RF-PSO: An Optimized Approach for Diabetes Prediction

Houda El Bouhissi[a,b], Amine Ziane[a,b], Lamia Rahmani[a], Meriem Medbal[a], Mariia Kostiuk[c]

[a] *LIMED Laboratory, Faculty of Exact Sciences, University of Bejaia, 06000, Bejaia, Algeria*
[b] *Laboratoire LITAN, Ecole Supérieure en Sciences et Technologies de l'Informatique et du Numérique, RN 75, Amizour, 06300 Bejaia, Algeria*
[c] *Khmelnytskyi National University, Institutska str., 11, Khmelnytskyi, 29016, Ukraine*

Abstract

Diabetes is a chronic disease due to a malfunction of the pancreas, which leads to high concentration of blood sugar in the blood and can affect the functioning of the body system. High blood sugar levels contribute to complications, over time as it can damage the heart, blood vessels, eyes, kidneys, and nerves...etc. Therefore, we need to develop a system capable of effectively diagnosing diabetic patients using medical details. There are various machine-learning techniques for diabetes prediction; this can help persons to prevent this disease or early detection to avoid its complications. This work is a continuation of our previous contribution for gestational diabetes prediction where we conducted a study demonstrating the effectiveness of the Random Forest algorithm in predicting gestational diabetes. In the present study, and to enhance performance, we use a swarm intelligence approach to extract best features for training Random Forest algorithm. The performance of the proposal was compared. The results demonstrate that the combination of the Random Forest algorithm with the Particle Swarm Optimization algorithm provides better accuracy.

Keywords

Prediction, Machine Learning, Diabetes, Feature extraction, Random Forest, PSO, Optimization.

1. Introduction and Motivation

Diabetes is a widely chronic malady that occurs when the pancreas does not produce enough insulin, or when the body cannot effectively use it has produced insulin. Insulin is a hormone that regulates blood sugar levels.

Hyperglycemia, also known as high blood sugar, is a common effect of uncontrolled diabetes and, over time, leads to considerable damage to many of the body's systems, particularly nerves and blood vessels.

There has been drastic increase in rate of people with diabetes since a decade. Current human lifestyle is the main reason behind growth in diabetes, due to unhealthy diet, lifestyle, stress, insufficient physical effort, and sport.

According to the World Health Organization [1] there are 400 million people with diabetes in the world. Each year, 1.5 million people die of diabetes, due to its complications and diabetes-related death rates rose 13% in lower-middle income nations.

In general, diabetics, live in low-and middle-income countries. Both the number of cases and the prevalence of diabetes have been steadily increasing over the past few decades.

ICST-2023: Information Control Systems & Technologies, September 21-23, 2023, Odesa, Ukraine.
EMAIL: houda.elbouhissi@gmail.com (H. El Bouhissi); amine.ziane@univ-bejaia.dz (A. Ziane); lamia.rahmani@univ-bejaia.dz (L. Rahmani); meriem.medbal@univ-bejaia.dz (M. Medbal), maria@khnu.km.ua (M. Kostiuk)
ORCID: 0000-0003-3239-8255 (H. El Bouhissi); 0009-0004-8368-6358 (A. Ziane); 0009-0009-1883-7451 (L. Rahmani); 0009-0002-6497-1503 (M. Medbal); 0000-0002-9559-4109 (M. Kostiuk)

In Algeria, the authorities are aware of the deadly clutch of diabetes and are implementing mechanisms to support diabetic patients through the media, care and follow-up centers and charity associations, however, people are not aware of the scale of this disease which is becoming rampant over time.

We distinguish two types of diabetes [2]: type 1 and type 2.

- Diabetes type 1: also called insulin-dependent, this disease affects 5 to 10% of all diabetics, is brought on by cellular-mediated autoimmune destruction of the pancreatic B-cells.
- Diabetes type 2: also called non-insulin-dependent diabetes, this kind of diabetes affects people who have insulin resistance and typically have a relative (as opposed to an absolute) insulin shortage. It accounts for 90–95% of people with diabetes.
- In addition, due to a variety of clinical, social, and lifestyle reasons, "*gestational diabetes*" is a different type of diabetes and one of the most common diseases in pregnant women.

With its growing prevalence in recent years, diabetes has become a disease of the century. Health professionals recommend that individuals take preventive measures by following certain guidelines to reduce risk factors and thus avoid the onset and complications of this disease.

However, as the number of individuals with this disease rises, so does the complexity of its therapy. To enable early diagnosis of infections and gain a better knowledge of the variables that significantly influence their occurrence, researchers have developed several methods.

Several machine-learning based systems have been implemented to support decreasing the infection risk. Risk prediction is useful in many domains in our daily life [3]. Early diabetes prediction makes to knowledge the most critical factors to control the disease.

The present study is a continuation of our previous contribution [4], which focused on gestational diabetes. In [4], we used three classifiers, and concluded that Random Forest (RF) performs better.

The purpose of this paper is to improve the results achieved in [4] and propose a novel approach for diabetes prediction. The proposal, called RF-PSO, involves four main steps and uses the PSO (Particle Swarm Optimization) algorithm and the RF classifier:

- RF [5] is belongs to the supervised ML techniques which is an ensemble learning method that combines multiple decision trees to make predictions.
- PSO [6] is a metaheuristic optimization algorithm inspired by the social behavior of bird flocking or fish schooling for feature selection and the for the prediction process.

Thus, the contributions of this paper are:

- Automatic prediction of diabetes in healthcare systems to early detection of the disease, therefore, healthcare systems can identify women at risk of diabetes before the disease’s beginning. In this sense, our work can be considered as a making-decision proposal in healthcare.
- Implementation of an effective diabetes tool for prediction to contribute of the population health, especially women in pregnancy.
- Highlighting the benefits of bioinspired algorithms for the optimization of machine leaning algorithms.

The paper is organized as follows. Main related works on diabetes prediction is discussed in Section 2. Section 3 describes the proposed approach based on PSO and RF, while experiments and results are reported in Section 4. Finally, Section 5 concludes the paper and present future works.

2. Related works

Various prediction models have been developed and implemented by researchers using different techniques such as machine-learning algorithms. The analysis of related work gives an insight of healthcare datasets used, where analysis and predictions have been realized with several methods and techniques.

Diabetes prediction is a significant area of research that has captured the attention of many researchers. While numerous systems have been developed, to the best of our knowledge, most of the works use RF algorithm, few works are dedicated to gestational diabetes and the utilization of swarm intelligence remains unexplored.

In this section, we present the main related works according to our research and attempt to explore the key features to improve the proposed methodology.

The aim of the study proposed in [7] is to build a model for early diabetes prediction with multiple machine-learning models. The authors use four supervised machine-learning algorithms: RF, Naïve Bayes (NB), Decision Tree (DT) and Support Vector Machine (SVM) and an unsupervised algorithm, K-means clustering to classify the patients into categories. The experiments were performed on the PIMA Indian Diabetes and the patient dataset provided by the hospital of Frankfurt in Germany. The authors start by implementing the k-means classifier simultaneously on the two databases to clean the columns where there is missing data and replace it with the cluster, and then they split the PIMA dataset into 80% for training and 20% for testing and the German dataset into 70% for training and 30% for testing. They tested each machine-learning algorithm on the datasets, next they calculated the accuracy, the precision, f1-measure and recall of each algorithm to evaluate and compare them. The results of the evaluation process showed that RF performed the best on the Frankfurt dataset with 97.60% of accuracy and that SVM was the best model on the PIMA dataset with an accuracy of 83.10%, the initialization of cluster centers and the number of clusters influenced the increase and decrease of the performance of algorithms.

The authors in [8] realized a system of early diabetes mellitus prediction based on artificial neural network and machine-learning algorithms. The test of the model has been done in the National Institute of Diabetes, Digestive, and kidney diseases. The authors proposed three models: artificial neural network (ANN), RF and k-means clustering. First, they implemented the ANN, the RF algorithm and finally the k-means clustering. ANN outperformed the two other models with an accuracy of 75.70%.

Another approach proposed in [9], which combines four algorithms: RF, J48, K-Nearest Neighbor (KNN) and NB, with WEKA hybrid system. The authors have built a hybrid model based on different machine-learning algorithms to increase the accuracy of the diabetes mellitus prediction. The approach uses the PIMA Indian diabetes database and involves different steps. The first step concerns the data processing, and then they split the dataset into two sets: 90% for training set and 10% for testing. The evaluation of the model was based on the accuracy, F-measure, Recall and Precision, the authors tested the four algorithms individually and the hybrid model. The authors claim that the combination of the four algorithms provided a better accuracy than each algorithm individually.

The authors in [10] use a variety of machine learning algorithms. The authors indicated that approaches based on machine learning provide better prediction outcomes than achieved by building models from patient-collected datasets. The main purpose of the proposal is to create a system, which, by combining several machine-learning approaches, can accurately conduct early diabetes prediction for a patient. The used algorithms are Gradient Boosting (GB), RF, DT, SVM, KNN and logistic Regression (LR). The emergence of different symptoms has been used to detect the presence of disease. The methodologies, metrics, and features that were employed affect the outcome of the prediction. A Disease Influence Measure (DIM) based diabetic prediction has been provided as a step toward diabetes prediction. The approach performs a preliminary processing on the input data set, removing the noisy records. The method calculates disease influence measure (DIM) in the second step using the characteristics of the input data point. The technique performs diabetic diagnosis depending on the DIM value. Various disease prediction methods have been taken into consideration, and their effectiveness has been compared. The analysis's results have been provided in-depth for development. The project work reveals that the model is capable of accurately predicting diabetes with an accuracy that exceeds 95%.

The authors in [11] built a model for diabetes prediction based on various machine-learning algorithms. The study was performed on the PIMA dataset. The authors implemented k-means clustering algorithm on two attributes of the dataset that are "age" and "Glucose", to classify each patient into a diabetic or non-diabetic class. The machine-learning model was built with SVM, RF, DT, Extra Tree Classifier, Ada Boost algorithm, Perceptron, Linear Discriminant Analysis algorithm, LR, KNN, Gaussian NB, Bagging algorithm and Gradient Boost Classifier, next created a pipeline for the algorithms that gave the highest accuracy and then evaluated the pipelines. The proposal evaluation showed that LR provide the highest accuracy of 96%, but using the pipeline Ada Boost algorithm was the best model with an accuracy of 98.80%.

The authors in [12] used various machine-learning techniques to predict type-2 diabetic mellitus disease. The authors' purpose was to create a predictive model that can accurately predict whether a person has diabetes or not. The chosen classifiers are logistic DT, ExtraTrees, RF, regression, XGBoost, gradient boosting, and the light gradient boosting machine (LGBM). Initially gathered and kept in the database are the data required for the investigation. The authors performed their tasks using the PIMA dataset. After that, the dataset is pre-processed using various strategies for exploratory data analysis. The dataset is splited into "training data" and "testing data." The best algorithm that works and has the highest accuracy is selected as the best model for predicting the disease after various experimentations and comparison of the obtained results. After evaluating each algorithm, the LGBM algorithm showed the best performance with an accuracy of 95.20 %.

Analysis of related work yields results on various healthcare datasets, where analysis and predictions were performed using a variety of methods and techniques. Some of these research utilized use of unique datasets or a mixture of other databases.

Various researchers have proposed different prediction models. These models use machine-learning, deep learning algorithms, data mining techniques or a combination of these techniques. All these works provide a promising outcome and differ in precision and accuracy according to used datasets. A review of the relevant papers leads to the conclusion that researchers have successfully combined several machine learning algorithms with various data preprocessing approaches for the automatic detection of diabetes.

Our work is inspired by related works and building on our initial contribution where we combined machine-learning algorithms with swarm intelligence optimization to achieve better accuracy.

3. Proposed Methodology

Over the past decade, the number of diabetic people has grown dramatically. The current human lifestyle is the main reason for the increase in diabetes, as are unhealthy eating habits, anxiety, use of tools that reduce physical effort [13], and many other reasons.

Our study focuses on gestational diabetes, which is common among women during pregnancy who do not already have diabetes and can lead to various complications.

Every year, gestational diabetes affects 2% to 10% of pregnancies in the United States [14]. Managing gestational diabetes is crucial to ensuring a healthy pregnancy and the well-being of your baby. In Algeria, the prevalence of gestational diabetes among pregnant women varies between 2% and 5%.

In our earlier research [4], we sought to predict early gestational diabetes more accurately. We chose the most significant machine learning (ML) algorithms that produce the greatest results in the literature, namely Deep Neural Network (DNN) [15], SVM, and RF, and we found that RF outperformed the other two algorithms and produced the best results.

The present study is a continuation of our previous contribution [4] and attempts to improve the RF algorithm by selecting the optimal features using the PSO algorithm for gestational diabetes prediction. In this sense, the present paper makes use of PSO and RF classifier for diabetes prediction with an improved accuracy.

The aim of our research is to predict in the future whether a healthy woman will become diabetic or not, i.e., affected by gestational diabetes, regarding woman data such as age, weight, personal information, and her history.

The Figure 1 depicted the overall architecture of the RF-PSO proposal, which involves four main steps:

- The first step concerns the data collection.
- The second step concerns data processing. This step involves many phases for treating and managing data.
- The third step is the main step which includes the selection of the best features for the RF algorithm; this step is performed using the PSO algorithm.
- Finally, the last step concerns the prediction process using the RF algorithm.

The proposal is supported by a software tool deployed on a smartphone.

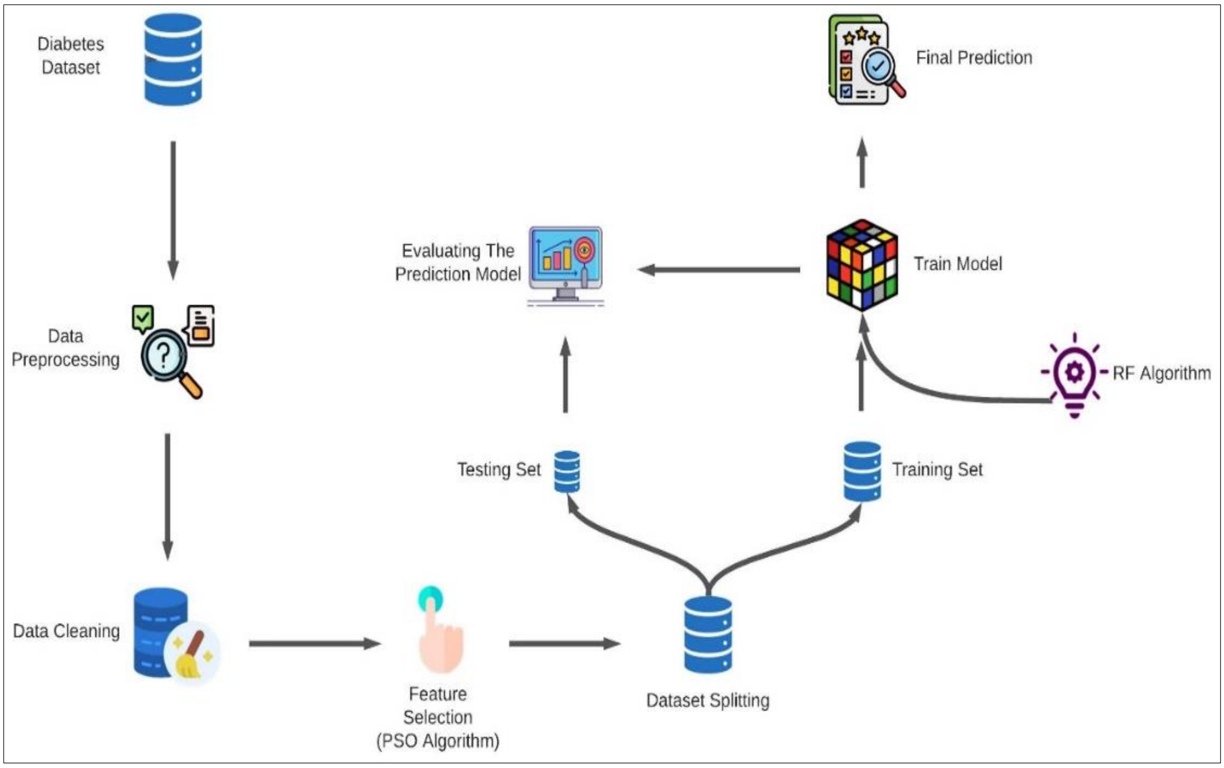

Figure 1: System architecture

These steps will be examined in detail in the following sections.

3.1. Data collection

Data collection is an essential initial step in the diabetes prediction. It enables proper treatment and accurate assessment of the study population. The purpose of this step is to ensure that the information gathered is complete and reliable, enabling better evaluation of results and more accurate anticipation of future probabilities and trends.

This step includes data gathering and understanding to study the patterns and trends, which helps in prediction, and evaluating the results.

The dataset, whose description is provided in table 1, is public and originated from Kaggle and is made up of several prognostic medical variables obtained from the hospital in Frankfurt, Germany [16]. The dataset includes 2000 person between diabetic and no diabetic and 9 attributes with a size 12.0 kB and two diabetes classes:

- Class 0 means healthy person.
- Class 1 means diabetic person.

It is important to note that this dataset was expanded by 200 records (patient cases), the persons' data were collected according to the same attributes. These additional records were gathered, with confidentiality, from the Internal Medicine Department of Khellil Amrane Hospital in Bejaia City, Algeria [17] during student internships at the hospital.

Table 1
Dataset information

N°	Attributes	Type
1	Number of pregnancies	Numerical
2	Glucose level	Numerical
3	Blood Pressure	Numerical
4	Skin Thickness	Numerical
5	Age	Numerical
6	Insulin	Numerical
7	Body Mass Index (BMI)	Numerical
8	DiabetesPedigreeFunction (family history)	Numerical
9	Outcome	Numerical

3.2. Data processing

This model phase manages inconsistent data to get more accurate and precise results, which is essential for our prediction process. The dataset contains missing values. Therefore, we delete missing values for few selected attributes like Glucose level, Blood Pressure, Skin Thickness, BMI, and Age because these attributes cannot be null. Then we scale the dataset to normalize all values.

The processing at all goes through different phases; first, the data was refined to eliminate errors. Second, a data cleaning and filtering process was conducted to avoid the formation of inappropriate rules and patterns, which included the removal of noisy data and missing values, as well as the removal of duplicates and irrelevant data. To resolve discrepancies, anomalies in noisy data were removed.

For example, if an attribute such as blood glucose had zero values (this is not possible in daily life), all such values were substituted by the median value of that attribute.

3.3. Feature selection

The purpose of this step is to achieve the first objective, which is to increase prediction accuracy by minimizing the feature selection subsets and select the best features from the disease dataset.

In this phase, feature selection was carried out using the PSO [18] algorithm to select the best features and reduce space calculation. The reduced subset includes significant features related to the dataset. Next, the RF algorithm was used for classification for a better prediction accuracy.

The fundamental concept behind the fusion of RF with PSO is to leverage PSO for parameter optimization within the RF algorithm. This optimization procedure aims to enhance the performance of the RF model by identifying the optimal parameter subset that minimizes prediction errors and maximizes accuracy.

In this way, the application of swarm intelligence optimization algorithms on machine learning has produced good results in other fields, such as image classification [19].

PSO is a swarm intelligence algorithm, which is advantageous in many aspects. PSO is very popular and usually used to resolve several optimization problems in different domains.

PSO is simple to implement and uses only few parameters. In addition, it is fast with simple complexity and includes an efficient global search process.

As depicted in figure 2, the PSO technique goes through various steps, starting from the initialization of all parameters, constants, values for position and velocity, then evaluate the fitness of each particle to update position and velocity of each particle and finally determine the best position (in our case, best features).

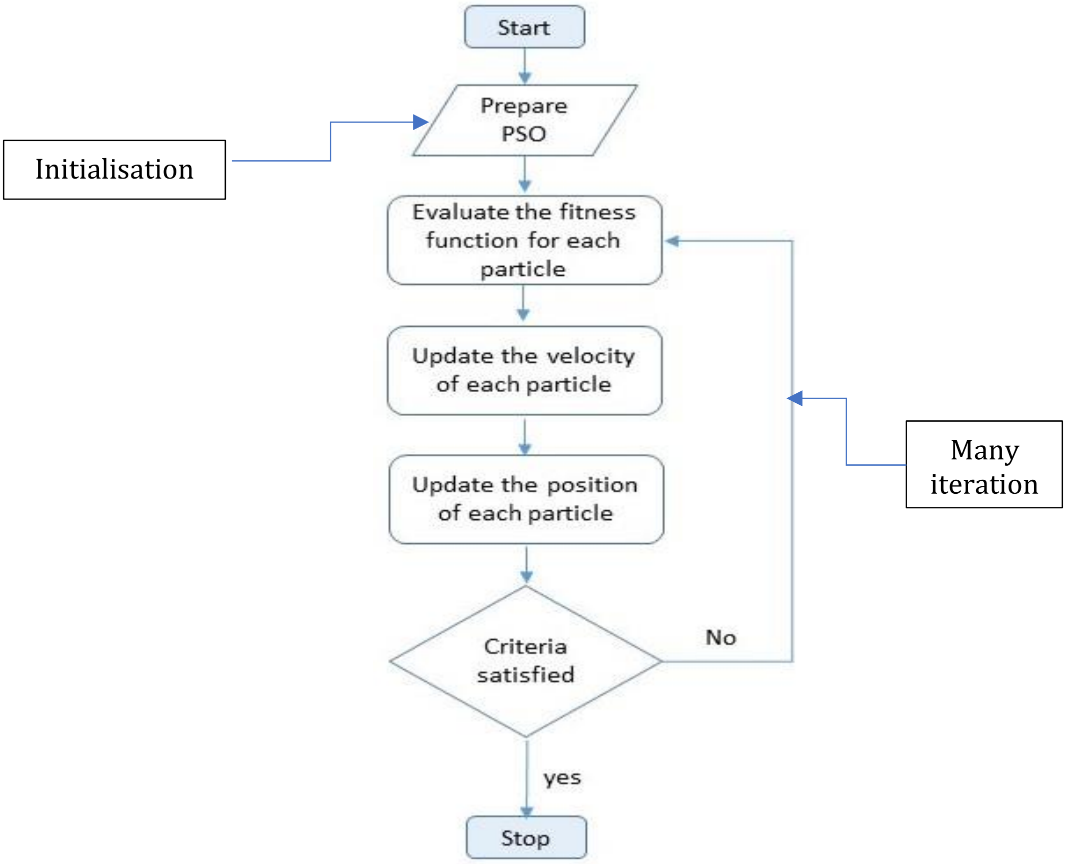

Figure 2: PSO flowchart

We note that choosing the best features improve substantially the model running time. For our approach, we use just a basic version of PSO.

The idea is first to see the impact of applying swarm intelligence algorithms on machine learning algorithms, considering the results obtained from our previous publication.

3.4. Prediction process

The aim of this step is to achieve the main objective, which is to predict whether a patient is diabetic or not, using the RF classifier. Previous research [4] has convincingly confirmed that RF has a better predictive accuracy. However, the evaluation of data collected from patients and the decisions of experts is an essential factor in the prediction process at all.

RF algorithm is a strong supervised machine-learning algorithm that is efficient of performing both Regression and Classification tasks. RF is very popular in medicine domain, through the application of RF in various healthcare applications, disease trends and risks of the disease can be identified.

In simple terms, RF works like this: we develop first different trees to classify a new object, next, based on the produced attributes, each tree gives a different classification and the tree votes for that class are saved. As a result, the forest chooses the classification with the higher number of votes for all trees.

During this step, we apply the RF classifier to the dataset, incorporating feature selection using PSO. This process aims to classify each patient as either diabetic or non-diabetic.

RF proceeds in two phases (see algorithm 1):

- The first phase involves creating the RF by combining N DT.
- The second makes predictions for each created.

At the end, we choose the class with the majority voting. The process at all is presented in the following algorithm:

Algorithm 1: Prediction process based on RF.
Input: Dataset, K
Output: predicted class

Begin
1: Select random K data points from the dataset.
2: Build the decision trees associated with the selected data points
3: Choose the number N for DT to build
4: Go to 1 then 2
5: For new data points, find the predictions of each decision tree, and assign the new data points to the class with the majority votes.
End

After implementation of this classification, we obtain class labels (0 or 1) for each of row of our dataset. Therefore, the proposed system will deal with extra-values according to the dataset format.

4. Experiment and evaluation

To evaluate and validate the proposed approach, we implemented a software tool in python with the environment JupyterLab. JupyterLab [20] is an interactive development environment for notebooks; it is very popular for data science, computing and artificial intelligence.

In this study, the main objective was to improve classification performance and increase diagnostic accuracy by reducing feature dimension using PSO. This section discusses the application of the RF-PSO Model on the dataset of the Frankfurt Hospital.

For this purpose, various scenarios (table 2) were considered in order to evaluate the proposed method. The distribution selection is arbitrary, and the idea is to see how the system reacts with several distributions and avoid overfitting and underfitting in machine learning.

The distribution of data gives a better insight into the behavior of the built module. The scenarios are presented as follows:

Table 2
Evaluation scenarios

Scenario	Data training (%)	Data testing (%)
1	80	20
2	70	30
3	60	40

Machine learning metrics are used to evaluate the performance and effectiveness of RF-PSO model [20]. These metrics provide insights into how well the models are performing and help in comparing different models or tuning the model's parameters.

The performance of the proposed method was evaluated using measures of accuracy (formula 1), and precision (formula 2), based on the following terms: true positive (TP), true negative (TN), false negative (FN) and false positive (FP). Precision is a well-known metric in the prediction field, measuring the accuracy with which a model can predict class membership. It therefore measures the quality of classification results. Accuracy is a metric for evaluating the performance of classification models with 2 or more classes.

These measures are calculated as follows:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ made\ predictions} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Every time, the techniques were executed, the scenarios proposed were selected for each run, thereby generating different results. To account for this variability, each proposal was executed 2 times for every scenario, once for RF and once for RF-PSO.

The table below shows the accuracy of the different experiments based on RF classifier and PSO according to the three scenarios presented before.

Table 3

Accuracy results of RF-PSO for the three scenarios

Scenario	Prediction result	Accuracy (%)
1	Positive	76.00
	Negative	87.00
2	Positive	95.00
	Negative	92.00
3	Positive	91.00
	Negative	88.00

The prediction column is the main column of interest and returns the result of the prediction: positive indicates that the person is likely to be diabetic and negative means that they are not. The accuracy column provides accuracy rate for each scenario. While the first column presents the different scenario of splitting dataset.

For a better analysis, we define a minimum accuracy value to achieve, which presents a tolerance threshold 75% for results comparison.

As Table 3 shows, all the accuracy results are higher than the threshold value, the combination of RF and PSO delivers encouraging results in terms of accuracy according to the three scenarios.

After applying RF and PSO on dataset, we obtain precision and accuracies as mentioned in table 4. The combination of RF and PSO gives highest accuracy of 90% with a 10% increase regarding RF model. introducing the PSO has contributed to the performance of the machine learning algorithm, as only attributes of interest to the machine learning algorithm are selected.

The We conclude that by using PSO for feature selection, we achieve significantly improved accuracy compared to our previous approach and the proposed state-of-the-art systems. However, data quality has a significant influence on the obtained results, especially with large datasets.

Table 4
Results of accuracy

Methodology	Accuracy (%)	Precision (%)
Random Forest	80.00	65.00
Random Forest + PSO	90.00	68.00

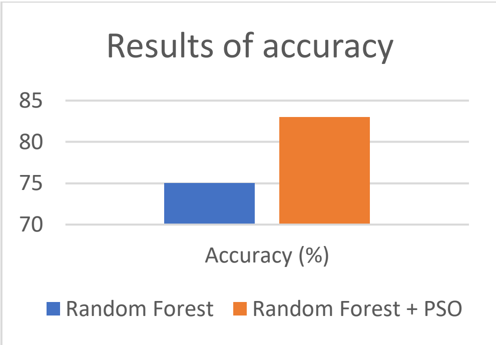

Figure 3: Comparison of the classifier

The performance comparison of the curve-based classification algorithms is shown in Figure 3. The area under a curve grows as a curve's values increase, and the classifier makes fewer errors as a result. The RF-PSO classifier performs 90% more accurately on average than RF.

However, there is a minor increase in precision, and this is due to the size of the data, so with a larger dataset, the results will be better.

Finally, we conclude from this study that the combination of the RF classification model for diabetes prediction with the PSO algorithm performs better for the dataset than the RF classifier. It is important to consider multiple datasets to have better accuracy for the diabetes prediction.

5. Conclusion

In this paper, we have explored and analyzed the main related works for diabetes diagnosis. In addition, we developed a hybrid model based on a combination of a swarm intelligence technique and a machine-learning algorithm. This model revealed a highest accuracy.

In this study, we used PSO and RF for feature selection and data classification, respectively. The aim was to achieve the highest accuracy for the early prediction of diabetes. Combining RF with PSO can be a powerful approach for solving various machine-learning problems.

To evaluate the model's performance, we used a dataset of 2000 patients gathered from the hospital of Frankfurt in Germany and improved by 200 patients gathered from the khelil Amrane hospital of Bejaia, Algeria.

We developed a software tool in Python. The experiments were performed in two distinct scenarios, The aim is to assess the contribution of the EHO algorithm to prediction in terms of speed, quality, and accuracy.

In the first scenario, we use only the RF classifier for diabetes prediction and in the second scenario we use the RF-PSO model.

The best results were obtained using RF-PSO model, which boosted the classification accuracy. The proposed approach increases accuracy with 10.00%, compared with other algorithms.

Regarding the promising results, we claim that this approach could be applied to the diagnosis of other diseases in different domains.

The use of machine learning for medical is increasingly needed in the future. Especially for the disease prediction, machine learning can speed up the process of diagnosing and triage patients.

In our future work, we intend to explore other swarm intelligence optimization algorithms such as Elephant Herding Optimization (EHO) and Grey Wolf Optimization (GWO) [22] with different dataset to assess the potential impact of the proposed approach.

6. Acknowledgements

The authors would like to express their gratitude to Dr. Djamel Eddine OUAIL, from the Internal Medicine Service at Khelil Amrane Hospital in Bejaia city, Algeria, for his unwavering support and encouragement since the beginning of this work.

7. References

[1] Diabetes. URL: https://www.who.int/health-topics/diabetes (Last accessed: June 2023).

[2] S. Roshi, S. Sanjay Kumar, G. Manali, A Comprehensive review of various diabetic prediction models: a literature survey. Journal of Healthcare Engineering, 8100697 (2022). DOI:10.1155/2022/8100697

[3] R. Bekka, S. Kherbouche, H. El Bouhissi, Distraction detection to predict vehicle crashes: a deep learning approach, Computación y Sistemas, 26 1 (2022) 373-387.

[4] H. El Bouhissi, R. E. Al-Qutaish, A. Ziane, K. Amroun, N. Yaya, M. Lachi, Towards Diabetes Mellitus Prediction Based on Machine-Learning, in: 2023 International Conference on Smart Computing and Application (ICSCA), 2023, pp. 1-6.

[5] E. Goel, Er. Abhilasha, Random Forest: A Review, International Journal of Advanced Research in Computer Science and Software Engineering, 7 1 (2017) 251-257.

[6] E. Ali, R. El-Sehiemy, A. El-Ela, S. Kamel, B. Baseem, Optimal planning of uncertain renewable energy sources in unbalanced distribution systems by a multi-objective hybrid PSO–SCO algorithm, IET Renewable Power Generation, 16 10 (2022) 2111-2124.

[7] M.O. Edeh, O.I. Khalaf, CA. Tavera, S. Tayeb, S. Ghouali, G.M. Abdulsahib, N.E. Richard-Nnabu NE, A. Louni, A Classification Algorithm-Based Hybrid Diabetes Prediction Model, Front. Public Health, (2022). DOI: 10.3389/fpubh.2022.82951.

[8] T. Alam, M. Atif Iqbal, Y. Ali, A. Wahab, S. Ijaz, T. Imtiaz Baig, A. Hussain, M. Awais Malik, M. Mehdi Raza, S. Ibrar, Zunish Abbas, A model for early prediction of diabetes, Informatics in Medicine Unlocked, 16 (2019). DOI:10.1016/j.imu.2019.100204.

[9] A. Minyechil, R. Joshi, M. Alehegn, Analysis and prediction of diabetes diseases using machine learning algorithm: Ensemble approach, International Research Journal of Engineering and Technology, 4 10 (2017) 426-436.

[10] P. Sonar, K. JayaMalini, Diabetes prediction using different machine learning approaches, in: 3rd International Conference on Computing Methodologies and Communication (ICCMC), 2019, pp. 367-371.

[11] M. Aishwarya, V. Vaidehi, Diabetes prediction using machine learning algorithms, Procedia Computer Science 165 (2019) 292-299.

[12] B. S. Ahamed, A.S. Meenakshi Sumeet, O. Auxilia, V. Nancy, Prediction of type-2 diabetes mellitus disease using machine learning classifiers and techniques, Frontiers in Computer Science, 4 (2022).

[13] O. Pavlova, O. Soltyk, V. Shvaiko, J. Ilchyshyna, H. El Bouhissi, Human Morphofunctional Indicators based Decision Support System for Choosing Kind of Sport, MoMLeT+DS, (2023) 322-333.

[14] American Diabetes Association. URL:https://diabetes.org/diabetes/gestational-diabetes (Last accessed: June 2023).
[15] S. Vivienne, C. Yu-Hsin, Y. Tien-Ju, J.S. Emer, Efficient processing of deep neural networks: A tutorial and survey, Proceedings of the IEEE, 105 12 (2017) 2295-2329.
[16] Frankfurt Hospital. URL:https://www.kaggle.com/datasets/johndasilva/diabetes.
[17] Diabetics Data, Khelil Amran Hospital, Bejaia, Algeria. Collected July 2023.
[18] M. Jain, V. Saihjpal, N. Singh, S.B. Singh, S. B, An overview of variants and advancements of PSO algorithm. Applied Sciences, 12 17 (2022) 83-92.
[19] M. Xu, C. L, D. Lu, Z. Hu, Y. Yue, Application of Swarm Intelligence Optimization Algorithms in Image Processing: A Comprehensive Review of Analysis, Synthesis, and Optimization, Biomimetics (Basel), 8 2 (2023) 235. DOI:10.3390/biomimetics8020235.
[20] JupyterLab. URL: https://jupyter.org/.
[21] A. Mahyar, A. Rahmani, Machine learning process evaluating damage classification of composites, International Journal of Science and Advanced Technology, 9 (2023) 240-250.
[22] A. Chakraborty, A.K. Kar, Swarm Intelligence: A Review of Algorithms, in: Patnaik, S., Yang, XS., Nakamatsu, K. (eds) Nature-Inspired Computing and Optimization. Modeling and Optimization in Science and Technologies, Springer, Cham, vol 10, 2017. DOI:org/10.1007/978-3-319-50920-4_19.