

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/357648143>

Diabetes prediction and analysis using medical attributes: A Machine learning approach

Research · January 2022

DOI: 10.37896/JXAT14.01/314405

CITATION

1

READS

1,160

2 authors:



Minakshi Ravindra Rajput
P.E.S College Of Engineering

14 PUBLICATIONS 17 CITATIONS

[SEE PROFILE](#)



Sushant Suhas Khedgikar
P.E.S College Of Engineering

1 PUBLICATION 1 CITATION

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Gender identification from human iris [View project](#)

Diabetes prediction and analysis using medical attributes: A Machine learning approach

Minakshi R. Rajput^{1*}, Sushant S. Khedgikar²

¹Department of Electronics & Telecommunication, P.E.S. College of Engineering, Aurangabad (MH)

²Department of Electronics & Telecommunication, P.E.S. College of Engineering, Aurangabad (MH)

Abstract

Diabetes is a serious health disease due to the presence of high glucose levels in the human body. If diabetes is untreated, it can cause other critical health issues in person. The main aim of the paper is to predict diabetes by analyzing different human body attributes. The proposed methodology adopts five different types of machine learning algorithms for diabetes prediction. This paper discusses important attributes responsible for inducing diabetes. It also computes the correlation between them. It also sheds light on the relative risk

I. INTRODUCTION

Diabetes is a chronic health condition in which the capacity of producing insulin decreases, which results in serious health problems like heart diseases, kidney diseases, vision loss, etc. Pre-diabetes is also called borderline diabetes characterized by blood glucose levels more than normal but not yet high enough to be classified as diabetes. Increasing cases of pre-diabetes is a global issue that will put more burdens on healthcare in near future. Despite this, pre-diabetes will completely make you unaware without giving any sign, it gradually leads to diabetes.

HbA1c (6.0-6.4%), fasting blood sugar (5.5mmol/L to 6.9mmol/L) tests are the indications of pre-diabetes. Making changes in diet and proper physical exercise will slow down the progress of entering into type 2 diabetes [1, 2]. Ability of artificial intelligence and machine learning algorithms to analyze complex datasets helps medical practitioners in the early prediction of diseases. This also helps in the advanced care of patients and improving outcomes of health care. Maximum utilization of AI for clinical decisions, risk scoring, and early alerting are the most promising areas of development of data analysis. Many researchers have proposed different machine learning and deep learning methods for the classification and prediction of early diabetes.

The main focus of this paper is predicting diabetes from patient's medical attributes and finding risk values for patients entering from no-diabetes to pre-diabetes and pre-diabetes to diabetes. The paper also throws light on significant machine learning methods that robustly classify patients into one of the three classes: 1. No-diabetes, 2. Pre-Diabetes, 3. Diabetes.

II. BACKGROUND AND RELATED WORK

As age grows, the diagnostic efficiency of HbA1c for diabetes decreases as a result of decreasing RBC count [3]. There is a likelihood of developing type 2 diabetes for age above 45 years. But in some cases, people suffer from diabetes without even knowing that they are suffering from pre-diabetic symptoms. Thus, there is a huge variation between age and diagnosis of diabetes [4]. Recent developments in diabetes studies concluded that

values of developing diabetes in a normal and pre-diabetic person. The paper also compares the performance of different classifiers based on different evaluation metrics. Experimental results show that decision tree and stochastic gradient boosting algorithms outperformed achieving better accuracy as compared to other machine learning algorithms.

Keywords

Diabetes prediction, machine learning, correlation analysis, Mendeley dataset

BMI and HbA1c have a positive correlation. The experimentations were conducted on 100 diabetic patients. Patients having BMI 25-29.9 and obese patients have very poorly controlled HbA1c (6.5-10%) which increases the probability of developing type 2 diabetes. It is essential to maintain HbA1c lower than 7% and BMI lower than 25 for preventing diabetic kidney disease (DKD), and cardiovascular disease (CVD) [5, 6]. Also, a high lean mass index accompanying a high-fat mass index induces type 2 diabetes. Type 1 diabetes is generally observed in children and adolescents who are often obese, overweight, and have a high percentage of lipid disorders [7].

In spite of having effective medications for diabetes, many patients continue to have elevated levels of HbA1c. The barrier to diabetic management includes factors like depression, physical disability/inactivity, lack of information or negligence of importance of HbA1c and LDL, medication side effects, social stigma, cognitive impairment, lack of knowledge in maintaining diet and weight loss, etc. These issues mainly lead to severe hyperglycemia. It also degrades life quality causing nephropathy, neuropathy, and retinopathy, etc. [8, 9]. Diabetes-related medical advice and discussions should be gender-specific. These discussions must include the importance of medication, the impact of lifestyle on diabetes-related risk factors. This will definitely help to slow down the spread of diabetes and also help to minimize the economic burden of disease on patients as well as the healthcare system [10].

Reviewing past literature, it is observed that various approaches have been proposed for the classification and early prediction of diabetes using artificial intelligence, machine learning and deep learning methods. Mitushi Soni and co-author experimented on the Pima Indian Diabetes dataset having attributes of 768 patients. They developed a model to predict diabetes using various machine learning techniques such as SVM, KNN, Decision tree, Random forest, logistic regression, and gradient boosting. They achieved 77% classification accuracy [11]. Asir Singh and co-author also worked on the Pima dataset. They have used a correlation-based feature selection technique to remove irrelevant features. They employed function-based MLP, probabilistic-based Naïve Bayes, decision tree-based random forest

algorithms. They got 79.69% prediction accuracy for the Naïve Bayes method [12]. Nnamoko, N et al. [13] utilized the capability of the ensemble learning method on the UCI diabetes dataset, and a meta-classifier was used to aggregate results obtained from the individual classifiers. They produced 83% accuracy with feature subset selection. T. Joshi and coauthor [14] implemented an early diabetes prediction model using logistic regression, SVM, and ANN networks. The authors claimed that they obtained better accuracy using these techniques. Deepti Sisodia and co-author performed experiments on the PIDD dataset using Naïve Bayes, SVM, and decision tree. Naïve Bayes outperformed with an accuracy of 76.30% [15].

Deeraj Shetty et al. [16] developed an intelligent diabetic prediction system using KNN and Bayesian algorithms. Nongyao Nai-arun and coauthor [17] experimented on a dataset of 30,122 patients consisting of 12 attributes per patient. They implemented web application using ANN, decision trees, logistic regression, Naïve Bayes and random forest algorithm. This application classifies patients' data into either diabetes or normal group. They got an accuracy of 85.55% with random forest. Author Nahla Barkat and Andrew Bradely [18] employed SVM based hybrid model for diagnosis and prediction of diabetes. They worked on a dataset consisting of 3014 patients with 11 different attributes per patient. They achieved a prediction accuracy of 94%, sensitivity of 93%, and specificity of 94%. They extracted rules by SQR ex- SVM for the diagnosis of diabetes. They predicted major risk factors in predicting diabetes. They are namely FBS (Fasting Blood Sugar), waist circumferences, and BPDIA (Diastolic Blood Pressure). Majji, Ramachandra, and co-authors [19] created a website that gives the probability of developing diabetes in near future based on the risk score. The different parameters used for risk calculation included Age, Smoking Status, Steroid usage, BMI (body mass index), Ethnicity, etc. They have used the WEKA tool for classification.

Few researchers have attempted to predict diabetes using deep learning methods. (Convolutional Long Short-term Memory) CLSTM is implemented for the early prediction of diabetes on the Pima Indian Diabetes Database (PIDD). Authors achieved 95.6% accuracy [20]. Comparing Deep learning CNN with machine learning approaches on PIDD database, authors got an accuracy of 76.81% with CNN which is quite lesser than the accuracy of 83.67% with the Random Forest method [21]. A deep neural network is trained on two datasets 1. Diabetes type

2. PIDD. Results obtained for Diabetes dataset are 94% and 99.4% for PIDD [22]. Huma Naz and Sachin Ahuja applied deep learning to PIDD. They obtained promising accuracy of 98.7% for diabetes prediction [23]. Artificial back propagation neural network (ABPNN) obtained 93% accuracy on PIDD [24].

Our proposed methodology is implemented with five machine learning algorithms. It computes a relative risk value of having diabetes in the near future for a normal and pre-diabetic person.

III. PROPOSED METHODOLOGY

Diabetic patients directly or indirectly become victims of other major health issues such as heart attack, vision loss, and kidney diseases. This is because of losing the ability of cells to respond to insulin and much of the blood sugar gets accumulated in the bloodstream. To

predict diabetes of the person using different machine learning methods, the system adopts a series of steps as shown in figure 1.

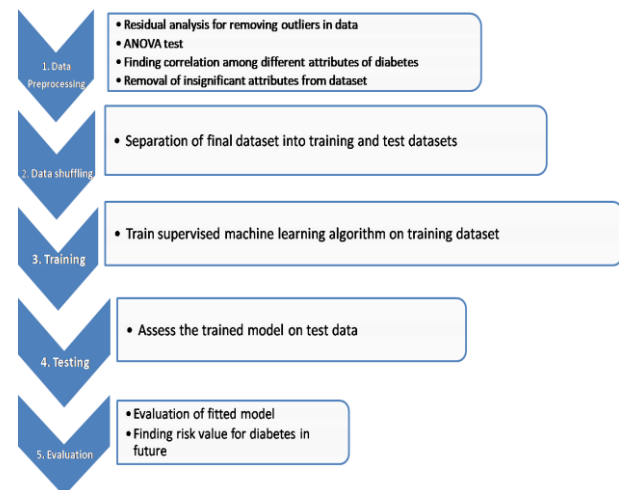


Figure 1. Sequential steps in the proposed system

Dataset collection and representation

The dataset utilized in the system is retrieved from the Mendeley Diabetes types dataset [25]. This data was collected from the Iraqi society from the laboratory of Medical City Hospital and (the Specialized Center for Endocrinology and Diabetes-Al-Kindy Teaching Hospital). Data is extracted from the patient's file and entered into the database. The data consist of medical information, laboratory analysis. The database includes 103 (no-diabetes), 53 (pre-diabetic) and 844 (diabetic) patients.

Table 1. Details of Diabetes type dataset

Sr. No.	Attribute	Description
1	Gender	Male or Female
2	Age	In years (min:20 , max : 79)
3	Urea	In Mg/dl(min: 0.5 , max:38.9)
4	Creatinine ratio	μmol/L (min:48, max:80)
5	Body Mass Index	(min:19 , max:47)
6	LDL	mmol/L (min:0.3, max:9.9)
7	Cholesterol	mmol/L (min:0.0, max:10.3)
8	VLDL	mmol/L (min:0.1, max:35)
9	Triglycerides(TG)	mmol/L (min:0.3, max:13.8)
10	HDL	mmol/L (min:0.2, max:9.9)
11	HBA1C	mmol/L (min:0.9, max: 16)
12	Class	Diabetic, Non-Diabetic, or Predict-Diabetic

Data Pre-processing

The dataset contains data for 1000 patients. It contains different diabetes attributes like the number of patients, Sugar Level Blood, Age, Gender, Creatinine ratio (Cr), Body Mass Index (BMI), Urea, Cholesterol (Chol), etc. Table 1 shows details about this dataset. Careful selection of these features should be done, as any irrelevant feature may mislead the results.

Removing outliers from data is a crucial part of data analysis as outliers give insignificant statistical results and can violate your assumptions. At first, Box plot is applied to a database for the removal of any outliers

present in the data. We removed all outliers which are falling outside the box (minimum and maximum values of Box). The outliers in this dataset are wrong data entered mistakenly in the database.

A. Correlation Analysis

The next step is to check whether diabetes features are correlated or not. Pearson's correlation analysis was performed to check if features are suitable for the prediction of diabetes or not. It is observed that most correlated features are age, HbA1c and BMI with Pearson's coefficient values $\rho_{\text{Age}} = 0.52$, $\rho_{\text{HbA1c}} = 0.51$, and $\rho_{\text{BMI}} = 0.55$. Also, other insignificantly correlated features are Urea ($\rho = 0.11$), Chol ($\rho = 0.12$), and TG ($\rho = 0.21$). Correlation between different attributes (between features correlation) was also analyzed. It showed positive correlation between Age and HbA1c ($\rho = 0.43$), Age and BMI ($\rho = 0.38$), Urea and Cr ($\rho = 0.64$), HbA1c and BMI ($\rho = 0.37$).

Table 2. F-Statistics results of ANOVA test

	Age	Urea	Cr	HbA1c	Chol	TG	LDL	HDL	VLDL	BMI
F-value	394.09	2.288	0.923	155.81	1.651	2.95	2.364	1.26	15.977	131.71
Pr(>F)	< 2e-16	0.131	0.337	< 2e-16	0.199	0.09	0.125	0.26	7.04E-05	< 2e-16

Further interaction ANOVA was done to see if significant features have any combined effect in determining diabetes. Results showed that the AGE: Gender variable has a high Sum Sq value (6.77) and low P-value (4.87e-07), which means there is much variation that can be explained by the interaction between AGE and Gender. In the Interaction plot for this test, gender is considered as a trace factor, and age and diabetes variables are on the X and Y-axis respectively. The interaction plot is shown in figure 2. As shown two lines in the plot are intersecting each other, which means that there is an interactive effect of gender and age on gradually developing diabetes in a patient. Therefore, gender attribute is also included during model training. It shows that there is a different effect of diabetes on males and females. In other words, diabetes does not equally affect males and females.

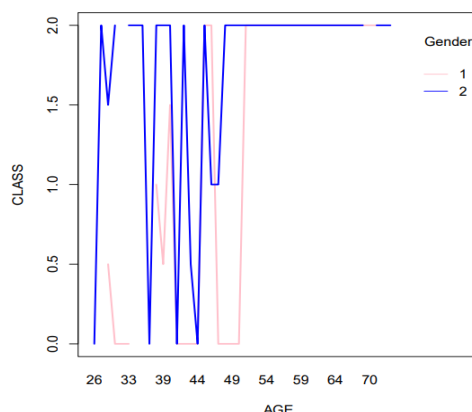


Figure 2. Interaction plot for ANOVA test

B. Feature selection by ANOVA

In the next step, ANOVA F-test was done, as the dataset contains two categorical variables 1. Diabetes class (response variable), and 2. Gender (predictor variable). ANOVA test states that, for a given predictor variable, if mean-variance exists between the groups, there is an impact of that variable on the response variable. This variable can be included for further model building. The table shows the results of multi-factor ANOVA on the diabetes dataset after excluding all outliers. Attributes like Age, HbA1c, VLDL, and BMI have Pr values less than that of alpha value, leading us to reject the null hypothesis. As compared to the remaining features, these four features have a major impact in determining diabetes class and are more significant. Thus we promoted these features for model training.

In the proposed system, five machine learning algorithms are implemented to train a diabetes prediction model. Implementation is done in the R programming environment (R studio version 1.14.1717). These ML algorithms are namely 1. Multinomial logistic regression, 2. Naïve Bayes, 3. Decision trees, 4. Random forest, and 5. Stochastic gradient boosting. The 80% of the total samples in each target class are considered for training and the remaining 20% are considered for testing the trained model.

C. Model building and deployment

1. Multinomial logistic regression

As the target class is non-ordinal and has three values, multinomial type of logistic regression was chosen for implementation. Out of three target classes, diabetes is chosen as the base value for two others (no-diabetes and pre-diabetes). Experimental output resulted in two models which are as below.

Model1: Log of Probability of Somebody having no diabetes over diabetes.

$$\ln\left(\frac{p(\text{No-diabetes})}{p(\text{Diabetes})}\right) = 40.608 + (-0.29)(\text{gender} = M) + (-0.022)(\text{AGE}) + (-1.7)(\text{HbA1c}) + (-0.238)(\text{VLDL}) + (-1.26)(\text{BMI})$$

Model2 : Log of Probability of Somebody having pre-diabetes over diabetes

$$\ln\left(\frac{p(\text{pre-Diabetes})}{p(\text{Diabetes})}\right) = 23.65 + (1.0)(\text{gender} = M) + (-0.047)(\text{AGE}) + (-0.043)(\text{HbA1c}) + (-0.49)(\text{VLDL}) + (-0.89)(\text{BMI})$$

The log function used in this technique is a linear function of five independent variables (6 coefficients = 1 intercept + 5 slope values). The exponential beta coefficients obtained with the log functions depict the 'relative risk'. For no diabetes class with a BMI coefficient of -1.2664647, this value is the odds or relative risk ratio for one unit increase in BMI value. It states that, if the BMI of a person will increase by one unit, the log of probability for somebody having no diabetes over diabetes will decrease by the amount 1.2664647. In other words,

the relative risk for that person (with no diabetes) is 1.2664647 times more likely to go to diabetes class. Similar conclusions can also be made for all the four attributes Age, HbA1c, VLDL, and BMI for the classes no diabetes and pre-diabetes with respect to diabetes class. From these results, it is clear that the dominant factors for the associated risk to go to diabetes class are BMI and HbA1c.

For the coefficient values of the Gender parameter, the relative risk for males in pre-diabetes class to go to diabetes class is less as compared to females by 1.0016127 units. The risk is more for males in no diabetes class to go to the next target class as compared to females by 0.29 units. From model 1 it is clear that log odds of being in no-diabetes versus in diabetes will decrease by 0.29 from male to female. Similarly, log odds of being in no-diabetes versus in diabetes will decrease by 0.022(2%), when age is increased by one unit. HbA1c and BMI are the most critical factors impacting diabetes.

2. Decision Tree

As shown in the decision tree model in the figure 3, the model splits diabetes data multiple times according to threshold values of BMI, HbA1c, and VLDL. Variable importance plot for decision tree shows that HbA1c is a primary important parameter in predicting diabetes. Subsequent important parameters are BMI, Age, and VLDL respectively.

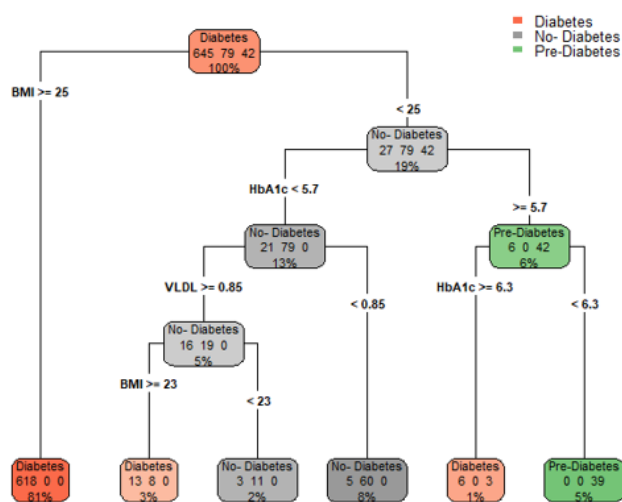


Figure 3: Decision tree plot

3. Random Forest

It chooses the features randomly during the training process. This method produces a forest of decision trees, giving better prediction results. During implementation the number of decision trees in the forest was kept 500 and the number of features randomly used for each split is 2. Variable importance plot for random forest classifier is shown in figure 4.

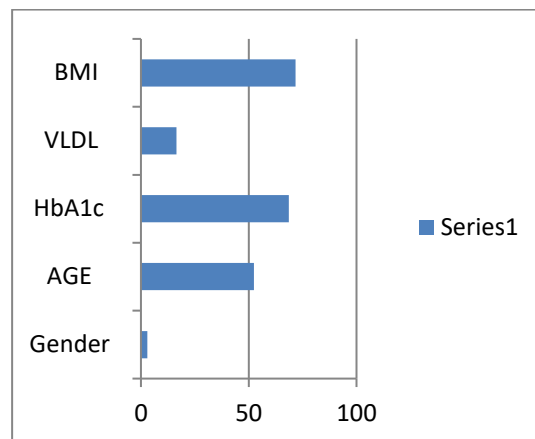


Figure 4. Variable Importance plot for Random Forest
As per the results obtained for this method, the most important feature parameter influencing diabetes is BMI followed by HbA1c and Age. The accuracy obtained is 90.64%.

4. Stochastic gradient boosting

Shortcomings of the decision tree are improved in stochastic gradient boosting by adding new models to the ensemble sequentially. We trained a model with 10 fold cross-validation approach. The variable importance plot obtained for this model is shown in figure 5. It is seen that BMI, HbA1c and Age are dominant attributes responsible for inducing diabetes in a person.

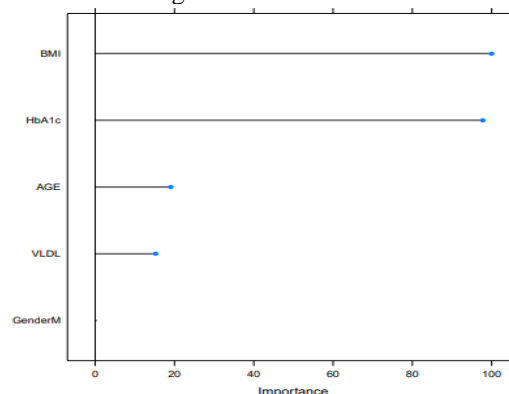


Figure 5. Variable Importance plot for Stochastic gradient boosting

5. Naïve Bayes

This algorithm calculates the posterior probability for each class considering features are not correlated with each other and are independent. The class with the highest probability is the outcome of the predictor. We trained the model using 5 fold cross-validation approach.

IV. RESULTS AND FINDINGS

As the problem is a multi-classification problem, accuracy alone cannot define the performance of the classifier. Hence we computed other performance metrics for all implemented classifiers. Table 3 lists metric values obtained for all implemented classifiers.

Table 3: Comparison of classifiers based on different performance metrics

Performance Metrics	Multinomial logistic Regression	Decision Tree	Random Forest	Stochastic gradient Boosting	Naïve Bayes
Overall Accuracy	0.867	0.9507	0.9064	0.9704	0.931
Recall	0.7	0.74621	0.78	0.81	0.7186
Precision	0.7	0.9812	0.75	0.9885	0.89
Balanced Accuracy	0.781	0.8258	0.8445	0.88	0.804
F1-Score	0.7	0.8967	0.764	0.89	0.795
No-information Rate	0.7882	0.8276	0.8276	0.8276	0.8276
P-Value	0.002664	1.10E-07	0.001067	1.78E-10	1.33E-05
Kappa	0.5969	0.8113	0.7053	0.8926	0.7364

Precision gives the proportion of predicted positives that are actually positive. Recall refers to the proportion of actual positives that are correctly classified. Precision and recall have a trade-off relationship, one has to be optimized at the cost of the other. F1-score is the best metric for evaluating the performance of the multiclass classifier as it is a harmonic mean of precision and recall. It equally minimizes both false positives and false negatives.

Stochastic gradient boost and decision tree classifier have achieved better accuracies of 97.04% and 95.07% respectively. Balanced accuracies obtained for these two classifiers are 88% and 82.58%. F1 scores are also satisfactory as compared to other methods. Kappa score measures the degree of agreement between two classifiers predicted labels and actual labels. It shows how good classifiers' predictions are matched actual class while controlling random accuracy. Kappa score obtained for decision tree and the stochastic boosting algorithm is above 0.8 and fairly good.

From the variable importance plots obtained in figure 4 and 5, it is clear that adult persons with high BMI and HbA1c values than normal range significantly increased risk of type 2 diabetes. These three are dominant and influencing attributes for increasing risk of diabetes. Risk of developing diabetes is more for persons having BMI above 23. Reducing body weight by appropriate body exercise and healthy diet lowers risk of diabetes and thereby preventing future critical health diseases related to it.

V. CONCLUSIONS AND FUTURE SCOPE

The main of the paper was to predict diabetes based on different human body attributes. We implemented five machine learning algorithms namely multinomial logistic regression, decision tree, random forest, stochastic gradient boosting, and Naïve Bayes. We obtained satisfactory results for the decision tree and stochastic gradient boosting algorithm. Our results led us to conclude that high Body Mass Index (BMI) and growing age are major factors in the development of risk for diabetes. Also, both of these are positively correlated with each other. The findings of this study indicate that with increasing BMI and age, other attributes like HbA1c, VLDL, Chol, etc. also cross their threshold values increasing chances of diabetes.

In spite of the results, the current study was limited to a small sample size and cannot be generalized.

We hope that further experimentations on a large and variety of diabetes datasets will confirm our findings.

We believe that our research will add to the growing body of literature on diabetes.

References

- [1] Diabetes, Centers for Disease Control and Prevention, <https://www.cdc.gov/diabetes/basics/getting-tested.html>, May 2019.
- [2] Prediabetes (Borderline Diabetes), <https://www.diabetes.co.uk/pre-diabetes.html>, Jan 2019.
- [3] Li Wu et al., "Effect of age on the diagnostic efficiency of HbA1c for diabetes in a Chinese middle-aged and elderly population: The Shanghai Changfeng Study", *LoS ONE* 12(9): e0184607. <https://doi.org/10.1371/journal.pone.0184607>
- [4] Kristeen Cherney, Age of Onset for Type 2 Diabetes: Know Your Risk, online article <https://www.healthline.com/health/type-2-diabetes-age-of-onset>.
- [5] Sisodia RK et al. The study of correlation between Body Mass Index and glycemic control-HbA1c in diabetes type 2 patients, *Int J Adv Med*. 2019 Dec;6(6):1788-1791
- [6] 19. Nakanishi, Shuhei et al., Comparison of HbA1c levels and body mass index for prevention of diabetic kidney disease: A retrospective longitudinal study using outpatient clinical data in Japanese patients with type 2 diabetes mellitus. *Diabetes Research and Clinical Practice*, Volume 155, 107807, august 2019.
- [7] Katarzyna Noras,1 Ewa Rusak,2 and Przemysław Jarosz-Chobot2 ,The of Abnormal Body Weight and Dyslipidemia as Risk Factors for Cardiovascular Diseases in Children and Adolescents with Type 1 Diabetes, *Journal of Diabetes Research*, August 2021.
- [8] Lynne Chepulis , 1 Brittany Morison,1 Shemana Cassim , 1 Kimberley Norman,1 Rawiri Keenan,1 Ryan Paul,1,2 and Ross Lawrenson1,2Barriers to Diabetes Self-Management in a Subset of New Zealand Adults with Type 2 Diabetes and Poor Glycaemic Control, *Journal of Diabetes Research*, August 2021.
- [9] Matloub F (2021) Barriers to Self-Care among Diabetic Patients and Ways to Recognize and Address. *J Fam Med Dis Prev* 7:135. ISSN: 2469-5793 doi.org/10.23937/2469-5793/1510135, January 2021.
- [10] Duncan LA, Hornsby P, Pannone A, Ebrahim OW, Ibrahim A (2021) Demographic, Behavioral, and Clinical Characteristics of the Ambulatory Diabetic Population of

the Dominican Republic, 2018. *Int J Diabetes Clin Res* 8:143. ISSN: 2377-3634, doi.org/10.23937/2377-3634/1410143, May 2021

[11] Mitushi Soni , Dr. Sunita Varma, 2020, Diabetes Prediction using Machine Learning Techniques, *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)* Volume 09, Issue 09 (September 2020).

[12] Dr. D. Asir Antony Gnana Singh, Dr. E. Jebamalar Leavline, B. Shanawaz Baig, "Diabetes Prediction Using Medical Data", *Journal of Computational Intelligence in Bioinformatics* ISSN 0973-385X Volume 10, Number 1 (2017) pp. 1-8.

[13] Nnamoko, N., Hussain, A., & England, D. (2018). *Predicting Diabetes Onset: An Ensemble Supervised Learning Approach*. 2018 *IEEE Congress on Evolutionary Computation (CEC)*. doi:10.1109/cec.2018.8477663.

[14] Tejas N. Joshi & Pramila M. Chawan, Diabetes Prediction Using Machine Learning Techniques, *Int. Journal of Engineering Research and Application* www.ijera.com ISSN: 2248-9622, Vol. 8, Issue 1, (Part - II) January 2018, pp.-09-13.

[15] Deepti Sisodia, Dilip Singh Sisodia, Prediction of Diabetes using Classification Algorithms, *Procedia Computer Science*, Volume 132, 2018, Pages 1578-1585, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2018.05.122>.

[16] Shetty, D., Rit, K., Shaikh, S., & Patil, N. (2017). Diabetes disease prediction using data mining. 2017 *International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*. doi:10.1109/iciiecs.2017.8276012.

[17] Nongyao Nai-arun, Rungruttikarn Moungrmai, Comparison of Classifiers for the Risk of Diabetes Prediction, *Procedia Computer Science*, Volume 69, 2015, Pages 132-142, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2015.10.014>.

[18] N. Barakat, A. P. Bradley and M. N. H. Barakat, "Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 4, pp. 1114-1120, July 2010, doi: 10.1109/TITB.2009.2039485.

[19] Majji, Ramachandro and Bhramaramba Ravi. "Type 2 Diabetes Classification and Prediction Using Risk Score.", *International Journal of Pure and Applied Mathematics* Volume 119 No. 15 2018, 1099-1111 (2018).

[20] P. Bharath Kumar Chowdary & Dr. R. Udaya Kumar , "An Effective Approach for Detecting Diabetes using Deep Learning Techniques based on Convolutional LSTM Networks", *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 12, No. 4, 2021.

[21] A. Yahyaoui, A. Jamil, J. Rasheed and M. Yesiltepe, "A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques," 2019 1st *International Informatics and Software Engineering Conference (UBMYK)*, 2019, pp. 1-4, doi: 10.1109/UBMYK48245.2019.8965556.

[22] Zhou, H., Myrzashova, R. & Zheng, R. Diabetes prediction model based on an enhanced deep neural network. *J Wireless Com Network* 2020, 148 (2020). <https://doi.org/10.1186/s13638-020-01765-7>.

[23] Naz, H., Ahuja, S. Deep learning approach for

diabetes prediction using PIMA Indian dataset. *J Diabetes Metab Disord* 19, 391–403 (2020). <https://doi.org/10.1007/s40200-020-00520-5>.

[24] Muhammad Mazhar Bukhari, Bader Fahad Alkhamees, Saddam Hussain, Abdu Gumaiei, Adel Assiri, Syed Sajid Ullah, "An Improved Artificial Neural Network Model for Effective Diabetes Prediction", *Complexity*, vol. 2021, Article ID 5525271, 10 pages, 2021. <https://doi.org/10.1155/2021/5525271>.

[25] Rashid, Ahlam (2020), "Diabetes Dataset", Mendeley Data, V1, doi: 10.17632/wj9rwkp9c2.1