An explainable analysis of diabetes mellitus using statistical and artificial intelligence techniques

William Hoyos[1,2,3*], Kenia Hoyos[4], Rander Ruiz[5] and Jose Aguilar[2,6,7]

Abstract

Background Diabetes mellitus (DM) is a chronic disease prevalent worldwide, requiring a multifaceted analytical approach to improve early detection and subsequent mitigation of morbidity and mortality rates. This research aimed to develop an explainable analysis of DM by combining sociodemographic and clinical data with statistical and artificial intelligence (AI) techniques.

Methods Leveraging a small dataset that includes sociodemographic and clinical profiles of diabetic and non-diabetic individuals, we employed a diverse set of statistical and AI models for predictive purposes and assessment of DM risk factors. The statistical tests used were Student's t-test and Chi-square, while the AI techniques were fuzzy cognitive maps (FCM), artificial neural networks (ANN), support vector machines (SVM), and XGBoost.

Results Our statistical models facilitated an in-depth exploration of variable associations, while the resulting AI models demonstrated exceptional efficacy in DM classification. In particular, the XGBoost model showed superior performance in accuracy, sensitivity and specificity with values of 1 for each of these metrics. On the other hand, the FCM stood out for its explainability capabilities by allowing an analysis of the variables involved in the prediction using scenario-based simulations.

Conclusions An integrated analysis of DM using a variety of methodologies is critical for timely detection of the disease and informed clinical decision-making.

Keywords Diabetes mellitus, Predictive models, Explainability, Artificial intelligence, Statistics

*Correspondence:
William Hoyos
william.hoyos@campusucc.edu.co
[1] Grupo de Investigación ISI, Universidad Cooperativa de Colombia, Montería, Colombia
[2] Grupo de Investigación en I+D+i en TIC, Universidad EAFIT, Medellín, Colombia
[3] GIMBIC, Universidad de Córdoba, Montería, Colombia
[4] Laboratorio Clínico Humano, Clínica Salud Social, Sincelejo, Colombia
[5] Grupo de Investigación Interdisciplinario del Bajo Cauca y Sur de Córdoba, Universidad de Antioquia, Medellín, Colombia
[6] CEMISID, Universidad de Los Andes, Merida, Venezuela
[7] IMDEA Networks Institute, Madrid, Spain

Background

Diabetes mellitus (DM) is a chronic metabolic disorder characterized by an increase in blood glucose levels because the body is unable to produce or use the hormone insulin, which is responsible for the regulation of glucose in the bloodstream [1, 2]. It has been associated with various disabling complications that increase the risk of mortality, such as kidney disease, cardiovascular disease, neuropathy, retinopathy, diabetic foot, and lower limb amputation [3].

According to the World Health Organization (WHO), the prevalence and mortality of diabetic patients have been progressively increasing in recent years [4]. By

Hoyos *et al. BMC Medical Informatics and Decision Making*   (2024) 24:383

Page 2 of 20

2021, there was a population of 529 million diagnosed with diabetes [5], estimating that by 2045 there will be an increase of 48% [6], with expected prevalence rates of 10% in several regions in the world such as North Africa, Middle East, Latin America, and the Caribbean [1, 5]. By 2021, an estimated 6.7 million adults aged 29–79 years died from DM or its complications, corresponding to 12.2% of deaths worldwide [1]. Each year, DM generates large economic losses, both for patients (loss of productivity, reduced quality of life, premature mortality) and health systems due to the high costs of care, diagnosis, and treatment [7]. The International Diabetes Federation estimates that health expenditures for 2021 resulted in US$966 billion worldwide, forecast to reach more than $1054 billion by 2045 [1]. Symptoms of type 2 DM can develop over years, people can live without realizing their health status [2], finding that about 50% of the population suffering from DM are not diagnosed in time [8]. Late diagnosis leads to numerous health problems and a large number of deaths each year, so the development of methods for early diagnosis and initiation of timely treatment is essential to improve the quality of life of patients and reduce the loss of productivity and healthcare costs.

In general, healthcare institutions are large generators of data, as led to an increase in research in the field of healthcare in recent years using computational techniques that exploit said data [9–12]. Specifically, several studies have been conducted to predict DM, involving artificial intelligence (AI) techniques, and datasets that include patients' clinical and sociodemographic information. For instance, Islam et al. [8] developed several models such as Naive Bayes, Decision Tree, Logistic Regression, Random Forest (RF), and Artificial Neural Networks (ANN). The results showed that ANN was the best model classifying 99% of the instances correctly. Ergün et al. [13] used eight AI techniques, where the highest accuracy rate was obtained by using Convolutional Neural Network (CNN) with 99.04%, followed by XGBoost and RF with accuracies of 97.89% and 97.69%, respectively. Chaves et al. [14] developed a model for early diagnosis of DM and they found in their results that ANN represented the best model with a correct prediction of 510 out of 520 instances, with an accuracy of 98.08%. García-Ordás et al. [15] used deep learning combined with augmentation techniques to address DM prediction, and compared the original PIMA dataset with an oversampled dataset with a variational automatic encoder (VAE); the best single classical model corresponded to multilayer perceptron on the VAE set with an accuracy of 79.22%. On the other hand, Reddy et al. [16] reported the use of genetic learning and chaotic features for non-invasive diagnosis of DM. In addition, they reported the use of an improved hybrid version of the Extreme Learning

Machine algorithm combined with an improved version of Particle Swarm Optimization (ELM-PSO) [17]. The results of these works showed good performance in detecting or diagnosing diabetes noninvasively. Finally, Swaroop et al. [18] used ensemble methods such as Ant Colony Optimization with Xgboost and Gray Wolf Optimization with Adaboost. The performance metrics showed that Ant Colony Optimization with Xgboost outperformed the other combinations tested.

Despite the great variety of articles reported in the literature on DM prediction, the works present some limitations. On one side, some studies did not use preprocessing techniques to improve data quality such as oversampling to balance classes; the use of unbalanced data can affect the quality of model results. On the other hand, several studies focus on the development of poorly understandable and complex models, which affect the interpretation by health professionals, who are the most interested in performing clinical follow-up of patients. The diagnosis of diseases such as DM is a complex process because it is a multifactorial problem where a professional must analyze all the factors to detect the disease early. The works reported in the literature focus on developing complex models that improve the accuracy to predict the disease sometimes sacrificing the interpretability of the results. The sole prediction of whether or not the patient has DM is insufficient for such a complex problem. Medical professionals are interested not only in tools that allow them to detect the disease with excellent predictive ability but also, in assessing the impact of each predictor variable on the presence of DM. In addition, the development of models that allow the behavior of each variable to be evaluated through simulated iterations is more useful for early detection of the disease. Also, another limitation of the studies reported in the literature is that they do not integrate different approaches of different nature for disease analysis. Using different modeling approaches could help to extract more complete information about the disease than if only one approach is used. The generation of knowledge from the data may be more complete when different approaches of different natures are integrated for the analysis of DM. Classical statistical techniques are excellent for analyzing linear relationships and simple patterns. However, DM is a complex disease with multiple interrelated factors. AI techniques can identify nonlinear and complex patterns in large datasets, providing a deeper understanding of the relationship between various variables. Nonetheless, a final limitation is that these approaches are data intensive. Thus, it is important to develop models on small datasets, because in many parts of the world data availability is low for this disease. Combining these two approaches on datasets with few or many data, depending

on the context, can provide a more complete and accurate view of the variables related to DM, which is crucial for informed health and medical decision-making.

Based on this context, the aim of our research was to carry out an explainable analysis of DM using statistical and AI models. The novelty of our work lies in the use of two different approaches, such as conventional statistics and AI, to develop a comprehensive in-depth analysis of diabetes. To this end, in this work we carry out a descriptive analysis, an analysis of associations between variables of interest, predictive models to classify individuals with and without diabetes; and finally, an explanatory analysis using AI techniques. Specifically, the contributions of our research to the state of the art are: i) an explainable analysis of DM that integrates several approaches, such as statistical tests and AI techniques in a small dataset; ii) several AI models that predict with high accuracy the presence of DM using sociodemographic and clinical information; iii) an analysis of the explainability and interpretability of all the models developed; iv) a quantitative comparison to evaluate the capacity of our proposed models and compare them with those reported in the literature.

The remainder of the paper is organized as follows: Material and methods section details the materials and methods, describing the dataset, preprocessing techniques, model development, and evaluation metrics. Results section shows the experimental results. Discussion section discusses the results and compares them with related work. Finally, Conclusions section concludes the paper, identifies limitations, and recommends future work.

Material and methods

This section describes the methodological framework of the research to meet the proposed objective. Figure 1 represents the outline of the main stages of the process from data collection to the evaluation of the proposed predictive models. We conducted a quantitative cross-sectional study, in which, we initially performed a descriptive analysis to analyze the distribution of the data [19]. Then, we performed an association analysis to find dependencies between each predictor variable and the presence of DM [12]. Subsequently, we used four AI techniques to develop DM prediction models with subsequent evaluation of the performance of each one. Finally, we performed an explainability analysis for each model using different methodologies such as SHAP (SHapley Additive exPlanations) values, feature importance, and scenario-based computational simulations.

Data source

An open-access dataset was used to develop several models to detect and analyze DM. We used the early stage DM risk prediction dataset, which is licensed under Creative Commons Attribution 4.0 International license
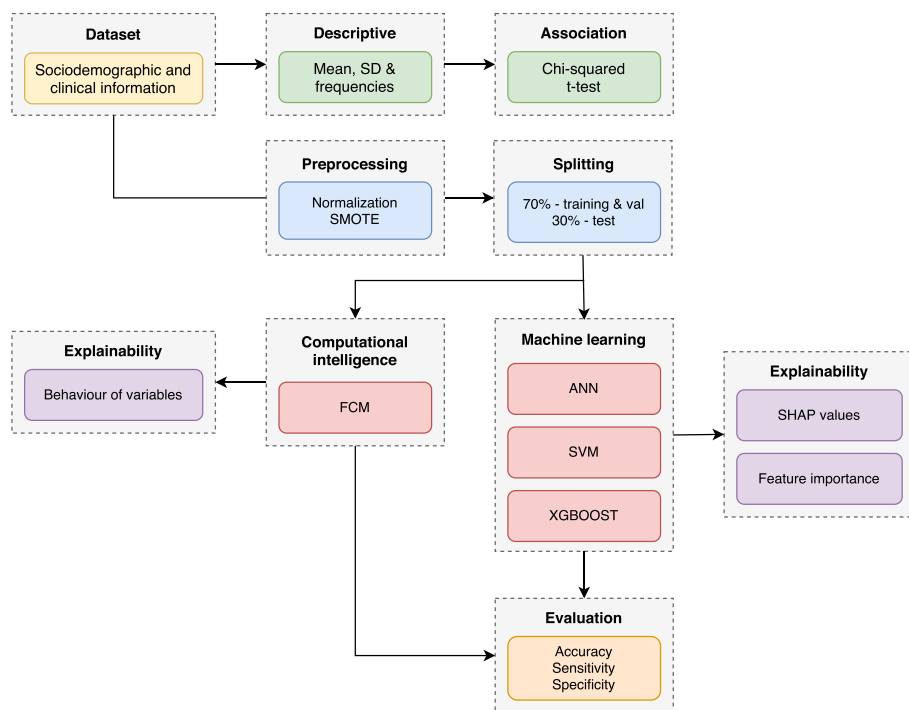

Fig. 1 Methodological framework of the research

Hoyos *et al. BMC Medical Informatics and Decision Making*   (2024) 24:383

Page 4 of 20

[8]. This dataset contains signs and symptoms information of 520 patients from Sylhet Diabetes Hospital, Bangladesh, which was collected by direct survey application to patients under medical supervision. This group of individuals consisted of 320 diabetic patients and 200 individuals without DM. Table 1 shows the set of the 17 variables, including a class variable corresponding to the presence or absence of DM. All variables, with the exception of age, are binary variables (Yes/No). We chose this dataset because we wanted to develop predictive models on a small DM dataset using statistical and AI techniques. The goal was to demonstrate that AI models can recognize patterns in small datasets.

Descriptive analysis

A descriptive analysis of the dataset was performed to evaluate its distribution. Descriptive analysis was used to determine statistical values that later are used in the following analyses (for example, the frequency of categorical variables was used in dependence analysis). It was also used to assess the quality of the dataset (for example, whether it had outliers, among other things). For numeric variables, we used measures of central tendency such as the mean, and measures of dispersion such as the standard deviation. For categorical variables, we used the distribution of absolute and relative frequencies with respect to class (absence/presence of DM).

Association or dependence analysis

Dependence analysis allows determining the association of each predictor variable in the dataset with the presence of diabetes. In particular, variables associated with diabetes are identified using statistical tests. To verify the dependence of age on DM, we used Student's t-test [20] to analyze significant differences between diabetics and non-diabetics. Previously, the Lilliefors test [21] was used to verify the normality of the data. For the comparison of age between the two study groups, we used the following hypotheses:

- $H_0 : \bar{e}_{diabetes} = \bar{e}_{non-diabetes}$
- $H_1 : \bar{e}_{diabetes} \neq \bar{e}_{non-diabetes}$

The previous hypothesis was used to evaluate whether the means of individuals with diabetes and those without diabetes are significantly different. For this, we used the p-value of the test and a significance level of 0.05. If the p-value is less than the significance level, the null hypothesis is rejected, which shows that there is evidence to say that the ages of the two groups are different. Otherwise, there are no significant differences between the two study groups.

To perform the analysis of the association between the categorical variables (sex, polyuria, polydipsia, etc.) and the target variable (diagnosis of DM), the chi-square test was used [22]. For example, if we want to assess the association between sex and DM, we first constructed a contingency table that stores the observed frequencies of individuals classified by sex and the absence/presence of DM. Subsequently, the expected frequencies are calculated under the null hypothesis that there is no association between sex and the presence of DM. Thus,

Table 1 Description of the variables included in the dataset

Variable	Type of variable	Concept ID	Brief description
Age	Sociodemographic	C1	Length of life since birth
Sex	Sociodemographic	C2	Male or female
Polyuria	Clinic	C3	Abnormal increase in the volume of urine eliminated
Polydipsia	Clinic	C4	Excessive thirst
Sudden weight loss	Clinic	C5	Decrease in body weight, when not seeking weight loss
Weakness	Clinic	C6	Lack of physical or muscular strength
Polyfagia	Clinic	C7	Imperious and uncontrollable sensation of hunger
Genital thrush	Clinic	C8	Genital infection caused by fungi
Visual blurring	Clinic	C9	Loss of visual acuity
Itching	Clinic	C10	Sensation causing desire to scratch the skin
Irritabilidad	Clinic	C11	Exacerbated reaction to negative changes in environment
Delayed healing	Clinic	C12	Alteration of the natural mechanisms of skin and tissue regeneration
Partial paresis	Clinic	C13	Mild to moderate loss of voluntary muscle movement anywhere in the body
Muscle stiffness	Clinic	C14	Muscle tension, which may produce spasms and jerky movements
Alopecia	Clinic	C15	Abnormal loss of part or all of the hair
Obesity	Clinic	C16	Condition characterized by excessive accumulation of fat in the body
Class	Diagnosis	C17	Presence or absence of DM

the Chi-square statistic is calculated using Eq. 1. Finally, statistical significance (p-value) was obtained, establishing the value of 0.05 as significant for the association between qualitative variables and the presence of DM. This same procedure was performed for each of the categorical variables in the dataset described in Table 1.

$$X^2 = \sum_{i=1}^{n} \frac{(X_i - E_i)^2}{E_i} \quad (1)$$

Where X_i is the observed frequency, E_i is the expected frequency and n is the number of categories in the contingency table. The hypotheses to be tested in the Chi-square test are as follows:

- H_0: There is no dependence or association between the evaluated variables.
- H_1: There is some kind of dependence or association between the evaluated variables.

Data preprocessing

Data preprocessing is an essential stage in data mining because different methods such as normalization or oversampling could improve data quality [23]. In this research, we used min-max normalization on the age variable (the only numerical variable) to sort the data between values of 0 and 1, in order to ensure faster model training. The formula for min-max normalization is expressed by the following equation [24]:

$$e_{norm} = \frac{e_i - e_{min}}{e_{max} - e_{min}} \quad (2)$$

where e_{norm} is the normalized age, e_i is the age of each individual, e_{min} is the minimum age and e_{max} the maximum age.

In the dataset, the distribution of classes was unbalanced, initially containing 61.53% records of patients with DM and 38.47% records without DM. The use of unbalanced data can bias the performance of the classifiers toward the majority class, affecting the accuracy of the minorities. Faced with this problem, we used the synthetic minority oversampling technique (SMOTE) [26], which consists of generating new instances from data of the minority class. SMOTE consists of a 4-stage process. First, the identification of instances of the minority class in the dataset is performed. Second, an instance and its nearest neighbors are selected using the Euclidean distance. Third, the algorithm generates new instances somewhere in the line connecting the identified instance and its selected neighbors. Finally, the fourth stage consists of adding the newly generated instances to the dataset. Figure 2 shows a schematic representation of the process performed by SMOTE. The instances of the minority class are represented by green circles. The circle with X_1 represents the selected instance and X_3 is the nearest neighbor. The red circle with X_n represents the new data generated. Thus, we increased the number of data for the minor label using SMOTE, obtaining a dataset with 640 records: 320 for DM and 320 without DM.
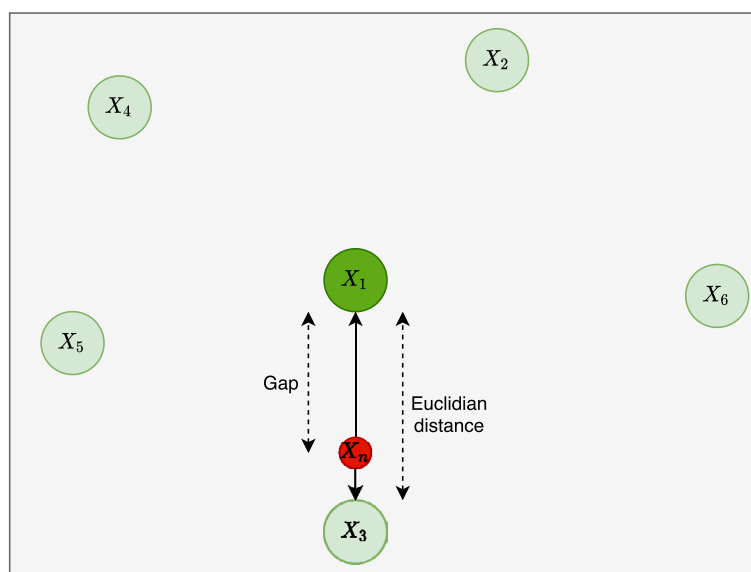

Fig. 2 Schematic representation of SMOTE. Green circles indicate the minority class. X_1 and X_3 represent the selected instance and the nearest neighbor, respectively. The red circle with X_n represents the new data generated (adapted from [25])

AI techniques

Fuzzy cognitive map

A Fuzzy Cognitive Map (FCM) is an AI technique that simulates human reasoning, through the graphical representation or modeling of complex systems, using their concepts and the interrelation between them [27]. FCMs are effective in modeling the uncertainty and imprecision present in many datasets. In the context of structured data, where ambiguity can arise, FCMs can better capture complexity. In addition, FCMs allow for clearer interpretation and explanation of model decisions. This is crucial in applications where understanding the reasoning behind a prediction is as important, especially in medical environments. Figure 3 illustrates in a simple way the representation of an FCM, a graphical structure with five concepts (C_1 to C_5). Each concept (C) represents a variable or characteristic of the system under study, for example, symptoms of a disease or laboratory tests. The influence of one concept on another is represented by a weight (W) on a directed edge [28].

Mathematically, the FCM in Fig. 3 can be represented by a matrix (see Eq. 3 [29]), called adjacency matrix [29], which contains information on the influences between concepts of an FCM:

$$\mathbf{W} = \begin{array}{c} \\ C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \end{array} \begin{array}{c} \begin{array}{ccccc} C_1 & C_2 & C_3 & C_4 & C_5 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & w_{14} & w_{15} \\ 0 & 0 & 0 & 0 & w_{25} \\ 0 & w_{32} & 0 & 0 & w_{35} \\ 0 & 0 & 0 & 0 & w_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \quad (3)$$

FCM models can be constructed in three ways [30]. In the first one, a group of experts select the concepts of interest and assign relationships between those concepts. In the second one, the relationships between concepts are determined using datasets and optimization algorithms [28]. Finally, the third option is a combination of the first two options, where experts define the concepts with their relationships, and the weights are optimized using data and algorithms for this purpose [24]. The resulting FCM models can be used for the description, prediction, or evaluation of the behavior of variables using computational simulations. In the present work, we used the third option, where first, three experts assigned relationships between concepts, and using a dataset we optimized the matrix with the available data. For this objective, we used the Particle Swarm Optimization (PSO) algorithm. This algorithm can be modeled with two equations, which represent the update of the velocity of a particle i and the position of the particle. Equation 4 allows the update of the velocity of particle i, which for our case is an FCM candidate, while Eq. 5 updates its position, which represents an optimal weight matrix for each FCM [24]:

$$v_i(t+1) = v_i(t) + s_1 \times r_1 \cdot (W_i^{best} - W_i(t)) + s_2 \times r_2 \cdot (W_i^{gbest} - W_i(t))$$
$$(4)$$

$$W_i(t+1) = W_i(t) + v_i(t) \quad (5)$$

where v_i is the particle velocity; r_1 and r_2 are random values with uniform distribution; s_1 is the cognitive coefficient, responsible for the particle tending to move toward the position where it has obtained the best results so far; s_2 is the social component, also known as collective behavior, responsible for the particle tending to move toward the best position found by the swarm so far; W_i^{best} is the best position obtained by a particular particle, while W_i^{gbest} is the best position obtained by any particle in the swarm. Recall that each particle is an FCM candidate, while the position is a weight matrix to construct each FCM [28].

For the construction of the FCM model, we first selected the variables described in Table 1. These variables define concepts within the FCM (e.g., C1: age; C2: sex, C3: Polyuria, etc.), which must be connected by arrows indicating the influence of one concept to
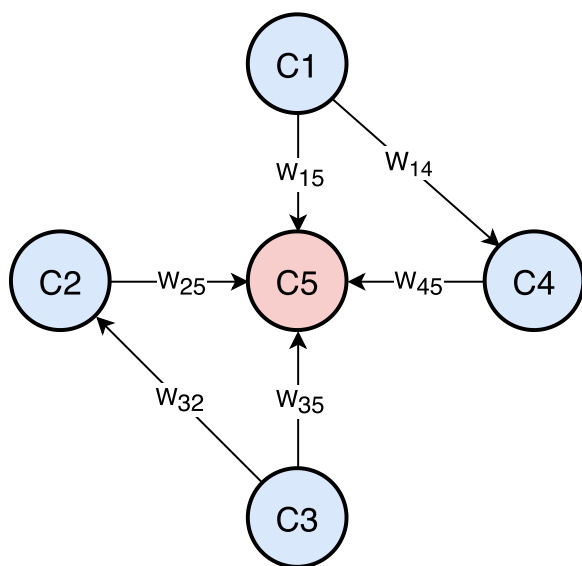

Fig. 3 Example of an FCM with five concepts and six relationships. Adapted from [28]

another. Three experts then proposed a preliminary connection of the concepts. Subsequently, using the PSO algorithm (see Eqs. 4 and 5), the relationships between the concepts are optimized from the data through a training process. In the case of PSO, a particle position is a weight matrix to build an FCM, such that it seeks to find the weight matrix that optimizes the learning error. After the FCM is built, it can be used to perform scenario-based simulations and thus evaluate the behavior of the variables used in the prediction. A grid search-based hyperparameter tuning was used to find the best combination of hyperparameters to generate the FCM with the best predictive performance. For each configured combination, the model was implemented and evaluated. In this way, the evaluation metrics were obtained allowing us to choose the model with the highest accuracy. Table 2 describes the different parameters, and their values, used in the adjustment of the hyperparameters in each of the developed models according to the technique used for their development.

To give more details, in the case of FCM, there are two types of parameters to optimize. Those linked to PSO, which is used in the learning phase of the FCM weights (in this case, the values of r_1, r_2; s_1, s_2 and the initial population are required to be determined), and those related to the behavior of the FCM (the activation function that determines the activation level of each concept and the inference function that describes its reasoning process). All these parameters were tuned using a grid

search-based hyperparameter optimization approach to find their best combination.

ANN

An ANN is an AI technique that aims to mimic the functions of the human brain to solve complex problems. ANNs are especially effective for modeling nonlinear relationships in structured data. They can learn complex patterns and hierarchical representations, which is beneficial when interactions between variables are not simple or straightforward. In more complex problems, deep learning architectures, such as deep neural networks, can automatically extract features and patterns from structured data, improving the predictive capability of the model. ANNs have the ability to generalize well from limited training data, making them suitable for predicting structured datasets of different sizes and complexities. Their applications are focused mainly on image and voice recognition and automatic translation [31]. The first neural model developed was the Perceptron, which consists of a fully interconnected direct-feed neural network, in which data are transmitted from the input to the hidden layer and from there to the output layer [32].

Figure 4 shows an example of an ANN. This neural network (MLP) integrates an input layer (visible layer) with n neurons $(x_1, \ldots x_n)$, a hidden layer $(h_1, \ldots h_o)$, and an output layer, which are the output variables of the model [33]. To define the final output, the model uses the fitted values of weights and biases, thus achieving a relationship between inputs and outputs [34]. We used an MLP

Table 2 Configuration of hyperparameters to tune in the developed models

| Technique	Hyperparameter	Options to configure
PSO-FCM	Initial population	50, 100 ,150 , 200, 250, 300
	Activation function	Sigmoid, tanh
	Inference function	Standard-Kosko, Modified-Kosko, Rescaled
ANN (MLP)	Hidden units	16, 32, 64, 128, 256
	Learning rate	0.0001, 0.001, 0.01, 0.05, 0.1, 0.5
	Activation function	Tanh, ReLU
	Optimizer	Stochastic gradient descent, Adam
	Type of learning	Constant, adaptive
SVM	Kernel	Lineal, radial, sigmoid
	C	0.0001, 0.001, 0.01, 0.1, 1.0, 10, 100, 1000
	Gamma	0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000
XGBoost	Predictors choose at random	Random values between 1 and 20
	Number of trees	10, 100, 1000
	Minimun node size	2, 5, 7, 11, 15, 20, 25, 30
	Depth of trees	1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25
	Learning rate	Random values between 0 and 1
	Minimum loss reduction	Random values between 0 and 0.01
	Percentage of sample	Random values between 0 and 1
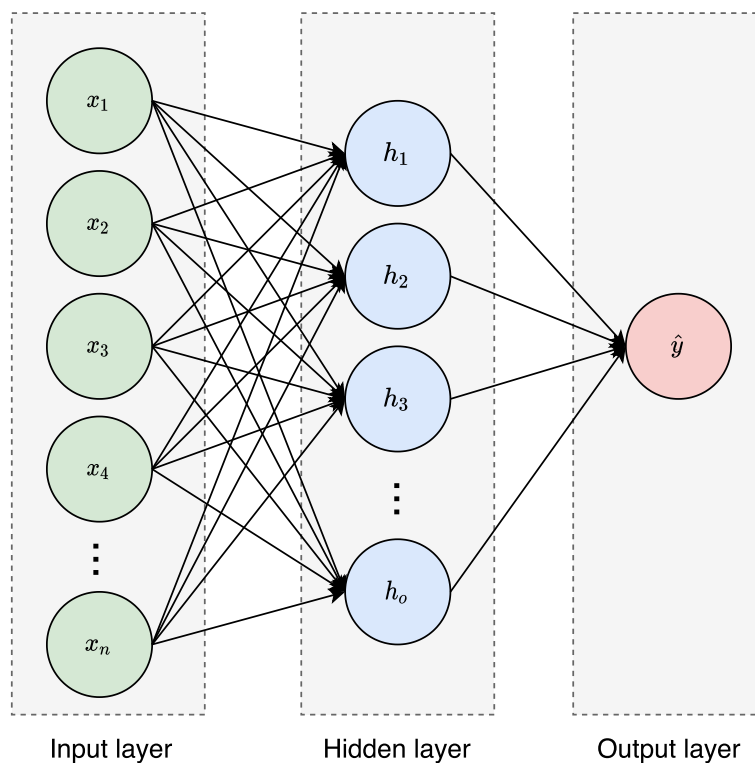
Fig. 4 Schematic representation of an ANN, specifically an MLP. Source: own elaboration

with different numbers of neurons in the hidden layer and learning of two types, constant and adaptive. All the hyperparameters used for optimization can be seen in Table 2.

SVM

The Support Vector Machine (SVM) is an AI technique that is suitable for relatively small datasets with few outliers [35]. SVMs are particularly useful in environments with high-dimensional structured data, as they can efficiently handle datasets with many predictor variables. SVMs are less prone to overfitting, making them suitable for problems where relatively small datasets are available. This is relevant in many real-world scenarios where data may be limited. SVMs tend to generalize well, providing robust predictions even on unseen data. This is crucial for the practical applicability of predictive models in real-world situations. Figure 5 shows a schematic representation of an SVM to classify instances. This strategy defines a cutting hyperplane, a line that separates categories of data and will divide the space into different domains containing each category of data [36]. This model was built using multiple combinations of the parameters to identify the

best configuration for addressing the DM classification problem. The hyperparameter configurations used for this technique are shown in Table 2.

XGBoost

XGBoost, or EXtreme Gradient Boosting, is a very effective AI technique, which is an end-to-end scalable gradient-boosted tree system modified from the Gradient Boosting Machine (GBM) technique [38]. XGBoost is widely used for prediction, classification, and regression [39]. Figure 6 shows the schematic representation of XGBoost, which is an iterative algorithm with multiple decision trees. Each tree learns from the residuals of all previous trees (f_k). In the end, the predicted output of XGBoost is the sum of all the results ($\hat{y}$) [39].

The XGBoost algorithm was used in the present study because of its performance characteristics. Initially, we can say that it is a fast classifier, with a high degree of learning and training speed, because of its parallel processing capability and optimization of new trees each time they are attached, it is effective in solving classification and data preprocessing problems, which increases accuracy [40]. The hyperparameters used to tune this model can be seen in Table 2.
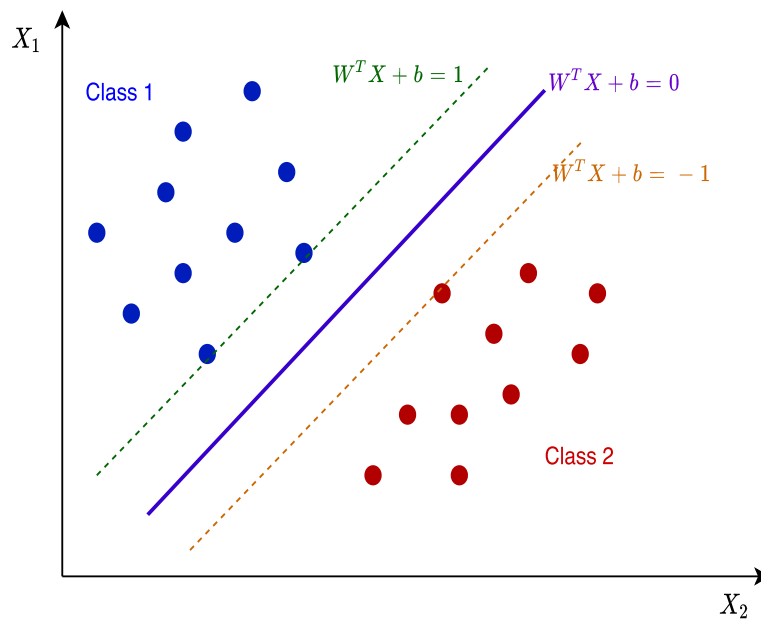
Fig. 5 Schematic representation of an SVM with a hyperplane of separation. Adapted from [37]
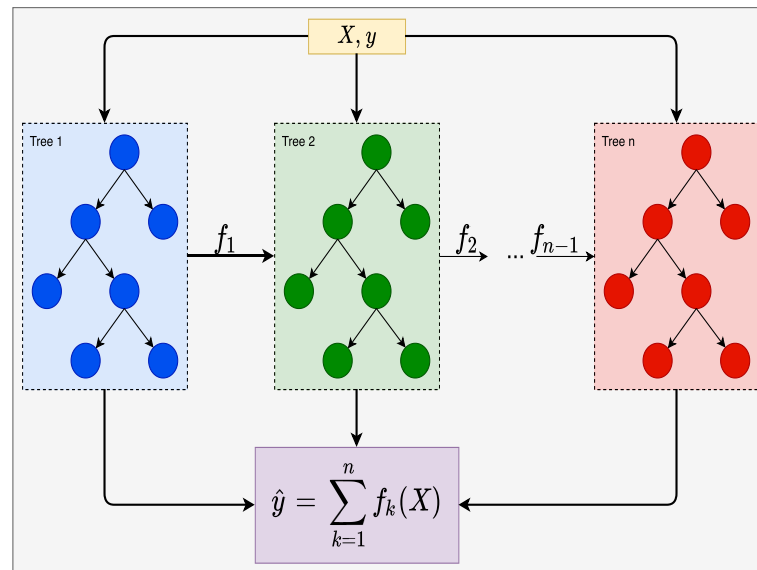

Fig. 6 Schematic representation of an XGBoost algorithm. Adapted from [39]

Experimental configuration

Training, validation and testing

For training, validation and testing, we have used the DM dataset described in the previous sections. We have used a 5-fold cross-validation method for training and validation of our AI models [14]. This method was executed using three subsets of data for training, validation, and testing of the model. 70% of the data was used for training and validation, and 30% for testing the model.

Figure 7 shows a schematic representation of the 5-fold cross-validation. First, we divided the training and validation data (70%) into five subsets, of which four were used for training and one for validation. The validation phase consists of applying the trained model on the validation set. Next, the process was repeated with a different subset from the previous one, and in the same proportion, i.e., four for training and the remaining subset for validation. After five-fold, the best model and its hyperparameters
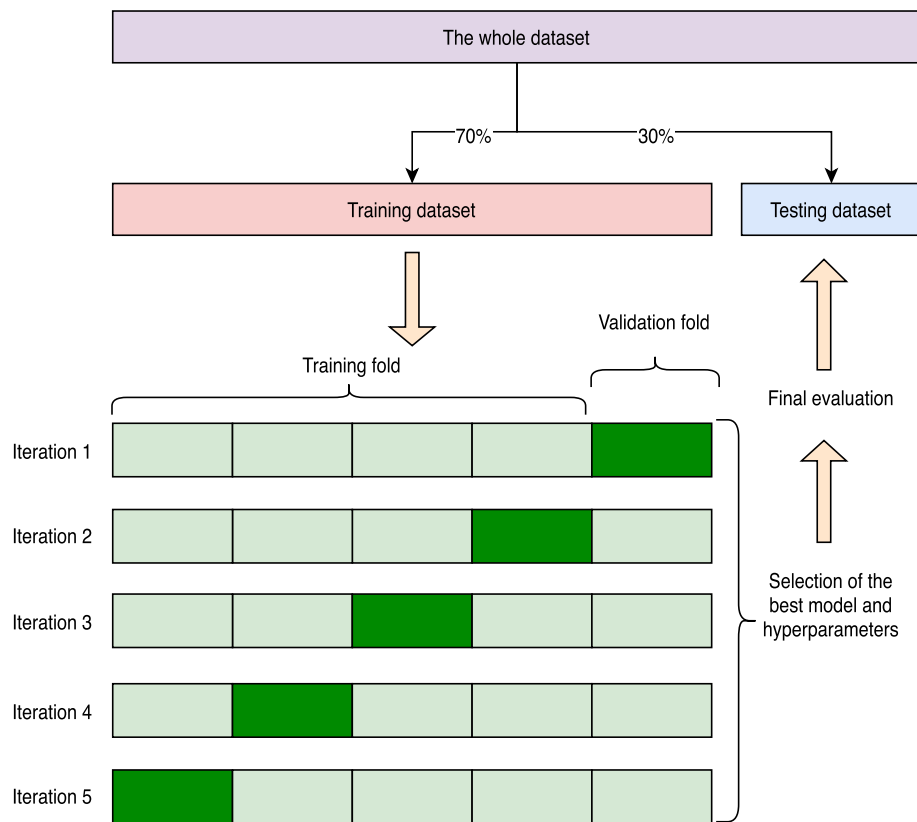
Fig. 7 Schematic representation of 5-fold cross-validation. Adapted from [24]

are selected to evaluate its performance by applying it to 30 percent of the test dataset.

Evaluation metrics

The developed models were evaluated using metrics to determine the ability to detect DM. Below, we briefly describe the estimation of each of these metrics:

- Accuracy: percentage of correctly classified examples among the total number of classified examples [41].

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \qquad (6)$$

  where TP are the true positives, FN are false negatives, FP are false positives and TN are true negatives.

- Sensitivity: measures the ability of the classifier to predict positive cases from those really positive [41].

$$Sensitivity : \frac{TP}{TP + FN} \qquad (7)$$

- Specificity: measures the ability of the classifier to predict negative cases from those really negative [41].

$$Specificity : \frac{TN}{TN + FP} \qquad (8)$$

Explainability analysis

To analyze the interpretability of the models, we used different methodologies. For the FCM-based model, computational simulations allow us to explore how certain changes in variables might affect the prediction of DM. This is crucial for assessing the impact of potential interventions. Additionally, since DM can be influenced by a variety of factors, the simulations allow to consider diverse scenarios and to evaluate personalized prevention strategies. For ANN and SVM-based models, we use Shapley values, which provide a measure of the individual contribution of each variable to the prediction. In the case of DM, this could help identify which specific factors are influencing the prediction of a particular risk. In addition, Shapley values not only allow an understanding of the overall impact of variables but also how they affect the prediction for individual cases. This is crucial for tailoring treatment strategies to specific patient needs. Finally, for XGBoost-based models, we use feature importance, which helps to identify the most influential

variables in DM prediction. This is essential for prioritizing efforts in disease management and prevention.

The importance of using different approaches for DM explainability analysis lies in the complementarity of their strengths. While Shapley values and feature importance provide a detailed understanding of the importance of variables, computational simulations allow the exploration of model behavior in hypothetical situations. By combining these approaches, a comprehensive analysis is achieved that not only highlights key variables but also provides valuable information on how the model reacts to changes in the environment and specific scenarios. This is essential to ensure that AI models applied to DM are not only accurate but also interpretable and useful in a clinical and decision-making context. The analysis of the variables in each model is shown in Discussion section.

Finally, we have carried out an ablation study to analyze the most important variables in the prediction of diabetes. This sensitivity analysis allows knowing the importance of each of the variables in the diagnosis by eliminating one by one the most important variables according to the previously performed explainability analysis. After the elimination of each variable, the performance of the model is observed. In this way, the impact of each characteristic on the performance of the model is evaluated.

Results

In this section, we present the results of exploratory data analysis, association tests, and the implementation of four AI techniques on the DM dataset described in the previous section.

Descriptive analysis

Table 3 shows the distribution of absolute and relative frequencies (percentages) for the characteristics concerning diagnosis. Regarding sex, it was observed that the frequency of DM is higher in women at 33.3%, while in non-diabetic patients men are found in higher frequency. Age was the only numerical variable within the dataset. The results showed that individuals ranged in age from 16 to 90 years. The mean age of all individuals was 48 years with a standard deviation of 12 years. About the study group, patients with DM had a higher mean age (mean = 49.1±12.1) than patients without DM (mean =46.4±12.1). Of all the clinical variables included in the dataset, we found that polyuria, polydipsia, sudden weight loss, weakness, polyphagia, visual blurring, and partial paresis were more frequent in individuals with DM.

Association results

To analyze the association between the predictor variables and the diagnosis of DM, we used the Chi-square

Table 3 Distribution of absolute and relative frequencies (%) of the characteristics present in the dataset

Characteristic	Category	Absolute frequency (%)		p
		Non-DM	**DM**	
Sex	Female	19(3.7)	173(33.2)	<0.001
	Male	181(34.8)	147(28.3)	
Polyuria	Absence	185(35.6)	77(14.8)	<0.001
	Presence	15(2.9)	243(46.7)	
Polydipsia	Absence	192(36.9)	95(18.3)	<0.001
	Presence	8(1.5)	225(43.3)	
Sudden weight loss	Absence	171(32.9)	132(25.3)	<0.001
	Presence	29(5.6)	188(36.2)	
Weakness	Absence	113(21.7)	102(19.3)	<0.001
	Presence	87(17.7)	218(41.3)	
Polyfagia	Absence	152(29.2)	131(25.2)	<0.001
	Presence	48(9.2)	189(36.4)	
Genital thrush	Absence	167(32.1)	237(45.6)	0.016
	Presence	33(6.3)	83(16.0)	
Visual blurring	Absence	142(27.3)	145(27.9)	<0.001
	Presence	58(11.2)	175(33.6)	
Itching	Absence	101(19.4)	166(31.9)	0.829
	Presence	99(19.1)	154(29.6)	
Irritability	Absence	184(35.4)	210(40.3)	<0.001
	Presence	16(3.1)	110(21.2)	
Delayed healing	Absence	114(21.9)	167(32.2)	0.326
	Presence	86(16.5)	153(29.4)	
Partial paresis	Absence	168(32.3)	128(24.6)	<0.001
	Presence	32(6.2)	192(36.9)	
Muscle stiffness	Absence	140(26.9)	185(35.6)	0.007
	Presence	60(11.5)	135(26.0)	
Alopecia	Absence	99(19.0)	242(46.5)	<0.001
	Presence	101(19.5)	78(15.0)	
Obesity	Absence	173(33.3)	259(49.8)	0.127
	Presence	27(5.2)	61(11.7)	

test. Table 3 shows the statistical significance (p-value) for the dependence between the characteristics and the presence of the disease. The results show that there is a dependence between all variables and the presence of DM, except itching, delayed healing, and obesity with p-values of 0.829, 0.326, and 0.127, respectively. Regarding age, there were significant differences between the ages of patients with DM and those without DM (p = 0.013). In this case, the average age of diabetics is higher than in non-diabetics.

Performance of the developed models

We applied four AI techniques to predict DM using sociodemographic and clinical information. Table 4 shows the performance results of each model expressed in

Hoyos *et al. BMC Medical Informatics and Decision Making*   (2024) 24:383

Page 12 of 20

Table 4 Performance results of the models developed in this research

Model	Optimal hyperparameters	Accuracy	Sensitivity	Specificity
PSO-FCM	Initial population = 200	0.95	0.96	0.94
	Activation function = sigmoid			
	Inference function = modified-Kosko			
ANN	Hidden layer units = 256	0.99	0.99	0.99
	Learning rate = 0.0001			
	Activation function = ReLU			
	Optimizer = Adam			
	Learning type = constant			
SVM	Kernel = radial	0.99	1.00	0.99
	C = 1.0			
	gamma = 1.0			
XGBoost	Randomly Selected Predictors = 5	1.00	1.00	1.00
	Trees = 1000			
	Minimal node size = 2			
	Tree depth = 10			
	Learning rate = 0.08321			
	Loss reduction = 7 x 10-9			
	Proportion observations sampled = 0.8852			

terms of accuracy, sensitivity, and specificity. The model built with XGBoost presented the best predictive performance with an accuracy, sensitivity, and specificity of 1.00. Of the four models developed, the FCM-based model is the only one that can be visualized due to its simplicity. Figure 8 shows a schematic representation of the FCM-based model for prediction. The FCM allows visualization of the influence of the concepts or predictor variables on the presence of DM.

Discussion

Characterization of the individuals included in the dataset

DM has increased in recent decades and has become one of the leading causes of mortality worldwide [1], due to reasons that include inadequate dietary habits due to unhealthy food consumption and sedentary lifestyle [2, 14]. In the present study, it was observed that variables such as age, polyuria, polydipsia, sudden weight loss, weakness, polyphagia, visual blurring, and partial paresis occurred more frequently in people with a positive diagnosis of DM. This finding is consistent with the frequencies reported by other authors for the variables in the dataset [8, 34]. In relation to the study group, patients with DM presented a higher average age. According to Nurjahan et al. [42], DM and other chronic diseases are very frequent and widespread, especially among the elderly. This may be associated with the fact that advanced age predisposes to the development of DM due to factors such as, decreased physical activity, increased adipose tissue, and increased insulin resistance

[1]. Indeed, the numbers of the aging population have increased in recent years [43], and several studies have related the risk of advanced age and associated complications [44, 45]. On the other hand, we have observed that there was a higher frequency of women with DM, according to Nipa et al. [34], gender, and other symptoms such as visual blurring, ictus, partial paresis, alopecia and weakness can be considered minor risk factors for DM, which increase if the person has a family history of the disease and habits such as, smoking [2]. Timely identification of these specific symptoms in patients can help the physician to detect DM more effectively.

Association analysis

Predicting the early onset of DM is a task that can be difficult due to the number of signs and symptoms to evaluate in each patient so that about 50% of these individuals are not diagnosed promptly [46]. The results showed that sex, polyuria, polydipsia, sudden weight loss, weakness, polyphagia, visual blurring, irritability, and partial paresis, presented significant statistical differences with respect to DM ($p<0.001$). Laila et al. [23] have stated that polyuria is the main indicator of DM risk. On the other hand, frequent urination during DM is generated because high blood sugar puts pressure on the kidneys, causing them to produce more urine to buffer the excess sugar, causing dehydration and a constant thirst signal [47]. Over time, these organs weaken moving towards progressive deterioration of their function, contributing to poor water and electrolyte compensation, given the
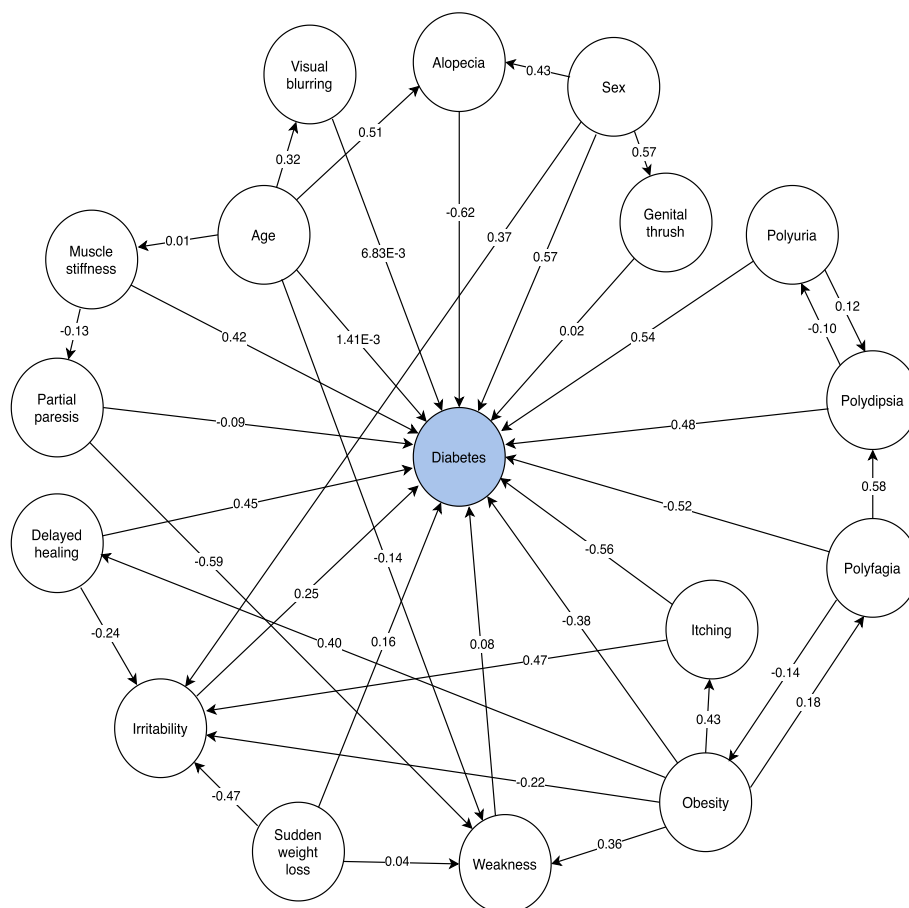
Hoyos *et al. BMC Medical Informatics and Decision Making*   (2024) 24:383

Page 13 of 20

Fig. 8 The FCM-based model developed to predict DM

sensitive losses during the disease and other over-aggravated causes (temperature, excessive sweating, diarrhea, fever, among others) [44]. For these reasons, Le et al. [33] have affirmed that polyuria and polydipsia are related characteristics, very important among patients with early manifestations of DM. Besides, symptoms such as sudden weight loss could also constitute an early signal for the onset of DM. This process is accompanied by an irrepressible and uncontrollable feeling of hunger (polyphagia) with a progressive reduction or gain of total body weight [48]. In this case, insulin insufficiency prevents glucose from reaching the body's cells through the bloodstream, so the body has to start burning fat and muscle to meet the daily energy demand. During this phenomenon, there may also be episodes characterized by muscle weakness or paralysis of any part of the body [34]. Other authors have also considered irritability and alopecia as early indicators of DM [8, 33].

In our study, characteristics such as delayed healing (p=0.326), genital thrush (p=0.016), itching (p=0.829), muscle stiffness (p=0.007), and obesity (p=0.127) did not present significant statistical differences concerning DM. Observing the behavior of these variables, we cannot directly relate them to DM, especially when the appearance of several of these symptoms can be attributable to other conditions, such as the environment [49, 50], heredity [51], hormonal regulation [52] or related to sex [53]. However, Lai et al. found a strong association between body mass index (BMI) and the prediction of DM [54].

Analysis of performance of the developed models

In this study, four AI techniques were used to build predictive models of DM: FCM, ANN, SVM, and XGBoost. The results showed excellent performance in all models; however, XGBoost achieved a perfect accuracy of 1.0; ANN and SVM achieved an accuracy of 0.99, and FCM obtained a slightly lower, but still significant, accuracy of 0.95.

The high accuracy achieved by the prediction models in this study can be attributed to several reasons. First, the AI techniques used, such as FCM, ANN, SVM, and XGBoost, are known for their ability to extract complex patterns and capture nonlinear relationships in the

Hoyos *et al. BMC Medical Informatics and Decision Making*   (2024) 24:383

Page 14 of 20

data allowing for excellent predictive performance and higher generalizability [55]. Second, the normalization technique applied before model training may have contributed to better comparability of the data and thus an improvement in overall accuracy. In addition, the SMOTE technique used to balance the classes may have been instrumental in improving the predictive ability of the models. Class imbalance in medical datasets, as in the case of DM, can negatively affect the model's ability to correctly recognize instances of the minority class.

On the other hand, the superior performance of XGBoost compared to the other techniques evaluated may be due to its ability to build deeper and more complex decision trees, as well as its ability to effectively handle nonlinear features and complex relationships in the data. Particularly, XGBoost is a boosting algorithm that focuses on iteratively improving model performance and can combine multiple weak models into a stronger one. This may have led to a better capture of the relationships between the most relevant variables/features, leading to improved prediction accuracy. This flexibility and superior predictive power may explain why XGBoost achieved perfect accuracy in this study. In terms of sensitivity and specificity, consistent performance was observed across all models, although again the XGBoost model showed the best results. This suggests that XGBoost was able to detect both positive and negative cases more accurately compared to the other models. Due to the very good results of XGBoost, we evaluated the ability of the XGBoost model to generalize and avoid overfitting by examining the training and validation losses. Training loss evaluates how well the model fits the training dataset. In contrast, validation loss evaluates the model's performance on the validation dataset. These two losses are usually visualized using curves to analyze the dynamics

of these two metrics. Figure 9 illustrates the behavior of the training loss and validation loss, where we can observe a remarkable decrease in the losses as the epochs increase. This behavior is observed for both subsets, indicating that there is no overfitting to the training dataset and the model can generalize to unseen data. This finding is important for DM detection because it strengthens the ability of our model to detect disease in previously unseen data.

Explanability analysis of the developed models
Model based on FCM
The results show that FCM can predict correctly DM in 95% of the cases. Despite the accuracy obtained by FCM, this model is simple to build, visualize and interpret. Additionally, it allows an evaluation of the variables involved in the prediction using scenario-based computational simulations. Figure 8 shows a graphical representation of the FCM developed where the predictor variables can be observed with their respective weights or influences on the class (presence or absence of DM). Due to the high complexity of the other models, it is not so easy to visualize them for the medical professional to analyze and interpret. In this case, FCM has an advantage over the other models.

Regarding the evaluation of factors over time, Fig. 10 shows an example of FCM simulations. The x-axis shows the simulated iterations and the y-axis shows the value of the variables or concepts. The simulation in Fig. 10 corresponds to a patient with polyuria, polydipsia, and polyphagia. After several simulated iterations, the system achieves an equilibrium state indicating that the concepts do not change value after iteration 72 (orange dotted line). In this plot, we can see how the model activates non-active variables from the beginning such
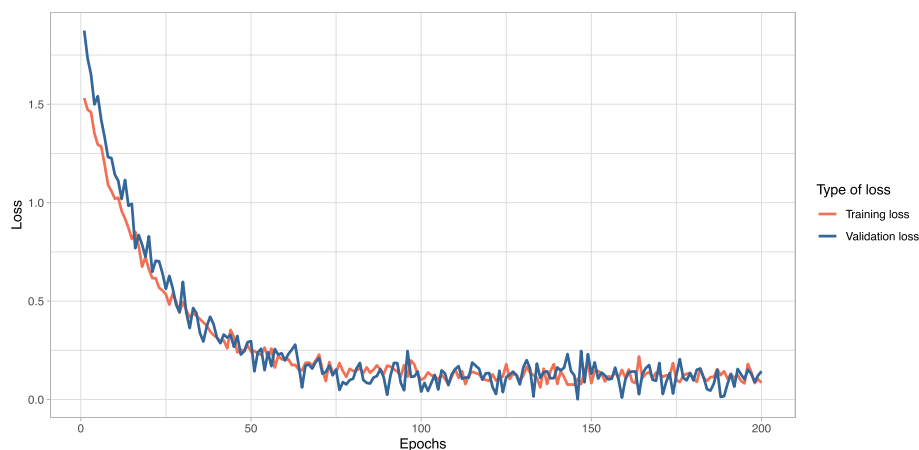

Fig. 9 Training and validation losses for the Xgboost model to predict DM
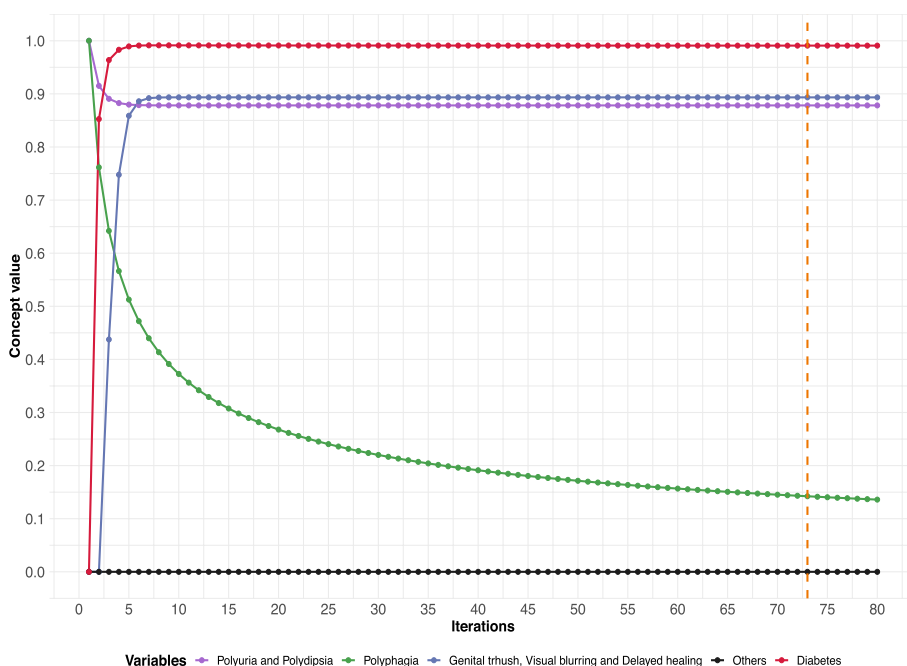
Fig. 10 Behavior of the variables for the prediction of DM using an FCM. The orange dotted line indicates when the system reaches equilibrium

as genital thrush, visual blurring and delayed healing (purple curve). On the other hand, we can see that the concept related to the diagnosis of DM (red curve) is activated from the first iteration, which indicates that the symptoms are characteristic of the disease. In this way, the FCM model not only makes a prediction but also shows the behavior of the variables that lead to that prediction. In this way, FCM allows building models in a simple and explainable way that allows medical professionals to have an overview of the problem and not only a prediction. Although XGBoost allows the evaluation of the importance of the features on the prediction, it does not allow an evaluation of factors over time (simulated iterations).

ANN and SVM-based models
To evaluate the impact of features on prediction for both ANN and SVM we use Shapley values. The objective of this methodology is to explain the prediction of an instance x by calculating the contribution of each feature to the prediction. Figure 11 shows a scatter plot of the SHAP values for ANN (see Fig. 11a) and SVM (see Fig. 11b), respectively. The X-axis of the plot represents the SHAP value; if the value moves to the right (positive value), then it indicates that it increases the final prediction, while if the value moves to the left (negative value), then it decreases the final prediction. Each SHAP value is color-coded: red represents the highest value of the attributes while blue represents the lowest. These two

figures show that the variables that most influence the prediction of DM are polyuria, polydipsia, and sex. Specifically, high values of polyuria, polydipsia, polyphagia, delayed healing, and blurred vision increase the prediction value, so these variables increase the risk of DM. On the other hand, the female sex (sex = 0) has a higher risk of presenting DM. Of the variables that have the least impact on the prediction of DM are genital thrush and obesity, perhaps because these are generated after presenting the disease.

Model based on XGBoost
One of the advantages of XGBoost is that it is a technique that allows us to identify the importance of each of the variables in the prediction. In Fig. 12, we observe the graphical representation of this importance which revealed that the most influential variables in the prediction of DM were polydipsia, polyuria, sex, and age. These results are consistent with the existing medical literature, where it has been established that these factors are significant indicators of the presence of DM [33, 56]. Some variables such as weakness, visual blurring, polyphagia, obesity and genital thrush, presented feature importance equal to 0, which may have a limited capacity to reduce the error in the tree partitions, perhaps due to the low correlation with the target variable. This behavior can also be explained by redundant information already explained by other more influential variables, such as polydipsia or polyuria.
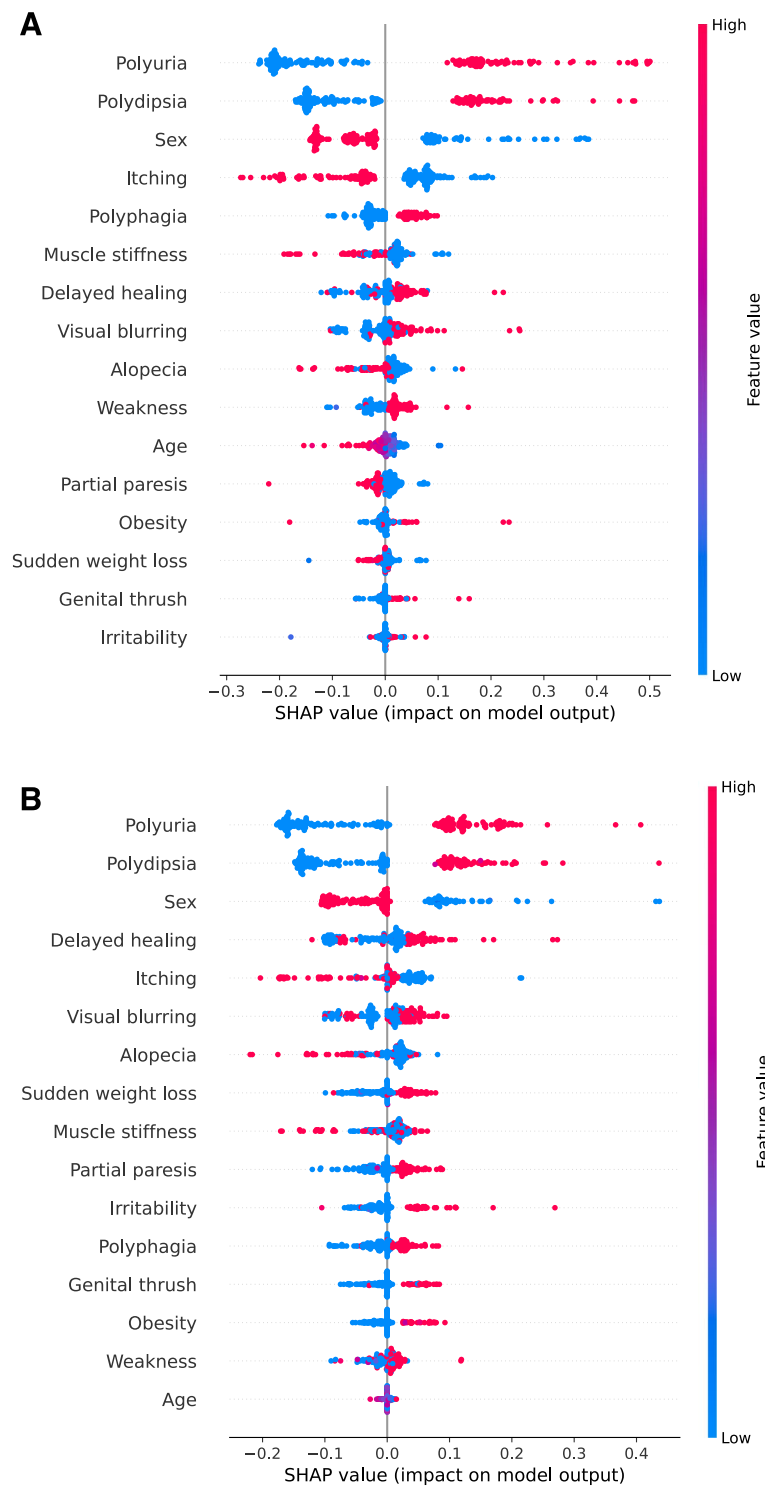
A

B

Fig. 11 SHAP values for ANN and SVM

Because Xgboost was the best model for predicting diabetes, we performed an *ablation study* to identify the most important features in the proposed classification model. We systematically eliminated one of the following characteristics: polyuria, polydipsia, sex, and age. These variables were chosen because they were the most
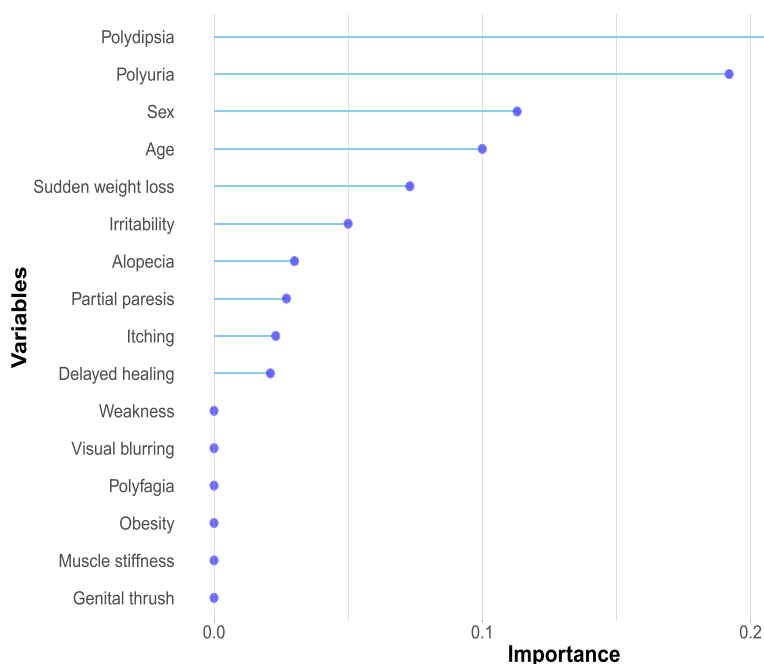
Hoyos *et al. BMC Medical Informatics and Decision Making*   (2024) 24:383

Page 17 of 20

Fig. 12 Feature importance to predict DM using XGBoost algorithm

important in the results according of the explainability analysis. The elimination of polydipsia and polyuria significantly reduces the accuracy of the model from 1.0 to 0.85 and 0.78, respectively. This suggests that polydipsia and polyuria are critical features for the prediction of DM in this dataset. Particularly, it is widely known in the literature that the presence of excessive thirst and excessive urine urination are strong indicators of the condition of patients with DM. On the other hand, removing sex and age has a moderate impact on accuracy, going from 1.0 to 0.90 and 0.91, respectively. This suggests that sex and age have a moderate relationship with the prediction of diabetes in this dataset. This could indicate that, in this particular case, polyuria and polydipsia are determining factors in the diagnosis of diabetes, however, age and sex are less so.

Quantitative comparison with previous studies
In this study, four models were developed to predict DM, of which the XGBoost-based model obtained an accuracy, sensitivity, and specificity of 1. These results indicate an outstanding performance of the model in accurately classifying positive and negative cases of DM. However, it is important to contextualize these findings in comparison to other previous studies that have also used AI-based models to address the same issue. To be fair, we collected studies that used the same dataset to build the models.

Table 5 Quantitative comparison of the best models reported in the literature for DM prediction using the same dataset

Study	Best model	Accuracy	Sensitivity	Specificity
Islam et al. [8]	RF	0.99	0.97	0.97
Chaves et al. [14]	ANN	0.98	0.98	0.97
Laila et al. [23]	RF	0.97	0.98	0.96
Ergün et al. [13]	CNN	0.99	0.98	-
Oladimeji et al. [57]	RF	0.98	-	-
Sadhu et al. [58]	RF	0.98	-	-
Hennebelle et al. [59]	RF	0.97	-	-
Our work	XGBoost	1.0	1.0	1.0

Table 5 shows a quantitative comparison of the performance of our best model with the best model from each study reported in the literature. This table compares the performance of the XGBoost model with other AI approaches such as RF, ANN, and CNN. Our XGboost model demonstrated an excellent ability to predict DM with 100% accuracy, indicating that the model correctly classifies 100% of the instances in the test set (data not previously seen by the model). The sensitivity and specificity of the model were 100%, indicating that the model generates neither false negatives nor false positives. Our best model outperforms models reported in the literature that used the same dataset. This could be because some previous studies did not

Hoyos *et al. BMC Medical Informatics and Decision Making*   (2024) 24:383

Page 18 of 20

employ data preprocessing techniques such as normalization and class balancing, which could have influenced the results and the comparison with XGBoost. Normalization standardizes features to ensure similar weight in model learning, and class balancing, using techniques such as SMOTE, is essential to address imbalances in the dataset. These techniques are crucial to improve the model's ability to correctly recognize and classify DM cases.

Conclusions

In this research, we set out to develop an explainability analysis of DM using statistical and AI techniques with a small dataset. We developed this analysis by combining these techniques with sociodemographic and clinical information related to DM. Initially, we carried out a descriptive analysis that allowed us to know the main characteristics of the individuals. Subsequently, an association analysis was performed to determine the association or relationship between the predictor variables and DM. Subsequently, prediction models built using AI techniques, such as FCM, ANN, SVM and XGBoost, showed excellent results in the prediction of DM. The XGBoost model stood out for superior performance in accuracy, sensitivity, and specificity compared to other works published in the literature. These results support the efficacy of AI techniques in the field of disease prediction and suggest their potential usefulness in clinical applications for DM. This work focused not only on disease prediction but also on a deep explainability analysis of the behavior of each of the variables used for prediction. Specifically, the four AI techniques used for building the models allowed studying the impact of the variables on the final prediction with different approaches to explainability analysis. Particularly, scenario-based computational simulations were used in the case of FCM, feature importance for XGboost, and SHAP values for ANN and SVM. In this way, we performed a deep explainability analysis of the MD, where we not only considered descriptive and inferential statistics but also the behavior of the variables involved in the prediction process.

This research has some limitations. First, only one dataset from a specific region was used. The results cannot be extrapolated to another region. Therefore, further research with larger and more varied datasets is recommended to confirm and validate these results. Second, only sixteen predictor variables were used, and no other variables of interest for the diagnosis of DM, such as physical exercise or laboratory test results. The development of models with these types of variables could provide a more robust analysis of DM. Finally, a limitation is that the data used for model construction were collected through a survey. Direct data collection with controlled measurements in patients with DM and in individuals without the disease would be interesting. Other future work derives from this work such as a deeper analysis of the explainability using different techniques existing in the literature. In addition, the development of hybrid models and the comparison of their performances and advantages with respect to the ML models used in this study. Of course, another future work is to use datasets with more data and variables as indicated above.

Abbreviations
DM	Diabetes mellitus
WHO	World health organization
AI	Artificial intelligence
RF	Random forest
ANN	Artificial neural networks
CNN	Convolutional neural networks
VAE	Variational automatic encoder
SHAP	SHapley additive exPlanations
SMOTE	Synthetic minority oversampling technique
FCM	Fuzzy cognitive map
MLP	Multilayer perceptron
SVM	Support vector machine
GBM	Gradient boosting machine
PSO	Particle swarm optimization

Authors' contributions
WH, KH, and RR conceived the idea. WH and JA developed the experiments and supervised the modeling results of this research. KH and RR interpreted the modeling results under the supervision of clinical experts. KH and RR drafted the manuscript under the supervision of WH and JA. All authors have read and approved the final manuscript.

Data availability
The dataset used in this research is available at https://doi.org/10.24432/C5VG8H.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1.   International Diabetes Federation. IDF Diabetes Atlas 10th edition. 2021. https://diabetesatlas.org/idfawp/resource-files/2021/07/IDF_Atlas_10th_Edition_2021.pdf. Accessed 25 Aug 2023.

2. Centers for Disease Control and Prevention. Type 2 Diabetes. 2021. https://www.cdc.gov/diabetes/basics/type2.html#print. Accessed 29 Aug 2023.

3. Elsayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, et al. 12. Retinopathy, Neuropathy, and Foot Care: Standards of Care in Diabetes-2023. Diabetes Care. 2023;46(1):S203– S215.

4. World Health Organization. Global report on Diabetes. 2016. https://www.who.int/publications/i/item/9789241565257. Accessed 10 Sep 2023.

5. GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet. 2023;402(10397203-234). https://doi.org/10.1016/s0140-6736(23)01301-6.

6. Standl E, Khunti K, Hansen TB, Schnell O. The global epidemics of diabetes in the 21st century: Current situation and perspectives. Eur. J. Prev. Cardiol. 2019;26(2_suppl):7–14. https://doi.org/10.1177/2047487319881021.

7. O'Connell JM, Manson SM. Understanding the economic costs of diabetes and prediabetes and what we may learn about reducing the health and economic burden of these conditions. Diabetes Care. 2019;42(9):1609–11. https://doi.org/10.2337/dci19-0017.

8. Islam MMF, Ferdousi R, Rahman S, Bushra HY. Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques. In: Gupta M, Konar D, Bhattacharyya S, Biswas S, editors. Advances in Intelligent Systems and Computing. Springer; 2020. pp. 113–25. https://doi.org/10.1007/978-981-13-8798-2_12.

9. Jothi N, Rashid NA, Husain W. Data Mining in Healthcare - A Review. Procedia Comput Sci. 2015;72(December):306–13. https://doi.org/10.1016/j.procs.2015.12.145.

10. Firdous S, Wagai G, Sharma K. A survey on diabetes risk prediction using machine learning approaches. J Fam Med Prim Care. 2022;11(11):6929. https://doi.org/10.4103/jfmpc.jfmpc_502_22.

11. Chaki J, Thillai Ganesh S, Cidham SK, Ananda Theertan S. Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review. J King Saud Univ Comput Inf Sci. 2022;34(6):3204–25. https://doi.org/10.1016/j.jksuci.2020.06.013.

12. Quintero Y, Ardila D, Camargo E, Rivas F, Aguilar J. Machine learning models for the prediction of the SEIRD variables for the COVID-19 pandemic based on a deep dependence analysis of variables. Comput Biol Med. 2021;134: 104500. https://doi.org/10.1016/j.compbiomed.2021.104500.

13. Ergün ÖN, O İlhan H. Early Stage Diabetes Prediction Using Machine Learning Methods. Eur J Sci Technol. 2021;(29):52–57. https://doi.org/10.31590/ejosat.1015816.

14. Chaves L, Marques G. Data Mining Techniques for Early Diagnosis of Diabetes: A Comparative Study. Appl Sci. 2021;11(5):2218. https://doi.org/10.3390/app11052218.

15. García-Ordás MT, Benavides C, Benítez-Andrades JA, Alaiz-Moretón H, García-Rodríguez I. Diabetes detection using deep learning techniques with oversampling and feature augmentation. Comput Methods Prog Biomed. 2021;202(105968). https://doi.org/10.1016/j.cmpb.2021.105968.

16. Reddy SS, Sethi N, Rajender R, Vetukuri VSR. Non-invasive Diagnosis of Diabetes Using Chaotic Features and Genetic Learning. In: Chen JIZ, Tavares JMRS, Shi F, editors. Lecture Notes in Networks and Systems. vol. 514 LNNS. Springer International Publishing; 2022. pp. 161–70. https://doi.org/10.1007/978-3-031-12413-6_13.

17. Reddy SS, Mahesh G. Risk Assessment of Type 2 Diabetes Mellitus Prediction using an Improved Combination of NELM-PSO. EAI Endorsed Trans Scalable Inf Syst. 2021;8(32). https://doi.org/10.4108/eai.3-5-2021.169579.

18. Swaroop CR, Jayamanasa V, Shankar RS, Babu MG, Shariff V, Kumar NSKM. Optimizing Diabetes Prediction through Intelligent Feature Selection: A Comparative Analysis of Grey Wolf Optimization with AdaBoost and Ant Colony Optimization with XGBoost. In: Algorithms in Advanced Artificial Intelligence. CRC Press; 2024. pp. 311–8. https://doi.org/10.1201/9781003529231-47.

19. Aguilar J, Salazar C, Velasco H, Monsalve-Pulido J, Montoya E. Comparison and Evaluation of Different Methods for the Feature Extraction from Educational Contents. Computation. 2020;8(2). https://doi.org/10.3390/computation8020030.

20. de Winter JCF. Using the student's t-test with extremely small sample sizes. Pract Assess Res Eval. 2013;18(10):1–12. https://doi.org/10.7275/E4R6-DJ05.

21. Lilliefors HW. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. J Am Stat Assoc. 1967;62(318):399–402. https://doi.org/10.1080/01621459.1967.10482916.

22. Tallarida RJ, Murray RB. Chi-Square Test. In: Manual of Pharmacologic Calculations. New York: Springer; 1987. pp. 140–2. https://doi.org/10.1007/978-1-4612-4974-0_43.

23. Laila UE, Mahboob K, Khan AW, Khan F, Taekeun W. An Ensemble Approach to Predict Early-Stage Diabetes Risk Using Machine Learning: An Empirical Study. Sensors. 2022;22(14):1–15. https://doi.org/10.3390/s22145247.

24. Hoyos W, Aguilar J, Toro M. Federated learning approaches for fuzzy cognitive maps to support clinical decision-making in dengue. Eng Appl Artif Intell. 2023;123(106371):1–15. https://doi.org/10.1016/j.engappai.2023.106371.

25. Schubach M, Re M, Robinson PN, Valentini G. Imbalance-Aware Machine Learning for Predicting Rare and Common Disease-Associated Non-Coding Variants. Sci Rep. 2017;7(1):2959. https://doi.org/10.1038/s41598-017-03011-5.

26. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–57. https://doi.org/10.1613/jair.953.

27. Kosko B. Fuzzy cognitive maps. Int J Man Mach Stud. 1986;24(1):65–75. https://doi.org/10.1016/S0020-7373(86)80040-2.

28. Hoyos W, Aguilar J, Toro M. PRV-FCM: An extension of fuzzy cognitive maps for prescriptive modeling. Expert Syst Appl. 2023;231: 120729. https://doi.org/10.1016/j.eswa.2023.120729.

29. Hoyos W, Aguilar J, Toro M. A clinical decision-support system for dengue based on fuzzy cognitive maps. Health Care Manag Sci. 2022;25(4):666–81. https://doi.org/10.1007/s10729-022-09611-6.

30. Aguilar J. Multilayer Cognitive Maps in the Resolution of Problems using the FCM Designer Tool. Appl Artif Intell. 2016;30(7):720–43. https://doi.org/10.1080/08839514.2016.1214422.

31. Shetty D, Varma J, Navi S, Ahmed M. Diving Deep into Deep Learning: History, Evolution, Types and Applications. Int J Innov Technol Exploring Eng. 2020;9(3):2835–2846. https://doi.org/10.35940/ijitee.A4865.019320.

32. Jahangir M, Afzal H, Ahmed M, Khurshid K, Nawaz R. An expert system for diabetes prediction using auto tuned multi-layer perceptron. In: 2017 Intelligent Systems Conference (IntelliSys). IEEE; 2017. pp. 722–8. https://doi.org/10.1109/IntelliSys.2017.8324209.

33. Le TM, Vo TM, Pham TN, Dao SVT. A Novel Wrapper-Based Feature Selection for Early Diabetes Prediction Enhanced with a Metaheuristic. IEEE Access. 2021;9:7869–84. https://doi.org/10.1109/ACCESS.2020.3047942.

34. Nipa N, Riyad MMH, Satu MS, Walliullah M, Howlader KC, Moni MA. Clinically Adaptable Machine Learning Model To Identify Early Appreciable Features of Diabetes In Bangladesh. Intell Med. 2023. https://doi.org/10.1016/j.imed.2023.01.003.

35. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–97. https://doi.org/10.1007/BF00994018.

36. Vapnik VN. An overview of statistical learning theory. IEEE Trans Neural Netw. 1999;10(5):988–99. https://doi.org/10.1109/72.788640.

37. Shrivastav SK, Ramudu PJ. Bankruptcy Prediction and Stress Quantification Using Support Vector Machine: Evidence from Indian Banks. Risks. 2020;8(2):52. https://doi.org/10.3390/risks8020052.

38. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM; 2016. pp. 785–94. https://doi.org/10.1145/2939672.2939785.

39. Wang L, Wang X, Chen A, Jin X, Che H. Prediction of Type 2 Diabetes Risk and Its Effect Evaluation Based on the XGBoost Model. Healthcare. 2020;8(3):247. https://doi.org/10.3390/healthcare8030247.

40. Dhaliwal S, Nahid AA, Abbas R. Effective Intrusion Detection System Using XGBoost. Information. 2018;9(7):149. https://doi.org/10.3390/info9070149.

41. Mago VK, Mehta R, Woolrych R, Papageorgiou EI. Supporting meningitis diagnosis amongst infants and children through the use of fuzzy cognitive mapping. BMC Med Inform Decis Making. 2012;12(1):98. https://doi.org/10.1186/1472-6947-12-98.

Hoyos *et al. BMC Medical Informatics and Decision Making*   (2024) 24:383

Page 20 of 20

42. Nurjahan, Rony MAT, Satu MS, Whaiduzzaman M. Mining Significant Features of Diabetes through Employing Various Classification Methods. In: 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD). Institute of Electrical and Electronics Engineers Inc.; 2021. pp. 240–4. https://doi.org/10.1109/ICICT4SD50815.2021.9397006.

43. World Health Organization. Ageing and health. 2022. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health. Accessed 03 Oct 2023.

44. Zhang J, Pan L, Guo Q, Lai Y, Liu T, Wang H, et al. The impact of global, regional, and national population ageing on disability-adjusted life years and deaths associated with diabetes during 1990–2019: A global decomposition analysis. Diabetes Metab Syndr Clin Res Rev. 2023;17(6): 102791. https://doi.org/10.1016/j.dsx.2023.102791.

45. Sattar N, Rawshani A, Franzén S, Rawshani A, Svensson AM, Rosengren A, et al. Age at Diagnosis of Type 2 Diabetes Mellitus and Associations With Cardiovascular and Mortality Risks. Circulation. 2019;139(19):2228–37. https://doi.org/10.1161/CIRCULATIONAHA.118.037885.

46. Oleiwi AK, Shi L, Wei L, Tao Y. A Comparative Analysis and Risk Prediction of Diabetes at Early Stage using Machine Learning Approach. Int J Futur Gener Commun Networking. 2020;13(3):4151–4163.

47. Timper K, Fenske W, Kühn F, Frech N, Arici B, Rutishauser J, et al. Diagnostic Accuracy of Copeptin in the Differential Diagnosis of the Polyuria-polydipsia Syndrome: A Prospective Multicenter Study. J Clin Endocrinol Metab. 2015;100(6):2268–74. https://doi.org/10.1210/jc.2014-4507.

48. Atrens DM. Schedule-induced polydipsia and polyphagia in nondeprived rats reinforced by intracranial stimulation. Learn Motiv. 1973;4(3):320–6. https://doi.org/10.1016/0023-9690(73)90022-2.

49. Giandoni MB, Grabski WJ. Cutaneous candidiasis as a cause of delayed surgical wound healing. J Am Acad Dermatol. 1994;30(6):981–4. https://doi.org/10.1016/S0190-9622(94)70122-9.

50. Xie Y, Thomas L, Johnston V, Coombes BK. Cervical and axioscapular muscle stiffness measured with shear wave elastography: A comparison between different levels of work-related neck disability. J Electromyogr Kinesiol. 2023;69: 102754. https://doi.org/10.1016/j.jelekin.2023.102754.

51. Dhurandhar NV, Petersen KS, Webster C. Key Causes and Contributors of Obesity: A Perspective. Nurs Clin North Am. 2021;56(4):449–64. https://doi.org/10.1016/j.cnur.2021.07.007.

52. Brenta G, Caballero AS, Nunes MT. Case finding for hypothyroidism should include type 2 diabetes and metabolic syndrome patients: A Latin American Thyroid Society (LATS) position statement. Endocr Pract. 2019;25(1):101–5. https://doi.org/10.4158/EP-2018-0317.

53. Sasani E, Rafat Z, Ashrafi K, Salimi Y, Zandi M, Soltani S, et al. Vulvovaginal candidiasis in Iran: A systematic review and meta-analysis on the epidemiology, clinical manifestations, demographic characteristics, risk factors, etiologic agents and laboratory diagnosis. Microb Pathog. 2021;154: 104802. https://doi.org/10.1016/j.micpath.2021.104802.

54. Lai H, Huang H, Keshavjee K, Guergachi A, Gao X. Predictive models for diabetes mellitus using machine learning techniques. BMC Endocr Disord. 2019;19(1):1–9. https://doi.org/10.1186/s12902-019-0436-6.

55. Hoyos W, Aguilar J, Toro M. An autonomous cycle of data analysis tasks for the clinical management of dengue. Heliyon. 2022;8(10): e10846. https://doi.org/10.1016/J.HELIYON.2022.E10846.

56. Haddad NG, Nabhan ZM, Eugster EA. Incidence of Central Diabetes Insipidus in Children Presenting with Polydipsia and Polyuria. Endocr Pract. 2016;22(12):1383–6. https://doi.org/10.4158/EP161333.OR.

57. Oladimeji OO, Oladimeji A, Oladimeji O. Classification models for likelihood prediction of diabetes at early stage using feature selection. Appl Comput Inform. 2021. https://doi.org/10.1108/aci-01-2021-0022.

58. Sadhu A, Jadli A. Early-Stage Diabetes Risk Prediction: A Comparative Analysis of Classification Algorithms. Int Adv Res J Sci Eng Technol. 2021;8(2):193–201. https://doi.org/10.17148/IARJSET.2021.8228.

59. Hennebelle A, Materwala H, Ismail L. HealthEdge: A Machine Learning-Based Smart Healthcare Framework for Prediction of Type 2 Diabetes in an Integrated IoT, Edge, and Cloud Computing System. Procedia Comput Sci. 2023;220:331–8. https://doi.org/10.1016/j.procs.2023.03.043.

Publisher's Note