Machine Learning-Based Prediction of Diabetes Using Multi-Classifiers: A Comprehensive Study

1st Alişan Köroğlu
The Graduate School of Natural and Applied Sciences
Dokuz Eylül University
İzmir, Turkey
alisan.koroglu@gmail.com

***Abstract*—Diabetes is a prevalent chronic disease characterized by elevated blood glucose levels resulting from insufficient insulin production or utilization. Early detection and effective management are critical to mitigating its adverse effects and preventing severe complications. This paper explores the application of machine learning algorithms for predicting diabetes and identifying significant risk factors. It is evaluated multiple classification models, including Random Forest, Logistic Regression and k-Nearest Neighbors, highlighting their accuracy. The findings demonstrate the effectiveness of ensemble methods like Random Forest in achieving high predictive accuracy. Additionally, feature selection techniques such as Principal Component Analysis (PCA) and Random Forest importance scores are employed to identify critical predictors, including glucose levels, age, and BMI. This study underscores the potential of machine learning in healthcare analytics for proactive diabetes prevention and personalized treatment strategies.**

***Keywords*—Diabetes prediction, machine learning, classification algorithms, Random Forest, Principal Component Analysis, feature selection, healthcare analytics, early diagnosis.**

I. Introduction (*Heading 1*)

Diabetes is a growing global health concern, affecting over 500 million individuals worldwide. This chronic condition arises from abnormalities in insulin production or function, leading to persistently high blood sugar levels. If unmanaged, diabetes can result in severe complications such as cardiovascular diseases, kidney failure, neuropathy, and vision loss. Early detection and management of diabetes are critical to mitigating these outcomes and improving the quality of life for patients.

The integration of machine learning (ML) in healthcare has revolutionized disease prediction and management, enabling the analysis of large datasets to identify patterns and risk factors. ML models can predict the likelihood of developing diabetes based on various demographic, clinical, and lifestyle features, offering opportunities for early intervention and personalized treatment plans. Furthermore, these models provide valuable insights into the interplay of factors influencing diabetes risk, paving the way for targeted prevention strategies.

In this study, it is investigated the efficacy of multiple machine learning classifiers in predicting diabetes. It is employed advanced preprocessing techniques, including feature selection, and cross-validation, to enhance model accuracy and reliability. This analysis focuses on comparing classifiers such as Random Forest, k-Nearest Neighbors, and Logistic Regression, with an emphasis on their predictive performance and computational efficiency. By identifying critical predictors and leveraging ensemble methods, this research contributes to the development of robust predictive models for diabetes management and prevention.

II. What is Diabetes?

Diabetes is a chronic condition characterized by high levels of sugar (glucose) in the blood. It occurs when the body either does not produce enough insulin, a hormone responsible for regulating blood sugar, or cannot effectively use the insulin it produces. Insulin helps glucose from the food we eat enter cells to provide energy. Without effective insulin function, glucose remains in the bloodstream, leading to various health complications over time.

There are several types of diabetes, each with unique causes and management approaches. Type 1 diabetes is an autoimmune condition where the body's immune system attacks insulin-producing cells in the pancreas. It typically develops in childhood or adolescence and requires insulin therapy for survival. Type 2 diabetes, the most common form, is often associated with lifestyle factors such as obesity, poor diet, and lack of physical activity, though genetics also play a role. It is characterized by insulin resistance and insufficient insulin production. Gestational diabetes occurs during pregnancy due to hormonal changes that affect insulin sensitivity and usually resolves after childbirth, although it increases the risk of type 2 diabetes later in life. Additionally, prediabetes is a condition where blood sugar levels are higher than normal but not yet in the diabetes range, serving as a warning sign for potential progression to type 2 diabetes.

The symptoms of diabetes can vary depending on the type and severity of the condition. Common symptoms include frequent urination, excessive thirst, extreme hunger, unexplained weight loss, fatigue, and blurred vision. In some cases, individuals may experience slow-healing wounds and tingling or numbness in the hands or feet, which are signs of nerve damage caused by prolonged high blood sugar levels. Early detection and management of these symptoms are crucial to prevent complications.

If left untreated or poorly managed, diabetes can lead to serious long-term complications. Cardiovascular diseases, such as heart attacks and strokes, are significantly more common in individuals with diabetes. Kidney damage, known as diabetic nephropathy, can progress to kidney failure, while eye damage, or retinopathy, may result in blindness. Nerve damage, or neuropathy, often affects the extremities and can cause pain, tingling, or loss of sensation. Diabetic foot problems, including ulcers and infections, can lead to severe outcomes like amputation. Additionally, poorly controlled diabetes can increase susceptibility to skin infections and other medical conditions.

Effective management of diabetes involves a combination of lifestyle changes, medications, and regular monitoring. Maintaining a healthy diet with an emphasis on low-glycemic foods, lean proteins, and whole grains helps control blood sugar levels. Regular physical activity improves insulin sensitivity and supports weight management, which is particularly beneficial for individuals with type 2 diabetes. Medications, including oral drugs like metformin and insulin injections, are prescribed based on the type and severity of diabetes. Blood glucose monitoring, both at home and through periodic HbA1c tests, helps track long-term blood sugar control. Diabetes education programs also play a vital role in empowering individuals to manage their condition effectively.

While type 1 diabetes cannot currently be prevented, type 2 diabetes and prediabetes are often preventable with lifestyle modifications. Maintaining a healthy weight, engaging in regular exercise, eating a balanced diet, and undergoing regular health screenings are key steps in reducing the risk of developing diabetes. Ongoing research continues to explore new treatments, technologies, and interventions to improve diabetes management and potentially cure the condition in the future. Early detection and proactive management remain critical to reducing the burden of diabetes on individuals and healthcare systems worldwide

A. *Importance of Understanding and Addressing Diabetes*

Diabetes is a major global health issue, affecting millions of people worldwide. Its importance lies not only in its prevalence but also in its profound impact on individuals, families, and healthcare systems. Understanding diabetes and addressing it effectively is crucial for several reasons.

Diabetes affects over 500 million people globally, a number that is expected to rise significantly in the coming decades. This chronic disease has become a leading cause of morbidity and mortality in both developed and developing countries. Recognizing its widespread nature is essential to implementing effective prevention, early detection, and management strategies to curb its rapid growth

If left unmanaged, diabetes can lead to a range of life-threatening complications, including cardiovascular disease, kidney failure, blindness, nerve damage, and lower-limb amputations. These complications not only reduce the quality of life for individuals but also place a significant burden on healthcare systems. Understanding diabetes is critical to minimizing its devastating consequences and improving patients' long-term outcomes

Diabetes imposes substantial economic costs on individuals, families, and healthcare systems. These include direct costs like medical care, medications, and hospitalizations, as well as indirect costs such as lost productivity and income due to illness. Preventing and managing diabetes effectively can help reduce these financial burdens at both personal and societal levels.

Unlike some chronic diseases, type 2 diabetes and prediabetes are largely preventable through lifestyle modifications. Even for those with type 1 diabetes, advancements in treatment and technology have made it possible to manage the condition effectively and lead a healthy life. Raising awareness about prevention and encouraging healthy habits, such as maintaining a balanced diet and engaging in regular exercise, are critical to controlling the diabetes epidemic.

Diabetes research and technological advancements, such as continuous glucose monitors, insulin pumps, and artificial pancreas systems, have significantly improved the lives of people with diabetes. Ongoing research holds promise for better treatments, early detection methods, and potentially a cure. Highlighting the importance of diabetes ensures continued investment in scientific and medical innovation

Diabetes affects not just individuals but also their families, caregivers, and communities. It requires lifestyle adjustments, emotional support, and resources from loved ones. Raising awareness about diabetes fosters understanding and empathy, enabling communities to support those affected and advocate for better healthcare policies

Understanding and addressing diabetes is essential for improving global health, reducing healthcare costs, and enhancing the quality of life for millions of people. By prioritizing prevention, education, and research, society can mitigate the profound impact of diabetes and build a healthier future for all.

B. *Machine Learning for Prediction of Diabetes*

Machine learning (ML) has proven to be a valuable tool in predicting diabetes, helping prevent complications and incidents by enabling earlier detection and personalized interventions.

Machine learning algorithms can analyze large volumes of patient data, including demographic, clinical, and lifestyle factors, to identify patterns associated with diabetes risk. By predicting the likelihood of developing diabetes (especially type 2 diabetes or gestational diabetes), ML allows for early preventive measures such as lifestyle changes or medical monitoring, reducing the onset of the disease.

For instance, ML models trained on datasets like the Pima Indian Diabetes dataset or real-world clinical data have shown high accuracy in predicting prediabetes and diabetes based on factors like BMI, glucose levels, and family history.

One-size-fits-all prevention strategies are often ineffective. Machine learning enables the development of personalized healthcare plans by analyzing an individual’s unique risk factors and health profile. ML models can recommend tailored interventions, such as dietary changes, exercise routines, or medications, based on a person’s specific needs and risk levels.

With advancements in wearable devices and continuous glucose monitoring systems, ML can analyze real-time data to predict and prevent critical incidents, such as hyperglycemia or hypoglycemia. Predictive alerts can warn patients or caregivers of potential blood sugar fluctuations, allowing for timely action.

Diabetes is influenced by a complex interplay of genetic, environmental, and lifestyle factors. Machine learning excels at processing and integrating such multidimensional data, uncovering insights that might be missed with traditional statistical methods. For example, deep learning models can analyze genomic data, medical imaging, or electronic health records to identify novel diabetes risk factors.

By predicting diabetes and its complications early, ML can reduce the need for expensive treatments, hospitalizations, and long-term care. Early intervention enabled by ML can help avoid severe complications like kidney failure or cardiovascular events, which are costly and difficult to manage.

C. *Challenges and Considerations*

While machine learning holds great potential, there are challenges to address:

- Data Quality: Predictions depend on the quality and diversity of the training data. Biases in datasets can lead to inequitable predictions.
- Integration into Healthcare: Implementing ML tools in clinical settings requires overcoming technological, regulatory, and ethical hurdles.
- Interpretability: Many ML models, especially deep learning, are "black boxes," making their predictions hard to explain to clinicians and patients.

Machine learning is a powerful tool for predicting diabetes and enabling preventive care. By identifying high-risk individuals early and providing actionable insights, ML has the potential to significantly reduce the incidence and burden of diabetes. However, its success relies on high-quality data, careful implementation, and collaboration between technology developers and healthcare providers.

III. Literature Review: Machine Learning in Diabetes Prediction

Diabetes prediction has emerged as a critical area of research, leveraging machine learning algorithms to improve early detection, intervention, and management strategies. Various studies have employed different datasets, techniques, and classifiers to enhance the accuracy and reliability of prediction models, addressing challenges like imbalanced data, missing values, and computational efficiency.

Soni (2020) evaluated several machine learning algorithms, including Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), and Gradient Boosting (GB), using the Pima Indian Diabetes Dataset (PIDD). The study highlights preprocessing techniques, such as handling missing values and normalizing the data, to improve model performance. Among the models, Random Forest demonstrated the highest accuracy due to its robustness in handling large datasets and reducing variance. Glucose levels were identified as the most critical predictor of diabetes. The research underscores the potential of ensemble methods like Random Forest in reducing prediction errors and aiding healthcare professionals in early diagnosis.[1]

Abnoosian (2023) proposed a novel ensemble framework integrating multiple classifiers to predict diabetes using the Iraqi Patient Dataset for Diabetes (IPDD). The study addressed challenges such as imbalanced datasets and missing data through advanced preprocessing methods, including feature selection techniques like Minimum Redundancy Maximum Relevance (MRMR) and dimensionality reduction using PCA and ICA. The ensemble model, combining classifiers like k-NN, AdaBoost, DT, and RF, achieved an accuracy of 99.87% and an AUC score of 0.999. This research highlights the effectiveness of ensemble approaches in overcoming limitations of individual classifiers and emphasizes the importance of robust preprocessing and feature selection in improving prediction accuracy.[2]

Lyngdoh (2020) explored the effectiveness of five supervised machine learning algorithms—KNN, Naïve Bayes (NB), DT, RF, and SVM—for diabetes prediction using the PIDD. The dataset underwent preprocessing steps such as normalization and handling missing values to ensure consistency and quality. KNN emerged as the most effective algorithm, achieving 76% accuracy with minimal overfitting. While DT and RF performed well, they showed slight overfitting tendencies, and SVM, though accurate, was computationally expensive. The study highlights the potential of machine learning in disease prediction and suggests exploring deep learning models for future enhancements.[3]

Daigavhane (2024) conducted a comparative analysis of machine learning classifiers, including RF, SVM, LR, KNN, and GB, using the PIDD. The study employed preprocessing steps such as normalization, handling missing values, and feature selection using correlation analysis and recursive feature elimination. Random Forest achieved the highest accuracy of 87.6%, followed by Gradient Boosting and Logistic Regression, which also demonstrated strong performance. Glucose levels, BMI, and age were identified as the most significant predictors of diabetes risk. The research underscores the reliability of tree-based ensemble methods like Random Forest and highlights the potential of integrating deep learning for improved accuracy in future studies.[4]

Hasan (2020) presented a framework for diabetes prediction using an ensemble of classifiers and boosting techniques. Using the PIDD, the study addressed challenges like missing values, outliers, and class imbalance through preprocessing steps such as outlier rejection, missing value imputation, and feature selection. The proposed ensemble, combining AdaBoost and XGBoost, achieved an AUC of 0.950, outperforming individual classifiers and previous models. The framework emphasized the importance of correlation analysis for feature selection and demonstrated the effectiveness of boosting algorithms in balancing sensitivity and specificity. The study suggested extending the framework to other medical datasets and integrating it into web-based diagnostic tools for broader applications.[5]

In summary, machine learning techniques have demonstrated significant potential in predicting diabetes, with ensemble methods and advanced preprocessing emerging as critical factors for improving accuracy and reliability. Studies consistently highlight the importance of feature selection, dimensionality reduction, and hyperparameter tuning in enhancing model performance. Future research should focus on integrating deep learning models, exploring larger and more diverse datasets, and developing real-world applications to further advance diabetes prediction and management.

IV. Dataset

The Diabetes Dataset available on Kaggle, contributed by Ankit Batra, is a comprehensive collection designed to facilitate the development of predictive models for diabetes diagnosis. The dataset is publicly accessible on Kaggle and can be downloaded for analysis. (https://www.kaggle.com/datasets/ankitbatra1210/diabetes-dataset)

The dataset encompasses a wide range of attributes, including genetic factors, environmental influences, and lifestyle choices, all of which are pertinent to diabetes risk assessment. This dataset is intended for use in machine learning projects aimed at predicting diabetes onset, understanding contributing factors, and enhancing diagnostic accuracy.

Researchers and data scientists can utilize this dataset to train and evaluate machine learning algorithms for early diabetes detection. The diverse set of features allows for in-depth analysis to identify the most significant predictors of diabetes, aiding in the development of targeted prevention strategies. The dataset serves as a valuable resource for educational purposes, enabling learners to practice data preprocessing, feature selection, and model evaluation in the context of healthcare analytics..

A. *Feature Analysis*

The dataset consists of 34 columns: one target variable and 33 features. Among the features, 20 are categorical, and 13 are numerical. It contains 70,000 samples and has no missing values. The target variable includes 13 distinct classes, representing 13 different types of diabetes. As shown in the fig. 1., the distribution of the target values is quite balanced.

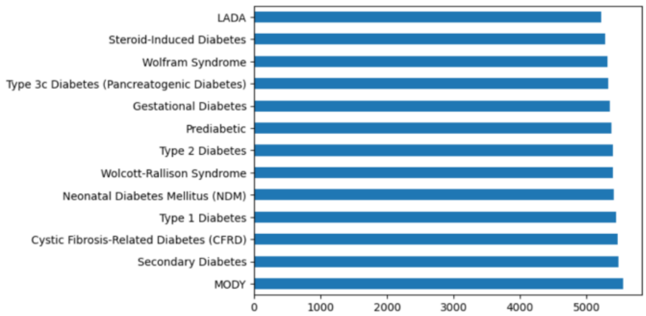

Fig. 1. The distribution of the target values

The fig. 2. illustrates the age distribution in the dataset, which spans from 0 to approximately 80 years and is divided into 20 bins, each representing an interval of about 4 years. Clear peaks are observed at specific age ranges, such as 0–5 years, indicating a large proportion of very young individuals, and 10–20 years, showing another significant group. The distribution becomes more even in the 30–50 age range, with no pronounced peaks, while the population density declines steadily for age groups over 50, suggesting fewer older individuals in the dataset. This trend may reflect a generally younger population or could be attributed to data collection bias if the dataset specifically targets younger age groups.

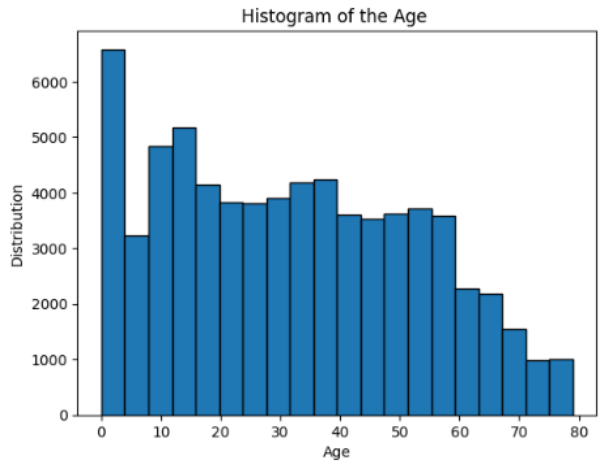

Fig 2. the distribution of the target values

The Fig. 3. illustrates the age distribution of observations in the dataset across different types of diabetes. It is evident from the graph that various types of diabetes are associated with distinct age ranges. This suggests that the age variable may serve as an important distinguishing factor in identifying different types of diabetes.

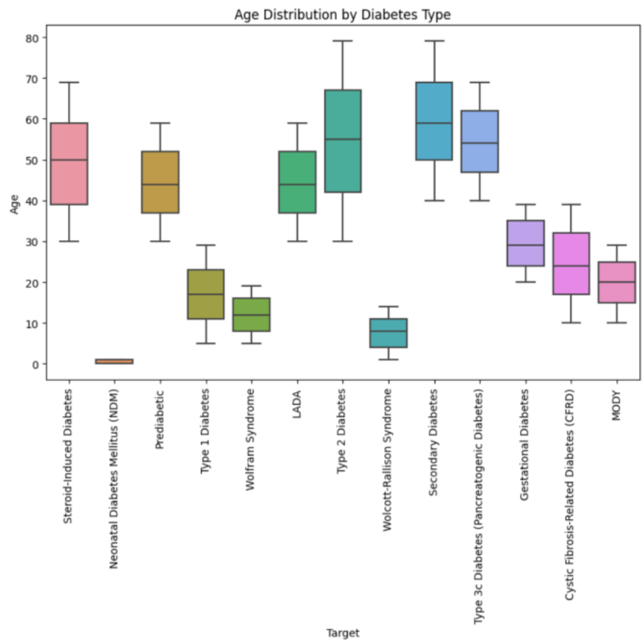

Figure 3 The age distribution by diabetes type

As shown in the Fig. 4., the majority of individuals in the dataset have a BMI ranging from 20 to 35, with the highest peak around 25, corresponding to the upper limit of the normal weight category. The distribution is slightly skewed to the right, with fewer individuals having a BMI greater than 30 (obese category). Additionally, there are relatively few individuals classified as underweight (very low BMI) or severely obese (very high BMI). Overall, the study population predominantly falls within the normal to slightly overweight categories, with fewer observations in the underweight or severely obese groups.

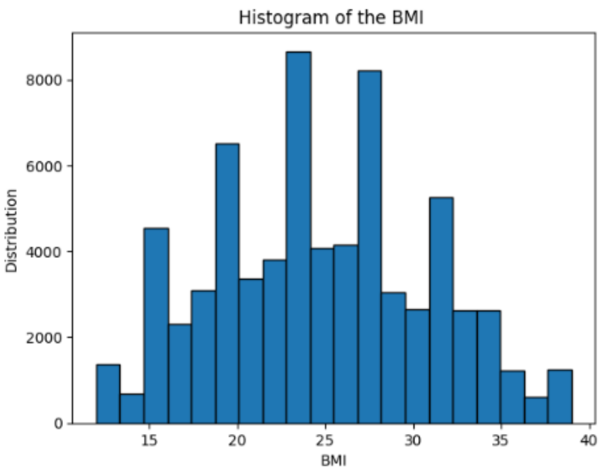

Figure 4 The histogram of the BMI

The Fig. 5. demonstrates correlation between numerical feature of the dataset. Age shows several strong correlations with other variables in the dataset, reflecting its significance as a key factor in health. For instance, Age and Blood Pressure (0.76) demonstrate a very strong correlation, indicating that blood pressure tends to increase with age due to factors like arterial hardening and other physiological changes. Similarly, Age and Waist Circumference (0.73) exhibit a strong relationship, suggesting that abdominal fat accumulation increases with age, raising the risk of metabolic diseases like type 2 diabetes. Age and Cholesterol Levels (0.73) also show a strong correlation, highlighting the natural rise in cholesterol levels as lipid metabolism becomes less efficient over time. Age and BMI (0.66) present a slightly less pronounced but still strong correlation, reflecting an age-related increase in BMI, influenced by factors like diet and physical activity. Lastly, Age and Insulin Levels (0.61) indicate a moderate-to-strong correlation, where insulin levels rise with age, potentially due to increased insulin resistance in older adults. These correlations collectively emphasize age as a crucial factor in metabolic health, as it strongly influences key indicators like blood pressure, waist circumference, cholesterol, and BMI, which worsen with age and elevate the risk of chronic diseases such as hypertension, diabetes, and cardiovascular conditions.

Interestingly, the correlation between Insulin Levels and Blood Glucose Levels is weak, suggesting variability in insulin response among individuals. Additionally, there is a moderate negative correlation between Pancreatic Health and Blood Glucose Levels, indicating that poorer pancreatic health is associated with higher blood glucose levels. A strong negative correlation between Pulmonary Function and Blood Glucose Levels (-0.60) suggests that better lung function is associated with lower blood glucose levels. Similarly, Pulmonary Function and Pancreatic Health (-0.60) show a strong negative correlation, implying an inverse relationship between lung and pancreatic health. These findings highlight the interconnectedness of metabolic and organ-specific health factors and underscore the need for monitoring these indicators to prevent chronic diseases effectively.

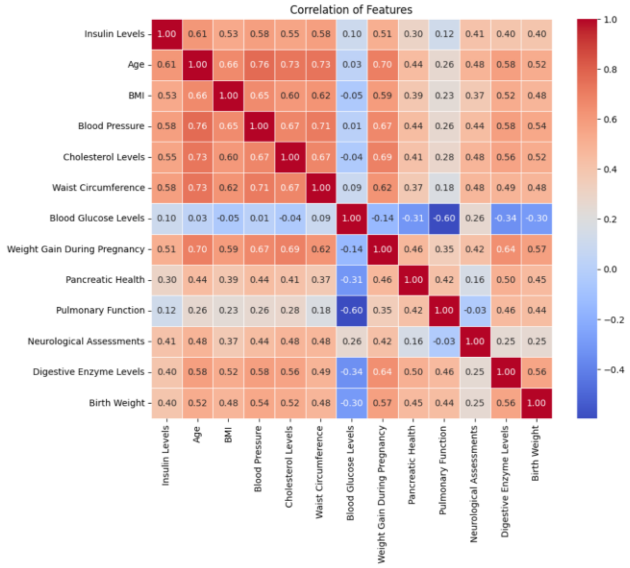

Figure 5 Corrolation Matrix of the features

The distribution of the numerical features illustureted in Fig. 6. by using boxplot graphs. Boxplot graphs are effective tools for outlier detection as they visually represent the data distribution and highlight potential outliers using the Interquartile Range (IQR) method. The figure below illustrates the IQR of the numerical features, and it can be observed from the graph that the numerical features do not require additional outlier handling.

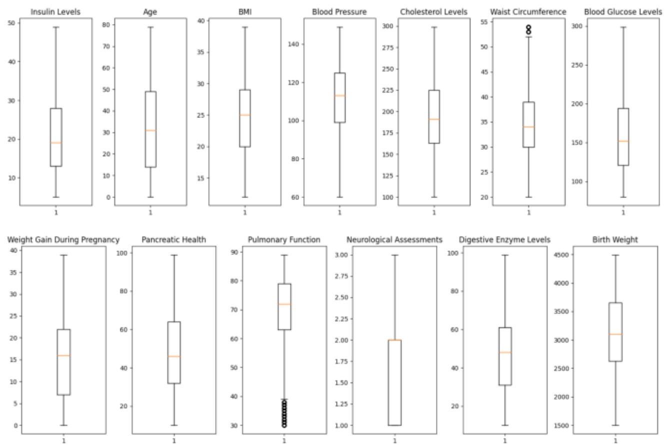

Fig. 6. Boxplot graphs of the numerical features

V. Feature Selection

A. *Principal Component Analisis(PCR)*

PCR combines Principal Component Analysis (PCA) and regression, and while its primary goal is dimensionality reduction rather than feature selection, it can be indirectly used for selecting the most important features. PCA transforms the original features into a new set of orthogonal components (principal components) ranked by the amount of variance they capture. PCR selects a subset of these principal components and uses them for regression.

PCR does not directly tell you which original features are important since principal components are linear combinations of all features. However, by examining the loadings (coefficients that relate the original features to the principal components), you can identify which features contribute most to the important components.

Fig. 7. represents the tuning process for a Principal Component Regression (PCR) model used for diabetes prediction. The x-axis shows the number of principal components, and the y-axis shows the Root Mean Square Error (RMSE), which measures the model's prediction error. The goal is to minimize the RMSE while balancing model complexity.

There is a significant reduction in RMSE as the number of components increases from 0 to around 10. This indicates that the first few principal components capture the majority of the variance in the data and significantly improve model performance.

The "elbow" of the graph, where the RMSE starts to stabilize (around 10–15 components), could indicate the optimal number of components to balance performance and complexity.

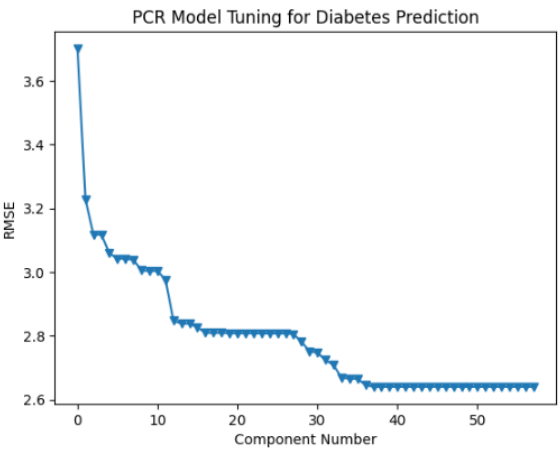

Fig. 7. PCR model performans according to component number

B. *Random Forest*

Random Forest is a much more direct and effective method for feature selection. Random Forest is an ensemble of decision trees that uses bootstrapping and random feature selection at each split. During training, it calculates metrics such as feature importance, which can be used to rank features. Feature importance scores measure the contribution of each feature to reducing impurity (e.g., Gini Impurity or Mean Squared Error) in the trees. Features with low importance scores can be excluded as they do not contribute significantly to the predictive performance. Random forest algorithm handles non-linear relationships and interactions between variables, provides clear and interpretable feature importance metrics and works well even with correlated features or high-dimensional datasets.

The chart in Fig. 8. represents the importance levels of features according to Randomforest model and it suggests that age and metabolic indicators such as blood glucose levels, blood pressure, and waist circumference play key roles in the model, likely due to their strong relationship with diabetes risk. Variables with lower importance may still be relevant in specific contexts but contribute less to the overall predictive accuracy of the model.

It is evident that numerical features are more influential than categorical features in the dataset. In this study, these 13 key features are utilized by the machine learning algorithm for classification purposes, highlighting their significance in predicting outcomes.

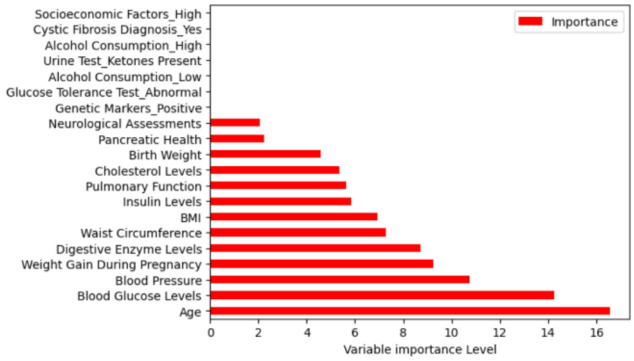

Fig. 8. Variable importance chart

VI. Model Comparison

In this study, five different machine learning models were implemented using the scikit-learn library. The models include Logistic Regression, k-Nearest Neighbors (KNN), Naive Bayes, Classification and Regression Tree (CART), and Random Forest. Hyperparameter tuning was conducted through a 10-fold cross-validation approach, utilizing the GridSearchCV method to optimize model performance. In Fig. 9, the classification models and their tuned hyperparameters are presented. All other parameters were set to their default values.

The dataset was split into training and testing sets, with 80% allocated for training and 20% for testing. Accuracy was used as the evaluation metric.

TABLE I. CLASSİFİCATİON MODELS AND HYPERPARAMETERS

Classification Models	**Hiperparameters**
LogisticRegression	Solver = "liblinear"
GaussianNB	
KNeighborsClassifier	'n_neighbors': 21
DecisionTreeClassifier	'max_depth': 9, 'min_samples_split': 7
RandomForestClassifier	'max_depth': 10, 'max_features': 8, 'min_samples_split': 5, 'n_estimators': 1000

Fig. 9. Classification model and hiperparameters

The bar chart as shown in Fig. 10., compares the accuracy of different machine learning models for classification. The x-axis represents accuracy, while the y-axis lists the models evaluated.

The Random Forest Classifier is the most effective model for this classification task, providing the highest accuracy. Models like Decision Tree and Gaussian Naive Bayes are also viable alternatives but with slightly lower performance. The results suggest that ensemble methods like Random Forest are well-suited for the problem at hand.

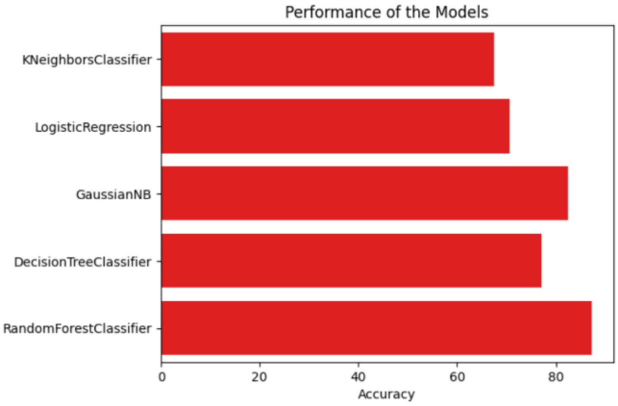

Fig. 10. Performans of the models

VII. Conclusion

This study evaluated five machine learning models for diabetes prediction using the scikit-learn library: Logistic Regression, k-Nearest Neighbors, Naive Bayes, Classification and Regression Tree, and Random Forest. Among the models, Random Forest achieved the highest accuracy, demonstrating its robustness for this task. Feature importance analysis identified age, blood glucose levels, and blood pressure as critical predictors. The findings highlight the effectiveness of machine learning, particularly ensemble methods, in early diabetes diagnosis and management.

Future research can explore larger datasets, deep learning models, and real-time data integration to enhance prediction accuracy and support proactive healthcare solutions. This work emphasizes the potential of machine learning in advancing personalized and preventive healthcare for diabetes.

Reference

[1] M. Soni and S. Varma, "Diabetes prediction using machine learning techniques," International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 09, Sept. 2020. [Online]. Available: https://doi.org/10.17577/IJERTV9IS090496.

[2] K. Abnoosian, R. Farnoosh, and M. H. Behzadi, "Prediction of diabetes disease using an ensemble of machine learning multi-classifier models," BMC Bioinformatics, vol. 24, no. 337, 2023. [Online]. Available: https://doi.org/10.1186/s12859-023-05465-z

[3] A. C. Lyngdoh, N. A. Choudhury, and S. Moulik, "Diabetes disease prediction using machine learning algorithms," in 2020 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2021, pp. 517–521. DOI: 10.1109/IECBES48179.2021.9398759.

[4] M. K. Daigavhane and S. Gundewar, "Diabetes prediction using different machine learning classifiers," in 2024 IEEE Parul International Conference on Engineering and Technology (PICET), 2024, pp. 1–6. DOI: 10.1109/PICET60765.2024.10716110.

[5] M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers," IEEE Access, vol. 8, pp. 76516–76530, 2020. https://doi.org/10.1109/ACCESS.2020.2989857.