A Novel MCDM-Based Framework to Recommend Machine Learning Techniques for Diabetes Prediction

Ajay Kumar[1,*], Kamaldeep Kaur[2]

[1]Department of IT, KIET Group of Institutions, Delhi-NCR, Ghaziabad, Uttar Pradesh, India

[2]USIC&T, Guru Gobind Singh Indraprastha University, New Delhi, India

Received 27 March 2023; received in revised form 04 June 2023; accepted 16 June 2023

DOI: https://doi.org/10.46604/ijeti.2023.11837

Abstract

Early detection of diabetes is crucial because of its incurable nature. Several diabetes prediction models have been developed using machine learning techniques (MLTs). The performance of MLTs varies for different accuracy measures. Thus, selecting appropriate MLTs for diabetes prediction is challenging. This paper proposes a multi-criteria decision-making (MCDM) based framework for evaluating MLTs applied to diabetes prediction. Initially, three MCDM methods—WSM, TOPSIS, and VIKOR—are used to determine the individual ranks of MLTs for diabetes prediction performance by using various comparable performance measures (PMs). Next, a fusion approach is used to determine the final rank of the MLTs. The proposed method is validated by assessing the performance of 10 MLTs on the Pima Indian diabetes dataset using eight evaluation metrics for diabetes prediction. Based on the final MCDM rankings, logistic regression is recommended for diabetes prediction modeling.

Keywords: diabetes prediction, machine learning techniques, WSM, TOPSIS, VIKOR

1. Introduction

Diabetes is among the most prevalent and severe health conditions worldwide. According to the International Diabetes Federation [1], 537 million individuals globally have diabetes. This number is projected to increase to 643 million by 2030. Over 18% of global deaths can be attributed to four primary diseases, namely chronic respiratory diseases, cancer, cardiovascular disease, and diabetes, emphasizing their significance as major public health concerns [2]. Obesity, poor dietary habits, elevated blood pressure, genetic predisposition to diabetes, advanced age, physical inactivity, and lifestyle factors all contribute to diabetes. As the disease progresses, patients with diabetes are more likely to develop health complications, such as heart disease, nerve damage, stroke, kidney failure, and vision problems.

Similar to many diseases, early diagnosis of diabetes is pivotal for managing diabetes and preventing its progression and severe symptoms. Diabetes is typically diagnosed either manually by medical professionals or through technology-driven methods. Each of these procedures has unique advantages and disadvantages. Although manual diagnosis by medical practitioners offers exceptional human insight, technological advancements have significantly improved this procedure, making it the predominant choice currently [1]. Technology-based approaches have the advantage of requiring less time and resources. In addition, in the early stage of the disease, technology can more efficiently identify the signs of diabetes than manual procedures while avoiding human error and complications. With the increasing availability of electronic health records, automated diabetes detection technologies have become increasingly appealing.

* Corresponding author. E-mail address: ajaygarg100@gmail.com

English language proofreader: Chih-Wen Teng

Studies have recently examined the performance of machine learning–based models in diabetes prediction and demonstrated their benefits in the healthcare industry [3-4]. Such models are widely used to predict not only diabetes but also other diseases. For example, Hussain et al. [5] examined the accuracy of four machine-learning models in the diagnosis of cardiovascular diseases.

Machine learning algorithms can be used to establish models based on training data to make predictions or judgments without requiring extensive programming. The predictive capability of machine learning techniques (MLTs) considerably varies across several competing model performance measures (PMs). Furthermore, the widely accepted no-free lunch theory in computational intelligence refutes the existence of a single prediction strategy that can outperform other techniques across all competing model PMs for a specific application domain [6]. Thus, selecting the appropriate technique for building accurate diabetes prediction models is challenging. To address this challenge, this paper proposed a method to identify the most suitable MLT for diabetes prediction among various available MLTs. This method involved optimizing the performance of these MLTs by considering various PMs altogether.

This research suggests an MCDM-based approach for choosing the most appropriate machine learning technique for diabetes prediction in the presence of several evaluation metrics. This study makes a variety of contributions, which can be summed up as follows:

(1) This study presents a novel MCDM-based approach to identify the most appropriate MLT among various available ones for diabetes prediction.

(2) In the proposed framework, three MCDM methods, namely the weighted sum method (WSM), a technique for order of preference by similarity to ideal solution (TOPSIS), and VlseKriterijumska optimizacija I kompromisno resenje (VIKOR), were initially applied to determine the individual ranking scores of MLTs for diabetes prediction performance based on multiple conflicting PMs.

(3) Next, a fusion approach (rank position method (RPM)) was employed to determine the final rank of each MLT because of variations in ranks evaluated using the three MCDM methods.

(4) To ensure the validity of the suggested framework, an experimental study was conducted to evaluate the diabetes prediction performance of 10 MLTs over the PID dataset, considering eight evaluation metrics.

(5) This paper also performed the Bayesian sign test to demonstrate that the MLT for diabetes prediction modeling recommended by the proposed MCDM-based approach significantly outperformed other MLTs.

The remainder of the paper is structured as follows. Section 2 presents the literature review. Sections 3 and 4 discuss the proposed framework and experimental study, respectively. Section 5 presents the results and discussion. Section 6 concludes the paper.

2. Literature Review

Numerous studies globally have utilized various strategies to improve and evaluate the capability of MLTs for predicting diabetes. Birjais et al. [7] applied three machine learning models, namely logistic regression (LR), Naive Bayes (NB), and gradient boosting, for diabetes prediction with the Pima Indian diabetes (PID) dataset. The authors used three PMs to measure the performance of these models: specificity, sensitivity, and accuracy. LR predicted diabetes with an accuracy, specificity, and sensitivity of 79.2%, 100%, and 77.8%, respectively. NB exhibited an accuracy, specificity, and sensitivity of 77%, 80.4%, and 66.6%, respectively. Gradient boosting demonstrated an accuracy, specificity, and sensitivity of 86%, 71.4%, and 81.5%, respectively.

Khanam and Foo [8] conducted a comparative study to examine the performance of the following seven machine learning models for diabetes prediction by using the PID dataset: linear regression, support vector machine (SVM), decision tree, k-nearest neighbor (KNN), random forest (RF), AdaBoost (AB), and NB. They calculated accuracy, F-measure, recall, and precision to examine the performance of these models.

Ramesh et al. [9] developed a model using an SVM for predicting diabetes. The authors used the PID dataset to validate their proposed model. Gupta et al. [10] developed two diabetes prediction models, namely quantum machine learning and deep learning models. An experimental study was performed using the PID dataset. The performance of the proposed models was examined by measuring the false detection rate, balanced accuracy, specificity, F1-score, recall, and precision. Houri et al. [11] developed diabetes prediction models using a decision tree and XGBoost.

Ganie and Malik [12] compared six MLTs, namely artificial neural network, RF, NB, decision tree, SVM, and LR, for diabetes prediction. They indicated that NB and RF outperformed other machine learning models. Panda et al. [13] developed diabetes prediction models by using four MLTs, namely SVM, gradient boosting, KNN, and LR. They compared the performance of these models for diabetes prediction over the PID dataset by calculating the F1-score, recall, precision, and accuracy.

Olisah et al. [14] proposed the 2GDNN framework for detecting diabetes. Their proposed framework adopts Spearman correlation for feature selection and polynomial regression for imputing missing values to increase predictive performance for diabetes detection. The authors compared the performance of the suggested model 2GDNN with RF and SVM over two diabetes datasets: the laboratory data of the medical city hospital (LMCH) and the PID dataset. The authors used four PMs for comparison: accuracy, F1-score, sensitivity, and precision. They reported that their proposed method, 2GDNN, outperformed RF and SVM.

Azit et al. [15] used SVM, artificial neural network, LR, and chi-square automatic interaction detection to predict diabetes. They compared the performance of these models in terms of six PMs, namely specificity, sensitivity, classification error, accuracy, negative predictive value, and positive predictive value. While no single machine learning model excelled in all six PMs, SVM emerged as the most suitable choice by performing best in five of them.

Tasin et al. [16] developed a system to automatically predict diabetes using different machine learning approaches, a private dataset of 203 female patients in Bangladesh, and a PID dataset. They addressed class imbalance using adaptive synthetic sampling (ADASYN) and synthetic minority oversampling technique (SMOTE). The authors used MLTs such as RF, LR, KNN, SVM, decision trees, and various ensemble techniques to ascertain which algorithm yields the best prediction outcomes. Recently, Aguilera-Venegas et al. [17] found that for diabetes patient classification, machine learning models performed better than traditional statistical methods. They compared four machine learning methods, KNN, RF, neural networks, and decision tree techniques, for diabetes prediction by measuring accuracy.

The following observations were made after thoroughly reviewing the literature for diabetes prediction using MLTs:

(1) Most researchers have demonstrated the use and efficiency of MLTs for diabetes prediction modeling.

(2) In some cases, researchers have used only single PMs to evaluate MLTs. Some studies have considered multiple PMs to evaluate MLTs for diabetes prediction. However, these studies have not considered the simultaneous optimization of all PMs.

(3) No study has recommended the most suitable diabetes prediction model considering various PMs.

This study proposed a multi-criteria decision-making (MCDM)-based framework to recommend MLTs for diabetes prediction in the presence of multiple PMs by considering simultaneous optimization of all PMs. To the best of the authors' knowledge, no study has identified the most suitable MLT for diabetes prediction using MCDM. However, MCDM has been

used to evaluate MLTs for solving other prediction problems. For example, in a recent study, Chowdhury et al. [18] utilized the concept of MCDM to select MLTs for predicting COVID-19 disease. Song and Peng [19] evaluated various machine-learning algorithms for predicting financial risk. Kumar and Kaur [20] proposed the MCDM-based evaluation of different machine-learning algorithms for software reliability prediction. Ali et al. [21] presented a precise MCDM method that empirically assesses and ranks classifiers, allowing end users to select the highest-ranked classifier for their application to train and create classification models.

3. Proposed Method

This paper proposes a novel MCDM-based approach to evaluate MLTs for diabetes prediction. Various PMs and accuracy were considered in this approach. MCDM is a well-known technique for selecting the most appropriate alternative among available ones based on various criteria [22]. Numerous MCDM methods have been used in the literature, and all of them use a decision matrix as the input for ranking alternatives. These matrices represent alternative performance relative to evaluation criteria.

Because the proposed approach involved various PMs (evaluation criteria), the selection of the most suitable machine learning model (alternative) for diabetes prediction was modeled as an MCDM problem. The proposed methods involve the following major steps:

(1) Train different MLTs for diabetes prediction by using the diabetes dataset.

(2) Measure the performance of MLTs for diabetes prediction by examining various PMs and present the results as a decision matrix.

(3) Apply three MCDM methods, namely WSM, TOPSIS, and VIKOR, on the decision matrix to obtain the individual rankings of diabetes prediction models.

(4) Apply the fusion approach (RPM) to determine the final rankings of diabetes prediction models.

(5) Recommend the MLT with the highest rank for diabetes prediction.

Fig. 1 presents an overview of the proposed approach followed by a detailed description.

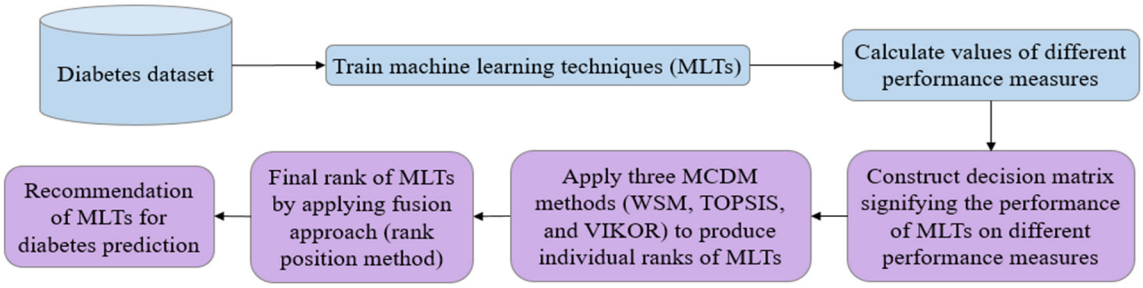


Fig. 1 Overview of the proposed approach

Step 1: Train various available MLTs (let us call it m) by using the diabetes prediction dataset.

Step 2: Measure the performance of MLTs in terms of various PMs (let us call it n.)

Step 3: Construct the decision matrix $D_{m\times n.}$, where each entry d_{ij} represents the predictive capability of diabetes detection for the i^{th} MLT on performance measure j.

Step 4: Apply the three MCDM methods on the decision matrix $D_{m\times n}$ to obtain the ranks of MLTs for diabetes prediction. To obtain reliable rankings, this study employed three MCDM methods, namely WSM, TOPSIS, and VIKOR, which are thoroughly discussed in Section 4.4.

Step 5: The individual ranks of MLTs obtained using the three MCDM methods were fused to determine the final ranks using the RPM. The RPM considers each alternative's position by each subordinate ranking technique [23]. In this study, alternatives were MLTs used for diabetes prediction modeling and subordinate ranking techniques were three MCDM methods. The RPM is based on the RPM score. The RPM score for each MLT (MLT_i) was employed to obtain the final ranks as follows:

$$\mathrm{RPM}\left(MLT_i\right)=\frac{1}{1/\mathrm{rank}\left(\mathrm{WSM}\right)+1/\mathrm{rank}\left(\mathrm{TOPSIS}\right)+1/\mathrm{rank}\left(\mathrm{VIKOR}\right)} \tag{1}$$

where rank (WSM), rank (TOPSIS), and rank (VIKOR) represent the individual rankings of MLT_i produced by three MCDM methods WSM, TOPSIS, and VIKOR respectively.

Step 6: Assign the final rank to the MLTs based on the RPM score determined in Step 5. Lower RPM scores result in higher ranks. The MLT with the highest rank is then recommended for diabetes prediction modeling.

4. Experimental Setup

To validate the proposed framework, an experimental study was conducted, and the effectiveness of 10 MLTs was compared for predicting diabetes by using a diabetes dataset, considering eight PMs. In the subsequent subsections, the diabetes dataset, PMs, MLTs, and MCDM methods are discussed in detail.

4.1. Diabetes dataset

This section is divided into two subsections. Section 4.1.1 presents a detailed description of the diabetes dataset, including the total number of observations, total number of attributes, and type of attributes. Section 4.1.2 describes the preprocessing of the dataset.

4.1.1. Description

This study used the PID dataset [16] that has been widely employed in previous studies [2, 7-10, 13-16] for developing diabetes prediction models. This dataset was used to identify American Pima Indians with diabetes. All patients in the PID dataset are women and at least 21 years old. The dataset contains the details of 768 patients, with their nine attributes (Table 1): number of pregnancies, age (years), diastolic blood pressure (mmHg), body mass index (BMI), skin fold thickness (mm), serum insulin level (μU/mL), glucose level (mg/dL), diabetes hereditary factor pedigree function, and outcomes. The attribute "outcome" was considered a dependent or target variable, whereas the remaining eight attributes were included as independent/feature variables. The diabetes attribute "outcome" is a binary value, where 0 indicates the absence of diabetes and 1 indicates the presence of diabetes.

Table 1 The attributes of the PID dataset

Attribute	Description	Type and measurement	Mean
Pregnancy	Number of times the female is pregnant	Numeric	3.8
Age	Age of the person	Numeric (years)	33
BP	Diastolic blood pressure	Numeric (mmHg)	69.1
BMI	Body mass index	Numeric (kg/m^2)	32
Skin thickness	The thickness of the triceps skin fold	Numeric (mm)	20.5
Insulin	Serum insulin level after 2 hours	Numeric (μU/mL)	79.8
Glucose	2-hour plasma glucose level during an oral glucose tolerance test	Numeric	120.8
Pedigree	Diabetes pedigree function	Numeric	0.47
Outcome	Target variable (0: non-diabetic, 1: diabetic)	Nominal	-

4.1.2. Preprocessing

Preprocessing aids in data transformation, facilitating the development of a more accurate machine-learning model. Preprocessing involves several tasks, such as filling in missing values, handling outliers, and feature selection. The likelihood of missing values or outliers in any secondary data retrieved from a repository is relatively high. The possibility of missing values in data increases for a medical dataset [7]. Before 2011, the PID dataset in the UCI machine learning repository had no missing values [7]. However, the zeros that have replaced missing values are biologically impossible to have various attributes at these places, such as age and blood pressure. The number of missing values in the PID dataset is listed in Table 2.

Table 2 The number of missing values in the PID dataset

Attribute	Pregnancy	Age	BP	BMI	Sick thickness	Insulin	Glucose	Pedigree
Number of missing values	0	0	35	11	227	374	5	0

Numerous approaches may be applied to handle missing values in the dataset. For dealing with missing values in the PID dataset, this study used KNN imputation. KNN imputation replaces missing values in a dataset by using the KNN method. The application of KNN imputation for discrete, continuous, categorical, and ordinal data makes it a superior technique for handling all types of missing data. Eliminating outliers when using distance-based algorithms, such as SVM and LR, is essential [2]. The outliers have been removed using the interquartile range (IQR) method.

4.2. Machine learning techniques

In this study, 10 MLTs were considered for diabetes prediction. These MLTs are NB, LR, SVM, KNN, AB, Bagging, decision table (DT), decision stump (DS), RF, and OneR. The performance of these MLTs was evaluated considering the eight PMs as described in the following section.

4.3. Performance measures

For the evaluation of MLTs, eight PMs were calculated: sensitivity, accuracy, F-measure, Matthews correlation coefficient (MCC), false positive rate (FPR), specificity, false negative rate (FNR), and receiver operating characteristic area under the curve (ROC-AUC). These PMs can be calculated in terms of true positive (T_rP_s), false positive (F_lP_s), true negative (T_rN_g), and false negative (F_lN_g). T_rP_s represent correctly classified positive class, F_lP_s denote incorrectly classified positive class, T_rN_g indicates correctly classified negative class, and F_lN_g represents incorrectly classified negative class.

Accuracy: Accuracy is the proportion of all correctly classified observations to the total number of observations [24].

$$\text{Accuracy} = \frac{T_rP_s + T_rN_g}{T_rP_s + F_lN_g + F_lP_s + T_rN_g} \tag{2}$$

Sensitivity: Sensitivity is the proportion of correctly classified positive observations to all observations in the actual class as positive [24].

$$\text{Sensitivity} = \frac{T_rP_s}{T_rP_s + F_lN_g} \tag{3}$$

False positive rate (FPR): FPR is the proportion of incorrectly classified positive observations to all observations in actual class as negative [25].

$$\text{FPR} = \frac{F_lP_s}{F_lP_s + T_rN_g} \tag{4}$$

F-measure: F-measure presents the balance between precision and recall [25].

$$\text{F-measure} = \frac{2 \times T_r P_s}{2 \times T_r P_s + F_l N_g + F_l P_s} \tag{5}$$

Matthews correlation coefficient (MCC): MCC is widely used for measuring the performance measure for measuring the performance of classifiers. The correlation between the target and prediction is measured using MCC [24].

$$\text{MCC} = \frac{T_r P_s \times T_r N_g - F_l P_s \times F_l N_g}{\sqrt{\left(T_r P_s + F_l P_s\right) \times \left(T_r P_s + F_l N_g\right) \times \left(T_r N_g + F_l P_s\right) \times \left(T_r N_g + F_l N_g\right)}} \tag{6}$$

Specificity: Specificity [25] is the ratio of predicted false observations to the total number of observations that are identified as false.

$$\text{Specificity} = \frac{T_r N_g}{T_r N_g + F_l P_s} \tag{7}$$

False negative rate (FNR): FNR is the proportion of incorrectly classified negative observations to all observations in actual class as positive [25].

$$FNR = \frac{F_l N_g}{F_l N_g + T_r P_s} \tag{8}$$

Receiver operating characteristic area under the curve (ROC-AUC): The ROC curve analysis compares the true positive rate and FPR in classification results, whereas AUC characterizes the ROC of a classifier. The classifier's performance is more effective when the ROC-AUC score is higher [5].

4.4. MCDM methods

When decisions must be made based on conflicting criteria, various MCDM methods are available. Every MCDM method has benefits and drawbacks. Currently, no approach permits the selection of a specific MCDM method. This study used three MCDM methods, namely WSM, TOPSIS, and VIKOR, instead of a single MCDM method to produce a more trustworthy ranking of MLTs for diabetes prediction modeling. In the following subsections, all three MCDM methods are explained in detail.

4.4.1. WSM

WSM is a widely used MCDM method for selecting the best alternative among different available alternatives in the presence of various criteria [26]. The steps for a detailed process are given below.

Step 1: Use the decision matrix $D_{m \times n.}$ as the input, where each entry d_{ij} represents the predictive capability of diabetes detection for the i^{th} MLT for performance measure j.

Step 2: Normalized decision matrix $ND_{m \times n}$ is obtained by,

$$ND_{m \times n} = \frac{d_{ij}}{\sqrt{\sum_{i=1}^{m} {d_{ij}}^2}},\ j = 1\ to\ n \tag{9}$$

Here, m represents the number of diabetes prediction models and represents the number of evaluation criteria.

Step 3: Determine the total benefit of machine learning-based diabetes prediction models by using the following equation:

$$A_i^{tot\text{-}benefit} = \sum_{j=1}^{bc} w_j nd_{ij},\ i = 1,\ 2,\ \ldots,\ m \tag{10}$$

Here, w_j represents the weight of the j^{th} criterion (performance measure), and *bc* denotes the benefit criterion (performance measure). The benefit performance measure is a measure for which a maximum value is desired. The value of *bc* is six in this study.

Step 4: Use the following equation to determine the total cost of machine learning-based diabetes prediction models.

$$A_i^{tot\text{-}cost} = \sum_{j=1}^{cc} w_j nd_{ij},\ i = 1,\ 2,\ \ldots,\ m \tag{11}$$

Here, w_j represents the weight of the j^{th} criterion (performance measure), and *cc* denotes the cost criterion (performance measure). The cost performance measure is a measure for which a minimum value is desired. The value of *cc* is two in this study.

Step 5: Calculate the score of WSM by using the following equation:

$$A_i^{ws} = A_i^{tot\text{-}benefit} - A_i^{tot\text{-}cost} \tag{12}$$

Here, A_i^{ws} represents the WSM score of the i^{th} diabetes prediction model.

Step 6: Use the WSM score to rank MLTs for diabetes prediction modeling. The highest rank is assigned to the MLT with the highest WSM score.

4.4.2. TOPSIS

TOPSIS is a well-known MCDM technique used for rating available methods to address a decision problem with competing criteria [27]. This technique chooses the alternative that is closest to the ideal alternative. An ideal alternative is defined as the alternative with the best possible criterion value. The Euclidean distance is utilized as the distance measure. Below is a step-by-step procedure.

Step 1: Use the decision matrix $D_{m\times n.}$ as the input, where each entry d_{ij} represents the predictive capability of diabetes detection for the i^{th} MLT for performance measure *j*.

Step 2: Normalized decision matrix $V_{m\times n}$ is obtained by,

$$v_{ij} = \frac{d_{ij}}{\sqrt{\sum_{i=1}^{m} d_{ij}^{\,2}}};\ j = 1,\ 2,\ \ldots,\ n \tag{13}$$

where each entry v_{ij} represents the normalized value of d_{ij}. Here, *m* represents the number of diabetes prediction models, and *n* represents the number of evaluation criteria.

Step 3: Weighted normalized decision matrix $T_{m\times n}$ is obtained by,

$$t_{ij} = v_{ij} \times w_j,\ i = 1\ to\ m\ and\ j = 1\ to\ n \tag{14}$$

where t_{ij} represents the weighted normalized value of d_{ij}. Here w_j is the weight assigned to criteria *j*.

Step 4: $PIS_{n\times 1}$ and $NIS_{n\times 1}$ (ideal solutions)

The positive ideal solution (*PIS*) has the highest value that each criterion may achieve. Calculating the negative ideal solution (*NIS*) involves determining the least/worst value that each criterion may attain. *PIS* and *NIS* can be determined as follows:

$$PIS = \left\{\left(\max\ T_{ij} / j \in z\right), \left(\min\ T_{ij} / j \in z'\right),\ i = 1\ to\ m\right\} = \left\{T_1^+,\ T_2^+,\ T_3^+,\ \ldots,\ T_n^+\right\} \tag{15}$$

$$NIS = \left\{\left(\min\ T_{ij} / j \in z\right), \left(\max\ T_{ij} / j \in z'\right),\ i = 1\ to\ m\right\} = \left\{T_1^-,\ T_2^-,\ T_3^-,\ \ldots,\ T_n^-\right\} \tag{16}$$

where z is related to benefit criteria and z' is related to cost criteria.

Step 5: The Euclidean distances ED^+ from *PIS* and ED^- by using the following equations:

$$ED_i^+ = \sqrt{\sum_{j=1}^{m}\left(T_{ij} - T_j^+\right)^2},\ i = 1\ to\ m \tag{17}$$

$$ED_i^- = \sqrt{\sum_{j=1}^{m}\left(T_{ij} - T_j^-\right)^2},\ i = 1\ to\ m \tag{18}$$

Step 6: Calculate $RC_{m\times 1}$ (relative closeness)

$$RC_i = \frac{ED_i^-}{\left(ED_i^- + ED_i^+\right)} \tag{19}$$

Eq. (19) can be used to determine *RC* (how closely each alternative comes to negative and PIS.)

Step 7: Ranking of alternatives (diabetes prediction models)

Arrange alternatives (machine learning techniques for diabetes prediction modeling) in the decreasing order of *RC* value, based on the value of *RC* obtained in Step 6. The MLT with the maximum value of *RC* will be suggested as the most suitable MLT for diabetes prediction modeling.

4.4.3. VIKOR

VIKOR is a widely used MCDM method. This method produces the ranking index of alternatives based on a particular measure of closeness to the ideal solution in the presence of conflicting criteria. Following is the detailed procedure of the VIKOR method [28].

Step 1: Construction of decision matrix as the input for the VIKOR method.

[Description] same as described in Step 1 of WSM and TOPSIS in previous Sections 4.4.1 and 4.4.2, respectively.

Step 2: Find the best d_j^+ and worst d_j^- values for each criterion by using the following equations:

$$For\ benefit\ criteria,\ d_j^+ = \max_i d_{ij},\ d_j^- = \min_i d_{ij};\ j = 1\ to\ n,\ i = 1\ to\ m \tag{20}$$

$$For\ cost\ criteria,\ d_j^+ = \min_i d_{ij},\ d_j^- = \max_i d_{ij};\ j = 1\ to\ n,\ i = 1\ to\ m \tag{21}$$

Step 3: Calculate the S_i (utility measure) and R_i (regret measure) by using the following equations:

$$S_i = \sum_{j=1}^{n}\frac{w_j\left(d_j^+ - d_{ij}\right)}{d_j^+ - d_j^-};\ j = 1\ to\ n,\ i = 1\ to\ m \tag{22}$$

$$R_i = \max_j\left[\frac{w_j\left(d_j^+ - d_{ij}\right)}{d_j^+ - d_j^-}\right];\ j = 1\ to\ n,\ i = 1\ to\ m \tag{23}$$

where w_j is the weight of criteria j.

Step 4: Calculate the values (S^-, S^*) and (R^-, R^*) by using the following relations.

$$S^- = \max_i S_i,\ S^* = \min_i S_i;\ i = 1\ to\ m \tag{24}$$

$$R^- = \max_i R_i,\ R^* = \min_i R_i;\ i = 1\ to\ m \tag{25}$$

Step 5: Compute the value of the VIKOR index Q_i for each alternative as follows:

$$Q_i = \frac{1}{2}\left(\frac{S_i - S^*}{S^- - S^*} + \frac{R_i - R^*}{R^- - R^*}\right);\ i = 1,\ 2,\ \ldots,\ m \tag{26}$$

Step 6: Rank alternatives (in this study, diabetes prediction models) in the order of the Q_i value, with a smaller Q_i value indicating a higher rank.

4.5. Experimental design

Fig. 2 presents a graphical representation of the experimental study design. The experimental study can be divided into two phases. In the first phase, diabetes prediction models were applied to obtain the results of eight performance measures. In the second phase, the MCDM ranking of MLTs was obtained. A detailed explanation is provided as follows.

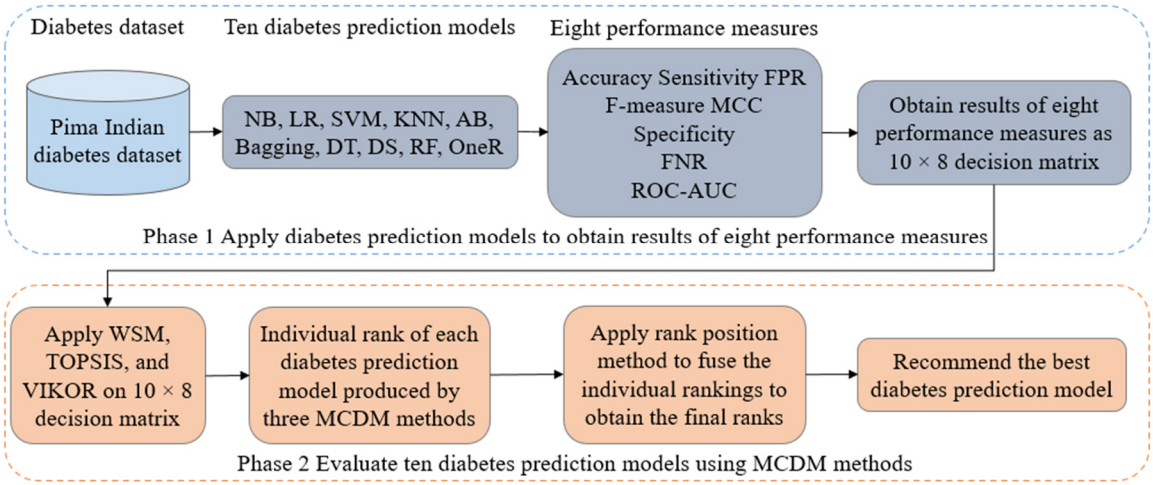


Fig. 2 Graphical representation of the experimental design

Phase 1: This phase involved building diabetes prediction models by applying 10 MLTs described in Section 4.2 on the PID dataset. For building all 10 diabetes prediction models, the open-source tool Waikato Environment for Knowledge Analysis (WEKA) version 3.8.3 was used [29]. Next, the performance of the models was examined in terms of eight PMs (accuracy, sensitivity, FPR, F-measure, MCC, specificity, FNR, and ROC-AUC) as described in detail in Section 4.3. The results are stored in a 10 × 8 matrix.

Phase 2: Ten diabetes prediction models were evaluated using three MCDM methods, as described in detail in Section 4.4. First, the 10 × 8 matrix obtained from Phase 1 was used as the input for all three MCDM methods for generating the individual ranks of all the ten diabetes prediction models. Next, the individual MLT ranks of MLTs obtained using the three MCDM methods were combined to obtain the final rank list by using the RPM described in Section 3. Finally, the diabetes prediction model with the highest rank was recommended.

5. Results and Discussions

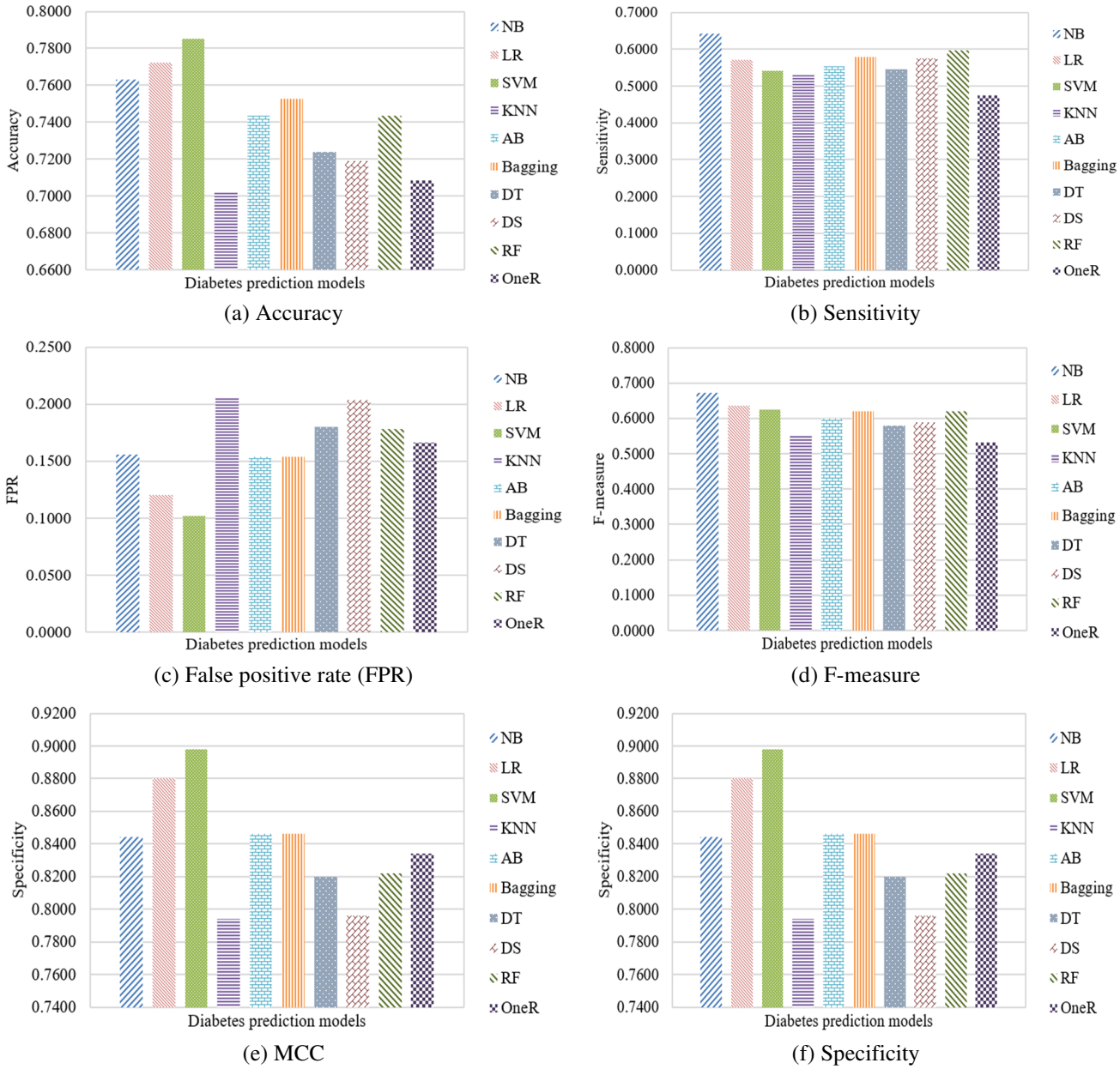
This section is divided into two subsections. The first subsection presents the diabetes prediction results of 10 MLTs concerning eight PMs. The proposed MCDM-based method was applied in the second subsection to identify the most suitable MLT for diabetes prediction modeling.

5.1. Results of diabetes prediction models for eight PMs

This subsection presents the diabetes prediction results of 10 MLTs in terms of eight PMs as the output of Phase 1 (Fig. 2) of the experimental study. The results are listed in Table 3 and presented in a pictorial form in Fig. 3, followed by a detailed discussion.

Table 3 Results of diabetes prediction for the Pima Indian diabetes dataset

Diabetes prediction models	Accuracy	Sensitivity	FPR	F-measure	MCC	Specificity	FNR	ROC-AUC score
Naive Bayes (NB)	0.7630	0.6425	0.1560	0.6713	0.4678	0.8440	0.3881	0.8010
Logistic regression (LR)	0.7721	0.5709	0.1200	0.6362	0.4801	0.8800	0.4291	0.8380
Support vector machine (SVM)	0.7852	0.5410	0.1020	0.6250	0.4800	0.8980	0.4590	0.7200
k-nearest neighbor (KNN)	0.7018	0.5299	0.2060	0.5536	0.3312	0.7940	0.4701	0.6500
AdaBoost (AB)	0.7435	0.5522	0.1540	0.6004	0.4171	0.8460	0.4478	0.7815
Bagging	0.7526	0.5784	0.1540	0.6200	0.4405	0.8460	0.4216	0.8110
Decision table (DT)	0.7240	0.5448	0.1800	0.5794	0.3768	0.8200	0.4552	0.7820
Decision stump (DS)	0.7188	0.5746	0.2040	0.5878	0.3747	0.7960	0.4254	0.6840
Random forest (RF)	0.7435	0.5970	0.1780	0.6190	0.4267	0.8220	0.4030	0.7901
OneR	0.7083	0.4739	0.1660	0.5314	0.3292	0.8340	0.5261	0.6540

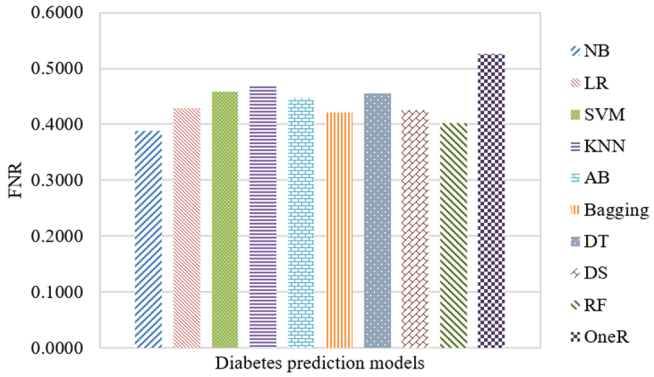


(a) Accuracy (b) Sensitivity

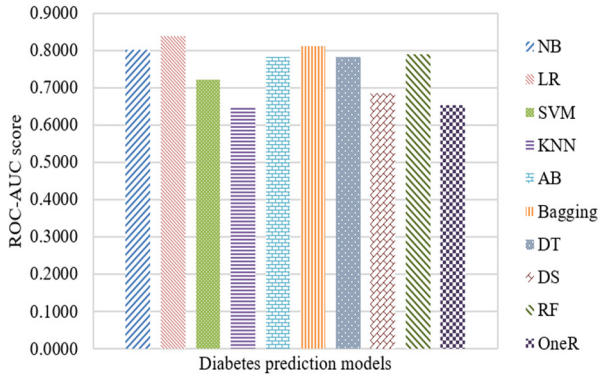
(c) False positive rate (FPR) (d) F-measure

(e) MCC (f) Specificity

Fig. 3 Performance of 10 diabetes prediction models



(g) FNR



(h) ROC-AUC

Fig. 3 Performance of 10 diabetes prediction models (continued)

As presented in Fig. 3(a), 3(c), and 3(f), the diabetes prediction model SVM outperformed other models in terms of accuracy, FPR, and specificity. Furthermore, as shown in Fig. 3(b), 3(d), and 3(g), the MLT NB outperformed all other MLTs in terms of accuracy measures, namely sensitivity, F-measure, and FNR. The LR and SVM exhibited comparable performance (Fig. 3(e)). The LR was identified as the most suitable model for diabetes prediction in terms of the ROC-AUC score (Fig. 3h). The results indicated that no diabetes prediction model demonstrated the best performance for all the eight PMs. This motivates the authors to use the MCDM methods for selecting the most suitable MLT for diabetes prediction modeling by considering eight PMs.

5.2. MCDM ranking

The proposed MCDM-based framework (described in Section 3) was used to identify the most suitable MLT for diabetes prediction. The performance of each diabetes prediction model was optimized in terms of eight PMs as the output of Phase 2 (Fig. 2) of the experimental study. Table 4 lists the ranks of 10 diabetes prediction models obtained after applying the proposed MCDM-based approach. WSM, TOPSIS, and VIKOR were applied to a 10 × 8 decision matrix (diabetes prediction results of 10 MLTs for eight PMs) to calculate WSM, TOPSIS, and VIKOR scores, respectively. Next, the individual rank of each diabetes prediction model was determined for each MCDM method. According to WSM and TOPSIS procedures (explained in Sections 4.4.1 and 4.4.2), a higher score of the alternative diabetes prediction model results in a higher rank.

In the case of VIKOR, a lower score of the alternative diabetes prediction model results in a higher rank. Using the RPM, the individual ranks of diabetes prediction models obtained using the three MCDM methods were combined to obtain the final rank list. The RPM [23] considers the position of each alternative according to each subordinate ranking technique. In this study, alternatives were diabetes prediction models and subordinate ranking techniques were WSM, TOPSIS, and VIKOR. The RPM score can be calculated using Eq. (1). Based on the calculated RPM score, the final ranks were assigned to diabetes prediction models. Consequently, the ranking index for diabetes prediction models was established, favoring those with lower RPM scores.

Based on the application of three MCDM approaches, the following inferences can be drawn (Table 4):

(1) The individual ranks of all diabetes prediction models were determined by applying three MCDM methods, namely WSM, TOPSIS, and VIKOR.

(2) Next, individual rankings were combined using the fusion approach to obtain the final ranks of diabetes prediction models.

(3) LR is recommended as the most suitable diabetes prediction model because it has the highest final rank.

A statistical test was conducted namely the Bayesian sign test to demonstrate that the diabetes prediction model: LR significantly outperformed other MLTs in diabetes prediction modeling. The entire methodology of the statistical test is presented in the subsequent subsection (Section 5.3).

Table 4 Recommendation of diabetes prediction models based on the MCDM-based proposed approach

Diabetes prediction models	Ranking produced by three MCDM methods						RPM score produced by rank fusion method	Final rank
	WSM score	WSM rank	TOPSIS score	TOPSIS rank	VIKOR score	VIKOR rank		
Naive Bayes (NB)	0.1813	2	0.6678	3	0.184	2	0.7500	2
Logistic regression (LR)	0.1869	1	0.8252	1	0	1	0.3333	1
SVM	0.1808	3	0.757	2	0.2908	4	0.9231	3
k-nearest neighbor (KNN)	0.1229	10	0.1665	10	0.9884	9	3.2143	10
AdaBoost (AB)	0.1627	6	0.5547	5	0.3274	5	1.7647	5
Bagging	0.1714	4	0.6217	4	0.2542	3	1.2000	4
Decision table (DT)	0.1468	7	0.3669	7	0.6232	7	2.3333	7
Decision stump (DS)	0.1397	8	0.3115	8	0.8344	8	2.6667	8
Random forest (RF)	0.1658	5	0.5185	6	0.4514	6	1.8750	6
OneR	0.1243	9	0.2513	9	1	10	3.1034	9

5.3. Statistical test

The Bayesian sign test was performed to validate whether LR significantly outperforms other MLTs in diabetes prediction modeling. The Bayesian sign test, proposed by Benavoli et al. [30], involves the calculation of posterior probabilities obtained for the pairwise comparison of two methods. Three regions are defined in the Bayesian test: left, region of practical equivalence (ROPE), and right. The probability of the left region indicates that the left method is better than the right method and vice versa. The probability of the ROPE region indicates that the performance of both methods is equivalent. The results of the Bayesian test for LR against other diabetes prediction models are listed in Table 5. The posterior probability p (ROPE) of the LR model against all other diabetes prediction models was approximately equal to 0 (< 0.05), indicating that the performance of LR against other diabetes prediction models was distinguishable. Furthermore, the posterior probability p (left) of the LR model against all other diabetes prediction models was considerably high (> 95%). Thus, the MLT recommended by the proposed MCDM-based approach significantly outperformed other MLTs for diabetes prediction modeling.

Table 5 Results of the Bayesian test in terms of posterior probabilities

Logistic regression (LR) against	P (left)	P (ROPE)	P (right)
Naive Bayes (NB)	0.9677	0.0323	0
Support vector machine (SVM)	0.9687	0.0313	0
k-nearest neighbor (KNN)	0.9691	0.0309	0
AdaBoost (AB)	0.9694	0.0306	0
Bagging	0.9690	0.0310	0
Decision table (DT)	0.9679	0.0321	0
Decision stump (DS)	0.9694	0.0306	0
Random forest (RF)	0.9693	0.0307	0
OneR	0.9689	0.0311	0

6. Conclusions

This study proposed a novel MCDM-based approach to identify the best diabetes prediction model considering various PMs. The selection problem was modeled as an MCDM issue. Initially, three MCDM methods—WSM, TOPSIS, and VIKOR—were used to rank MLTs based on PMs. Next, a fusion approach (RPM) was employed to establish the final rank of each MLT. The proposed approach was tested using 10 MLTs, 8 PMs, 3 MCDM methods, and a diabetes dataset. The experimental results yielded the following conclusions:

(1) Diabetes prediction performance results (as described in Section 5.1) of the 10 MLTs for the eight PMs indicated that no single MLT exhibited the best performance in terms of the eight PMs. Thus, it is necessary to evaluate diabetes prediction models by optimizing the eight PMs.

(2) Based on the final ranking obtained by the proposed MCDM-based method, LR was recommended as the most suitable diabetes prediction model because of its highest rank.

(3) The proposed method can be extended as a future scope by applying MLTs over a large number of diabetes datasets. Moreover, hybridizing different MCDM methods may be another future research direction.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] V. Chang, M. A. Ganatra, K. Hall, L. Golightly, and Q. A. Xu, "An Assessment of Machine Learning Models and Algorithms for Early Prediction and Diagnosis of Diabetes Using Health Indicators," Healthcare Analytics, vol. 2, article no. 100118, November 2022.

[2] N. Ahmed, R. Ahammed, M. M. Islam, M. A. Uddin, A. Akhter, M. A. Talukder, et al., "Machine Learning Based Diabetes Prediction and Development of Smart Web Application," International Journal of Cognitive Computing in Engineering, vol. 2, pp. 229-241, June 2021.

[3] K. D. Silva, W. K. Lee, A. Forbes, R. T. Demmer, C. Barton, and J. Enticott, "Use and Performance of Machine Learning Models for Type 2 Diabetes Prediction in Community Settings: A Systematic Review and Meta-Analysis," International Journal of Medical Informatics, vol. 143, article no. 104268, November 2020.

[4] V. Jaiswal, A. Negi, and T. Pal, "A Review on Current Advances in Machine Learning Based Diabetes Prediction," Primary Care Diabetes, vol. 15, no. 3, pp. 435-443, June 2021.

[5] A. Hussain, H. Malik, and M. U. Chaudhry, "Supervised Learning Based Classification of Cardiovascular Diseases," Proceedings of Engineering and Technology Innovation, vol. 20, pp. 24-34, January 2022.

[6] D. H. Wolpert and W. G. Macready, "No Free Lunch Theorems for Optimization," IEEE Transactions on Evolutionary Computation, vol. 1, no. 1, pp. 67-82, April 1997.

[7] R. Birjais, A. K. Mourya, R. Chauhan, and H. Kaur, "Prediction and Diagnosis of Future Diabetes Risk: A Machine Learning Approach," SN Applied Sciences, vol. 1, no. 9, article no. 1112, September 2019.

[8] J. J. Khanam and S. Y. Foo, "A Comparison of Machine Learning Algorithms for Diabetes Prediction," ICT Express, vol. 7, no. 4, pp. 432 439, December 2021.

[9] J. Ramesh, R. Aburukba, and A. Sagahyroon, "A Remote Healthcare Monitoring Framework for Diabetes Prediction Using Machine Learning," Healthcare Technology Letters, vol. 8, no. 3, pp. 45-57, June 2021.

[10] H. Gupta, H. Varshney, T. K. Sharma, N. Pachauri, and O. P. Verma, "Comparative Performance Analysis of Quantum Machine Learning with Deep Learning for Diabetes Prediction," Complex & Intelligent Systems, vol. 8, no. 4, pp. 3073-3087, August 2022.

[11] O. Houri, Y. Gil, R. Chen, A. Wiznitzer, A. Hochberg, E. Hadar, et al., "Prediction of Type 2 Diabetes Mellitus According to Glucose Metabolism Patterns in Pregnancy Using a Novel Machine Learning Algorithm," Journal of Medical and Biological Engineering, vol. 42, no. 1, pp. 138-144, February 2022.

[12] S. M. Ganie and M. B. Malik, "Comparative Analysis of Various Supervised Machine Learning Algorithms for the Early Prediction of Type-II Diabetes Mellitus," International Journal of Medical Engineering and Informatics, vol. 14, no. 6, pp. 473-483, September 2022.

[13] M. Panda, D. P. Mishra, S. M. Patro, and S. R. Salkuti, "Prediction of Diabetes Disease Using Machine Learning Algorithms," IAES International Journal of Artificial Intelligence, vol. 11, no. 1, pp. 284-290, March 2022.

[14] C. C. Olisah, L. Smith, and M. Smith, "Diabetes Mellitus Prediction and Diagnosis from a Data Preprocessing and Machine Learning Perspective," Computer Methods and Programs in Biomedicine, vol. 220, article no. 106773, June 2022.

[15] N. A. Azit, S. Sahran, V. M. Leow, M. Subramaniam, S. Mokhtar, and A. M. Nawi, "Prediction of Hepatocellular Carcinoma Risk in Patients with Type-2 Diabetes Using Supervised Machine Learning Classification Model," Heliyon, vol. 8, no. 10, article no. e10772, October 2022.

[16] I. Tasin, T. U. Nabil, S. Islam, and R. Khan, "Diabetes Prediction Using Machine Learning and Explainable AI Techniques," Healthcare Technology Letters, vol. 10, no. 1-2, pp. 1-10, February-April 2023

[17] G. Aguilera-Venegas, A. López-Molina, G. Rojo-Martinez, and J. L. Galán-Garcia, "Comparing and Tuning Machine Learning Algorithms to Predict Type 2 Diabetes Mellitus," Journal of Computational and Applied Mathematics, vol. 427, article no. 115115, August 2023.

[18] N. K. Chowdhury, M. A. Kabir, M. M. Rahman, and S. M. S. Islam, "Machine Learning for Detecting COVID-19 from Cough Sounds: An Ensemble-Based MCDM Method," Computers in Biology and Medicine, vol. 145, article no. 105405, June 2022.

[19] Y. Song and Y. Peng, "A MCDM-Based Evaluation Approach for Imbalanced Classification Methods in Financial Risk Prediction," IEEE Access, vol. 7, pp. 84897-84906, 2019.

[20] A. Kumar and K. Kaur, "SOM-FTS: A Hybrid Model for Software Reliability Prediction and MCDM-Based Evaluation," International Journal of Engineering and Technology Innovation, vol. 12, no. 4, pp. 308-321, October 2022.

[21] R. Ali, S. Lee, and T. C. Chung, "Accurate Multi-Criteria Decision Making Methodology for Recommending Machine Learning Algorithm," Expert Systems with Applications, vol. 71, pp. 257-278, April 2017.

[22] A. Sotoudeh-Anvari, "The Applications of MCDM Methods in COVID-19 Pandemic: A State of the Art Review," Applied Soft Computing, vol. 126, article no. 109238, September 2022.

[23] S. Altuntas, T. Dereli, and M. K. Yilmaz, "Evaluation of Excavator Technologies: Application of Data Fusion Based MULTIMOORA Methods," Journal of Civil Engineering and Management, vol. 21, no. 8, pp. 977-997, 2015.

[24] R. K. Bania and A. Halder, "R-HEFS: Rough Set Based Heterogeneous Ensemble Feature Selection Method for Medical Data Classification," Artificial Intelligence in Medicine, vol. 114, article no. 102049, April 2021.

[25] U. Ahmed, G. F. Issa, M. A. Khan, S. Aftab, M. F. Khan, R. A. T. Said, et al., "Prediction of Diabetes Empowered with Fused Machine Learning," IEEE Access, vol. 10, pp. 8529-8538, 2022.

[26] P. C. Fishburn, "Additive Utilities with Incomplete Product Sets: Application to Priorities and Assignments," Operations Research, vol. 15, no. 3, pp. 537-542, May-June 1967.

[27] C. L. Hwang and K. Yoon, Multiple Attribute Decision Making: Methods and Applications: A State-of-the-Art Survey, Berlin: Springer-Verlag, 1981.

[28] S. Opricovic and G. H. Tzeng, "Compromise Solution by MCDM Methods: A Comparative Analysis of VIKOR and TOPSIS," European Journal of Operational Research, vol. 156, no. 2, pp. 445-455, July 2004.

[29] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," ACM SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 10-18, June 2009.

[30] A. Benavoli, G. Corani, J. Demšar, and M. Zaffalon, "Time for a Change: A Tutorial for Comparing Multiple Classifiers Through Bayesian Analysis," The Journal of Machine Learning Research, vol. 18, no. 1, pp. 2653-2688, January 2017.