Research Article

Leveraging SHAP for Interpretable Diabetes Prediction: A Study of Machine Learning Models on the Pima Indians Diabetes Dataset

Ismail Kırbas and Ahmet Cifci

***Abstract*—This paper investigates the application of machine learning (ML) models for predicting diabetes using the Pima Indians Diabetes Database, with a focus on enhancing model interpretability through the use of SHapley Additive exPlanations (SHAP). The study evaluates eight ML models, including Adaptive Boosting (AdaBoost), k-Nearest Neighbors (k-NN), Logistic Regression (LR), Multi-layer Perceptron (MLP), Naive Bayes (NB), Random Forest (RF), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost), utilizing both test/train split and 10-fold cross-validation methods. The RF model demonstrated superior performance, achieving an accuracy of 82% and an F1-score of 0.83 in the test/train split, and an accuracy of 83% and an F1-score of 0.84 in the 10-fold cross-validation. SHAP analysis was employed to identify the most influential predictors, revealing that glucose, BMI, pregnancies, and insulin levels are the key factors in diabetes prediction, aligning with established clinical markers. Additionally, the use of the Synthetic Minority Over-sampling TEchnique (SMOTE) for class balancing and data scaling contributes to robust model performance. The study emphasizes the necessity for interpretable ML in healthcare, proposing SHAP as a valuable tool for bridging predictive accuracy and clinical transparency in diabetes diagnostics.**

Index Terms—**Diabetes Prediction, Explainable Artificial Intelligence, Machine Learning Models, Model Interpretability, SHapley Additive exPlanation.**

I. Introduction

DIABETES IS a chronic and increasingly prevalent condition with profound implications for public health worldwide [1-3]. According to recent estimates, the global prevalence of diabetes among adults continues to rise, creating a significant burden on healthcare systems and underscoring the urgent need for effective preventive and diagnostic tools [4]. Early prediction of diabetes can enable timely interventions, reducing the likelihood of severe complications, improving patient outcomes, and potentially decreasing healthcare costs. ML models have shown promise in predicting diabetes by identifying patterns within clinical and demographic data, facilitating early detection [5-9]. However, despite advances in predictive accuracy, traditional ML models often lack interpretability, which is a critical limitation in clinical settings where transparency is paramount.

The interpretability of a model is especially crucial in healthcare, as it provides clinicians with insights into the decision-making process, enhances trust in model predictions, and supports more informed and individualized patient care. Conventional ML models, such as neural networks and ensemble methods, typically operate as "black boxes," yielding high predictive accuracy but offering limited understanding of how predictions are derived [10, 11]. This opacity creates challenges in clinical applications, as healthcare providers require an explanation of model decisions to comply with ethical standards, support diagnostic conclusions, and facilitate shared decision-making with patients. The field of XAI aims to address these challenges by developing methods that enhance the transparency and interpretability of ML models, making them more suitable for sensitive applications such as diabetes prediction [12-14].

One promising XAI method is SHAP [15], which assigns importance values to each feature in a model's prediction process, helping to elucidate the contribution of specific patient characteristics to the overall prediction. SHAP is based on Shapley values, a concept from cooperative game theory [16], and provides consistent, theoretically grounded explanations that allow clinicians to understand which features most strongly influence the likelihood of diabetes in individual cases. By incorporating SHAP, healthcare providers can make more confident decisions, potentially identifying high-risk patients based on meaningful patterns in data [17].

The utilization of ML models for diabetes prediction has garnered significant attention in research, driven by the rising global prevalence of diabetes and the pressing need for early diagnostic solutions. A substantial body of studies has concentrated on evaluating and comparing various ML algorithms, with a particular emphasis on the Pima Indians Diabetes Database [18]. Verma and Khatoon [19] compared LR, SVM, k-NN, and RF models, identifying RF as the best

İsmail Kırbaş, is with Department of Computer Engineering Burdur Mehmet Akif Ersoy University, Burdur, Türkiye, (e-mail: ismailkirbas@mehmetakif.edu.tr).
https://orcid.org/0000-0002-1206-8294

Ahmet Çifci, is with Department of Electrical-Electronics Engineering Burdur Mehmet Akif Ersoy University, Burdur, Türkiye, (e-mail: acifci@mehmetakif.edu.tr).
https://orcid.org/0000-0001-7679-9945

Manuscript received Nov 1, 2024; accepted Dec 27, 2025.
DOI: 10.17694/bajece.1577929

performer with an accuracy of 80.08%. Similarly, Xie [20] demonstrated that LR slightly outperformed RF and SVM, achieving a prediction accuracy of 79.13%. Chang et al. [21] emphasized the importance of interpretable models in the context of the Internet of Medical Things (IoMT), exploring NB, RF, and J48 Decision Tree (DT) models. RF was particularly effective for datasets with more features, while NB excelled in simpler configurations. Sahoo et al. [22] conducted a comparative analysis of supervised classification algorithms, including LR, SVM, and RF. Their study highlighted LR and DT classifiers as achieving the highest accuracy, demonstrating their suitability for diabetes prediction tasks. Similarly, You and Kang [23] identified glucose, BMI, and age as the most significant predictors using correlation analysis and employed SVM and DT models, achieving an accuracy of 70%. Ashour et al. [24] evaluated feedforward neural networks (FNN) and convolutional neural networks (CNN), with the former achieving the highest accuracy of 82%. Akyol and Şen [25] examined ensemble learning methods such as AdaBoost, Gradient Boosted Trees, and RF, reporting that AdaBoost combined with stability selection achieved the best accuracy of 73.88%. Reza et al. [26] achieved the highest accuracy of 79.33% using the RF model for the Pima Indian Diabetes Dataset, highlighting its effectiveness in diabetes classification. Pyne and Chakraborty [27] implemented an artificial neural network (ANN) without feature extraction, achieving a classification accuracy of 80.79%.

Efficient preprocessing plays a critical role in enhancing model performance. Jain et al. [28] analyzed imputation techniques such as Multivariate Imputation by Chained Equations (MICE), k-NN, and mean/mode replacement, finding that k-NN based imputation improved the predictive accuracy of RF models. Karatsiolis and Schizas [29] proposed a region-based SVM approach that integrated clustering and kernel selection, achieving an accuracy of 82.2%.

While previous studies have primarily focused on predictive accuracy and have implemented only a limited selection of ML classifiers, they often lack the integration of XAI methods necessary for clinical application. This study addresses this gap by systematically integrating SHAP into the classification process, thereby enhancing interpretability without sacrificing accuracy. Our objective is to improve transparency in diabetes prediction models, making them more useful for healthcare providers and ultimately contributing to better patient care through informed decision-making.

In this study, we employ a variety of ML models, including AdaBoost, k-NN, LR, MLP, NB, RF, SVM, and XGBoost. These models were selected for their diverse characteristics, which range from linear to non-linear and from probabilistic to tree-based approaches. The diversity of model types allows for a comprehensive evaluation of classification performance and interpretability when XAI methods are applied. Furthermore, we utilize the Pima Indians Diabetes Database [18], a widely referenced dataset in diabetes prediction research, which includes relevant clinical and demographic variables.

II. Materials and Method

This section provides a comprehensive explanation of the dataset, the application of SHAP for model interpretability, the performance evaluation metrics utilized to assess the models, and the ML algorithms implemented in this study.

Fig. 1 outlines the process for predicting diabetes using ML models. It begins with data collection, where the Pima Indians Diabetes Database is utilized as the dataset. This is followed by data normalization, ensuring all features are scaled to a consistent range for unbiased model training. Next, class balancing is performed, employing techniques like SMOTE to address any class imbalances in the dataset, such as unequal representation of diabetic and non-diabetic cases. In the test-train splitting step, the data is divided into training and testing sets (80% training, 20% testing) to facilitate model evaluation. For enhanced reliability, 10-fold cross validation is applied, splitting the training data into ten subsets to train and validate the model iteratively. During model training, ML algorithms are used to learn patterns from the training data. The trained model's effectiveness is then assessed through performance evaluation, using metrics like accuracy, precision, recall, and F1-score. Finally, SHAP analysis is conducted to interpret the model's decision-making process, identifying the relative importance of features in predicting diabetes.

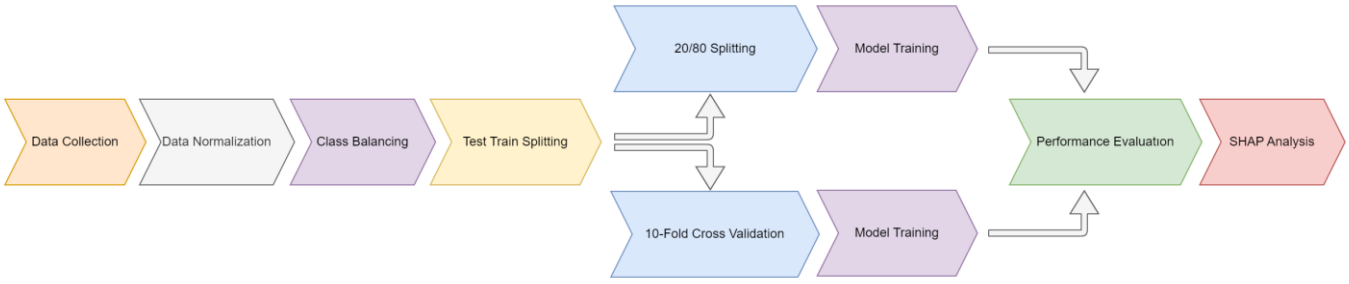


Fig.1. Workflow of the proposed system

A. Dataset

The dataset employed in this study is the Pima Indians Diabetes Database [18], sourced from the National Institute of

Diabetes and Digestive and Kidney Diseases repository. This dataset is frequently utilized in diabetes research due to its comprehensive inclusion of health-related attributes that are predictive of diabetes onset. It is widely available for research and has become a standard benchmark in diabetes prediction modeling studies.

The Pima Indians Diabetes Database comprises 768 instances, each representing an individual of Pima Indian heritage who is 21 years or older. The dataset includes a total of eight predictive features, each capturing an essential physiological or clinical variable. These features include:

1. Number of pregnancies
2. Plasma glucose concentration (measured two hours post-oral glucose tolerance test)
3. Diastolic blood pressure (mm Hg)
4. Triceps skinfold thickness (mm)
5. Serum insulin level (μU/mL)
6. Body mass index (BMI) (weight in kg/height in m^2)
7. Diabetes pedigree function (a score representing the genetic predisposition to diabetes)
8. Age (years)

In addition to the eight features, there is a binary target variable, representing whether a subject is diagnosed with diabetes (1) or not (0). The collection of these variables provides a multidimensional view of the factors associated with diabetes risk, facilitating the development of predictive models in epidemiological studies.

The data was initially collected as part of a longitudinal study aimed at understanding the prevalence of diabetes and its related risk factors within the Pima Indian population, which has a historically high prevalence of Type 2 diabetes. Variables were measured through clinical tests and self-reported metrics under controlled conditions, ensuring consistency and reliability in the dataset. This database remains a valuable resource for diabetes research, especially in exploring predictive analytics and the relationship between physiological markers and diabetes onset.

The dataset contains no null or missing values. However, based on domain knowledge [21], certain features—blood pressure, BMI, glucose, insulin, and skin thickness—have inconsistent values. Specifically, zero values for these features are inaccurate as they fall outside the normal range (see Table 1).

Table 1 presents the descriptive statistics of eight features used in the classification task. The descriptive statistics for the eight features in the diabetes dataset provide a comprehensive view of each variable's central tendency, variability, and range. The age feature has a mean of 33.24 years, with a mode of 22, suggesting a concentration of younger individuals. The median age of 29, combined with a dispersion of 0.35, indicates a relatively balanced distribution, covering a broad age range from 21 to 81. Blood pressure shows a mean of 69.11 mm Hg, with central measures closely aligned (mode of 70 and median of 72), suggesting symmetry in the distribution. However, the minimum value of 0 may indicate missing or erroneous data, given that blood pressure values typically exceed zero, and the maximum value of 122 reflects a wide range, with a dispersion of 0.28. For BMI, the mean and mode both stand at approximately 32 kg/m^2, suggesting a balanced distribution, yet the minimum of 0 and maximum of 67.1 kg/m^2 indicate considerable variability, which might signal the need for data cleaning or further investigation, given its dispersion of 0.25. The diabetes pedigree function, with a mean of 0.47, mode of 0.25, and median of 0.37, reveals significant variability, indicated by a high dispersion of 0.70 and a range from 0.08 to 2.42. This variability likely reflects diverse genetic or familial risks for diabetes across the dataset. Glucose levels have a mean of 120.89 mg/dL and display moderate dispersion (0.26), with values spanning from 0 to 199 mg/dL. A zero glucose value may point to missing data, as this measure is generally non-zero. Insulin levels, with a mean of 79.80 mu U/ml, show a high dispersion of 1.44, and values range widely from 0 to 846, suggesting individual differences or data collection issues given the unusually high mode of 0. Pregnancies exhibit a mean of 3.85 and a mode of 1, showing a right-skewed distribution with a dispersion of 0.88 and a range from 0 to 17 pregnancies, aligning with typical reproductive variability. Lastly, Skin thickness has a mean of 20.54 mm and shows a right-skewed distribution with zeros appearing as the mode, likely indicating missing or incomplete data, and a median of 23. The range extends from 0 to 99 mm, with a relatively high dispersion of 0.78.

1) Data scaling

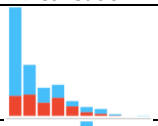
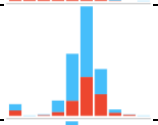
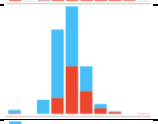
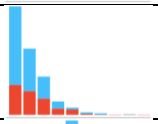
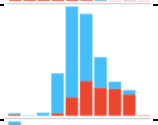
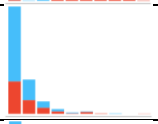
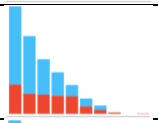
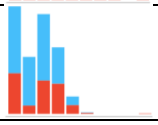
In ML, scaling is a crucial preprocessing step that standardizes the range of independent features to ensure consistent and effective model performance. Data often contains features with varying ranges and units, which can introduce bias during the training process, particularly when models rely on distance calculations, such as in k-NN and SVM [30]. By normalizing or standardizing features, scaling brings data into a uniform range, reducing the influence of features with larger numerical values and ensuring that each feature contributes equitably to the model. This step is especially beneficial in gradient-based algorithms, where unscaled data may lead to suboptimal convergence or slower learning as the model becomes prone to oscillating toward larger-scale features. Beyond enhancing model efficiency, scaling offers several key advantages, such as improved algorithm accuracy and increased computational speed. Models trained on scaled data demonstrate better generalization and tend to avoid overfitting by reducing variance related to magnitude differences between features [31]. As ML solutions are increasingly applied to varied datasets across domains, implementing scaling is vital for achieving reliable, reproducible, and high-performance models.

2) Class balancing

In this study, the Pima Indian Diabetes dataset, characterized by two classes—diabetic and non-diabetic individuals—exhibits a marked class imbalance, with a significantly higher number of non-diabetic cases relative to diabetic ones. Class imbalance is a critical issue in ML and statistical modeling, as it can lead to biased models that disproportionately favor the majority class, consequently compromising the predictive performance, especially for the minority class [32]. To mitigate this imbalance and enhance model efficacy, we employed SMOTE [33]. SMOTE is a sophisticated resampling method that generates synthetic examples in the minority class by interpolating between existing samples, thus balancing the dataset without simply duplicating minority instances. By

equalizing the sample size across classes, SMOTE promotes a more representative learning process, enabling the model to better capture patterns pertinent to both diabetic and non-diabetic individuals. This process not only improves classification accuracy but also ensures that the model's performance is more robust and reliable, particularly in real-world applications where balanced prediction accuracy across classes is crucial.

TABLE I
DESCRIPTIVE STATISTICAL VALUE OF THE DATASET

Feature	Distribution	Mean	Mode	Median	Dispersion	Min.	Max.
Age		33.24	22	29	0.35	21	81
Blood pressure		69.11	70	72	0.28	0	122
Body mass index (BMI)		31.99	32	32	0.25	0	67.1
Diabetes pedigree function		0.47	0.25	0.37	0.70	0.08	2.42
Glucose		120.89	99	117	0.26	0	199
Insulin		79.80	0	30.50	1.44	0	846
Pregnancies		3.85	1	3	0.88	0	17
Skin thickness		20.54	0	23	0.78	0	99

B. Machine Learning Models

To predict diabetes diagnoses, we evaluated multiple ML classification models, each with distinct principles and methodologies tailored to enhance predictive performance and interpretability. Below, we detail the models considered for this study.

Adaptive Boosting (AdaBoost) is a ML algorithm that combines multiple weak learners to create a strong learner [34]. It works by iteratively weighting the training data, giving more weight to misclassified instances in each iteration. This process results in a final model that is more accurate and robust than any individual weak learner.

k-Nearest Neighbors (k-NN) works by measuring the distances between a test data instance and all instances in the training dataset. It then identifies the k nearest training instances to classify the test instance [35]. The model is advantageous in scenarios with well-defined clusters and is non-parametric, requiring minimal assumptions, making it adaptable for varying diabetes-related datasets.

Logistic Regression (LR) is based on a statistical model that predicts binary outcomes by estimating probabilities through a logistic function [36]. In the context of diabetes prediction, LR is favored for its simplicity and interpretability, particularly in assessing linear relationships between predictors and the likelihood of disease presence.

Multi-Layer Perceptron (MLP) is a neural network model that consists of multiple layers of interconnected nodes, where each node represents a neuron [37]. MLPs are characterized by their ability to capture non-linear relationships in data through backpropagation and activation functions. This model is advantageous for its flexibility in learning complex patterns, which is beneficial when diagnosing diabetes based on various patient features.

Naive Bayes (NB) assumes independence among predictor features and calculates the probability of class membership using Bayes' theorem [38]. Despite the simplicity of this

independence assumption, NB often performs well in medical contexts where conditional probabilities are informative, thus providing a quick and computationally efficient option for diabetes prediction.

Random Forest (RF) enhances the decision tree method by constructing an ensemble of multiple trees, each trained on different data subsets and feature splits [39]. This model improves generalization and reduces overfitting, making it robust for the variability present in medical data, such as diverse patient demographics and health indicators relevant to diabetes.

Support Vector Machine (SVM) attempts to find an optimal hyperplane that maximizes the margin between classes [40], effectively separating diabetic and non-diabetic instances in high-dimensional spaces. SVM is particularly suited for datasets where feature dimensions are high, offering strong performance with appropriate kernel selection, especially for complex, non-linear decision boundaries.

eXtreme Gradient Boosting (XGBoost) optimizes the Gradient Boosting algorithm with techniques such as regularization, parallelization, and efficient handling of sparse data [41]. It has demonstrated success in improving both speed and accuracy, which is beneficial in a diabetes diagnosis setting where quick, reliable predictions are essential for timely patient interventions.

Each of these models was selected for its unique properties, strengths, and applicability to the problem of diabetes diagnosis, providing a comprehensive approach to exploring the predictive capacity of different ML techniques.

C. SHapley Additive exPlanation (SHAP)

SHAP is a model-agnostic interpretability approach rooted in cooperative game theory [16]. SHAP aims to enhance model transparency by assigning importance scores to features based on their contribution to prediction, enabling researchers to gain insights into the influence of each feature. By calculating Shapley values, SHAP helps to decompose the model output in a way that considers all possible feature interactions, making it a robust tool for feature interpretability in complex ML models [15]. This subsection provides a detailed methodology on two SHAP visualizations: SHAP Feature Importance (meanSHAP) and SHAP Summary Plot (Beeswarm Plot), each of which plays a distinct role in elucidating model behavior.

1) SHAP feature importance (meanSHAP)

SHAP feature importance, commonly expressed as meanSHAP, quantifies the average effect of each feature on the model output by calculating the mean of the absolute SHAP values for each feature across all instances in the dataset. Mathematically, for a given feature f, the mean SHAP value is computed as [42]:

$$meanSHAP(f) = \frac{1}{N}\sum_{i=1}^{N}\left|\emptyset_{f,i}\right| \quad (1)$$

where N represents the number of instances, and $\phi_{f,i}$ denotes the SHAP value for feature f in instance i. This aggregation of absolute SHAP values provides a singular, intuitive metric that ranks features by their average importance, capturing both the magnitude and frequency of their impact on model predictions.

The meanSHAP metric serves as a foundational interpretability measure, offering a clear, quantitative assessment of feature relevance. By focusing on absolute values, meanSHAP accounts for both positive and negative contributions of each feature, facilitating a comprehensive view of feature importance. Unlike traditional importance measures that may overlook feature interactions or nonlinear effects, meanSHAP is based on Shapley values, which incorporate the complete range of feature interdependencies, thus offering a reliable and interpretable importance ranking that aligns closely with model behavior [15, 43].

2) SHAP summary plot (beeswarm plot)

The SHAP summary plot, also referred to as the beeswarm plot, visually represents the distribution of SHAP values across all instances for each feature. In this plot, each feature is displayed along the vertical axis, while the horizontal axis represents the range of SHAP values, indicating the magnitude and direction of feature impact on model predictions. Each point on the plot corresponds to the SHAP value for an individual instance, with the points color-coded to represent the feature values, typically on a blue-to-red gradient, where red denotes higher feature values and blue denotes lower ones. The beeswarm plot is particularly useful in examining how each feature affects the model output, providing insights into the distribution of feature impact. For instance, a wide horizontal spread of points suggests that a feature has variable importance across instances, while clustering around zero indicates minimal influence. Additionally, by examining the color gradients, researchers can infer the relationship between feature values and their corresponding SHAP values, revealing patterns such as whether higher feature values lead to increased or decreased predictions [15, 44].

D. Performance Evaluation Metrics

Evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix are employed to assess the models. These statistical measures are derived from ground truth values, namely True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). The calculations for accuracy, precision, recall, and F1-score can be found in Eqs. (2), (3), (4), and (5), respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - score = \frac{2\ x\ Precision\ x\ Recall}{Precision\ +\ Recall} \quad (5)$$

These variables are calculated using the confusion matrix, which is a tabular representation showing the values of the actual outcome classes and the predicted outcome classes on the testing dataset as shown in Fig. 2.

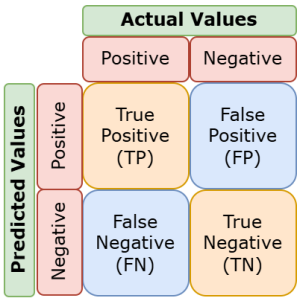


Fig. 2. Confusion matrix

III. Results

This section provides a comprehensive analysis of the study's findings, including the evaluation of eight ML models. Additionally, the application of XAI method, specifically SHAP, is detailed to elucidate the models' internal mechanisms and decision-making processes, enhancing interpretability and transparency. To evaluate model performance, the dataset was divided into training (80%) and testing (20%) sets and applied a 10-fold cross-validation.

Table 2 presents confusion matrices for several ML models tested to predict diabetes in women of Pima Indian heritage. Each confusion matrix displays the performance of a specific model in classifying individuals as either having diabetes (positive) or not having diabetes (negative). The rows of the matrix represent the true class (actual diabetes status), while the columns represent the predicted class assigned by the model.

TABLE II
CONFUSION MATRICES OF ML MODELS USING TEST/TRAIN SPLIT

AdaBoost

Actual		*Predicted* Diabetic	*Predicted* Non-Diabetic
	Diabetic	70	29
	Non-Diabetic	21	80

NB

Actual		*Predicted* Diabetic	*Predicted* Non-Diabetic
	Diabetic	79	20
	Non-Diabetic	26	75

k-NN

Actual		*Predicted* Diabetic	*Predicted* Non-Diabetic
	Diabetic	75	27
	Non-Diabetic	12	89

RF

Actual		*Predicted* Diabetic	*Predicted* Non-Diabetic
	Diabetic	76	23
	Non-Diabetic	13	88

LR

Actual		*Predicted* Diabetic	*Predicted* Non-Diabetic
	Diabetic	72	27
	Non-Diabetic	23	78

SVM

Actual		*Predicted* Diabetic	*Predicted* Non-Diabetic
	Diabetic	71	28
	Non-Diabetic	12	89

MLP

Actual		*Predicted* Diabetic	*Predicted* Non-Diabetic
	Diabetic	68	31
	Non-Diabetic	14	87

XGBoost

Actual		*Predicted* Diabetic	*Predicted* Non-Diabetic
	Diabetic	71	28
	Non-Diabetic	17	84

In Table 2, the AdaBoost algorithm demonstrates a balanced performance in terms of sensitivity and specificity, producing 70 true positives (TP) and 80 true negatives (TN), with 21 false positives (FP) and 29 false negatives (FN). The NB model performs well in identifying diabetic individuals with 79 true positives, though it shows a slight decline in specificity with 75 true negatives. The k-NN algorithm exhibits high specificity, achieving 89 true negatives and only 12 false positives, but relatively lower sensitivity with 27 false negatives. The RF model delivers balanced performance, achieving 76 true positives and 88 true negatives. LR provides an acceptable balance between sensitivity and specificity, with 72 true positives and 23 false positives, but yields 27 false negatives, indicating limitations in identifying diabetic cases. The SVM model shows a comparable performance to LR but stands out with higher specificity, achieving 89 true negatives. The MLP model, with 68 true positives and 87 true negatives, demonstrates slightly lower sensitivity and yields 31 false negatives. Finally, the XGBoost model strikes a balance between sensitivity and specificity, achieving 71 true positives and 84 true negatives.

Table 3 provides a performance comparison of eight ML models based on accuracy, precision, recall, and F1-score using two approaches: the test/train split and 10-fold cross-validation.

For the test/train split (20/80), the RF model stands out with the highest performance metrics: an accuracy of 0.82, precision of 0.82, recall of 0.82, and F1-score of 0.83. This indicates that RF is particularly adept at capturing complex, non-linear relationships in the dataset, resulting in balanced and robust predictions. Following RF, the k-NN model demonstrates competitive performance, achieving an accuracy of 0.81, precision of 0.81, recall of 0.80, and F1-score of 0.82. Similarly, SVM also shows strong results with an accuracy of 0.80, precision of 0.81, recall of 0.80, and F1-score of 0.82, indicating its effectiveness in handling high-dimensional data and distinguishing between diabetic and non-diabetic cases. In contrast, models such as AdaBoost, LR, and NB exhibit moderate performance, with accuracy and F1-scores ranging between 0.75 and 0.77. While these models offer interpretability and computational efficiency, their lower sensitivity suggests potential limitations in identifying diabetic cases.

TABLE III
COMPARATIVE PERFORMANCE ANALYSIS OF ML MODELS FOR TEST/TRAIN AND CROSS-VALIDATION

Data Splitting	Model	Accuracy	Precision	Recall	F1-score
Test/Train (20/80)	AdaBoost	0.75	0.75	0.75	0.76
	k-NN	0.81	0.81	0.80	0.82
	LR	0.75	0.75	0.75	0.76
	MLP	0.78	0.78	0.77	0.79
	NB	0.77	0.77	0.77	0.77
	RF	**0.82**	**0.82**	**0.82**	**0.83**
	SVM	0.80	0.81	0.80	0.82
	XGBoost	0.78	0.78	0.77	0.79
Cross-validation (10-fold)	AdaBoost	0.79	0.79	0.79	0.79
	k-NN	0.81	0.82	0.81	0.82
	LR	0.76	0.76	0.76	0.75
	MLP	0.80	0.80	0.80	0.81
	NB	0.74	0.74	0.74	0.72
	RF	**0.83**	**0.83**	**0.83**	**0.84**
	SVM	0.79	0.79	0.79	0.80
	XGBoost	0.81	0.81	0.81	0.81

Cross-validation results further reinforce RF's superior performance, with an accuracy of 0.83, precision of 0.83, recall of 0.83, and F1-score of 0.84. The consistency of RF's performance across both evaluation methods underscores its reliability and robustness in diabetes prediction. The k-NN model also performs exceptionally well, with an accuracy of 0.81, precision of 0.82, recall of 0.81, and F1-score of 0.82. Its strong results are indicative of its effectiveness in leveraging the local relationships among data points, particularly after the dataset has been balanced and scaled. The XGBoost model achieves an accuracy of 0.81 and maintains precision, recall, and F1-scores at 0.81 as well. The MLP model achieves comparable results, with accuracy, precision, and recall scores of 0.80, and an F1-score of 0.81. The slight improvement in F1-score compared to the test/train split suggests that MLP benefits from the diversified training subsets in cross-validation, allowing it to better generalize its predictions. The SVM model also performs well, achieving an accuracy and precision of 0.79, with recall and F1-scores matching at 0.79 and 0.80, respectively. SVM's performance underscores its ability to construct decision boundaries effectively, especially in high-dimensional spaces, though its metrics are slightly lower compared to RF and k-NN. The AdaBoost model demonstrates moderate performance during cross-validation, with accuracy, precision, recall, and F1-scores all at 0.79. While its performance is slightly better than that in the test/train split, it still lags behind ensemble methods like RF and XGBoost in capturing the dataset's complexity. The LR model maintains consistent but relatively lower results compared to more advanced models, achieving an accuracy and precision of 0.76, recall of 0.76, and an F1-score of 0.75. Its simplicity and interpretability remain its key advantages, though its limited ability to handle non-linear relationships constrains its performance. The NB model exhibits the weakest performance among all models, with accuracy, precision, and recall scores at 0.74, and an F1-score of 0.72. Its simplistic assumption of feature independence may not align well with the real-world interactions within the dataset, leading to suboptimal results in distinguishing diabetic and non-diabetic cases.

Figs. 3 and 4 visually compare the performance of various ML models under two distinct evaluation methods: the test/train split and 10-fold cross-validation.

Fig. 3 provides a comparative visual analysis, showing that models such as RF, k-NN, and SVM consistently achieve superior performance across metrics, demonstrating their robustness and suitability for diabetes prediction. Simpler models, like LR and NB, exhibit moderate performance, which, while computationally efficient, reveal limitations in handling the dataset's complexity.

Fig. 4, on the other hand, presents the performance results obtained through 10-fold cross-validation. The results further validate the consistency and robustness of RF and k-NN models, which maintain high accuracy and F1-scores, underscoring their reliability in diabetes prediction tasks. Additionally, cross-validation highlights improvements in performance for models such as AdaBoost, MLP, and XGBoost.

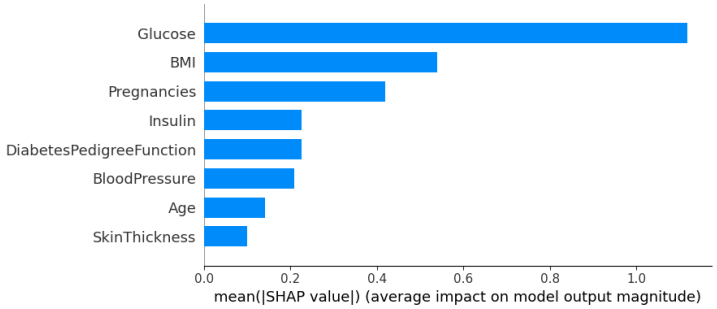


Fig. 5. Importance ranking of the model prediction features

The ranking indicates that glucose level is the most critical feature, suggesting that higher glucose levels strongly correlate with diabetes likelihood. This aligns with clinical expectations, as elevated glucose is a primary indicator of diabetes. BMI ranks as the second most important feature, highlighting the strong association between obesity and diabetes risk. This finding underscores the importance of body weight relative to

height in assessing metabolic health. Pregnancy is also identified as a critical factor, likely due to the increased metabolic stress during pregnancy, which can elevate the risk of developing diabetes. Insulin levels demonstrate a considerable impact on the model's predictions, capturing the intricate relationship between insulin metabolism and diabetes. Less influential features include diabetes pedigree function, blood pressure, age, and skin thickness, although they still contribute to the predictive model. The relatively lower importance of these features may indicate that while they provide valuable context, they are not as directly correlated with diabetes onset as glucose levels and BMI.

Fig. 6. is a visual representation that shows the importance of various features in predicting diabetes, using SHAP values. In this plot, each feature is displayed on the y-axis, ordered by its significance in the prediction, with the SHAP value distribution across instances plotted horizontally on the x-axis. The SHAP values indicate the magnitude and direction (positive or negative impact) of each feature on the model's prediction outcome for diabetes risk. Each dot represents an individual instance, with the color gradient (from blue to red) corresponding to the feature value's magnitude, where red signifies higher values and blue lower ones.

Glucose often appears as the most significant predictor in diabetes-related models. Higher glucose levels, represented by red-colored points on the positive side of the SHAP values, usually increase the prediction probability for diabetes, reflecting the well-established clinical link between elevated blood glucose and diabetes risk. BMI, which indicates body weight relative to height, typically ranks high in importance. Higher BMI values (marked in red) are often associated with a higher probability of diabetes due to the strong association between obesity and diabetes risk. Pregnancies, which represent the number of times a patient has been pregnant, play an important role. Higher values (red) generally increase the SHAP values, indicating a higher likelihood of diabetes, potentially due to physiological changes associated with multiple pregnancies. Insulin demonstrates a moderate level of importance. Higher insulin levels (red) are positively associated with increased diabetes prediction scores, reflecting the body's compensatory response to insulin resistance. Lower values (blue), however, have a varying impact, suggesting a more complex relationship.

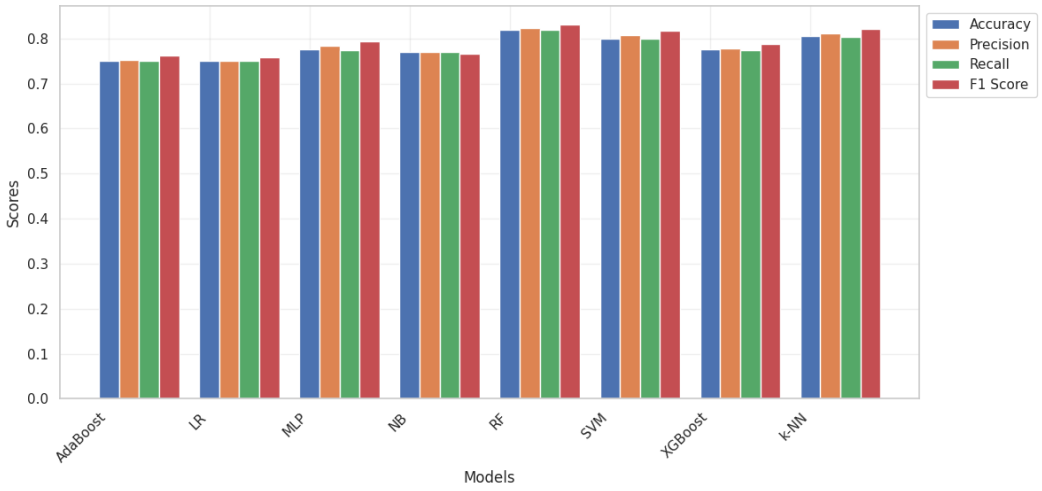


Fig. 3. Visualization of the performance metrics results for ML models using test/train split

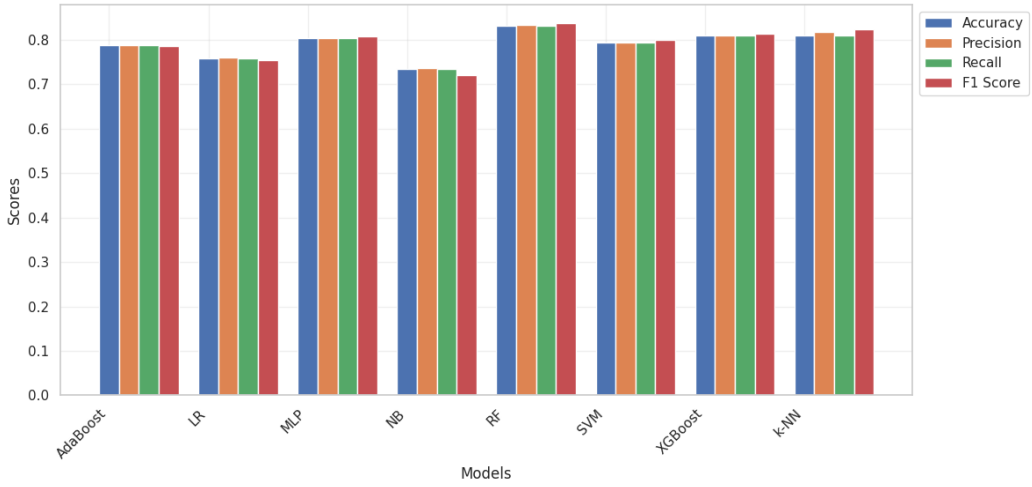


Fig. 4. Visualization of the performance metrics results for ML models using 10-fold cross-validation

RF was selected for SHAP analysis due to its superior predictive performance. Fig. 5 illustrates the significance of

various features in predicting diabetes using SHAP values. In this plot, features are ranked based on their average SHAP values, highlighting their relative impact on the model's predictions.

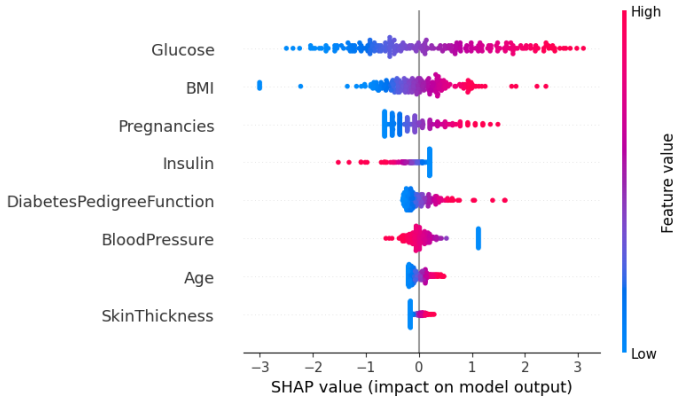


Fig. 6. SHAP summary plot

Diabetes pedigree function, blood pressure and age also contribute to the model, albeit with less impact compared to glucose and BMI. Diabetes pedigree function quantifies the genetic predisposition to diabetes. High values (red), indicating a stronger family history, consistently increase the SHAP values, reinforcing the heritability of diabetes risk. Low values (blue) contribute negatively, reducing the prediction probability. Older individuals (red dots for age) tend to show a higher likelihood of diabetes due to age-related declines in metabolic health, while lower ages (blue) reduce the risk. Similarly, higher blood pressure values (red) are modestly associated with increased predictions, consistent with the link between hypertension and diabetes risk. Skin thickness, the least impactful feature, shows a nuanced pattern.

IV. Discussions

This study reveals critical insights into the application of ML models for predicting diabetes, particularly highlighting the efficacy of ensemble methods and the importance of model interpretability. The most significant finding is the superior performance of the RF model, which achieved an accuracy of 82% and an F1-score of 0.83 in the test/train split evaluation, and an even higher accuracy of 83% and an F1-score of 0.84 in the 10-fold cross-validation (Table 3, Fig. 3, and Fig. 4). These results underscore RF's ability to capture complex, non-linear relationships within the Pima Indians Diabetes Database. Furthermore, the SHAP analysis identified glucose, BMI, pregnancies, and insulin as the most influential predictors, aligning with established clinical markers of diabetes (Fig. 5 and Fig. 6). The prominence of glucose levels, as indicated by the highest mean SHAP value, reinforces its well-documented role as a primary indicator of diabetes risk, while the importance of BMI reflects the known association between obesity and metabolic disorders.

TABLE IV
COMPARISON OF DIABETES PREDICTION STUDIES

Study	Models	Best Model	Accuracy
Verma and Khatoon [19]	LR, SVM, k-NN, RF	RF	80.08%
Xie [20]	k-NN, LR, SVM, RF	LR	79.13%
Chang et al. [21]	NB, RF, J48 DT	RF	79.57%
Sahoo et al. [22]	NB, LR, DT, RF, SVM, XGBoost	LR	74.03%
You and Kang [23]	SVM, DT	SVM	70.40%
Ashour et al. [24]	FNN, CNN	FNN	82%
Akyol and Şen [25]	AdaBoost, Gradient Boosted Trees, RF	AdaBoost	73.88%
Reza et al. [26]	Stacking Ensemble (Classical + Deep)	Stacking Ensemble (Deep NN)	75.03% (train/test), 77.10% (5-fold cross-validation)
Pyne and Chakraborty [27]	ANN	ANN	80.79%
Jain et al. [28]	DT, RF, SVM, NB	RF	79.08%
Karatsiolis and Schizas [29]	Modified SVM with RBF and Polynomial Kernel	Modified SVM	82.2%
This Study	AdaBoost, k-NN, LR, MLP, NB, RF, SVM, XGBoost	RF	82% (Train/Test), 83% (10-fold cross-validation)

The findings of this study contribute to a growing body of literature that evaluates ML models for diabetes prediction, as summarized in Table 4. Our results demonstrate that the RF model, particularly when combined with SHAP analysis, outperforms previously reported as applied to the same dataset. For instance, Akyol and Şen [25] reported an accuracy of 73.88% using AdaBoost, while Verma and Khatoon [19] achieved 80.08% accuracy with RF. Our study's RF model surpasses these benchmarks, achieving 83% accuracy in the 10-fold cross-validation. This improvement can be attributed to our

use of the SMOTE for class balancing and the integration of SHAP values for enhanced interpretability. However, our results are comparable to those of Xie [20], who reported 79.13% accuracy with LR, and Jain et al. [28], who achieved 79.08% accuracy with Random Forest. These variations highlight the influence of different preprocessing techniques and model configurations on predictive performance.

A key strength of this study lies in its integration of XAI methods, specifically SHAP, to enhance model interpretability. While many previous studies have focused on predictive accuracy, the clinical applicability of ML models hinges on their transparency and the ability to provide actionable insights. By incorporating SHAP values, we provide a clear, quantitative assessment of feature importance, bridging the gap between predictive accuracy and clinical utility. This approach not only elucidates the model's decision-making process but also builds trust among healthcare providers by offering a deeper understanding of the factors driving predictions. Additionally, the use of SMOTE to address class imbalance ensures that the models are trained on a representative dataset, thereby enhancing their robustness and reliability in real-world scenarios.

Despite the strengths, this study has certain limitations. The reliance on the Pima Indians Diabetes Database, while a widely used benchmark, may introduce biases related to the specific population studied, potentially limiting the generalizability of the findings to other ethnic groups. Additionally, the study identified inconsistencies in the dataset, particularly in insulin, skin thickness, and blood pressure values, which could affect model performance. Although SMOTE was employed to mitigate class imbalance, the inherent limitations of the dataset cannot be entirely overcome. Furthermore, while the SHAP analysis enhances interpretability, it is essential to acknowledge that model interpretability is a complex and evolving field, and the explanations provided by SHAP, while valuable, may not fully capture the intricate decision-making processes of the models.

This study makes a significant contribution to the field of diabetes prediction by demonstrating the effectiveness of advanced ML models, particularly RF and k-NN, and by enhancing model interpretability through SHAP analysis. The findings underscore the importance of integrating XAI methods in healthcare applications to foster trust and facilitate clinical adoption. Future research should focus on validating these models with more diverse datasets and refining feature engineering to address the identified inconsistencies. Additionally, exploring the integration of other XAI techniques and investigating the longitudinal performance of these models in real-world clinical settings could further enhance their applicability. The insights gained from this study pave the way for developing more transparent, reliable, and clinically relevant predictive tools for diabetes, ultimately contributing to improved patient outcomes and more effective healthcare strategies. The study's findings open new research avenues, particularly in the development of personalized medicine approaches, where individual risk factors can be evaluated with greater precision and transparency.

V. Conclusion

This study demonstrates the successful integration of ML models and XAI techniques to enhance the predictive accuracy and interpretability of diabetes diagnosis using the Pima Indians Diabetes Database. The RF model emerged as the most effective classifier, achieving an accuracy of 83% and an F1-score of 0.84 in 10-fold cross-validation, underscoring its capability to model complex, non-linear relationships within the dataset. The incorporation of SHAP values provided critical insights into the contributions of various predictors, with glucose, BMI, pregnancies, and insulin identified as the most influential features. These findings align with established clinical markers of diabetes, affirming the validity of the model's decision-making process. This study, therefore, not only bridges the gap between predictive accuracy and clinical transparency but also provides a methodological framework for leveraging XAI to enhance the interpretability of ML models in healthcare. The incorporation of the SMOTE for class balancing further contributed to the robustness of the models, ensuring their reliability across diverse datasets and real-world scenarios.

The contributions of this research are multifold, extending the frontier of knowledge in both data mining and artificial intelligence applications within the healthcare domain. The integration of SHAP values into the diabetes prediction process is demonstrated to enhance transparency and trustworthiness in AI systems, facilitating their adoption in clinical practice. However, this study acknowledges its limitations, including the reliance on a single dataset, which may constrain the generalizability of the findings to other populations and clinical settings. Additionally, while SHAP analysis enhances interpretability, the inherent complexities of ML models mean that complete transparency remains an elusive goal. Future research should endeavor to validate these models with more diverse datasets and explore the integration of additional XAI techniques to further enhance model interpretability. A speculative, yet promising, direction could involve the development of longitudinal studies that track model performance and interpretability over time, providing insights into the dynamic nature of diabetes risk factors.

References

[1] P. David, S. Singh, R. Ankar, "A comprehensive overview of skin complications in diabetes and their prevention," Cureus, vol. 15, no. 5, p. e38961, 2023.

[2] A. F. Walker et al., "Interventions to address global inequity in diabetes: international progress," Lancet, vol. 402, no. 10397, 2023, pp. 250-264.

[3] M. Zakir et al., "Cardiovascular complications of diabetes: From microvascular to macrovascular pathways," Cureus, vol. 15, no. 9, p. e45835, 2023.

[4] A. Avogaro, M. Rigato, E. di Brino, D. Bianco, I. Gianotto, G. Brusaporco, "The socio-environmental determinants of diabetes and their consequences," Acta Diabetol., vol. 61, no. 10, 2024, pp. 1205-1210.

[5] S. Gowthami, R. Venkata Siva Reddy, M. R. Ahmed, "Exploring the effectiveness of machine learning algorithms for early detection of Type-2 Diabetes Mellitus," Measur. Sens., vol. 31, no. 100983, p. 100983, 2024.

[6] A. A. L. Ahmad, A. A. Mohamed, "Artificial intelligence and machine learning techniques in the diagnosis of type I diabetes: Case studies," in

Studies in Computational Intelligence, Singapore: Springer Nature Singapore, 2024, pp. 289-302.
[7] T. Althobaiti, S. Althobaiti, M. M. Selim, "An optimized diabetes mellitus detection model for improved prediction of accuracy and clinical decision-making," Alex. Eng. J., vol. 94, 2024, pp. 311-324.
[8] R. F. Albadri, S. M. Awad, A. S. Hameed, T. H. Mandeel, R. A. Jabbar, "A diabetes prediction model using hybrid machine learning algorithm," Math. Model. Eng. Probl., vol. 11, no. 8, 2024, pp. 2119-2126.
[9] S. Buyrukoğlu, A. Akbaş, "Machine learning based early prediction of type 2 diabetes: A new hybrid feature selection approach using Correlation Matrix with Heatmap and SFS," Balkan Journal of Electrical and Computer Engineering, vol. 10, no. 2, 2022, pp. 110-117.
[10] A. Adadi, M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," IEEE Access, vol. 6, 2018, pp. 52138-52160.
[11] F. Bodria, F. Giannotti, R. Guidotti, F. Naretto, D. Pedreschi, S. Rinzivillo, "Benchmarking and survey of explanation methods for black box models," Data Min. Knowl. Discov., vol. 37, no. 5, 2023, pp. 1719-1778.
[12] A. Barredo Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Inf. Fusion, vol. 58, 2020, pp. 82-115.
[13] W. Ding, M. Abdel-Basset, H. Hawash, A. M. Ali, "Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey," Inf. Sci. (Ny), vol. 615, 2022, pp. 238-292.
[14] V. Hassija et al., "Interpreting black-box models: A review on explainable Artificial Intelligence," Cognit. Comput., vol. 16, no. 1, 2024, pp. 45-74.
[15] S. Lundberg, S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan et al., Eds. Curran Associates, Inc., 2017.
[16] L. S. Shapley, "Stochastic games," Proc. Natl. Acad. Sci. U.S.A., vol. 39, no. 10, 1953, pp. 1095-1100.
[17] K. Aliyeva, N. Mehdiyev, "Uncertainty-aware multi-criteria decision analysis for evaluation of explainable artificial intelligence methods: A use case from the healthcare domain," Information Sciences, vol. 657, no. 119987, p. 119987, 2024.
[18] Kaggle Dataset, "Pima Indian Diabetes Database," 2017. [Online]. Available: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database
[19] P. Verma, A. Khatoon, "Data mining applications in healthcare: A comparative analysis of classification techniques for diabetes diagnosis using the PIMA Indian diabetes dataset," in 2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM), 2024.
[20] L. Xie, "Pima Indian diabetes database and machine learning models for diabetes prediction," Highlights in Science, Engineering and Technology, vol. 88, 2024, pp. 97-103.
[21] V. Chang, J. Bailey, Q. A. Xu, Z. Sun, "Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms," Neural Comput. Appl., vol. 35, no. 22, 2022, pp. 1-17.
[22] S. Sahoo, T. Mitra, A. K. Mohanty, B. J. R. Sahoo, and S. Rath, "Diabetes prediction: A study of various classification based data mining techniques," International Journal of Computer Science and Informatics, vol. 4, no. 3, 2022, pp. 1-13.
[23] S. You, M. Kang, "A Study on Methods to Prevent Pima Indians Diabetes using SVM," Korean Journal of Artificial Intelligence, vol. 8, no. 2, 2020, pp. 7-10.
[24] A. F. Ashour, M. M. Fouda, Z. M. Fadlullah, M. I. Ibrahem, "Optimized neural networks for diabetes classification using Pima Indians diabetes database," in 2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI), 2024.
[25] K. Akyol, B. Şen, "Diabetes mellitus data classification by cascading of feature selection methods and ensemble learning algorithms," Int. J. Mod. Educ. Comput. Sci., vol. 6, 2018, pp. 10-16.
[26] M. S. Reza, R. Amin, R. Yasmin, W. Kulsum, S. Ruhi, "Improving diabetes disease patients classification using stacking ensemble method with PIMA and local healthcare data," Heliyon, vol. 10, p. e24536, 2024.
[27] A. Pyne, B. Chakraborty, "Artificial Neural Network based approach to Diabetes Prediction using Pima Indians Diabetes Dataset," in *2023 International Conference on Control, Automation and Diagnosis (ICCAD)*, Rome, Italy, 2023.
[28] A V. Jain, S. Shukla, N. Khare, "Analysis of various data imputation techniques for diabetes classification on PIMA dataset," in *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, 2024, pp. 1–6.
[29] S. Karatsiolis, C. N. Schizas, "Region based Support Vector Machine algorithm for medical diagnosis on Pima Indian Diabetes dataset," in *2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE)*, 2012.
[30] M. Bilal, G. Ali, M. W. Iqbal, M. Anwar, M. S. A. Malik, R. A. Kadir, "Auto-Prep: Efficient and Automated Data Preprocessing Pipeline," IEEE Access, vol. 10, 2022, pp. 107764-107784.
[31] L. B. V. de Amorim, G. D. C. Cavalcanti, R. M. O. Cruz, "The choice of scaling technique matters for classification performance," Appl. Soft Comput., vol. 133, no. 109924, p. 109924, 2023.
[32] A. D. Amirruddin, F. M. Muharam, M. H. Ismail, N. P. Tan, M. F. Ismail, "Synthetic Minority Over-sampling TEchnique (SMOTE) and Logistic Model Tree (LMT)-Adaptive Boosting algorithms for classifying imbalanced datasets of nutrient and chlorophyll sufficiency levels of oil palm (Elaeis guineensis) using spectroradiometers and unmanned aerial vehicles," Comput. Electron. Agric., vol. 193, no. 106646, p. 106646, 2022.
[33] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, 2002, pp. 321-357.
[34] T. Yilmaz, "Microwave spectroscopy based classification of rat hepatic tissues: On the significance of dataset," Balkan Journal of Electrical and Computer Engineering, vol. 8, no. 4, 2020, pp. 307-313.
[35] T. Tulgar, A. Haydar, İ. Erşan, "A distributed K Nearest Neighbor classifier for Big Data," Balkan Journal of Electrical and Computer Engineering, vol. 6, no. 2, 2018, pp. 105-111.
[36] T. Pala, A. Y. Camurcu, "Design of decision support system in the metastatic colorectal cancer data set and its application," Balkan Journal of Electrical and Computer Engineering, vol. 4, no. 1, 2016, pp. 12-16.
[37] C. Greco, P. Pace, S. Basagni, G. Fortino, "Jamming detection at the edge of drone networks using Multi-layer Perceptrons and Decision Trees," Appl. Soft Comput., vol. 111, no. 107806, p. 107806, 2021.
[38] İ. Kırbaş, A. Çifci, "Machine Learning-Based Rice Grain Classification Through Numerical Feature Extraction from Rice Image Data." in 9th International Zeugma Conference on Scientific Research. Gaziantep, Türkiye, 2023.
[39] A. Çifci, M. İlkuçar, "Analysis of window sizes in prediction of daily COVID-19 cases using machine learning models," International Journal of Mechatronics, Electrical and Computer Technology (IJMEC), vol. 12, no. 45, 2022, pp. 5208-5217.
[40] G. Bilgin, A. Çifci, "Eritematöz skuamöz hastalıkların teşhisinde makine öğrenme algoritmaları performanslarının değerlendirilmesi," Journal of Intelligent Systems: Theory and Applications, vol. 4, no. 2, 2021, pp. 195-202.
[41] C. Bentéjac, A. Csörgő, G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," Artif. Intell. Rev., vol. 54, no. 3, 2021, pp. 1937-1967.
[42] C. Molnar, Interpretable machine learning: a guide for making black box models interpretable. Morisville, North Carolina: Lulu, 2019.
[43] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," Nat. Mach. Intell., vol. 2, no. 1, 2020, pp. 56-67.
[44] S. M. Lundberg, G. G. Erion, S.-I. Lee, "Consistent individualized feature attribution for tree ensembles," arXiv [cs.LG], 2018.

BIOGRAPHIES

İsmail Kırbaş completed his Ph.D. in Electronics and Computer Education at Sakarya University and currently serves as the Dean of the Faculty of Computer and Informatics at Burdur Mehmet Akif Ersoy University, where he also holds the position of Director at the HAYTEK Collaborative Application Research Center. Dr. Kırbaş's research focuses on wireless body area network protocols, artificial intelligence, the Internet of Things,

and machine learning. With a strong academic background, including a bachelor's and master's degree from Kocaeli University, he has led a variety of innovative projects throughout his career.
Dr. Kırbaş's work significantly contributes to digital transformation across sectors such as agriculture and healthcare, utilizing artificial intelligence and data mining techniques. Actively involved as both a researcher and project leader in numerous national and international projects, Dr. Kırbaş has a robust portfolio of scientific publications and patents, underscoring his impactful presence in the field.

Ahmet Çifci received the B.S., M.S. and Ph.D. degrees in Electrical and Electronics Engineering from Sakarya University, Türkiye, in 2007, 2011 and 2015, respectively.
From 2009 to 2016, he was a Lecturer with the Vocational School of Burdur within Burdur Mehmet Akif Ersoy University. Since 2016, he has been an Assistant Professor with the Department of Electrical-Electronics Engineering, Faculty of Engineering and Architecture, Burdur Mehmet Akif Ersoy University, Burdur, Türkiye. Throughout his career, Dr. Çifci has held various administrative roles, including Senate Membership, Department Head of Electrical-Electronics Engineering, Vice-Director of the Institute of Science, and Vice-Dean. He currently serves as the Vice-Dean of the Faculty of Engineering and Architecture. Additionally, he is the Deputy Director at the Digital Technologies in Livestock Sector Joint Application and Research Center at Burdur Mehmet Akif Ersoy University. He has authored and co-authored numerous scholarly papers in peer-reviewed journals, books, and national/international conferences. His academic and research interests include high voltage engineering, power systems, and artificial intelligence.
Dr. Cifci is a member of the Chamber of Electrical Engineers of Türkiye.