Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/index
Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
e-ISSN: 2685-0893

Design of Diabetes Prediction Interface Using E-ss and Classification Tree Algorithm

DOI: http://dx.doi.org/10.35889/jutisi.v14i3.3370

Venecia[1*], Genrawan Hoendarto[2], Tony Darmanto[3]
[1,3]Information System, Universitas Widya Dharma, Pontianak, Indonesia
[2]Informatics, Universitas Widya Dharma, Pontianak, Indonesia
*e-mail *Corresponding Author:* venecia0302@gmail.com

Abstract

Diabetes was a chronic disease that continued to increase globally, making early detection essential to reduce long-term complications. This study aimed to develop a desktop-based diabetes prediction system that provided fast and simple classification results for medical personnel and individual users. The system used the entropy-based subset selection (E-ss) method to choose the most relevant attributes and a classification tree to classify the risk. The dataset from the National Institute of Diabetes and Digestive and Kidney Diseases, contained 768 patient records with attributes such as number of pregnancies, glucose level, blood pressure, and other risk factors. The E-ss process produced three attributes with the highest information scores, namely body mass index (BMI), blood pressure, and triceps skinfold thickness. These three attributes were then used as input to the classification tree model to generate diabetes risk predictions. Cross-validation testing showed an accuracy of up to 78.95%. These findings indicated that E-ss feature reduction helped maintain prediction performance while improving computational efficiency. This system was expected to serve as a practical and reliable diagnostic tool.

Keywords: *Diabetes; Risk Prediction; Entropy-based subset selection; Classification tree; Information system*

1. Introduction

Diabetes is a chronic, metabolic disease characterized by elevated levels of blood glucose (or blood sugar), which leads over time to serious damage to the heart, blood vessels, eyes, kidneys and nerves. About 830 million people worldwide have diabetes, the majority of the population comes from low-and middle-income countries [1], [2]. Early detection to decrease the risk of death from diabetes is very crucial to prevent long-term complications. Without proper early detection, patients have the potential to experience a decreased quality of life and increased medical costs. Therefore, an approach is needed that can support the early detection process more effectively, quickly, and be easily accessed.

The process of examining and diagnosing diabetes generally still relies on conventional medical methods such as fasting plasma glucose (FPG), the oral glucose tolerance test (OGTT), and hemoglobin A1c (HbA1c). Although these methods are widely used, they are often considered complicated, time-consuming, and not patient-friendly [3]. Some studies show that these approaches are still not effective for early detection of the risk [4]. This is because most of these methods only show significant results after the metabolic disorder has progressed quite far. This condition has the potential to delay treatment and increase the risk of complications. Therefore, an alternative, technology-based approach is needed that can predict diabetes risk more quickly and practically at an early stage.

As a solution to this problem, a desktop-based diabetes risk prediction system is considered capable of providing a more adaptive alternative for the early detection of diabetes risk and is designed to be used by both medical personnel and the general public as a tool for early detection. One of the well-known methods for its interpretability is the classification tree, which can provide visual decision support through a logical and easy-to-understand tree structure. The main advantage of this method is its ability to generate predictive models that can be directly explained to non-technical users, such as medical professionals and patients [5], [6], [7]. However, the complexity and redundancy of data in medical variables often become obstacles to building and efficient model. To overcome this problem, the E-ss approach is used to filter the most relevant features. This technique evaluates the level of irregularity (entropy) of each attribute to select a subset of features with the highest information, thereby reducing noise and improving the performance of the classification model [8], [9], [10]. The combination of E-ss and the classification tree is believed to be capable of producing a decision support system that is transparent and reliable in supporting the early detection of diabetes.

Various studies have examined diabetes prediction using machine learning techniques. Previous research by [8] developed a diabetes prediction method using E-ss mixed with principal component analysis (PCA) to filter and reduce input variables. This approach has been proven to improve the accuracy of classification models by removing less relevant features and reducing the dimensionality of the data. The authors in [11] used an entropy analysis approach on continuous glucose monitoring (CGM) data to understand glucose dynamics in patients with type 2 diabetes. The results of the study show that entropy analysis can provide deeper insights into daily glucose variability and can be used as a biomarker for the risk of diabetes complications. In [12], they used the decision tree C4.5 to classify diabetes disease using data from patients at a hospital in Bangladesh. This study shows that the symptom polydipsia was a dominant factor in determining the possibility of diabetes, with an accuracy of up to 90.38%. The study emphasizes the interpretability of decision trees in analyzing medical data. The authors in [13] applied the classification and regression tree (CART) approach to group children with type 1 diabetes. This study shows that decision trees have high potential in understanding clinical patient heterogeneity. However, there is no implementation of a user-based system or consideration of the user experience (UX) aspect in any of these studies. And finally, the authors in [14] proposed a web-based prediction system with the integration of several learning algorithms. This application enables users to input their health data to receive an instant prediction. However, this study has not combined explicit feature selection techniques such as E-ss. Based on this study, it can be concluded that most research still focuses on developing predictive models, while implementation in user-friendly information systems with optimal feature selection support remains limited. Therefore, the state of the art in this study lies in the combination of the E-ss method and classification trees implemented directly in the form of a desktop-based information system. Thus emphasizing not only model accuracy but also aspects of actual use by end users. This represents the novelty of this research.

This paper provides two main contributions to the health information system, especially in predicting the risk of diabetes based on technology:

1. This study aims to develop a desktop-based information system for predicting the potential risk of diabetes based on user input of symptoms and risk factors. This system includes an interface designed to be easy to navigate, allowing both medical professionals and individuals without a background in medicine or technology to operate it effectively.
2. This study further integrates an E-ss and classification tree model to validate the effectiveness of the methods used in previous research, applying them in a more realistic end-user context. Here, machine learning serves as a supportive component of the system rather than the primary contribution.

This paper begins with an introduction that explains the background and context of the research. The next section describes the research flow, including the approach and main steps. After that, we present the system design, focusing on the interface planning, button functions, and user interactions with the prototype. The final user interface and system accuracy testing are presented in the following section to ensure that the prediction model integrated into the system can produce classifications that align with the reference data. Finally, in the last section, we conclude the paper with our findings, limitations, and directions for future improvements.

2. Research Flow Method

This research was carried out through several systematic stages to ensure that the process runs in a focused and efficient manner, in line with the objectives that have been set. The stages of the research flow are designed as shown in Figure 1. With this flow, the diabetes prediction system is expected not only to provide accurate predictions but also to be easy to use by real target users.

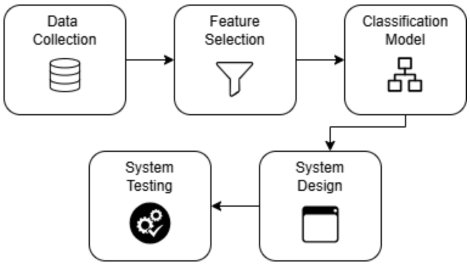

Figure 1. Research Flow

2.1 Data Collection

We collected data from [15], which mentions important parameters for diagnosing diabetes. These parameters consist of 768 patient records with attributes including the number of pregnancies, glucose level, blood pressure, skin thickness, insulin level, body mass index (BMI), diabetes pedigree function, and age. This data will then be used as the basis for the feature selection process.

2.2 Feature Selection (Entropy-Based Subset Selection)

The E-ss method was used to filter the parameters that have the most significant contribution to diabetes risk classification. Even though this process is not explicitly displayed to the end user, it is important for reducing data noise, simplifying the model, and improving the efficiency and accuracy of the calculation results. Regarding the E-ss method, please refer to [8].

2.3 Model Development (Classification Tree)

The classification tree was chosen because of its ability to build a structured and interpretable prediction model. Although the visual representation of the decision tree is not directly presented in the application interface, the use of the classification tree still plays an important role in producing a prediction system that is reliable and technically traceable. Regarsing the Classification Tree method, please refer to [12].

2.4 System Design

In this stage, we planned the interface components such as edit fields, dropdowns, spinners, and buttons – all of which were structured according to the principles of user-centered design (UCD). The application of UCD focuses not only on aesthetics, but also on the context of use and the needs of the end user. With this approach, the user is not only the target but also the main actor that affects the form and behavior of the system [16]. The main focus of the design is to ensure that the system operates according to its objectives with an efficient and easy-to-use workflow.

2.4.1 Overview of the System

This desktop application is designed to predict the risk of diabetes by providing results that are fast and easy for users to understand. Users only need to input the required data, and the system will automatically process the information and display the classification result. The calculation and prediction modelling processes are performed transparently in the background, so user interaction remains focused on data input and result interpretation. With this flow, the system can be used both by medical personnel as an initial diagnostic tool and by individual users to monitor personal risk based on known health parameters.

To make the interaction flow between users and the system more understandable, we designed a use case diagram in Figure 2 that describes the relationship between the actor and the main system functions.

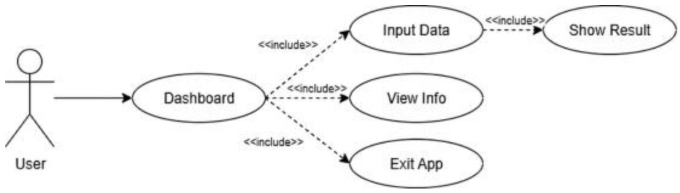

Figure 2. Use Case Diagram

In this use case, there is only one actor –the user– who can be either medical personnel or a general individual. With this use case, all the functions in the system have a clear and focused interaction flow, ensuring that the entire process runs according to the design objectives.

Users access the dashboard, which serves as the main page for navigating to other features. From the dashboard, users can select prediction form, which involves entering medical parameters. Once the data is complete, the system automatically processes the information and displays the diabetes risk prediction results. Users can also access information about the parameters and the diabetes disease. Finally, users have the option to Exit App to safely close the program.

2.4.2 User Interaction Flow

Based on the previously defined use case, the interface is separated into several functional forms, namely:

1) Dashboard, which is the home page and the user's entry point to access all features. When opening the application, the user will first see the dashboard form. From the dashboard, users can choose a menu to start a prediction, view parameter guide, or exit the application.
2) Prediction Form. When users click the prediction form, they will be asked to enter health parameters to obtain the prediction result. In the first line, the user can click the gender dropdown, and when choosing "female", the number of pregnancies spinner will be enabled so the user can input it, while choosing "male" will disable the spinner. In addition, in the BMI fields, the user only needs to input the weight and height fields, and the backend will process them so the BMI field will be automatically filled.
3) Once the data is completed, the user clicks the "Predict Now" button, and the system will perform attribute selection and classification steps in the background. The classification results are displayed immediately below the "Predict Now" button in a clear text format.
4) View the information about parameters and diabetes disease. This feature displays explanations of the medical parameters used in the prediction process and general information about diabetes. The goal is to educate users so that they not only understand the meaning of each parameter entered but also gain a deeper understanding of diabetes and the importance of maintaining good health.
5) Exit application. Users can close the application to end all running processes via the "Exit App" button.

2.5 System Testing

At this stage, system testing focused on evaluating the performance and accuracy of the prediction algorithm used in the system. The evaluation process was carried out by comparing the system's prediction results with actual data to obtain the prediction accuracy value. The results of this test were used to assess the extent to which the developed model was able to provide reliable predictions and was suitable for use as an early diabetes detection tool.

To clarify how the performance metrics are obtained, the following standard evaluation formulas are used:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \qquad \text{...........(1)}$$

The evaluation of model performance uses a confusion matrix consisting of four components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), which represent correctly and incorrectly predicted positive and negative data points. The detailed calculation can be found in Chapter 3.

3. System Implementation and Testing

This chapter presents the results of the diabetes prediction system's design, including the realized user interface display and system testing results. This is intended to demonstrate how the design outlined in Chapter II has been implemented into a ready-to-use application. This chapter also discusses the results of tests conducted to verify system performance, both in terms of functionality and prediction accuracy. Functional testing focuses on checking the conformity of each feature with the specified requirements, while accuracy testing is used to assess the model's ability to accurately classify diabetes risk based on test data.

3.1 Data Collection

The dataset consists of 768 patient records with 8 input variables and 1 output variable. The input variables represent clinical factors related to diabetes risk, while the output variable indicates whether the patient has diabetes or not. All data samples are fully labeled and formatted in numerical form, making them suitable for direct processing by the prediction system. The dataset serves as the primary reference for system testing and validation to ensure that the implemented prediction model operates according to its intended design. A small sample of the patient data used in this study is shown in Table 1 as a representation of the entire dataset.

Table 1. Sample Patient Data from the Diabetes Dataset

Pregnancies	Glucose	BloodPressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1

In Table 1, the outcome attribute represents the class label, where a value of 0 indicates a normal patient, while a value of 1 indicates a diabetic patient.

3.2 Feature Selection

The process begins with the E-ss method, which calculates the entropy and information gain values for each attribute, as described in previous studies. Based on the reference study, the three attributes with the highest information values are BMI, blood pressure, and triceps skinfold thickness. The results of the E-ss process show that not all initial attributes have the same contribution in determining diabetes risk. Some attributes have a higher level of relevance to the output class than others. These selected attributes are then used in the classification model development stage. The selected features used in the prediction process are presented in Table 2.

Table 2. Selected Features for Diabetes Prediction

Risk factor	Selection Status
Pregnancy	Not Selected
Glucose	Not Selected
Blood pressure	Selected
Tricep skinfold thickness	Selected
Insulin	Not Selected
Body Mass Index (BMI)	Selected
Diabetic pedigree function	Not Selected
Age	Not Selected

As shown in Table 2, only three attributes were selected for the prediction process based on their high relevance, while the remaining features were excluded to reduce data redundancy and improve system efficiency.

3.3 Classification Model

At this stage, the system builds a decision tree model that sequentially divides the data based on the attribute with the highest information gain at each node [17]. The tree formation process continues until it reaches the leaf node condition, which represents the final category—either normal or diabetes. The extracted rules are presented in Table 3.

Table 3. Rule-Based Classification for Diabetes Risk Prediction

Rule	Decision Rule	Predicted Class
R1	If BMI < 27.85	Normal
R2	If BMI ≥ 27.85 and Blood Pressure < 42	Diabetes
R3	If BMI ≥ 27.85 and Blood Pressure ≥ 42 and BMI < 32.85 and Blood Pressure < 59.3246	Normal
R4	If BMI ≥ 27.85 and Blood Pressure ≥ 42 and BMI < 32.85 and Blood Pressure ≥ 59.3246	Diabetes
R5	If BMI ≥ 27.85 and Blood Pressure ≥ 64.1908 and BMI ≥ 40.85	Diabetes
R6	If BMI ≥ 27.85 and Blood Pressure ≥ 64.1908 and BMI < 38.0782 and Blood Pressure < 67.999	Diabetes
R7	If BMI ≥ 27.85 and Blood Pressure ≥ 64.1908 and BMI < 38.0782 and Blood Pressure ≥ 67.999 and Skin Thickness ≥ 39.5	Normal
R8	If BMI ≥ 27.85 and Blood Pressure ≥ 64.1908 and BMI < 38.0782 and Skin Thickness < 33.0869	Diabetes
R9	If BMI ≥ 27.85 and Blood Pressure ≥ 64.1908 and Skin Thickness ≥ 14.8427 and Blood Pressure ≥ 80.4105	Diabetes
R10	If BMI ≥ 27.85 and Blood Pressure ≥ 64.1908 and Skin Thickness ≥ 14.8427 and BMI < 33.2315	Normal

Table 3 presents the classification process in the proposed system, which is implemented based on a rule-based mechanism derived from a previously published decision tree model.

3.4 System Implementation

In the system implementation stage, the user interface plays an important role as the main medium that directly connects the user and the system. A good user interface design not only makes it easier for users to operate the system but also improves the overall user experience. In this subsection, we present screenshots of the end-user interface of all the forms available in the system.

Figure 3. Dashboard Interface

Figure 3 shows the dashboard interface, which is the main page that appears when users open the system. This page has three navigation menus in the center of the page to access other features. The dashboard is designed to be simple and intuitive, so users can easily choose the menus they want. From the dashboard, user can choose the "Prediction Form" to open the prediction page. In Figure 4, users are required to input the data needed to process the prediction. Each input element is designed using easy-to-use interface components such as dropdowns, spinners, and edit fields.

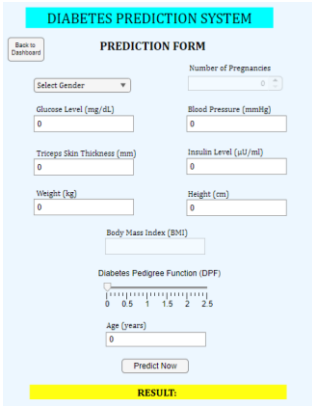

Figure 4. Prediction Form Interface

After all data has been filled in, users can click the "Predict Now" button to run the process using a predetermined algorithm. The system then displays the prediction results in text format below the button, as shown in Figure 5 and Figure 6. There is also a "Back to Dashboard" button that can be used to return to the dashboard.

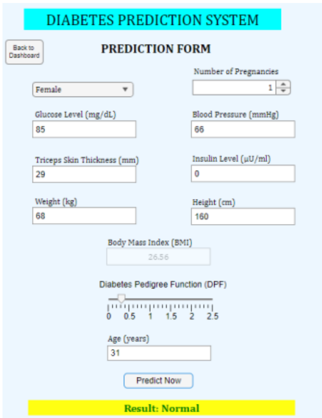

Figure 5. Normal Prediction Result Interface

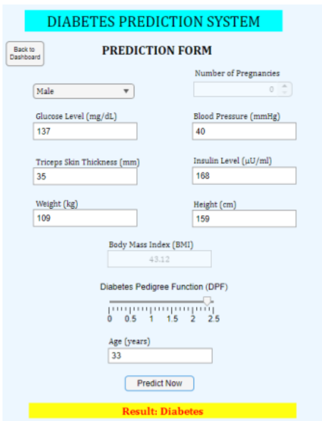

Figure 6. Diabetes Prediction Result Interface

In Figure 7 and Figure 8, an interface is shown where the user clicks the "Parameter Guide" in the dashboard. In this form, there are two pages: the parameter explanation and a little bit of information about diabetes.

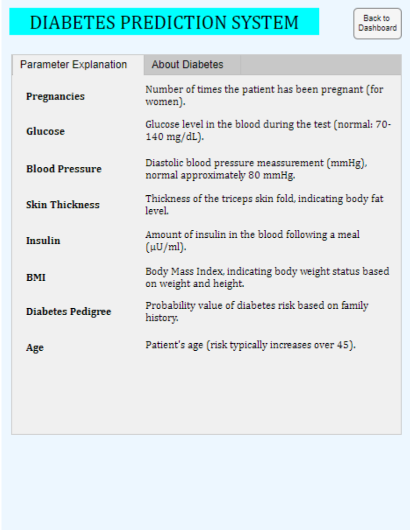

Figure 7. Parameter Explanation Interface Design

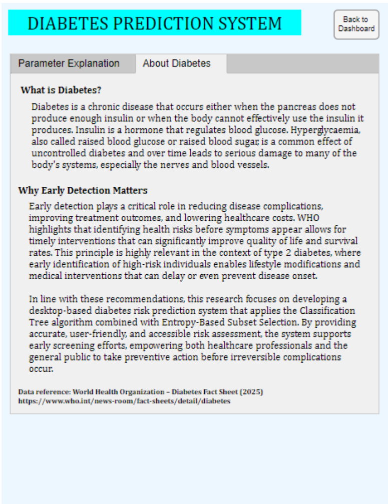

Figure 8. About Diabetes Interface Design

Lastly, in Figure 9 we present the "Exit App" page to close the application completely. When this button is clicked, the system will close all windows and stop running processes. Before that, a confirmation modal is displayed to prevent accidental closure and ensure that the user is sure about closing the application.

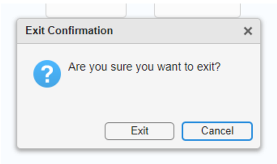

Figure 9. Exit Confirmation Interface Design

3.5 System Testing
3.5.1 Functional Testing

Based on the results of functional testing, all system components have been running according to the planned specifications. No errors were found in the navigation process, input validation, or prediction results display. This indicates that the system is ready for accuracy testing, where the main focus shifts from functionality to evaluating the performance of the prediction model in generating accurate diabetes risk classification based on the test data.

Table 4. Functional Testing

Testing Scenarios	Test Steps	Test Results	Status
Access the Dashboard	Run the application and go to the main page	The system displays the Dashboard page with the main menu	Valid
Input Medical Data Parameters	Input all the required data, then click the "Predict Now" button	The system validates the data, processes the prediction, and displays the diabetes risk results	Valid
View Information	Click the "Parameter Guide" menu	The system displays an explanation of each parameter and general information about diabetes	Valid
Exit the Application	Click the "Exit App" button	The system closes the application correctly without the errors	Valid

3.5.2 Model Accuracy Testing

Accuracy testing is needed to evaluate the performance of the prediction model in classifying diabetes risk based on selected features using the E-ss method. After the feature selection process, the three attributes with the highest information values were used as the main inputs for the classification tree model.

To illustrate the system testing mechanism, several sample test data are presented in Table 5. This table displays only a small portion of the total test data used. Overall, the system was tested on 38 test datasets, consisting of 24 non-diabetic datasets and 14 diabetic datasets, as summarized in the confusion matrix in Figure 10.

Table 5. Sample Test Cases

Blood Pressure	Skin Thickness	BMI	Prediction	Actual Results	Status
72	35	33.6	1	1	Valid
92	-	37.6	1	0	Invalid
82	19	22.2	0	0	Valid
40	35	43.1	1	1	Valid
50	32	31	0	1	Invalid

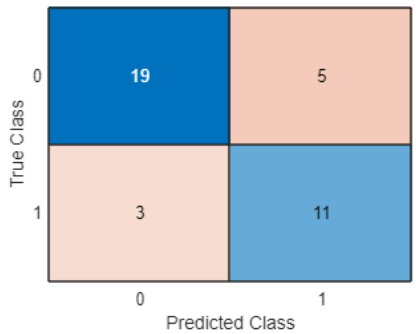

Figure 10. Accuracy Results

Based on the results of the confusion matrix in Figure 10, it is shown that using the cross-validation method, the model achieved an accuracy of 78.95%. This result indicates that the feature selection process not only succeeded in simplifying the model but also maintained the system's predictive capability at an adequate level.

As an illustrative example based on the confusion matrix presented (TN = 19, FP = 5, FN = 3, TP = 11; total N = 38), the calculated values are as follows: Accuracy = (11+19)/11+19+5+3 = 30/38 = 78.95%.

In this study, the system implementation follows the same feature selection and classification configuration as reported in the reference to ensure methodological consistency. The adopted configuration enables the system to perform automated risk classification based on clinically relevant attributes while maintaining an acceptable level of predictive reliability.

4. Discussion

The main contribution of this study does not lie in proposing a new classification method or claiming improved accuracy over existing works, but rather in the successful integration of the E-ss based feature selection a decision tree classification into a functional and user-friendly desktop-based decision support system. While many previous studies focused primarily on model development and accuracy reporting [8], [18], [19], the implementation aspect in an operational information system environment remains limited.

By embedding a previously validated model directly into a desktop-based application, this study strengthens earlier findings by demonstrating that high-performing machine learning models can be effectively translated into practical tools for early diabetes screening. The adoption of only three dominant features (BMI, blood pressure, and triceps skinfold thickness) also supports findings from [8] that emphasize the importance of feature parsimony for improving usability and reducing data collection complexity without significantly degrading predictive performance.

Accordingly, the contribution of this study can be positioned as a system level validation and practical reinforcement of prior methodological findings, rather than a statistical performance improvement. This integration confirms that the method proposed previously is not only effective in controlled experimental settings, but also feasible for deployment in real applications for early diabetes risk detection.

5. Conclusion

Based on the results of research and implementation that have been carried out, it can be concluded that this diabetes risk prediction system based on a desktop application has successfully fulfilled the research objectives. The results indicate that the E-ss method is able to select a subset of relevant attributes from the dataset, which contributes to reducing model complexity and improving computational efficiency. While the E-ss approach has already been shown in the literature to effectively optimize attribute selection and computational efficiency, our focus is on integrating such methods within a user-friendly information system rather than advancing the algorithms itself. In this research, the most significant factors in determining risk were BMI, blood pressure, and triceps skinfold thickness. These attributes then became the input for the classification tree algorithm, which built a structural classification model to differentiate individuals at risk and not at risk of diabetes. Based on functional testing, all the system components can run as designed, from data input, prediction processing, to presenting the results in the interface. The success of this system comes from its ability to provide classification results quickly and in a way that is understandable to non-technical users. With its simple user interface, this system can be used both by medical personnel as an initial diagnostic tool and by individual users to monitor personal risk based on known health parameters.

However, this system has some limitations. First, this model was trained and tested using only a single diabetic dataset, so its accuracy may decrease if used on populations with different characteristics. Second, this system does not consider non-medical risk factors, such as lifestyle or dietary habits, which can also affect diabetes risk. Last but not least, as a desktop application, the system limits user mobility compared to web or mobile platforms.

Given this system's potential, future developments could consider integration with wider and more diverse health databases, the addition of comparison algorithms for a more comprehensive model performance evaluation, as well as the development of a prediction history feature to monitor risk changes periodically. Moreover, adapting the system to a web or mobile platform also has the potential to expand user reach and improve accessibility, along with the addition of interactive educational modules on diabetes prevention and management so that the system can play a dual role as both a prediction tool and a health learning tool.

Reference

[1] W. H. Organization, “Diabetes,” WHO. Accessed: Aug. 02, 2025. [Online]. Available: https://www.who.int/health-topics/diabetes
[2] I. D. Federation, “The Diabetes Atlas,” IDF. Accessed: Aug. 02, 2025. [Online]. Available: https://diabetesatlas.org/
[3] Y. Liu *et al.*, “Performance of a prediabetes risk prediction model: A systematic review,” *Heliyon*, vol. 9, no. 5, p. e15529, 2023, doi: 10.1016/j.heliyon.2023.e15529.
[4] J. Zhang, Z. Zhang, K. Zhang, X. Ge, R. Sun, and X. Zhai, “Early detection of type 2 diabetes risk: limitations of current diagnostic criteria,” *Front. Endocrinol. (Lausanne).*, vol. 14, p. 1260623, 2023, doi: 10.3389/fendo.2023.1260623.
[5] B. T. Jijo and A. M. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” *J. Appl. Sci. Technol. Trends*, vol. 2, no. 1, pp. 20–28, 2021, doi: 10.38094/jastt20165.
[6] T. Dudkina, I. Meniailov, K. Bazilevych, S. Krivtsov, and A. Tkachenko, “Classification and prediction of diabetes disease using decision tree method,” in *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 163–172.
[7] O. O. Oladimeji, A. Oladimeji, and O. Oladimeji, “Classification models for likelihood prediction of diabetes at early stage using feature selection,” *Appl. Comput. Informatics*, vol. 20, no. 3–4, 2024, doi: 10.1108/ACI-01-2021-0022.
[8] V. Pratama and J. Tjen, “Entropy-based subset selection principal component analysis for diabetes risk factor identification,” *J. Emerg. Investig.*, vol. 6, no. 1, pp. 1–4, 2023, doi: 10.59720/23-015.
[9] J. Tjen, F. Smarra, and A. D’Innocenzo, “An entropy-based sensor selection algorithm for structural damage detection,” in *IEEE International Conference on Automation Science and Engineering*, 2020, pp. 1566–1571. doi: 10.1109/CASE48305.2020.9216828.
[10] F. Smarra, J. Tjen, and A. D’Innocenzo, “Learning methods for structural damage detection via entropy-based sensors selection,” *Int. J. Robust Nonlinear Control*, vol. 32, no. 10, pp. 1566–1571, 2022, doi: 10.1002/rnc.6124.
[11] Y. Sun *et al.*, “Analysis of continuous glucose monitoring data in patients with type 2 diabetes mellitus based on entropy analysis,” in *Proceedings of SPIE*, SPIE – The International Society for Optics and Photonics, 2024, p. 132701C. doi: 10.1117/12.3045353.
[12] B. A. C. Permana, R. Ahmad, H. Bahtiar, A. Sudianto, and I. Gunawan, “Classification of diabetes disease using decision tree algorithm (C4.5),” *J. Phys. Conf. Ser.*, vol. 1869, no. 1, p. 012082, 2021, doi: 10.1088/1742-6596/1869/1/012082.
[13] P. Achenbach *et al.*, “A classification and regression tree analysis identifies subgroups of childhood type 1 diabetes,” *eBioMedicine*, vol. 82, p. 104118, 2022, doi: 10.1016/j.ebiom.2022.104118.
[14] N. Ahmed *et al.*, “Machine learning based diabetes prediction and development of smart web application,” *Int. J. Cogn. Comput. Eng.*, vol. 2, no. December, pp. 229–241, 2021, doi: 10.1016/j.ijcce.2021.12.001.
[15] M. Akturk, “Diabetes Dataset,” Kaggle. Accessed: Aug. 07, 2025. [Online]. Available: https://www.kaggle.com/datasets/mathchi/diabetes-data-set/data
[16] K. Larsen *et al.*, “Developing a User-Centered Digital Clinical Decision Support App for Evidence-Based Medication Recommendations for Type 2 Diabetes Mellitus: Prototype User Testing and Validation Study,” *JMIR Hum. Factors*, vol. 9, no. 1, p. e33470, 2022, doi: 10.2196/33470.
[17] A. M. Posonia, S. Vigneshwari, and D. J. Rani, “Machine learning based diabetes prediction using decision tree J48,” in *Proceedings of the 3rd International Conference on Intelligent Sustainable Systems, ICISS 2020*, 2020, pp. 498–502. doi: 10.1109/iciss49785.2020.9316001.

[18] R. Saxena, S. K. Sharma, M. Gupta, and G. C. Sampada, “A Novel Approach for Feature Selection and Classification of Diabetes Mellitus : Machine Learning Methods,” *Comput. Intell. Neurosci.*, vol. 2022, p. 3820360, 2022, doi: 10.1155/2022/3820360.
[19] R. Sistem, M. D. Anasanti, K. Hilyati, and A. Novtariany, “Exploring feature selection techniques on Classification Algorithms for Predicting Type 2 Diabetes at Early Stage,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 5, pp. 832–839, 2022, doi: 10.29207/resti.v6i5.4419.