[53]. Kavitha K, Sarojamma RM, Monitoring of Diabetes with Data Mining via CART Method, International Journal of Emerging Technology and Advanced Engineering, Website: www.ijetae.com ISSN 2250-2459, Volume 2, Issue 11, November 2012.
[54]. Karegowda AG, Jayaram MA, Manjunath AS. Cascading K-means clustering and K-nearest neighbor classifier for categorization of diabetic patients. *Int J Eng Adv Technol*. 2012;1(3):147–151.
[55]. Lakshmi KR, Kumar SP. Utilization of data mining techniques for prediction of diabetes disease survivability. *Int J Sci Eng Res*. 2013;4(6):933–942.
[56]. Vijayan VV, Ravikumar V, Study of data mining algorithms for prediction and diagnosis of diabetes mellitus. Int J Comput Applications. 95(17), 12-16.
[57]. Saxena K, Khan Z, Singh S. Diagnosis of diabetes mellitus using K nearest neighbor algorithm. *Int J Comput Sci Trends Technol*. 2014;2(4):36–43.
[58]. Zangooei MH, Habibi J, Alizadehsani R. Disease diagnosis with a hybrid method SVR using NSGA-II. *Neurocomputing*. 2014;136:14–29.
[59]. Breault JL, Data mining diabetic databases: are rough sets a useful addition? Proceedings of the 33rd Symposium on the Interface. Computing Science and Statistics, Fairfax, 2001.
[60]. Farahmandian M, Lotfi Y, Maleki I, 2015. Data mining algorithms application in diabetes diseases diagnosis: a case study. MAGNT Research Report. 3(1), 989-997.
[61]. Ramezankhani A, Pournik O, Shahrabi J, Azizi F, Hadaegh F. An application of association rule mining to extract risk pattern for type 2 diabetes using Tehran lipid and glucose study database. *Int J Endocrinol Metabolism*. 2015;13 (2). https://doi.org/10.5812/ijem.25389.
[62]. Choubey DK, Paul S. GA_J48graft DT: a hybrid intelligent system for diabetes disease diagnosis. *Int J Bio-Sci Bio-Technol*. 2015;7(5):135–150.
[63]. Mohtaram Mohammadi, Mitra Hosseini, Hamid Tabatabaee, Using Bayesian Network for the prediction and Diagnosis of Diabetes, MAGNT Research Report, vol.2(5), pp. 892-902.
[64]. Krishnaiah VVJR, ChandraShekar DV, Prasad RS, Rao RH. An empirical study about type2 diabetics using duo mining approach. *Intl J Comput Eng Res*. 2012;2(6):33–42.
[65]. Priyadarshini K, Lakshmi I, 2018. Predictive Analysis of Diabetes Using Bayesian Network and Naive Bayes Techniques. International Conference on Advancements in Computing Technologies. 4(2), 71-74.
[66]. Wei S, Zhao X, Miao C. (2018). A comprehensive exploration to the machine learning techniques for diabetes identification. 2018 IEEE 4th World Forum on Internet of Things (WF-IoT). doi: 10.1109/wf-iot.2018.8355130.
[67]. Polat K, Günes S. An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. *Digital Signal Process*. 2007;17(4):702–710.
[68]. Lee Chang-Shing, Wang Mei-Hui. A fuzzy expert system for diabetes decision support application. *IEEE Trans Systems Man Cybernetics, Part B (Cybernetics)*. 2011;41(1):139–153.
[69]. Rajeswari K, Vaithiyanathan V. Fuzzy based modeling for diabetic diagnostic decision support using Artificial Neural Network. Int J Comput Sci Network Security, 11(4), 126-130.
[70]. Gomathi C, Rajamani V, Jeya K, Prediction of diabetes using fuzzy ontology approach. Int J Eng Res Technol. 3(16), 1-4.
[71]. Nnamoko N, Arshad F, England D, Vora J. Fuzzy Expert System for Type 2 Diabetes Mellitus (T2DM) Management Using Dual Inference Mechanism. In: 2013 AAAI Spring Symposium Series (2013).
[72]. Reddy GT, Khare N. FFBAT-optimized rule based fuzzy logic classifier for diabetes. *Int J Eng Res Afr*. 2016;24:137–152.
[73]. Rajeswari AM, Sidhika MS, Kalaivani M, Deisy C. (2018). Prediction of Prediabetes using Fuzzy Logic based Association Classification. 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT). doi: 10.1109/icicct.2018.8473159.
[74]. Benamina M, Atmani B, Benbelkacem S. Diabetes diagnosis by case-based reasoning and fuzzy logic. *Int J Interactive Multimedia Artificial Intelligence*. 2018;5(3):72–80.
[75]. Pandya R, Nadiadwala S, Shah R, Shah M. Buildout of methodology for meticulous diagnosis of K-complex in EEG for aiding the detection of Alzheimer's by artificial intelligence. *Augmented Human Res*. 2019. https://doi.org/10.1007/s41133-019-0021-6.
[76]. Kahramanli H, Allahverdi N. Design of a hybrid system for the diabetes and heart diseases. *Expert Syst Appl*. 2008;35(1–2):82–89.
[77]. Ganji MF, Abadeh MS. A fuzzy classification system based on Ant Colony Optimization for diabetes disease diagnosis. *Expert Syst Appl*. 2011;38 (12):14650–14659.
[78]. Kalpana M, Kumar AVS. Diagnosis of diabetes using correlation fuzzy logic in fuzzy expert system. *Int J Adv Res Comput Sci*. 2012;3(1):244–250.
[79]. Luukka Pasi. Feature selection using fuzzy entropy measures with similarity classifier. *Expert Systems with Applications, Elsevier*. 2011;38:4600–4607.
[80]. Ephzibah E. Cost effective approach on feature selection using genetic algorithms and fuzzy logic for diabetes diagnosis. *Int J Soft Comput*. 2011;2 (1):1–10.
[81]. Pradhan M, Kohale K, Naikade P, Pachore A, Palwe E. Design of classifier for detection of diabetes using neural network and fuzzy k-nearest neighbor algorithm. *Int J Comput Eng Res*. 2012;2(5):1384–1387.
[82]. Thirugnanam M, Kumar P, Srivatsan SV, Nerlesh CR. Improving the prediction rate of diabetes diagnosis using fuzzy, neural network, case based (FNC) approach. *Procedia Eng*. 2012;38:1709–1718.
[83]. Fayssal B, Chikh MA. Design of fuzzy classifier for diabetes disease using Modified Artificial Bee Colony algorithm, Computer methods and programs in biomedicine, No. 1, pp.92-103, 2013. Pdf.
[84]. Rahman RM, Afroz F. Comparison of various classification techniques using different data mining tools for diabetes diagnosis. *J Software Eng Appl*. 2013;2013(6):85–97.
[85]. Sahebi HR, Ebrahimi S. A Fuzzy Classifier Based on Modified Particle Swarm Optimization for Diabetes Disease Diagnosis. Geoinformatica 2015; 5: 11-17. Sahebi HR and Ebrahimi S: A Fuzzy Classifier Based on Modified Particle Swarm Optimization for Diabetes Disease Diagnosis. Geoinformatica 2015; 5: 11-17.
[86]. Chandgude N, Pawar S. (2016). Diagnosis of diabetes using Fuzzy inference System. 2016 International Conference on Computing Communication Control and Automation (ICCUBEA). doi: 10.1109/iccubea.2016.7860001.
[87]. Vaishali R, Sasikala R, Ramasubbareddy S, Remya S, Nalluri S. (2017). Genetic algorithm based feature selection and MOE Fuzzy classification algorithm on Pima Indians Diabetes dataset. 2017 International Conference on Computing Networking and Informatics (ICCNI). doi: 10.1109/iccni.2017.8123815.
[88]. Edla DR, Cheruku R. Diabetes-finder: a bat optimized classification system for type-2 diabetes. *Procedia Comput Sci*. 2017;115:235–242.
[89]. Prajapati H, Jain A, Pal SK (2017). An enhance expert system for diagnosis of diabetes using fuzzy rules over PIMA dataset. 4(9), 225-230.