Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, No.2, June : 2024

TOWARDS ACCURATE AND NON-INVASIVE DIABETES DETECTION: ML INSIGHTS

Dr. Mrunal Pathak, Assistant Professor, Dept.Of Information Technology, AISSMS IOIT Pune College, SPPU University.
Dr. Meenakshi Thalor, Head of the Department, Dept. Of IT, AISSMS IOIT Pune, SPPU University.
Prajwal A. Joshi, Student at IT, AISSMS IOIT Pune, SPPU University.
Nehal J. Pawar, Student at IT Dept, AISSMS IOIT Pune, SPPU University.
Ninad V. Futane, Student at IT Dept, AISSMS IOIT Pune, SPPU University.
Gauri R. Patil, Student at IT Dept, AISSMS IOIT Pune, SPPU University.

ABSTRACT
Diabetes mellitus poses a significant global health burden, necessitating efficient and timely detection methods to mitigate its adverse effects. This project explores the feasibility and effectiveness of utilizing Photoplethysmography (PPG) signals, in conjunction with machine learning algorithms, for early diabetes detection. PPG offers a non-invasive means of assessing blood volume changes in microvascular tissue, holding promise for diagnosing various health conditions, including diabetes.In this study, we undertake a thorough investigation into the application of machine learning models, with a focus on XGBoost, for analyzing PPG signals and predicting diabetes risk. We begin by collecting a diverse dataset, complemented by relevant patient information and diagnostic labels. Next, we preprocess the data, including cleaning, feature extraction, and normalization, to ensure optimal model performance.XGBoost (Extreme Gradient Boosting) is a powerful and efficient machine learning algorithm known for its high performance and scalability, particularly in classification and regression tasks. In our diabetes classification project, XGBoost demonstrated impressive accuracy, achieving a rate of 95%. This high accuracy underscores XGBoost's capability to handle complex datasets and deliver robust predictive models.

Keywords:
Diabetes, Machine Learning, Logistic Regression, Support Vector Machine, Random Forest, Diagnosis Technique, Prediction.

I. Introduction
Diabetes is a serious worldwide health issue that can cause a large number of fatalities annually. We are concentrating on Type 2 Diabetes, which results from improper insulin utilization by the body, raising blood sugar levels. This illness typically has moderate symptoms at first and is discovered only until complications like renal problems, heart disease, visual problems, or amputations happen. People at risk frequently put off medical testing because they are too busy or fear the expense, despite the fact that early detection and monitoring are critical.Our goal is to develop a method for Type 2 Diabetes prediction that makes use of readily available population-level data. We use data about your body and a quick (2.1 seconds) blood flow measurement with a basic gadget called a sensor for photoplethysmography (PPG). PPG is a non-invasive, low-cost way to monitor blood flow in your vessels, providing information on blood pressure, heart rate, and oxygen saturation. According to recent research, PPG's ability to carry significant information in its signal makes it potentially useful for early disease detection. Thanks to technological advancements like PPG sensors in phones and smartwatches, we can quickly monitor our health and identify early signs of diseases like diabetes. Unlike previous studies, our aim is to predict diabetes based just on a single brief PPG signal and basic personal data. We're concentrating on a straightforward method, while others have examined PPG signals and other user data for diabetes prediction. utilizing a two- to three-second PPG signal and

some basic personal data. If we are successful, wearable technology and healthcare could both get better.

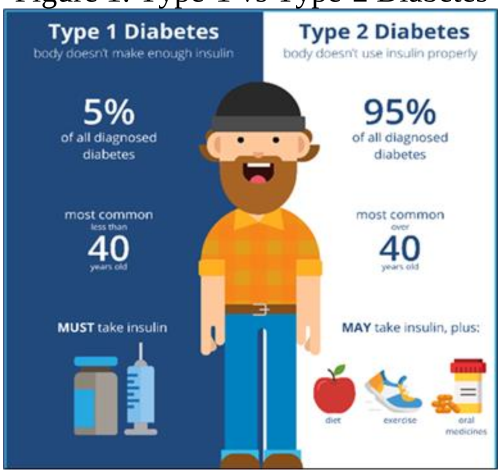
1.1 Diabetes Type I

The misdirected death of insulin-producing beta cells in the pancreas by the immune system is the hallmark of type 1 diabetes, a chronic autoimmune illness. This leads to a serious insulin shortage, a hormone that is essential for controlling blood sugar levels. Type 1 diabetes, which is usually discovered in childhood or adolescence, requires a lifetime reliance on exogenous insulin in order to properly control blood glucose levels. Symptoms of Type 1 diabetes include extreme thirst, frequent urination, exhaustion, and unexplained weight loss. In contrast to Type 2 diabetes, which is frequently linked to lifestyle factors, The main cause of type 1 diabetes is heredity and it cannot be avoided. It is still unknown what precisely sets off the autoimmune reaction that destroys beta cells; it seems to be a complicated interaction between genetic predisposition and environmental circumstances. In order to manage Type 1 diabetes and avoid consequences including cardiovascular disease, kidney failure, and nerve damage, continuous glucose monitoring, insulin pumps, and routine blood glucose testing are essential. Research on improving insulin delivery strategies, comprehending the underlying reasons, and investigating novel therapies to lessen the burden on people with Type 1 diabetes is still ongoing.

1.2 Diabetes Type II

Insulin resistance and decreased insulin secretion are the hallmarks of type 2 diabetes, a chronic metabolic disease that raises blood glucose levels. The majority of occurrences of diabetes worldwide are of this kind, which is frequently linked to lifestyle factors like obesity, poor eating habits, and sedentary behavior. Hyperglycemia is exacerbated by poor glucose uptake caused by insulin resistance, a condition in which cells do not respond to insulin as well. The pancreas may find it more difficult to generate enough insulin over time to keep blood glucose levels within normal ranges. The symptoms of type 2 diabetes, which might include increased thirst, frequent urination, exhaustion, and blurred vision, can first go unrecognized as the condition often develops gradually. If untreated, it might result in serious side effects like cardiovascular disease neuropathy, renal failure, and visual impairment. A balanced diet, consistent exercise, and weight control are all essential lifestyle changes for the management of type 2 diabetes. To regulate blood sugar levels, doctors may also prescribe medications and, in certain situations, insulin therapy. Minimizing the effects of type 2 diabetes on a person's health and general well-being requires thorough care, early detection, and routine monitoring.

Figure 1: Type-1 vs Type-2 Diabetes



II. Literature

Various methods have been used to detect changes in photoplethysmography (PPG) waveforms, including machine learning-based, traditional, and deep learning-based approaches. Traditional approaches use the Pointcaré plot of finger PPG signals to distinguish between healthy and diabetic subjects, while machine learning methods use primary PPG features and demographic data to simplify pre-processing. Monte-Moreno et al. used computer models like random forest and gradient boosting, along with important PPG and demographic information, to identify diabetes, achieving an AUC of 0.7. Hettiarachchi employed a Decision Tree model to identify Type 2 Diabetes even when hypertension or pre-hypertension was present, achieving an AUC score of 0.83. Nirala et al. obtained an AUC score of 0.83 while detecting DT2 in healthy and diabetic subjects without hypertension using a support vector machine learning algorithm. Deep learning has revolutionized the analysis of biomedical signals, with convolutional neural networks (CNNs) being frequently used in PPG analysis. Srinivasan and Forozan developed a CNN model using 30-second finger PPG signals to detect type 2 diabetes in a MIMIC-III dataset of 808 subjects. However, this approach is not suitable for portable devices due to its complexity. Avram et al. introduced a CNN model to determine the presence of diabetes using three different sets of data, achieving an average AUC of 0.69 when tested on three different datasets. They also suggested a combination method, using both CNN model results and demographic information as input for a logistic regression model, which improved the model's performance.

Table 1: Literature Survey Summary

Sr. No.	Ref.	Objective	Data Type	Approach	Feature	Main Outcome
1	Bagus et al. [1]	Non-diabetic OR diabetic OR not healthy diabetic	PPG|-] [22/23/17) In-house	Multiscale pointcaré analysis	PPG amplitude	SSR: ratio between long and short variations
2	Pilt et al. [2]	Non-Diabetic or Diabetic	PPG|-1 (24/20) In-house	Statistical analysis	PPG Augmentation Index	PPGAI
3	Moreno et al. [3]	Non-Diabetic or Diabetic	PPG[1min] [340/830] In-house	RF GBoost	Photoplethysmography features + physical data	Accuracy:70 %
4	Hettiarachchi et al. [4]	Non-Diabetic Or Diabetic	PPG[2sec] [83/52] Public	LDA	Photoplethysmography features + physical data	Accuracy:83 %
5	G. Fagherazzi et al. [5]	Non-Diabetic OR diabetic	PPG [30sec] [1467/3411] Public	2D-CNN	PPG scalogram + physio data	Accuracy: 76.34%
6	S. Stern et al. [6]	Non-Diabetic or diabetic	ECG data sampled at 500 Hz	5-layer CNN, LSTM, and SVM	Heart rate variability (HRV)	Accuracy: 95.7%:

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, No.2, June : 2024

7	S. Zanelli et al. [7]	Non-Diabetic or diabetic	PID dataset	Softmax layer	Photoplethysmography features + physical data	Accuracy: 86.26%
8	F. Esgalhado et al. [8]	Non-Diabetic OR diabetic	PPG (pulse) [44/97] In-house	SVM	PPG signal, first and second derivative parameters+ eigenvalues	Accuracy: 97.87%

2.1 Dataset Handling:
The process of developing a robust diabetes prediction model involves acquiring a comprehensive dataset with critical clinical features and labels, such as gender, age, height, weight, blood pressure readings, heart rate, BMI, and hypertension. Data partitioning is then crucial, dividing the dataset into subsets for training, validation, and testing. The training subset is the largest and serves as the foundation for model learning. The validation subset fine-tunes the model, identifying potential overfitting or underfitting issues. The testing subset is the ultimate benchmark for evaluating the model's predictive performance. By leveraging diverse health metrics and strategic data partitioning techniques, healthcare practitioners can enhance diabetes management and improve patient outcomes.

Table 2: Sample Dataset

Sr. No.	Gender	Age	Height	Weight	Systolic Blood Pressure	Diastolic Blood Pressure	Heart Rate	BMI	Hypertension	Diabetes
1	0	45	152	63	161	89	97	27.27	2	0
2	0	50	157	50	160	93	76	20.28	2	0
3	1	80	170	70	173	90	73	24.22	2	0
4	0	66	150	57	182	102	81	25.33	2	1
5	1	44	170	65	110	64	66	22.49	0	1
6	0	59	151	48	139	85	80	21.05	3	1
7	0	48	153	49	126	78	84	20.93	3	0
8	0	53	160	70	108	73	84	27.34	0	0
9	1	53	155	85	141	72	87	35.38	1	1
10	0	47	150	47	98	56	69	20.89	0	0

2.2 Data Cleaning and Preprocessing :
The process of developing a robust diabetes prediction model involves acquiring a comprehensive dataset containing critical clinical features and labels indicating the presence or absence of diabetes. This dataset includes various health indicators such as gender, age, height, weight, blood pressure

readings, heart rate, BMI, and indications of hypertension and diabetes status. Data partitioning is crucial for robust model development and evaluation. The dataset is partitioned into distinct subsets for training, validation, and testing purposes. Data preprocessing involves checking for null values, removing duplicate rows, generating descriptive statistics, and counting unique values in the 'Diabetes' column. A heatmap is created to visualize the correlation between numerical features in the DataFrame df. The training subset serves as the foundation for model learning, allowing the model to discern patterns and relationships between health indicators and diabetes status. The validation subset fine-tunes the model and optimizes hyperparameters, while the testing subset serves as the ultimate benchmark for evaluating the model's predictive performance. By leveraging diverse health metrics and employing strategic data partitioning techniques, healthcare practitioners can harness the power of predictive analytics to enhance diabetes management and improve patient outcomes

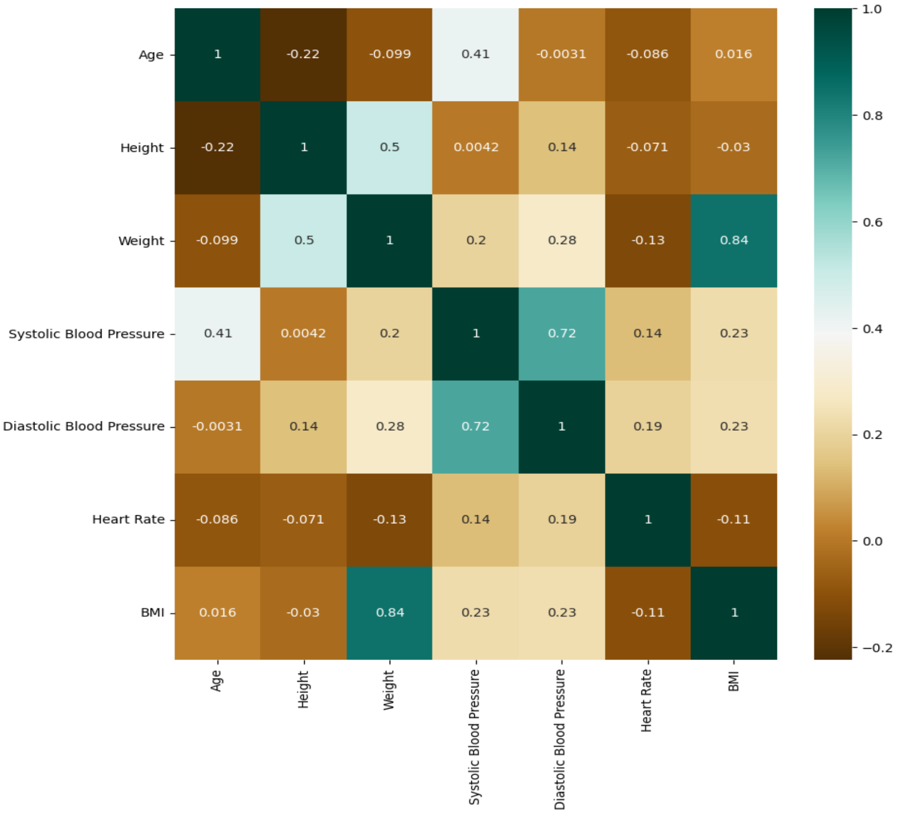


Figure 2: Heatmap for Various Parameter

It is heatmap for various physical parameters. It offers valuable insights into the interdependencies among different variables in the dataset. Each cell in the heatmaprepresents the correlation coefficient between two features, ranging from -1 to 1.

Correlation Formula

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(\bar{x} - x_i)^2 \, \Sigma(\bar{y} - y_i)^2}}$$

Where:

- xi and yi are the individual sample points.
- $\bar{x}$ and $\bar{y}$ are the means of the x and y variables, respectively.
- $\sum$ denotes summation over all sample points.

2.3 Hyper parameter Tuning

A number of hyperparameter tuning procedures for diabetes prediction are described in the work. The first step is to initialize a Logistic Regression model with a fixed random state to ensure reproducibility. The second procedure employs the K-Nearest Neighbors (KNN) model to determine the optimal number of neighbors using 5-fold cross-validation. The third method employs a Decision Tree Classifier to discover the best combination of splitting criterion, maximum tree depth, and minimum samples per leaf via 5-fold cross-validation. The fourth procedure employs a Random Forest Classifier to determine the ideal number of trees and the minimal amount of samples per leaf through 5-fold cross-validation. The fifth process use AdaBoost Classifier to determine the ideal number of boosting stages and learning rate using 5-fold cross-validation.

2.4 Model Selection & Training

The XGBoost classifier is trained using the RandomizedSearchCV object, which performs a randomized search over specified hyperparameter values to find the best combination of hyperparameters for maximum performance. The XGBClassifier is created with specific hyperparameters, such as **n_estimators=50 and learning_rate=0.5.**

xgb.fit(X_train, y_train)

xgb = XGBClassifier(n_estimators= 50, learning_rate= 0.5)

The model is then evaluated and tested using the Accuracy Score Metric, which measures the proportion of correctly predicted observations out of the total observations. The classifier is then fitted on the training data, and predictions are made on the test data. The accuracy of the model is calculated by comparing the true labels (y_test) with the predicted labels (y_pred), providing a measure of the classifier's performance. The actual and predicted labels are printed for visual comparison, allowing for a systematic evaluation and comparison of the classifier's performance. This process allows for the evaluation of different classifiers based on their accuracy scores.

2.5 Implementation Of XGBoost Classifier

params_xgb = {'n_estimators': [50, 100, 250, 400, 600, 800, 1000], 'learning_rate': [0.2, 0.5, 0.8, 1]}

params_xgb is a dictionary that defines the hyperparameters and their respective values to be explored during the tuning process.

n_estimators represents the number of decision trees (estimators) to be built in the XGBoost model. The values to be explored range from 50 to 1000.

learning_rate is a hyperparameter that controls the step size at each iteration. The values to be explored range from 0.2 to 1.

rs_xgb = RandomizedSearchCV(xgb, param_distributions=params_xgb, cv=5)

RandomizedSearchCV is a class from scikit-learn that performs a randomized search over the specified hyperparameters.

xgb is the XGBoost classifier model created in the previous code snippet.

param_distributions=params_xgb specifies the hyperparameters and their respective values to be explored during the tuning process.

cv=5 indicates that a 5-fold cross-validation will be performed during the tuning process**.**

2.6 System Testing For Model Evaluation

Sr No.	Sample Data	Expected	Actual	Test Case Result
1	[45, 152, 63, 161, 89, 97, 27.27, 2]	1	1	Pass
2	[50, 157, 50, 160, 93, 76, 20.28, 2]	1	1	Pass
3	[47, 150, 47, 101, 71, 79, 20.89, 0]	0	0	Pass
4	[45, 172, 65, 136, 93, 87, 21.97, 3]	1	1	Pass
5	[46, 155, 65, 123, 73, 73, 27.06, 3]	0	0	Pass

Table 3 System Testing with 80% Train Data

Table 4 System Testing with 70% Train Data

Sr No.	Sample Data	Expected	Actual	Test Case Result
1	[45, 152, 63, 161, 89, 97, 27.27, 2]	1	1	Pass
2	[50, 157, 50, 160, 93, 76, 20.28, 2]	1	1	Pass
3	[47, 150, 47, 101, 71, 79, 20.89, 0]	0	1	Fail
4	[45, 172, 65, 136, 93, 87, 21.97, 3]	1	1	Pass
5	[46, 155, 65, 123, 73, 73, 27.06, 3]	0	0	Pass

Table 5 System Testing with 50% Train Data

Sr No.	Sample Data	Expected	Actual	Test Case Result
1	[45, 152, 63, 161, 89, 97, 27.27, 2]	1	1	Pass
2	[50, 157, 50, 160, 93, 76, 20.28, 2]	0	1	Fail
3	[47, 150, 47, 101, 71, 79, 20.89, 0]	0	0	Pass
4	[45, 172, 65, 136, 93, 87, 21.97, 3]	1	0	Fail
5	[46, 155, 65, 123, 73, 73, 27.06, 3]	0	0	Pass

Splitting data into appropriate subsets is a fundamental step in preparing a dataset for machine learning. The primary goal is to have separate sets for training, validation, and testing. A common approach is the 70/30 or 80/20 split, where the majority of the data is used for training and the rest for validation and testing. We have divided the dataset into 80/20 ratios for training and testing respectively.

2.7 Experimental Results.

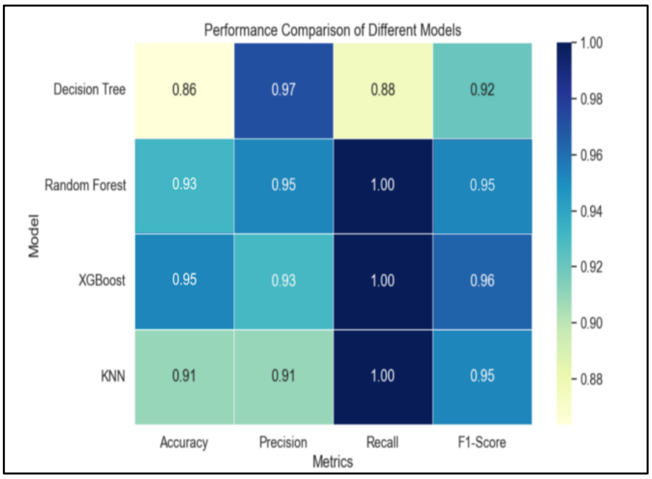


Figure 4: Performance Comparison of Different Models

The above figure is the heatmap, depicting the performance comparison of Decision Tree, Random Forest, XGBoost, KNN. It compares the various metrics like accuracy, precision, recall and F1-score.

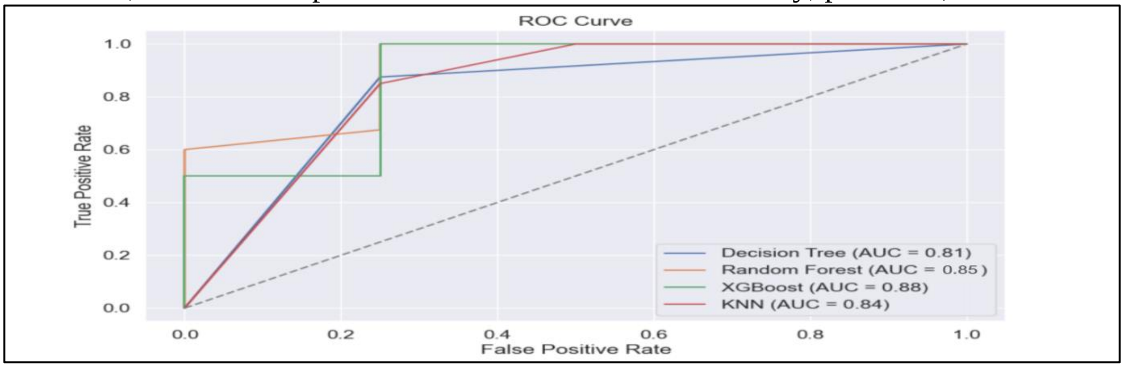


Figure 5: ROC Curve

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, No.2, June : 2024
ROC curve shows the performance of four different machine learning models in classifying diabetic patients. Each model is represented by a different coloured line on the graph. The area under the ROC curve (AUC) is a metric that indicates the overall performance of a classifier. A larger AUC indicates better performance.

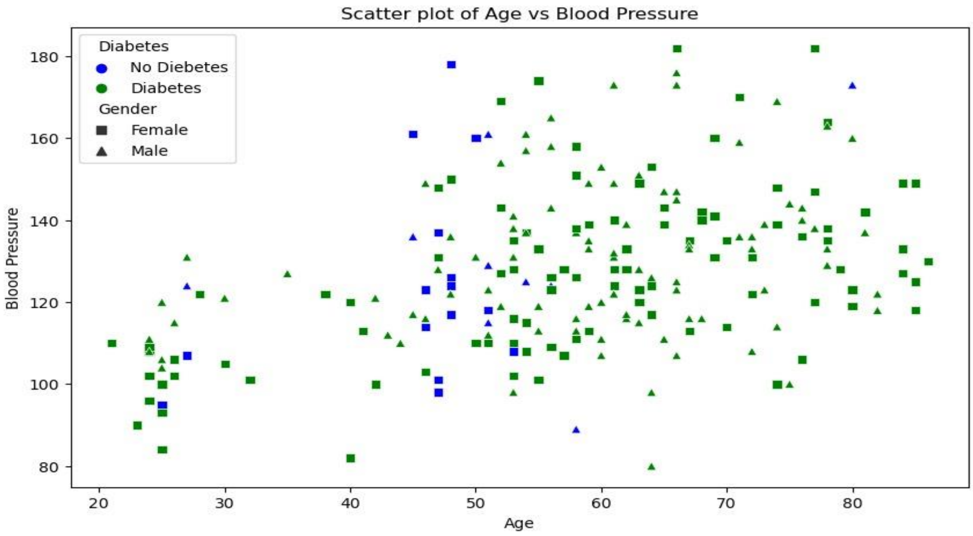


Figure 6: Scatter plot reflecting densely diabetic people at age in 50-70

This is a scatter plot showing the relationship between age and blood pressure. It appears to focus on blood pressure readings of people with and without diabetes. The x-axis of the graph represents age, ranging from 20 to 80 years old. The y-axis represents blood pressure, ranging from 80 to 180. There are two separate data series plotted on the graph, with an “x” symbol for people without diabetes and a circle symbol for people with diabetes. We can see a general upward trend in both sets of data points, which suggests that both with and without diabetes, blood pressure tends to increase with age. The prevalence of diabetes is highest in the age group of 50 to 70 years.

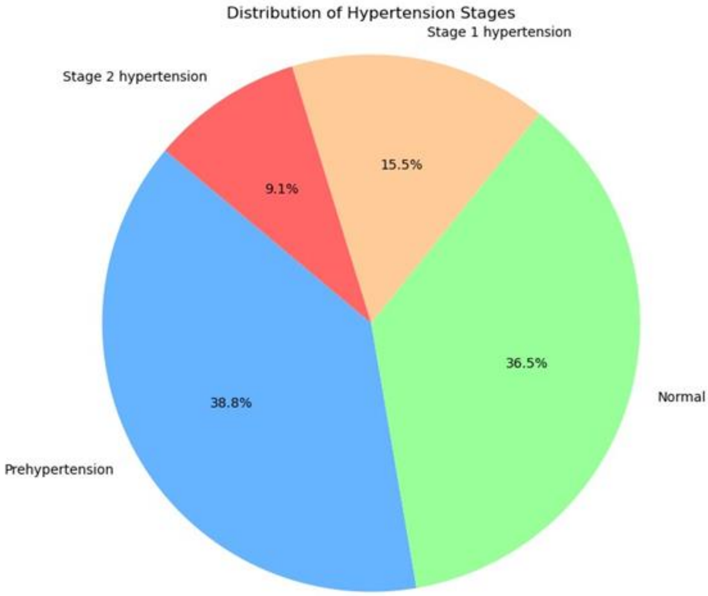


Figure 7: Pie Chart Reflecting Preventive measure

The pie chart shows the distribution of four stages of hypertension: normal, prehypertension, stage 1 hypertension, and stage 2 hypertension. Stage 1 hypertension is the most common stage, accounting for 38.8% of the population. Stage 2 hypertension is the second most common stage, accounting for

36.5% of the population. Prehypertension accounts for 15.5% of the population, and normal blood pressure accounts for 9.1% of the population.

III. Conclusion

This study achieved its objectives of exploring various methodologies for early-stage diabetes detection using PPG signals and classifying diabetic and non-diabetic patients. We successfully investigated the application of four machine learning algorithms: Decision Tree, Random Forest, XGBoost, and KNN for diabetes detection. The Random Forest model achieved the highest accuracy among the models tested, indicating it is the most effective at correctly identifying cases of diabetes. Its ensemble nature, which combines multiple decision trees, likely contributes to its robustness and high performance. The KNN model also performed well, with an accuracy of 0.91. This suggests it is a reliable model for diabetes detection, though slightly less effective than the Random Forest. The performance of KNN can be influenced by the choice of 'k' and the distance metric used. The Decision Tree model had the lowest accuracy at 0.86. While still a useful model, it may be more prone to overfitting compared to the Random Forest. Its simpler structure makes it easier to interpret but less powerful in this context. XGBoost emerged as the most effective algorithm, achieving a remarkable accuracy of 95% in classifying patients based on PPG data. This indicates that PPG signals hold promise for non-invasive diabetes detection. While this study focused on machine learning-based detection, early detection paves the way for timely intervention. This can involve lifestyle changes and medication to manage blood sugar levels and prevent complications associated with diabetes Overall, this study highlights the potential of machine learning, particularly XGBoost, for accurate diabetes detection using PPG signals. Early detection through such methods can empower individuals to take control of their health and prevent complications

References

[1] M. A. Atkinson, G. S. Eisenbarth, and A. W. Michels, ‘‘Type 1 diabetes,’’ Lancet, vol. 383, no. 9911, pp. 69–82, 2014.
[2] S. Chatterjee, K. Khunti, and M. J. Davies, ‘‘Type 2 diabetes,’’ Lancet, vol. 389, no. 10085, pp. 2239–2251, 2017.
[3] T. A. Buchanan and A. H. Xiang, ‘‘Gestational diabetes mellitus,’’ J. Clin. Invest., vol. 115, no. 3, pp. 485–491, 2005.
[4] A. Esteghamati, B. Larijani, M. H. Aghajani, F. Ghaemi, J. Kermanchi, A. Shahrami, M. Saadat, E. N. Esfahani, M. Ganji, S. Noshad, E. Khajeh, A. Ghajar, B. Heidari, M. Afarideh, J. I. Mechanick, and F. Ismail-Beigi, ‘‘Diabetes in Iran: Prospective analysis from first nationwide diabetes report of national program for prevention and control of diabetes (NPPCD-2016),’’ Sci. Rep., vol. 7, no. 1, pp. 1–10, Oct. 2017.
[5] G. Fagherazzi and P. Ravaud, ‘‘Digital diabetes: Perspectives for diabetes prevention, management and research,’’ Diabetes Metabolism, vol. 45, no. 4, pp. 322–329, Sep. 2019.
[6] A. N. Long and S. Dagogo-Jack, ‘‘Comorbidities of diabetes and hypertension: Mechanisms and approach to target organ protection,’’ J. Clin. Hypertension, vol. 13, no. 4, pp. 244–251, Apr. 2011.
[7] S. Zanelli, M. Ammi, M. Hallab, and M. A. El Yacoubi, ‘‘Diabetes detection and management through photoplethysmographic and electrocardiographic signals analysis: A systematic review,’’ Sensors, vol. 22, no. 13, p. 4890, Jun. 2022.
[8] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, ‘‘Deep learning for healthcare applications based on physiological signals: A review,’’ Comput. Methods Programs Biomed., vol. 161, pp. 1–13, Jul. 2018.
[9]] F. Esgalhado, B. Fernandes, V. Vassilenko, A. Batista, and S. Russo, ‘‘The application of deep learning algorithms for PPG signal processing and classification,’’ Computers, vol. 10, no. 12, p. 158, Nov. 2021.

[10] C. C. Hettiarachchi ansd C. Chitraranjan, ''A machine learning approach to predict diabetes using short recorded photoplethysmography & physiological characteristics,'' in Proc. 17th Conf. Artif. Intell. Med. (AIME). Poznan, Poland: Springer, Jun. 2019, pp. 322–327.
[11] K. Pilt, K. Meigas, R. Ferenets, K. Temitski, and M. Viigimaa, ''Photoplethysmographic signal waveform index for detection of increased arterial stiffness,'' Physiolog. Meas., vol. 35, no. 10, pp. 2027–2036, Oct. 2014.
[12] E. M. Moreno, M. J. A. Lujan, M. T. Rusinol, P. J. Fernandez, P. N. Manrique, C. A. Trivino, M. P. Miquel, M. A. Rodriguez, and M. J. GonzalezBurguillos, ''Type 2 diabetes screening test by means of a pulse oximeter,'' IEEE Trans. Biomed. Eng., vol. 64, no. 2, pp. 341–351, Feb. 2017.
[13] H. Bagus, L. Gen-Min, Y. Chieh-Ming, C. Shiao-Chiang, and W. Hsien-Tsai, ''Poincaré plot of fingertip photoplethysmogram pulse amplitude suitable to assess diabetes status,'' in Proc. 4th Int. Conf. Electr. Eng., Comput. Sci. Informat. (EECSI), Sep. 2017.
[14] V. R. Reddy, A. D. Choudhury, S. Jayaraman, N. K. Thokala, P. Deshpande, and V. Kaliaperumal, ''PerDMCS: Weighted fusion of PPG signal features for robust and efficient diabetes mellitus classification,'' in Proc. 10th Int. Joint Conf. Biomed. Eng. Syst. Technol., 2017, pp. 553–560.
[15] A. Prabha, J. Yadav, A. Rani, and V. Singh, ''Design of intelligent diabetes mellitus detection system using hybrid feature selection based XGBoost classifier,'' Comput. Biol. Med., vol. 136, Sep. 2021, Art. no. 104664.
[16] S. Masaoka, A. Lev-Ran, L. Robert Hill, G. Vakil, and E. H. G. Hon, ''Heart rate variability in diabetes: Relationship to age and duration of the disease,'' Diabetes Care, vol. 8, no. 1, pp. 64–68, Jan. 1985.
[17] S. Stern and S. Sclarowsky, ''The ECG in diabetes mellitus,'' Circulation, vol. 120, no. 16, pp. 1633–1636, Oct. 2009. [27] H. Njoum and P. A. Kyriacou, ''Photoplethysmography for the assessment of haemorheology,'' Sci. Rep., vol. 7, no. 1, pp. 1–11, May 2017.
[18] N. R. Anderson, E. S. Lee, J. S. Brockenbrough, M. E. Minie, S. Fuller, J. Brinkley, and P. Tarczy-Hornoch, ''Issues in biomedical research data management and analysis: Needs and barriers,'' J. Amer. Med. Informat. Assoc., vol. 14, no. 4, pp. 478–488, Jul./Aug. 2007.
[19] R. Avram, G. Tison, P. Kuhar, G. Marcus, M. Pletcher, J. E. Olgin, and K. Aschbacher, ''Predicting diabetes from photoplethysmography using deep learning,'' J. Amer. College Cardiol., vol. 73, no. 9, p. 16, Mar. 2019.
[20] M. Zheng, X. Zhang, S. Chen, Y. Song, Q. Zhao, X. Gao, and S. Wu, ''Arterial stiffness preceding diabetes: A longitudinal study,'' Circulat. Res., vol. 127, no. 12, pp. 1491–1498, Dec. 2020 [21] A. Kamal, J. Harness, G. Irving, and A. Mearns, ''Skin photoplethysmography—A review,'' Comput. Methods Programs Biomed., vol. 28, no. 4, pp. 257–269, 1989.
[22] J. Allen, ''Photoplethysmography and its application in clinical physiological measurement,'' Physiolog. Meas., vol. 28, no. 3, pp. R1–R39, Mar. 2007.
[23] V. B. Srinivasan and F. Foroozan, ''Deep learning based non-invasive diabetes predictor using photoplethysmography signals,'' in Proc. 29th Eur. Signal Process. Conf. (EUSIPCO), Aug. 2021, pp. 1256–1260.
[24] C. Irace, C. Carallo, F. Scavelli, M. S. De Franceschi, T. Esposito, and A. Gnasso, ''Blood viscosity in subjects with normoglycemia and prediabetes,'' Diabetes Care, vol. 37, no. 2, pp. 488–492, Feb. 2014.
[25] Y. Liang, Z. Chen, G. Liu, and M. Elgendi, ''A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China,'' Sci. Data, vol. 5, no. 1, pp. 1–7, Feb. 2018
[26] N. Nirala, R. Periyasamy, B. K. Singh, and A. Kumar, ''Detection of type-2 diabetes using characteristics of toe photoplethysmogram by applying support vector machine,'' Biocybern. Biomed. Eng., vol. 39, no. 1, pp. 38–51, Jan. 2019.
[27] H. Njoum and P. A. Kyriacou, ''Photoplethysmography for the assessment of haemorheology,'' Sci. Rep., vol. 7, no. 1, pp. 1–11, May 2017.