Performance Comparison of Machine Learning Models Powered by SHAP and LIME Based Explainability Techniques on Diabetes Dataset

Hakan Güler [a], Derya Avcı [b], Mustafa Ulaş [c], Tülay Omma [d]

[a]*Software Engineering, Faculty of Technology, Fırat University, Elazığ, Türkiye*
[b]*Computer Technologies, Vocational School of Technical Sciences, Fırat University, Elazığ, Turkey*
[c]*Artificial Intelligence and Data Engineering, Faculty of Engineering, Fırat University, Elazığ, Turkey*
[d]*Department of Endocrinology and Metabolic Diseases, Faculty of Medicine, Lokman Hekim University, Ankara, Turkey*

Abstract
Early diagnosis of diabetes can increase patients' quality of life and improve treatment processes. In this context, this article focuses on the early diagnosis and prediction of diabetes, addressing the performance of various machine learning models and the role of explainable artificial intelligence (XAI) techniques. With the rise of large datasets in the healthcare industry, data mining and machine learning techniques have become an important tool for the discovery and analysis of diabetes datasets spanning healthcare systems. This study investigates a diabetes dataset that includes healthcare systems. Various machine learning models such as K-NN, SVM, Naive Bayes, CNN, Decision Tree, Random Forest and XGBoost were evaluated on this data set and their performances were compared. Visualizing the overall structure of the data set is important for analyzing relationships between diabetes-related features. The article starts with cleaning the dataset and preprocessing steps, followed by the training and testing phases of each model on the dataset. Each model was evaluated based on success criteria such as accuracy, F1 score, sensitivity, and specificity. In addition, the understandability of the model's decisions was increased by applying explainable artificial intelligence (XAI) methods, SHAP (Shapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to the outputs of the most successful model. These techniques explain the internal working mechanism of the model by determining which features have the most impact on model outputs. The analyzes were supported by expert doctor's comments and the potential of the models in real world applications was highlighted. When the models and results are examined, respectively; it can be seen that the results of K-NN: 81.18%, SVM: 75.38%, Naïve Bayes: 75.49%, CNN: 74.83%, Decision Tree: 76.91%, Random Forest: 91.68%, XGBoost: 98.91% are obtained. As a result, machine learning models effectively demonstrate early detection and diagnosis of diabetes. The explainability of these applied models is emphasized and their effects on real life are shown.

Keywords: Diabetes, Explainable Artificial Intelligence, SHAP, XGBoost

1. Introduction

Today, data science and artificial intelligence technologies [1] are pioneering a significant transformation in the healthcare industry to improve the diagnosis, treatment and management processes of diseases. This transformation involves the development and application of various machine learning models in the field of clinical data analysis [2]. In this context, data-based approaches have gained importance for early diagnosis and effective treatment of chronic diseases such as diabetes. The disease called Diabetes Mellitus (DM) is a type of chronic metabolic disease. It is a chronic disease that occurs under the influence of insulin or caused by various factors. There are types called type 1 and type 2. Type 1 is characterized by insulinopenia. Type 2 appears to occur with insulin resistance and insufficiency. However, there are also specific types of diabetes caused by genetic defects in insulin secretion, pancreatic diseases, endocrinopathies, infections, drugs and chemical agents. For the more common Type 2 DM, risk factors such as age, gender, family history of diabetes, gestational diabetes, hypertension, sedentary life, and body mass index have been identified. Fasting and postprandial blood sugar levels are used for the diagnosis of DM. Early diagnosis and treatment of DM is vital in terms of macro and microvascular complications that may develop in the future. It is known that prediabetes causes a significant increase in cardiovascular risk [3]. Busy lifestyle choices have an impact on the development of diabetes. It has been observed that while it delays at certain times, it is prevented at other times. Despite the existence of these options, it is said that they have an effect on the risks of cardiovascular disease. In large-scale screening studies conducted in our country, it has been determined that nearly half of diabetic and prediabetic patients have not yet been diagnosed.

For this reason, risky individuals must be identified correctly and diagnostic screening tests must be performed on these individuals at regular intervals. Risk scoring tests such as the FINDRISK survey are used to identify risky individuals.

Figure 1. Diabetes

Diabetes has become a growing health problem globally. Managing the effects of the disease and providing personalized treatment plans to individuals requires an accurate and reliable data analysis infrastructure. Targets were set in the study using the data set related to diabetes. Performance evaluations were made by working on the dataset with the help of machine learning models. The main purpose of the study is to comprehensively examine and evaluate the effectiveness of different machine learning methods used in the classification of diabetes. The models used in the study are as follows; SVM, Naive Bayes, Artificial Neural Networks, K-NN, XGBoost, Decision Trees and Random Forest are model types. The capabilities of each of these methods were analyzed separately and the results were shared. Then, LIME and SHAP methods were applied to the XGBoost model to obtain the highest data. By applying the methods to the model, the data obtained was compared and analyzed. These explainable artificial intelligence (XAI) [4] techniques helped us understand model decisions and identify which features have the most impact on model outputs. These evaluations were compared with the feedback we received from doctors who are experts in their fields.

The remainder of the article presented a detailed evaluation along with the metrics used to determine the performance of each model. We will also discuss feature importance ranking to understand which features and parameters affect model performance. In this context, we aim to provide a perspective on how machine learning models can be used to better understand health data and effectively manage diseases. This study aims to create a useful resource for professionals, researchers and data science experts in the healthcare industry.

The unique contributions of the study are as follows:

- **Optimized Deep Learning Architectures:** Our study includes deep learning architectures designed specifically for diabetes prediction and aims for maximum prediction accuracy by comparing with models such as K-NN, SVM, Naive Bayes, Random Forest and XGBoost. This aims to increase their ability to understand and effectively predict the complexity of diabetes.
- **Holistic Evaluation of Prediction Models:** The study aims to comprehensively evaluate traditional and modern machine learning models on diabetes prediction and distinguish the strengths and weaknesses of each model. This paves the way for important recommendations for future research and practice.
- **Robust Feature Importance Assessment:** Beyond understanding model performance, the study also elaborates on the importance of individual features in diabetes prediction. Feature importance analysis applied to Random Forest and XGBoost models highlights critical factors, making the model's decisions more understandable.
- **Health Perspective for Practical Applications**: Going beyond technical methods, this study focuses on how models can be used in real-world health scenarios, bridging theoretical developments to tangible benefits in health applications. It aims to provide practical information for healthcare professionals and decision makers.
- **Examination of Future Directions:** The study not only focuses on current findings but aims to explore potential future directions in the field of diabetes prediction. It discusses model adaptability and scalability and provides suggestions for future research and development.

These unique contributions provide a comprehensive perspective that informs not only current knowledge but also future research and applications.

2. Literature Review

The relevant part of the study focuses on studies conducted in similar fields and their methods. It is an important step to present a different approach from the existing ones or an innovative approach, considering the methods used. Studies on diabetes have been examined due to these factors. This chapter provides an extensive review of previous studies to provide a comprehensive understanding of diabetes. Information compiled from current articles on health data analysis reveals the latest advances in diabetes diagnosis and management. In this context, how the literature review can contribute to the existing knowledge in the health sector is discussed in detail.

Mirza and his colleagues conducted a study investigating how fast and accurate prediction of diabetes, known as Diabetes Mellitus (DM), can contribute to the development of health technology and the creation of e-health systems. Mirza et al. proposed an effective data processing method in their study. They said that they achieved high accuracy with this method they suggested. They emphasized that the ability of machine learning algorithms to make effective diabetes predictions has the potential to significantly reduce annual mortality rates, especially in developing countries such as Bangladesh [5].

In their study, Hassan and colleagues suggested that diabetes can be predicted more accurately in various health data sets with the help of an intelligent health recommendation system using models of machine learning algorithms. Technology is developing every day. This developing technology is also reflected in machine learning models. However, they stated that these systems cannot adequately handle large, multi-feature health data sets. They proposed an intelligent health recommendation system to predict diabetes using machine learning. They observed that the machine learning model provided a robust evaluation using the K-Fold Cross Validation test when trained for diabetes prediction. Hassan and his colleagues first evaluated the results of their study themselves. Then, they compared it with similar studies and recent developments [6].

Morteze and colleagues reported a detailed review of how existing data mining techniques are used to estimate the impact on diabetes. They also revealed ways to use the RapidMiner data mining tool to predict the severity of diabetes. Diagrams, models, etc. containing various models and parameters that can be tested/operationalized with the help of RapidMiner. They provided a preview process with. This preview indicates that the same dataset is retrieved from the Kaggle website and the same dataset can be used by any user [7].

Kebira and his team collected patient data to predict the incidence of severe diabetes. While doing this, they proposed an IoT-based system architecture. They introduced a statistical method to analyze the data and used Blockchain and IPFS to secure the data. First, they collected data from IoT devices. They then created a data set using IPFS. The data was then subjected to different methods. First of all, they scaled with data expansion. They then trained the data using their proposed model and created an adaptive random forest algorithm to classify people with diabetes. They used three different data sets. They created and used a dataset that was the Pima Indian diabetes dataset, the Frankfurt Hospital diabetes dataset, and a combination of the two. Finally, they compared the performance of the method with similar prediction methods [8].

Lukmanto et al. aimed to classify diabetes in their study. To achieve the goal of the study, F-Score Feature Selection and Fuzzy Support Vector Machine methods were preferred. They used the feature selection feature to specify valuable features in data sets. Then, they trained the data sets using the SVM model. They created fuzzy rules with this trained data set. In the later process, they made the classification based on these processes. They applied this method to the Pima Indian Diabetes dataset. They said that their results were promising in predicting diabetic patients with an accuracy rate of 89.02% [9].

Changsheng and his colleagues conducted a study on the early diagnosis and prediction of diabetes. They conducted their study using the Pima Indian Diabetes data set. They proposed an approach based on data mining. Its primary targets included the k-means clustering algorithm and the logistic regression accuracy algorithm. They said that the results were improved with the models they suggested. In addition to these proposed models, they also conducted principal component analysis. As a result, they said that these methods were successful in these methods [10].

Huma et al. He emphasized the importance of early diagnosis, as untreated diabetes poses serious health risks. In their research, they adopted advanced machine learning algorithms that use large health datasets to uncover hidden patterns for early diabetes diagnosis and used the PIMA dataset. They used it as a goal. Huma and her colleagues preferred to use artificial intelligence models and classifiers such as Naive Bayes, Artificial Neural Network, Decision Tree, Deep Learning in their studies [11].

Zeeshan and colleagues focused on the problem of "Pre-Diabetes Detection" and addressed this challenging problem using various deep learning and machine learning algorithms. They stated that the main purpose of their work is to detect diabetes at an early age, to enable patients to diagnose this disease correctly and to take better care of their health. Zeeshan and colleagues used various machine learning approaches. First of all, they used deep learning techniques such as ANN, LSTM, MLP. After applying these approaches, they have also applied various machine learning approaches. Among the approaches they apply, there are also machine learning methods such as random forest and logistic regression. They stated that among all these models applied, the model they proposed achieved better results. The reason for this good performance is that it is effective in LSTM and deep learning methods [12].

Muhammed and his colleagues aimed to develop an artificial intelligence model that can predict diabetes. They used machine learning methods in line with these goals. Their proposed method involves a comparison between K-Nearest Neighbors (KNN) and Naive Bayes algorithms to determine the most suitable algorithm for predicting diabetes. Based on the results of their study, they tested two different K-Nearest Neighbors algorithms and Naive Bayes algorithms to predict diabetes based on various health characteristics in the study data set. They said that their results were successful [13].

Mohammed and his team aimed to predict diabetes. They used Saudi Arabia hospital dataset. There are three different types of diabetes sets in this data set. They used machine learning methods. They used different methods such as SVM, Random Forest, KNN, Decision Tree, Bagging and Partitioning. They conducted four different experiments on the data set with these algorithms. To address the balance issue in the dataset, they resorted to the Synthetic Minority Oversampling Technique method. They said that they were successful after all these procedures. They stated that their work would help other studies in the relevant field [14].

Aishwariya and her colleagues studied diabetes. They accessed a diabetes-related data set from the South Asian country of Bangladesh. They studied this data set using machine learning methods. They used various machine learning (ML) methods such as Naive Bayes, Random Forest, Decision Tree, XGBoost and LightGBM (LGB). Different methods are used to make critical adjustments in machine learning. Aishwariya and his colleagues also used Grid Search Hyperparameter Optimization, one of these methods. They also added missing value imputation, feature selection, and K-fold cross-validation to the framework design. They shared their accuracy rates after their studies. They stated that they achieved a high accuracy rate of 73.5% in these rates they shared. They also shared the ROC curve result. Therefore, they argued that their rates were good [15].

Ayşe and her colleagues proposed a new collective learning model for the early diagnosis of diabetes. In their proposed method, they aimed to make better predictions by combining the prediction results of more than one machine learning algorithm. In this direction, they aimed for an approach based on cross-validation. They suggested a model. This model they propose consists of different machine learning methods. They proposed a model by combining basic learner methods and meta learner methods. The results were analyzed using three different data sets to measure the robustness of the proposed model [16].

Raja and his team conducted a study on diabetes prediction. They carried out their operations using common types of machine learning. The most well-known among these are SVM, decision tree and random forest method. Leveraging machine learning, they proposed to develop a unique intelligent diabetes prediction framework (IDMPF). They carefully reviewed existing forecasting models in the literature. They then talked about their recommendations after evaluating its applicability to diabetes. Using this framework, they described training procedures, model evaluation strategies, problems and solutions for diabetes prediction. Raja and his colleagues said that they achieved a high accuracy rate in their study and said that it was a guiding source in future studies [17].

Monalisa and colleagues have studied an effective model with high sensitivity for diabetes prediction. They used the K-nearest neighbor algorithm to reduce the processing time. They used support vector machine to determine the class corresponding to each data sample. They predicted whether an individual had diabetes using four different algorithms and shared their results [18].

Prasannavenkatesan and colleagues wanted to perform predictions on the diabetes data set. They have benefited from traditional classification algorithms. Neural network-based machine learning methods were also used in their studies. They evaluated various performance methods in different directions for k-nearest neighbor, Naive Bayes, decision trees, radial basis function and multilayer perceptron algorithms. They shared the results of their work. They stated that multilayer perceptron methods work at a high accuracy rate [19].

Mahesh et al conducted a study to diagnose diabetes. In this study, machine learning approaches were chosen as a guide. They made comparisons with clinical results using these algorithms in the data set they received. They showed the best values as a result of the comparison. They have tried various machine learning approaches in their work. They thus aimed to obtain different results by choosing many methods. They said that the purpose of using multiple methods was to see different results. They shared the results and comparisons of these techniques used [20].

There are many studies conducted in the field of health with explainable artificial intelligence [21-25].

3. Materials and Methods

This study aims to create an effective prediction model for the early diagnosis of diabetes. Many types of chronic diseases exist around the world. Diabetes is among the most important of these. In this section, sections such as used data settings, data pre-processing steps, and machine operating models used are explained in detail.

3.1. Data Set

Researchers focusing on diabetes risk assessment and prediction have chosen the dataset in the study as an important resource on the Kaggle platform for data scientists and medical professionals [26]. This data set includes various health-related functions. On the other hand, there are various fields that are different from each other at each recording point of the data set. The illustration and presentation of the relevant parts in the data set are available in table 1. When the areas are examined, the basic effects of diabetes can be seen together. We see the characteristics of the patient. There are many values in this individual, from his identity information to his age, his family tree, blood pressure and values. Each of these values causes different characteristics and results. Their association with diabetes is seen as the most important factor. These features need to be extracted and turned into meaningful features. Afterwards, it is effectively classified and developed using machine learning methods. An overview of the data set is made in table 1.

Table 1. Dataset Summary

Field	Description
Id	Field that is unique for each record in the data set
Pregnancies	Number of times getting pregnant in the data set
Glucose	Number of glucose in the data set
Blood Pressure	Blood pressure field in the data set in mm Hg type
Skin Thickness	The area with skin thickness on the data set is taken in mm
Insulin	area called insulin in mu U/ml
BMI	Body mass index area in kg
Diabetes Pedigree Function	Genealogy field to find genetic problem of diabetes
Age	Age field in the dataset
Outcome	Diabetes or not diabetes field in the data set (0 and 1 output)

3.2. Explore Data Analysis

In this part of our study, we examined our data set in detail. The biggest factor in doing the relevant part is to understand and make sense of the data set. It was done to understand the numerical values on the data set and to identify the fields that need to be associated. The structure of the parts and variables in the data set is understood. Afterwards, it is shown which fields they may be related to.

Table 2. Explore Data

Label	Count	Mean	Std	Minimum	% (25)	% (50)	% (75)	Max
Id	2768	384.50	799.19	1.00	692.75	1384.50	2076.25	2768.00
Pregnancies	2768	3.74	3.32	0.00	1.00	3.00	6.00	17.00
Glucose	2768	121.10	32.03	0.00	99.00	117.00	141.00	199.00
Blood Pressure	2768	69.13	19.23	0.00	62.00	72.00	80.00	122.00
Skin Thickness	2768	20.82	16.05	0.00	0.00	23.00	32.00	110.00
Insulin	2768	80.12	112.30	0.00	0.00	37.00	130.00	846.00
BMI	2768	32.13	8.07	0.00	27.30	32.20	36.62	80.60
Diabetes Pedigree Function	2768	0.47	0.32	0.07	0.24	0.37	0.62	2.42
Age	2768	33.13	11.77	21.00	24.00	29.00	40.00	81.00
Outcome	2768	0.34	0.47	0.00	0.00	0.00	1.00	1.00

We need to convert some parts of the data set into numerical data. We see that that field is introduced as 0 to find the number of missing values. Likewise, the number of duplicate records is seen as 0. The dual chart function of the Seaborn library is used to visualize diabetes status based on the "Result" column in the dataset. This graph visually reveals possible relationships on diabetes status by plotting a point cloud plot for each pair of features. This graph provides visual support for the analysis results in the article. Images of some features are given below.

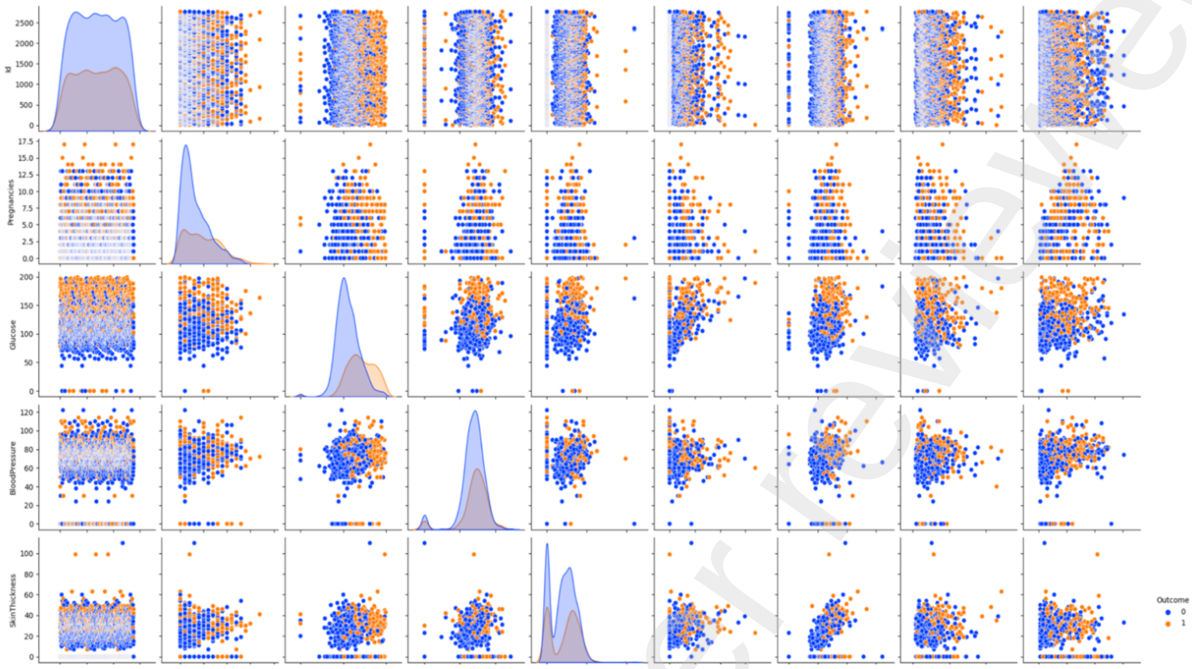

Fig. 1. Colorized Dual Feature Chart

Then, the distributions of the classes and the relationships of the features were analyzed. Various analyzes were made on the data set. The main purpose of these analyzes is to show and understand the diversity between classes in the data set. In addition, it is aimed to show the differences between the variables. First, the class distribution was shown. In this feature, the distribution of class labels is displayed as a percentage through the "Result" feature. Additionally, a distribution was created to determine the relationship between “Age” and “Insulin” features. In the result section, the output result of the features is shown. In addition, a summary containing average values according to classes was created for each numerical feature. This summary provided a detailed understanding of the distribution of numerical features across classes. As a result of the analysis, a better understanding of the data was achieved. In this way, transactions were made in a more meaningful way.

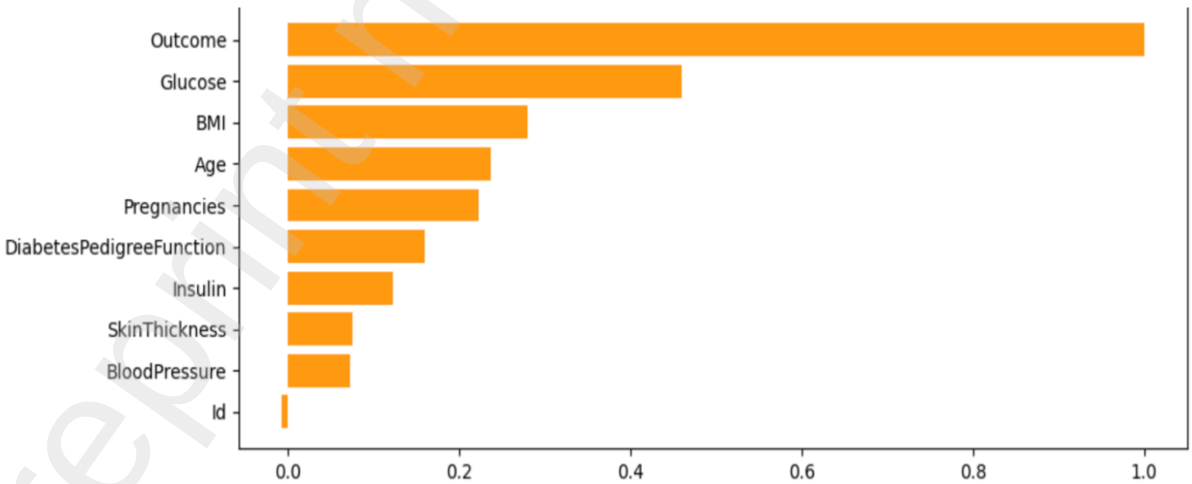

Fig. 2. Correlation with Outcome

Correlation analysis is a widely used technique for understanding relationships between numerical variables in a data frame. The heat map below visualizes the correlations between the “Result” variable and other numerical variables.

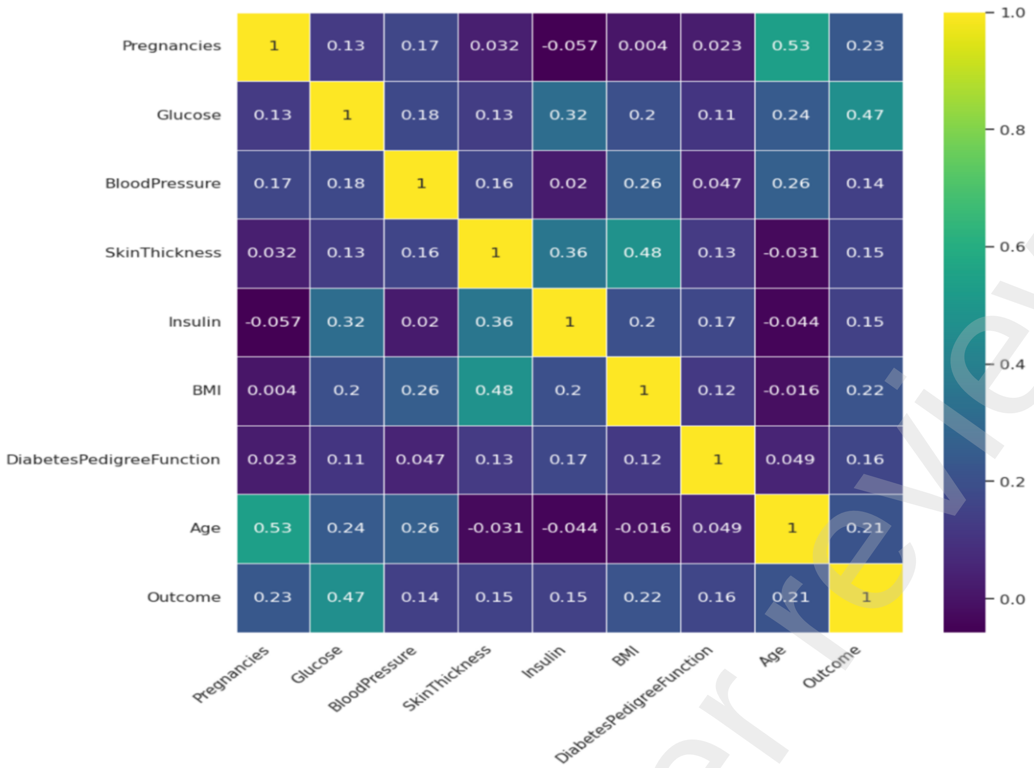

Fig. 3. Correlation Analysis

Each of the numbers in the relevant cell means something. It appears as the coefficient between variables or features. Illustration of positive and negative relationships is also provided. The closer the values, the stronger the relationship. Values close to 0 indicate low correlation or no correlation. Therefore, each cell of the table makes important points to explain the relationship between variables.

Kernel density estimation plots and boxplots were drawn to examine the distribution of various numerical features of the data set. Removal of outliers was also performed for each numerical feature. In particular, a cleaning process was applied against outliers by keeping the data points within the limits determined according to the lower and upper percentiles for each feature. These visual and statistical analyzes provided important information for understanding the distributions in the data set and the process of removing outliers.

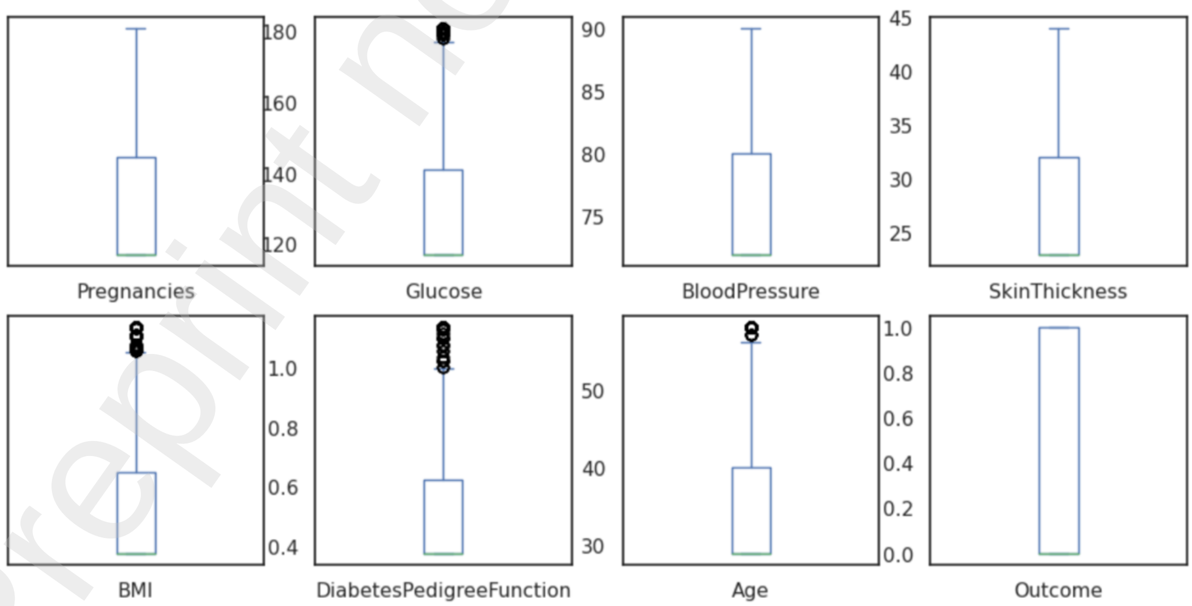

Fig. 4. Box Plot Analysis of Numerical Features

3.3. Machine Learning Models

In this section, the methods used in the study were examined in detail. Machine learning methods and methods were examined. After the data set analysis process was completed, the relevant data set was applied by different methods. When we look at these methods; Methods such as KNN, SVM, CNN, Random Forest, Decision Tree and XGBoost are available. Each algorithm has been carefully evaluated for their ability to understand the data set and identify potential cases of diabetes.

The K-nearest neighbor (KNN) classifier is a robust algorithm that has long been used in the field of classification as both a base model and a reference point. Despite its simplicity and applicability, the KNN classifier often achieves competitive results compared to more complex machine learning methods. Traditionally, without relying on prior knowledge, KNN classifiers use Euclidean distances to measure the closeness between examples [27].

Support Vector Machine is a form of supervised machine learning. Although it is statistically based, it is a widely used classification type. It was released by Cortes and Vapnik in 1995. They said that the main purpose of its removal was to increase the separation between classes and minimize the prediction error. The support vector machine algorithm has various capabilities. The most important of these is that it can work with linear and non-linear data. It is a type of algorithm that focuses on the problems in the sizing section. It is effective on high-dimensional spaces. In addition, it also plays a role in the classification of small data sets [28].

One of the commonly used methods is Naive Bayes classification. This method appears to be a practical method. Its basic working principle is to estimate probability. He finds the probabilities by making predictions based on his past experiences. It is a method that appears as a situation related to the existence or non-existence of a situation. It is the most preferred method type in supervised learning methods. The reason for this is its probability estimation and flexible structure. Another advantage of the Naive Bayes classifier is its minimal need for training data to predict parameters (averages and variances of variables) required for classification [29].

Decision tree algorithms start with many situations or examples and create a tree data structure that can be used to classify new situations. In each case, it is defined by various properties that can have numerical or symbolic values. In the working method of the decision tree, each training condition has a label corresponding to it. This label is associated with the relevant class by representing the name of the class. Additionally, one of its most important features is the presence of a decision-making mechanism [30].

Random Forest is a supervised learning technique used for classification and regression. Based on statistical learning theory, the algorithm uses bootstrapping with random sampling to create multiple replicates from initial training datasets. The logic of the algorithm selects a value representing the number of trees in the forest, creates bootstrap samples using the bagging method, and determines the optimal split using the Gini criterion [31]. RF effectively combines multiple decision trees and the predictions made through each of them to create a robust prediction model, making it suitable for complex datasets and classification problems.

The CNN algorithm is a powerful method for classifications. It has a successful algorithm that includes different classes, achieving highly accurate classification results. Additionally, CNN minimizes the hyperparameters used in the algorithm and does not require excessive training time. The training and test accuracy rates of CNN are high, making it more user-friendly for working with small datasets compared to other deep learning algorithms [32]. With these features, CNN has become a preferred algorithm, especially in fast, efficient, and high-accuracy classification tasks.

XGBoost is one of the most popular optimization advancements for gradient boosting machines (GBM). Widely used in the industry due to its minimal complexity in problem-solving, performance, and feature engineering. When compared to deep learning solutions, XGBoost is considered easier for working with small data on CPUs. The advantage of XGBoost lies in its reliable objective function for graphical representation [33].

These models offered different approaches to the complexity of the dataset. We evaluated the various models we selected in detail. Having a large number of models was seen as an important step in comparing the results. It was stated that the data set allows for more robust investigation of patterns and behaviors. In order to increase the general understandability of the model, various explanations and comments were added, the reliability of the models was emphasized and future predictions were made.

Suggestions for studies were made. This study aims to provide an understanding of the complexity of machine learning models as well as a comprehensive look at how these models can perform in real-world applications.

4. Result and Analysis

The focus of the study was diabetes. Diagnosis and diagnosis of the disease are of great importance. While doing this, different machine learning models were used. Models such as KNN, CNN, SVM, Naive Bayes, Random Forest, Decision Tree and XGBoost have been used. Each software was evaluated to create a diabetes diagnostic model using selected aggregated health features comprising a large health data set. During this evaluation process, the objectives of each program were examined separately. The results obtained will help us understand the advantages and limitations of each application and provide important information on how these models can be used in clinical applications. In the study, it was observed that the highest accuracy rate was on the XGBoost model. Explainable artificial intelligence methods were also applied to the model. These strategies are called SHAP and LIME. Both methods made different contributions to the model. These techniques operate to identify the features of programs' decisions and which features contribute most to their predictions. By including expert doctor services in the field and evaluating these services, the results obtained are supported by doctor opinions, providing more insight into how these models can be used in real-world applications. As a result, this revised study provides a comprehensive analysis file that includes explainable artificial intelligence techniques and doctor recommendations, as well as evaluations of machine options used in diabetes prediction.

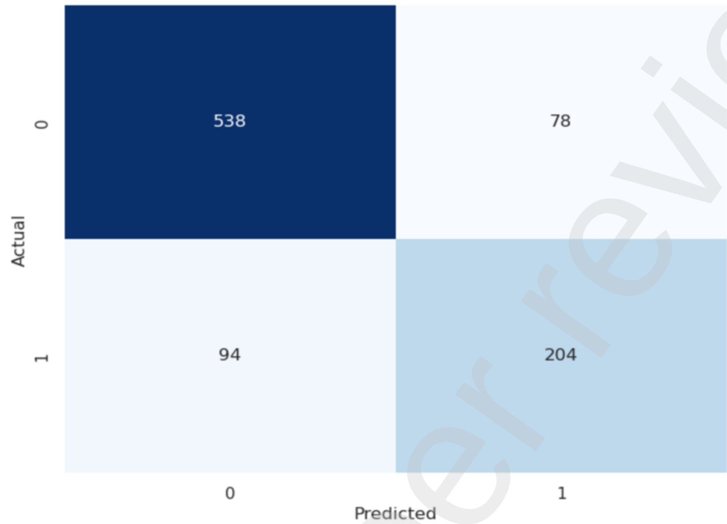

Fig. 5. KNN Confusion Matrix

Additionally, for each confusion matrix, let's focus on the confusions between classes to better understand the success of the model and identify possible errors. Confusion matrices helped us understand in which classes the model performed better or worse by showing the actual and predicted values of each class. Let's examine the confusion matrices of some models.

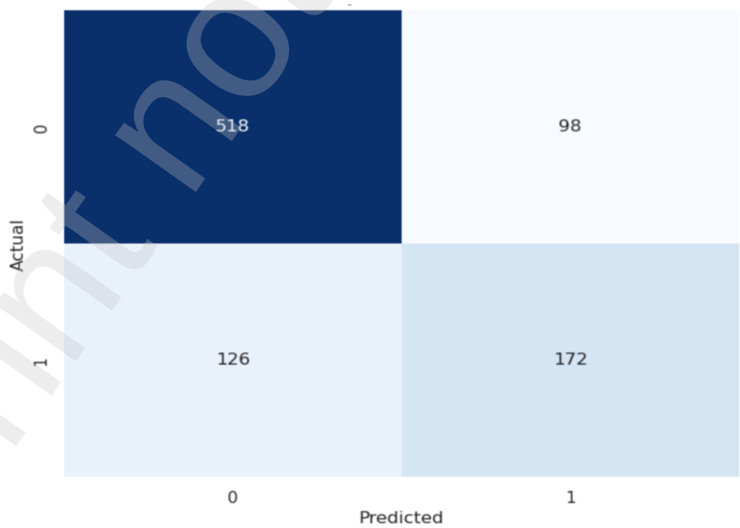

Fig. 6. Naïve Bayes Confusion Matrix

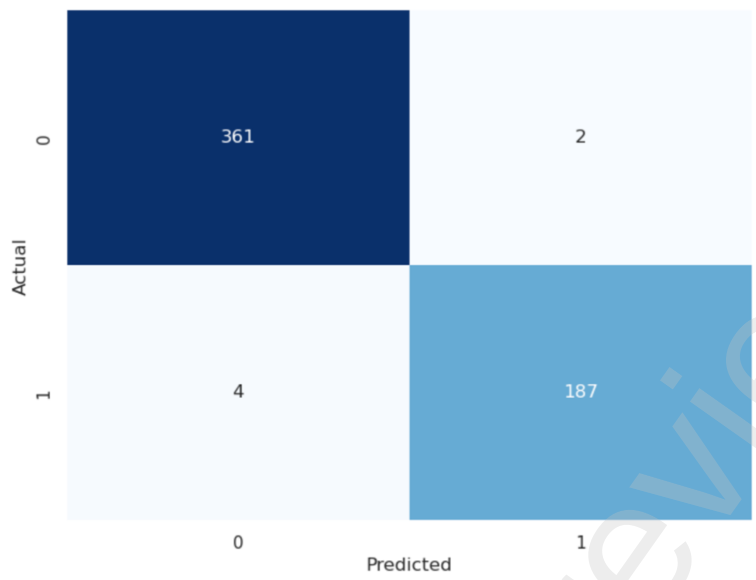

Fig. 7. XGBoost Confusion Matrix

Among the models used, KNN showed 81.18% accuracy. In addition, it was observed that it showed good results when looking at the F1 score, recall and precision values. SVM was used as another method. The accuracy of the SVM model is 75.38%. It showed a similar performance in other metrics, with F1 score, recall and precision values being close to each other. The Naive Bayes model performed similarly to other models with an accuracy rate of 75.49%. F1 score, recall and precision values are also close to each other. The Convolutional Neural Network (CNN) model was evaluated with an accuracy rate of 74.84%. It performed similarly in other metrics. The accuracy rate of the Decision Tree model is 80.53%. F1 score, recall and precision values were also evaluated in a balanced manner. The Random Forest model achieved the highest success with an accuracy rate of 91.58%. F1 score, recall and precision values are quite high and balanced. The XGBoost model has an outstanding accuracy rate of 98.92%. It showed similar high performance in other metrics.

During the comprehensive evaluation of each model's performance, meticulous attention was paid not only to accuracy but also to important metrics such as F1 score, recall and precision values. These metrics play a crucial role in providing a nuanced understanding of the overall classification capabilities of the models. F1 score is formed by the average of precision value and sensitivity values. The F1 score is based on false positive values and false negative values. Both values help evaluate performance. How much success is achieved in positive situations is called recall. Precision is the part that calculates the accuracy rates of classes defined as positive.

Table 3. Results

Model Name	Accuracy	F1 Score	Recall	Precision
K-NN	81.18	81.04	81.18	80.95
SVM	75.38	73.53	75.38	74.50
Naïve Bayes	75.49	75.16	75.49	74.97
CNN	74.83	71.47	74.83	75.03
Decision Tree	76.91	76.20	76.91	76.16
Random Forest	91.68	91.54	91.68	91.70
XGBoost	98.91	98.91	98.91	98.91

In our study, we focused on the results of our study on diagnosing diabetes using various machine learning models. Among these models, the XGBoost model stood out with its high accuracy rates and performance metrics. However, due to the complexity of the model, it is important to understand which attributes XGBoost focuses on when making a particular prediction. As a result of the study, LIME and SHAP methodologies were preferred to explain a specific prediction of the XGBoost model. At this stage, it was seen that the complex version of the model was reduced to a simple one. Detailed examinations were made on the simple model.

LIME (Local Interpretable Model-Agnostic Explanations) is among the explainable models artificial intelligence methods. The working logic works by changing the inputs of the queries. When it works like this, it works on certain parts of the model rather than explaining the whole model [34].

SHAP (SHapley Additive exPlanation) method, presented by Lundberg and Lee in 2017, is another explainable artificial intelligence method. The SHAP method works by assigning importance values to each feature rather than a specific part [35].

In the study, the internal decision process of the model was evaluated by applying explainability Methods such as LIME and SHAP are taught on a pattern model separated by XGBoost applications. The results obtained were aimed to show which features both explainability techniques focus on when creating a specific prediction for the model and the effects of these features on the prediction.

Explanations produced by LIME show the local effects of features that affect the probability of a particular prediction. For example, while the Pregnancy feature has a positive effect, the blood sugar (Glucose) and Age features have a negative effect. The reason for these operations is to ensure that the prediction probability is understood. Factors that increase and decrease the probability come first.

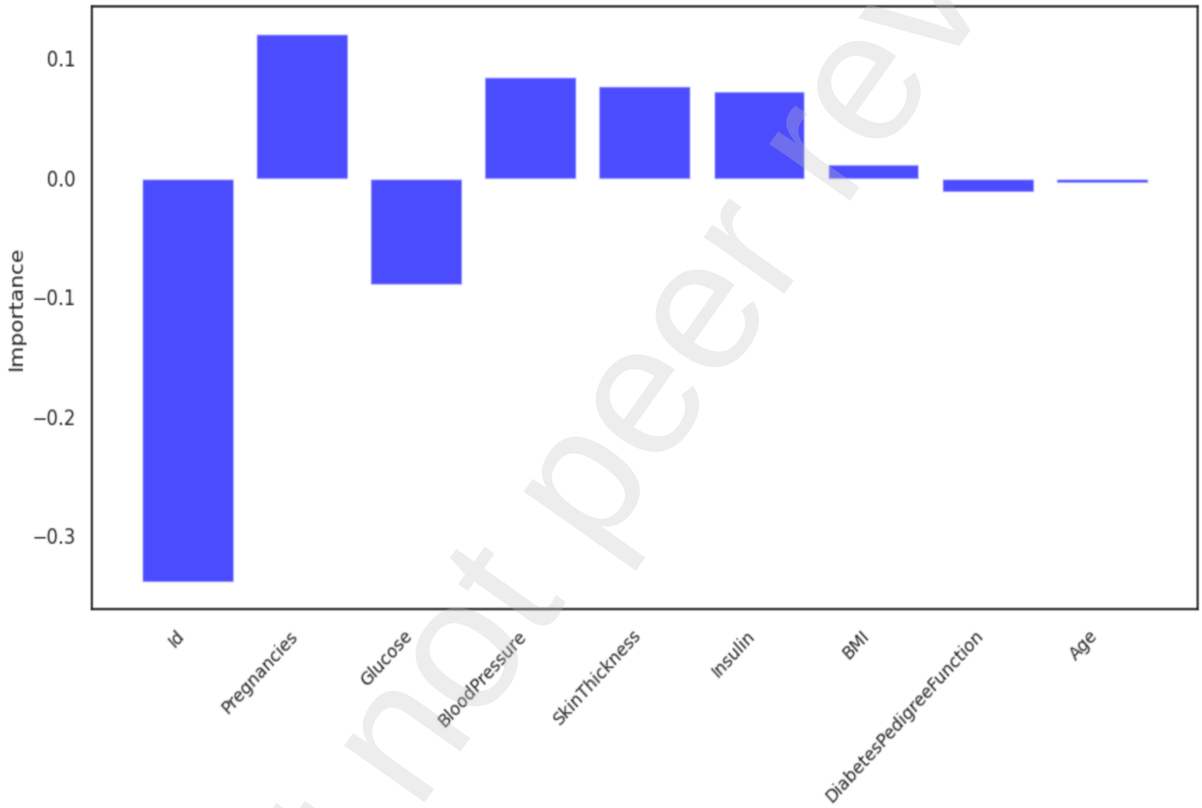

Fig. 8. LIME Feature Importance

SHAP ranks first among effective explainable artificial intelligence methods. When the SHAP study is examined, the contribution of the features to the prediction is clearly seen. The effect of variables such as body mass index and pregnancy on the model prediction is shown. Another variable, blood sugar, has been shown to have a negative effect. Thanks to these results and the results of the SHAP feature, decision processes have become easier. A broader perspective is presented and performance analyzes are provided.

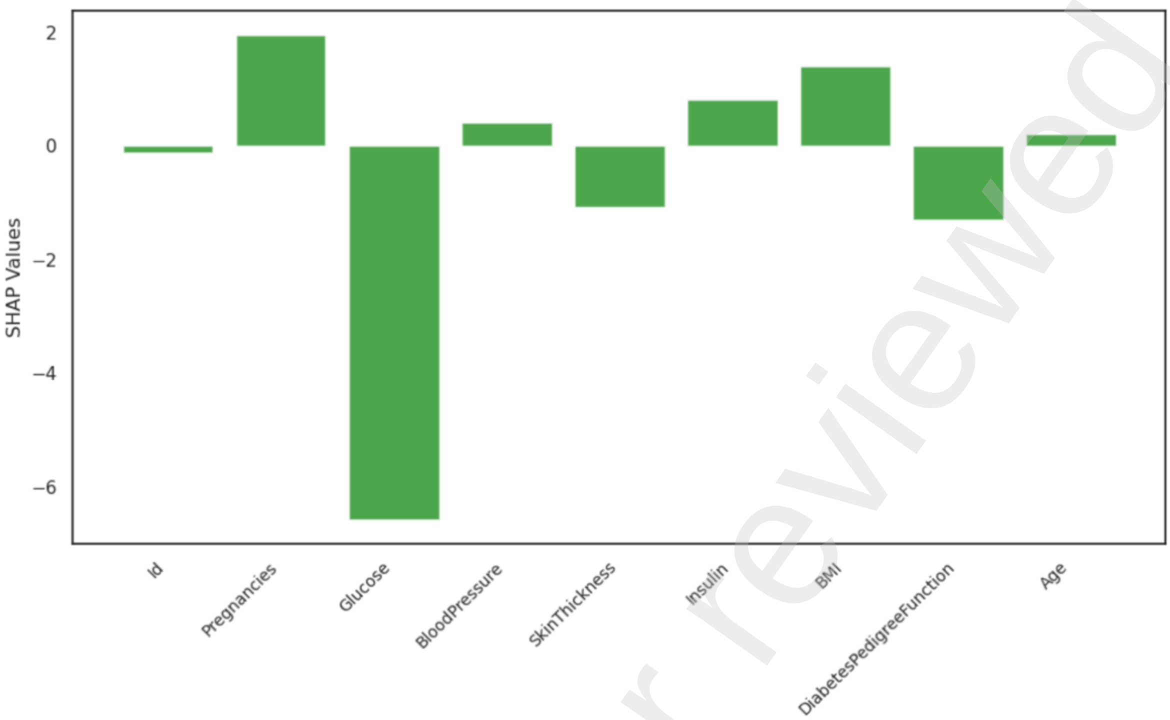

Fig. 9. SHAP Summary

The graph representing the SHAP (SHapley Additive exPlanations) summary graph showing the contribution of the XGBoost model to the class prediction of each feature is shown below. Each bar shows the impact of a feature on the model output, while bar colors represent the values of the feature. In particular, the 'Glucose' feature plays an important role in the prediction of the model by affecting the prediction of the model by pulling down the class label with its negative SHAP value. The effects of other features were evaluated similarly.

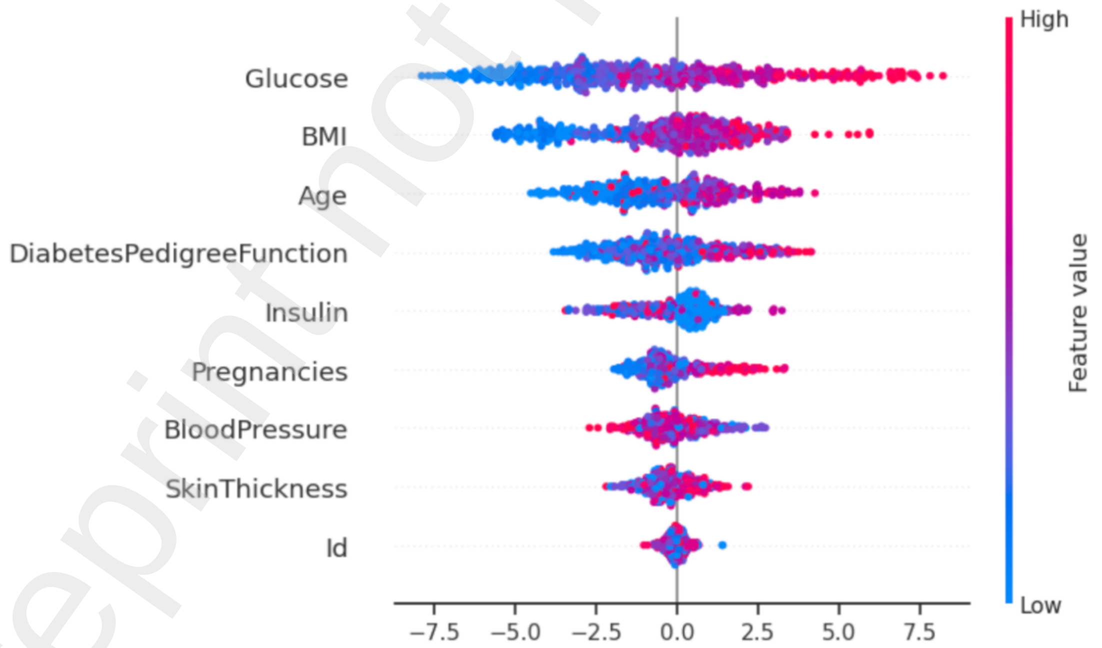

Fig. 10. SHAP Feature Weights of the XGBoost Model

A comparison between the shap evaluation in the study and the ranking of the expert doctor in the field was taken into account. Both rankings reveal that Glucose, BMI and Age stand out in determining important factors in diagnosing diabetes. However, it is noticeable that Pedigree plays an important role in both rankings. Factors such as Insulin, Pregnancies, Blood Pressure and Skin Thickness differ slightly between the two rankings. This comparison highlights that both the machine learning model and the specialist doctor rank similar importance, but have different priorities on some features.

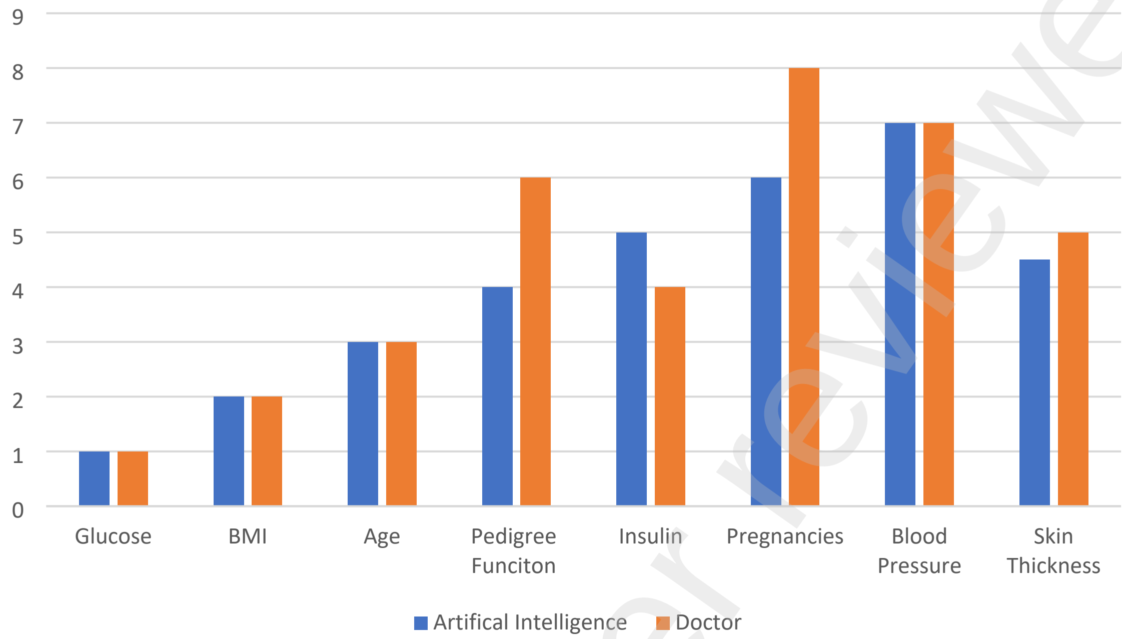

Fig. 11. Comparison of Artificial Intelligence and Doctor Rankings: Priority Factors in Diabetes Diagnosis

In this part of the study, explainable artificial intelligence methods such as LIME and SHAP are mentioned. These methods are needed to understand the decision structures in the methods used and to explain why certain predictions are made. These targeted features were achieved through explainable artificial intelligence methods. Explanations have a reliability-increasing feature on the study, as in many areas. Thus, the performance and confidence coefficients of the model also increase. In addition, it also enables detection of exceptions on the model. As a result of all these, the model is evaluated in terms of performance. Deficiencies or excesses are detected and processed accordingly.

The results obtained with these products demonstrate progress in diabetes diagnosis across a variety of featured machine models. The successful performance of the model we developed can be evaluated more broadly by comparing it with other disease results in the literature. Especially the high accuracy rate of our XGBoost model, 98.92%, is remarkable when compared to the results obtained from other studies. These results provide a framework for public registry review of various machine parameter models relevant to diabetes diagnosis. However, when making model comparisons, it is important to consider the relationships between the operations performed, methodological data and the characteristics of the data sets used and to monitor these operations. The results provide a new perspective on diabetes diagnostic research that can guide behavior and glean information from it.

Table 4. Comparison of Models in Various Studies

Article	**Model**	**Accuracy (%)**
Proposed Method	XGBoost	98.91
The study of Rastogi et al. [36]	Logistic Regression	82.46
The study of Reddy et al. [37]	K-NN	80.80
Bansal et al.'s study [38]	Logistic Regression	75.32
Nass et al.'s study [39]	Support Vector Machine	78.00
Khanam et al.'s study [40]	K-NN	79.42

The results of this study reveal the impact of machine learning models in diabetes diagnosis. High accuracy rates and balanced performance metrics show that these models can be used effectively in clinical applications. Comprehensive studies with larger data sets and various methodologies in future research may further increase the reliability and general applicability of these models. These developments may provide significant advances in early diagnosis and effective treatment of diabetes.

5. Discussion

The results demonstrate the effectiveness of various machine learning models in diagnosing diabetes and demonstrate the diverse abilities of these models to address this complex problem. He emphasizes that what is important is the choice of the most appropriate model in the context of a particular problem. By examining not only accuracy but also these complex metrics, we gain insights into the strengths and weaknesses of the models. This nuanced model evaluation approach contributes to a more robust understanding of their performance and guides us in selecting an optimal solution that suits the unique characteristics of the diabetes diagnosis problem.

In DM guidelines, it is recommended to calculate people's risks of developing DM in the future by using questionnaires containing risk factors, and lifestyle change treatment is planned before complications begin. In the FINDRISK questionnaire, which is the most commonly used for this purpose, the total risk score can be calculated by giving points to questions such as age, body mass index, waist circumference, exercise level, fruit and vegetable consumption, blood pressure level, history of DM during pregnancy and history of DM in the family.

Environmental factors have effects on diabetes. These effects are seen on type 2. Societies' adoption of many critical points has led to these. There are many factors, ranging from a modern lifestyle to the decrease in people's active movements. When we look at the third quarter of the century, we see differences in food. The importance of quickly prepared and consumed foods that are low in fiber and high in fat and calories has increased. This increase has become inevitable for the incidence and frequency of diabetes to increase. The occurrence of these reasons also affected type 2 diabetes. As a result of these reasons, extensive studies have been carried out to take some precautions. The study titled 'Da Qing Diabetes and IGT' is included in the field of prevention studies. This 1997 study was conducted on people with IGT in China. In the study, 40% said that diet and exercise had effects on diabetes [41]. Another study, the Finnish Diabetes Prevention Study (FDPS), was conducted and published in 2001. This program is based on obese or overweight patients with IGT. They showed that the risk of type 2 diabetes would be reduced by 58% [42]. In the Diabetes Prevention Program (DPP) study, whose results were published in the United States and Canada in 2003 and included overweight or obese individuals with IAD or IGT, the risk of diabetes was 58% with intensive life changes, while this rate was 31% with metformin. It has been reported that the rate decreases [43]. Long-term follow-up reports of these studies published later showed that the protective effect continued even after years. In the 30th year of the Da Qing study, it was observed that the protection continued to be 30% in the lifestyle change group, 43% in the 7th year of FDPS, 34% in the 10th year of DPP (DPPOS), and 27% in the 15th year [44]. The two main elements targeted by lifestyle change in these studies are 7% weight loss in the first 6 months and at least 150 minutes of moderate-paced exercise (such as brisk walking) per week. Again, calorie targets are planned to be 500-1000 calories below the total calorie intake, taking into account the individual's daily eating habits and body mass index. As an exercise, it was applied minimum 3 days a week and for a total of 150 minutes (provided not to go below 10 minutes), in order to lose an average of 700 kcal per week. In order to achieve metabolic goals, studies have been conducted on a number of nutritional models such as the Mediterranean diet, low-carbohydrate diet, plant-based diet and Dietary Approaches to Stop Hypertension diet, and similar positive results have been obtained in most of them. Again, there are studies showing that technology-assisted lifestyle change programs, whose benefits have been proven through studies based on patient preference, can be effective, especially during the current pandemic period.

This research evaluated the adequacy of various machine learning models for diagnosing diabetes. As a result of the study, it was revealed that each model had similar accuracies. There are different models; K-NN, SVM, Naive Bayes, CNN, Decision Tree, Random Forest and XGBoost were compared in terms of performance and the strengths and weaknesses of each were determined. As a result, the Random Forest model was found to have the highest accuracy rate; However, the XGBoost model performed better than expected in other measurements.

The results show that various machine learning models can be used effectively to diagnose diabetes. However, as each model has its own advantages, it is important to choose the most appropriate model in the context of the application. Future studies may further improve the overall applicability and reliability of these models in larger and more diverse data sets.

6. Conclusion

Machine learning models were used in the study. With these models used, diagnosis and diagnosis of diabetes is made. The performance and analysis of these results were evaluated. Looking at the results, high performance rates were achieved in Random Forest and XGBoost models. When we look at other models, we see that different models stand out in different conditions. When we look at the processes and results in general, it can be seen that machine learning models are effective. An important process has been followed in the diagnosis of diabetes with the models used. The fact that machine learning models show such effective results has been a stepping stone to important developments in the field of health. Early detection and diagnosis of different types of diseases can be made. In this way, the most appropriate treatment processes and practices can be determined. However, future studies on larger and more diverse data sets may further improve the overall applicability and reliability of these models.

References

[1] Fetzer, J. H., & Fetzer, J. H. (1990). What is Artificial Intelligence? (pp. 3-27). Springer Netherlands.

[2] Song, C., Ristenpart, T., & Shmatikov, V. (2017, October). Machine learning models that remember too much. In Proceedings of the 2017 ACM SIGSAC Conference on computer and communications security (pp. 587-601), https://doi.org/10.1145/3133956.3134077.

[3] Egan, A. M., & Dinneen, S. F. (2019). What is diabetes?. Medicine, 47(1), 1-4, https://doi.org/10.1016/j.mpmed.2018.10.002.

[4] Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. Science robotics, 4(37), eaay7120, https://doi.org/10.1126/scirobotics.aay7120.

[5] Nishat, M. M., Faisal, F., Mahbub, M. A., Mahbub, M. H., Islam, S., & Hoque, M. A. (2021). Performance assessment of different machine learning algorithms in predicting diabetes mellitus. Biosc. Biotech. Res. Comm, 14(1), 74-82, http://dx.doi.org/10.21786/bbrc/14.1/10.

[6] Kaleem, H., Liaqat, S., Hassan, M. T., Mehmood, A., Ahmad, U., & Ditta, A. (2022). An Intelligent Healthcare system for detecting diabetes using machine learning algorithms. Lahore Garrison University Research Journal of Computer Science and Information Technology, 6(03), 1-11, https://doi.org/10.54692/lgurjcsit.2022.0603327.

[7] Shabtari, M. M., Shukla, V. K., Singh, H., & Nanda, I. (2021, March). Analyzing PIMA Indian Diabetes Dataset through Data Mining Tool 'RapidMiner'. In 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 560-574). IEEE, https://doi.org/10.1109/ICACITE51222.2021.9404741.

[8] Azbeg, K., Boudhane, M., Ouchetto, O., & Jai Andaloussi, S. (2022). Diabetes emergency cases identification based on a statistical predictive model. Journal of Big Data, 9(1), 1-25, https://doi.org/10.1186/s40537-022-00582-7.

[9] Lukmanto, R. B., Nugroho, A., & Akbar, H. (2019). Early detection of diabetes mellitus using feature selection and fuzzy support vector machine. Procedia Computer Science, 157, 46-54, https://doi.org/10.1016/j.procs.2019.08.140.

[10] Zhu, C., Idemudia, C. U., & Feng, W. (2019). Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. Informatics in Medicine Unlocked, 17, 100179, https://doi.org/10.1016/j.imu.2019.100179.

[11] Naz, H., & Ahuja, S. (2020). Deep learning approach for diabetes prediction using PIMA Indian dataset. Journal of Diabetes & Metabolic Disorders, 19, 391-403, https://doi.org/10.1007/s40200-020-00520-5.

[12] Aamir, Z., & Murtza, I. (2023). Pre-Diabetic Diagnosis from Habitual and Medical Features using Ensemble Classification. Journal of Computing & Biomedical Informatics, 5(01), 283-294, https://doi.org/10.56979/501/2023.

[13] Febrian, M. E., Ferdinan, F. X., Sendani, G. P., Suryanigrum, K. M., & Yunanda, R. (2023). Diabetes prediction using supervised machine learning. Procedia Computer Science, 216, 21-30, https://doi.org/10.1016/j.procs.2022.12.107.

[14] Gollapalli, M., Alansari, A., Alkhorasani, H., Alsubaii, M., Sakloua, R., Alzahrani, R., ... & Albaker, W. (2022). A novel stacking ensemble for detecting three types of diabetes mellitus using a Saudi Arabian dataset: Pre-diabetes, T1DM, and T2DM. Computers in Biology and Medicine, 147, 105757, https://doi.org/10.1016/j.compbiomed.2022.105757.

[15] Dutta, A., Hasan, M. K., Ahmad, M., Awal, M. A., Islam, M. A., Masud, M., & Meshref, H. (2022). Early prediction of diabetes using an ensemble of machine learning models. International Journal of Environmental Research and Public Health, 19(19), 12378, https://doi.org/10.3390/ijerph191912378.

[16] Doğru, A., Buyrukoğlu, S., & Arı, M. (2023). A hybrid super ensemble learning model for the early-stage prediction of diabetes risk. Medical & Biological Engineering & Computing, 61(3), 785-797, https://doi.org/10.1007/s11517-022-02749-z.

[17] Krishnamoorthi, R., Joshi, S., Almarzouki, H. Z., Shukla, P. K., Rizwan, A., Kalpana, C., & Tiwari, B. (2022). A novel diabetes healthcare disease prediction framework using machine learning techniques. Journal of Healthcare Engineering, 2022, https://doi.org/10.1155/2022/1684017.

[18] Panda, M., Mishra, D. P., Patro, S. M., & Salkuti, S. R. (2022). Prediction of diabetes disease using machine learning algorithms. IAES International Journal of Artificial Intelligence, 11(1), 284, http://doi.org/10.11591/ijai.v11.i1.pp284-290.

[19] Theerthagiri, P., Ruby, A. U., & Vidya, J. (2022). Diagnosis and classification of the diabetes using machine learning algorithms. SN Computer Science, 4(1), 72, https://doi.org/10.1007/s42979-022-01485-3.

[20] Mahesh, T. R., Vivek, V., Kumar, V. V., Natarajan, R., Sathya, S., & Kanimozhi, S. (2022, January). A comparative performance analysis of machine learning approaches for the early prediction of diabetes disease. In 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (pp. 1-6). IEEE, https://doi.org/10.1109/ACCAI53970.2022.9752543.

[21] Nazar, M., Alam, M. M., Yafi, E., & Su'ud, M. M. (2021). A systematic review of human–computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques. IEEE Access, 9, 153316-153348, https://doi.org/10.1109/ACCESS.2021.3127881.

[22] Mohanty, A., & Mishra, S. (2022). A comprehensive study of explainable artificial intelligence in healthcare. In Augmented Intelligence in Healthcare: A Pragmatic and Integrated Analysis (pp. 475-502). Singapore: Springer Nature Singapore.

[23] Payrovnaziri, S. N., Chen, Z., Rengifo-Moreno, P., Miller, T., Bian, J., Chen, J. H., ... & He, Z. (2020). Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review. Journal of the American Medical Informatics Association, 27(7), 1173-1185, https://doi.org/10.1093/jamia/ocaa053.

[24] Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., ... & Deveci, M. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. Information Fusion, https://doi.org/10.1016/j.inffus.2023.03.008.
[25] Lauritsen, S. M., Kristensen, M., Olsen, M. V., Larsen, M. S., Lauritsen, K. M., Jørgensen, M. J., ... & Thiesson, B. (2020). Explainable artificial intelligence model to predict acute critical illness from electronic health records. Nature communications, 11(1), 3852, https://doi.org/10.1038/s41467-020-17431-x.
[26] https://www.kaggle.com/datasets/nanditapore/healthcare-diabetes
[27] Song, Y., Huang, J., Zhou, D., Zha, H., & Giles, C. L. (2007, September). Iknn: Informative k-nearest neighbor pattern classification. In European conference on principles of data mining and knowledge discovery (pp. 248-264). Berlin, Heidelberg: Springer Berlin Heidelberg.
[28] Gollapalli, M., Alansari, A., Alkhorasani, H., Alsubaii, M., Sakloua, R., Alzahrani, R., ... & Albaker, W. (2022). A novel stacking ensemble for detecting three types of diabetes mellitus using a Saudi Arabian dataset: Pre-diabetes, T1DM, and T2DM. Computers in Biology and Medicine, 147, 105757, https://doi.org/10.1016/j.compbiomed.2022.105757.
[29] Kadar, J. A., Agustono, D., & Napitupulu, D. (2018). Optimization of candidate selection using naïve Bayes: case study in company X. In Journal of Physics: Conference Series (Vol. 954, No. 1, p. 012028). IOP Publishing, https://doi.org/10.1088/1742-6596/954/1/012028.
[30] Quinlan, J. R. (2014). C4. 5: programs for machine learning. Elsevier.
[31] Hassan, M. M., Mollick, S., & Yasmin, F. (2022). An unsupervised cluster-based feature grouping model for early diabetes detection. Healthcare Analytics, 2, 100112, https://doi.org/10.1016/j.health.2022.100112
[32] Risdin, F., Mondal, P. K., & Hassan, K. M. (2020). Convolutional neural networks (CNN) for detecting fruit information using machine learning techniques. IOSR Journal of Computer Engineering (IOSR-JCE), 22(2), 01-13.
[33] Dong, W., Huang, Y., Lehane, B., & Ma, G. (2020). XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring. Automation in Construction, 114, 103155, https://doi.org/10.1016/j.autcon.2020.103155.
[34] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144), https://doi.org/10.1145/2939672.2939778.
[35] Antwarg, L., Miller, R. M., Shapira, B., & Rokach, L. (2021). Explaining anomalies detected by autoencoders using Shapley Additive Explanations. Expert systems with applications, 186, 115736, https://doi.org/10.1016/j.eswa.2021.115736.
[36] Rastogi, R., & Bansal, M. (2023). Diabetes prediction model using data mining techniques. Measurement: Sensors, 25, 100605, https://doi.org/10.1016/j.measen.2022.100605.
[37] Reddy, S. K., Krishnaveni, T., Nikitha, G., & Vijaykanth, E. (2021, September). Diabetes prediction using different machine learning algorithms. In 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 1261-1265). IEEE, https://doi.org/10.1109/ICCMC.2019.8819841.
[38] Bansal, A., & Singhrova, A. (2021, March). Performance analysis of supervised machine learning algorithms for diabetes and breast cancer dataset. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 137-143). IEEE, https://doi.org/10.1109/ICAIS50930.2021.9396043.
[39] Nass, L., Swift, S., & Al Dallal, A. (2019). Indepth analysis of medical dataset mining: a comparitive analysis on a diabetes dataset before and after preprocessing. KnE Social Sciences, 45-63, https://doi.org/10.18502/kss.v3i25.5190.
[40] Khanam, J. J., & Foo, S. Y. (2021). A comparison of machine learning algorithms for diabetes prediction. Ict Express, 7(4), 432-439, https://doi.org/10.1016/j.icte.2021.02.004.
[41] Pan, X. R., Li, G. W., Hu, Y. H., Wang, J. X., Yang, W. Y., An, Z. X., ... & Howard, B. V. (1997). Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance: the Da Qing IGT and Diabetes Study. Diabetes care, 20(4), 537-544, *https://doi.org/10.2337/diacare.20.4.537*.
[42] Tuomilehto, J., Lindstrom, J., Eriksson, J., Valle, T., Hamalainen, H., Ilanne-Parikka, P., ... & Uusitupa, M. (2001). Finnish Diabetes Prevention Study. Group, 344, 1343-1350.
[43] Knowler, W. C., Barrett-Connor, E., Fowler, S. E., Hamman, R. F., Lachin, J. M., Walker, E. A., Nathan, D. M., & Diabetes Prevention Program Research Group (2002). Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. The New England journal of medicine, 346(6), 393–403, https://doi.org/10.1056/NEJMoa012512.
[44] Herman W. H. (2015). The cost-effectiveness of diabetes prevention: results from the Diabetes Prevention Program and the Diabetes Prevention Program Outcomes Study. Clinical diabetes and endocrinology, 1, 9, https://doi.org/10.1186/s40842-015-0009-1.