Predicting the Onset of Diabetes Using Multimodal Data and a Novel Machine Learning Method

Manjushree Nayak and Jagannath Tiyadi

May 13, 2023

Predicting the Onset of Diabetes using Multimodal Data and a Novel Machine Learning Method

Dr.Manjushree Nayak,Jagannath Tiyadi
Department of Computer Science and Engineering,
NIST Institute Of Science and Technology(Autonomous,Berhampur,
Odisha-761008)

drmanjushreemishra@gmail.com

jtc020@nist.edu

Abstract: Diabetes is a chronic or Deep-seated, Diabetes is a disease that occur when blood glucose is too high. Blood glucose is the source of Energy and it comes from food. As per world health organization, in 2019 Diabetes caused 1.5 million deaths.Most of these deaths occurred in low- and middle-income countries. As per MedicalNewsToday report more than 37 million Trusted Source adults are living with diabetes in the United States andthat has more than doubled in the last two decades.To reduce large Scale of death rate from diabetes a quick and efficient technique is to be deserved. Machine learning has a very crucial role in healthcare industry for prediction and analyse made by using machine learning requires different medical datasets.

The PIMA Indians diabetes dataset is used in this paper, which contain the information of patients affected and non-affected with diabetes.Our approach involves feature selection and data pre-processing, followed by the training of various ML models, including Random Forest(Rf), Support Vector Machine(SVM), Logistic Regression(LR), K-nearest neighbour, Decision Tree.

The main objective of this research project is to predict the diabetes of a patient using machine learning algorithm. The early detection and prevention of diabetes complications are critical for improving health outcomes and reducing healthcare costs. Machine learning algorithms have shown promising results in predicting the risk of complications in patients with diabetes.

Keywords:ML, KNN, SVM,LR,RF

1.Introduction:

Diabetes mellitus is a metabolic disease that causes high blood sugar. Diabetes complications can lead to heart stroke, kidney failure, lower limb amputation etc.Retinal damage is incurable in the advanced stage of diabetes [15].About 75% of deaths in people with diabetes are due to coronary artery disease [16].The adverse microvascular and macrovascular effects of elevated plasma glucose levels behave in a linear or exponential manner and do not change dramatically when patients' FPG or OGTT results pass the threshold we have defined as “diabetes” [6]. Hypertension is a common comorbid condition of diabetes, affecting ~ 20-60% of patients with diabetes, depending on ethnicity, age, and obesity [10]. Eating white rice excessively may increase the risk of diabetes, especially in Chinese and Japanese people [18]. The main cause of diabetes varies by type, no matter what type of diabetes you have affected. Rosiglitazone reduces the incidence of diabetes and mean FPG levels but is associated with a small excess risk of heart failure and with weight gain [5].Diabetes is a growing public health concern, with over 460 million people living with the disease worldwide. Early detection and prevention of diabetes are crucial for reducing the risk of complications and improving patient outcomes. Machine learning has shown promising results in predicting the risk of diabetes, with algorithms able to identify. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC-7) recommends that blood pressure in diabetic patients be controlled to levels ≤ 130/80 mmHg [11]. In the setting of poor glycaemic control, it has been previously shown that there is an augmented HF response to amino acids which can be reversed by intensive glucose control in patients with T1DM or type 2 diabetes (T2DM) [17].Fasting glucose is the standard measure used for the diagnosis of diabetes in the United States.

Chronic diabetes conditions include type-1, type-2, and gestational diabetes.

Type-1: Diabetes occurs when the does not produce Insulin. In type 1 diabetes, a T-cell–mediated autoimmune response against beta cells appears to be the main disease mechanism [1]. The geographic variation in the incidence of Type 1 diabetes in children is extraordinary [2]. it can start at any age but most of the cases it starts during childhood. It is partly inherited, with multiple genes including certain HLA genotypes. The oral route of insulin delivery may be more likely to modify the immunological environment in the islets in such a way as to stop or slow the progression of β-cell destruction [3]. Intensive glycaemic therapy reduces the risk of CVD in patients with type 1 diabetes [4].

<u>Symptoms of Type -1 Diabetes are</u>

- Having blurry vision
- Feeling tired and weak.
- Urinating often.
- Losing Weight.
- Issue with the Kidney.
- Foot Problem, including numbness.

Type-2: Diabetes occurs when body does not use insulin well. Insulin plays a critical role in regulating glucose levels in the body.Type 2 diabetes patients (NIDDM) are not autoimmune and a decreased sensitivity to insulin action is the main abnormality [8]. It is the most common types of diabetes accounting 95% as per WHO.It can be treated with dietary therapy or oral hypoglycaemic agents, and infrequently requires exogenous insulin [7].Conditions such as high blood pressure or high cholesterol can increase the risk of complications in people with type 2 diabetes. Managing these conditions through medication, lifestyle changes, or a combination of both can help reduce the risk of complications.The diagnosis of type 2 diabetes mellitus has traditionally been made in individuals with hyperglycaemia with or without symptoms like polyuria and polydipsia, who are generally obese, older than 40 years of age, with family history of diabetes, and sometimes accompanied by hypertension or dyslipidaemia [9].

Gestational diabetes: This type of diabetes, that can develop during pregnancy due to inadequate insulin secretion. A blood sugar level of 190(10.6 mmol/L) or higher than that indicates gestational diabetes. Gestational diabetes is increasing in prevalence coincidently with the dramatic increase in the prevalence of overweight and obesity in women of childbearing age [12]. While pregnancy changes maternal metabolism, GDM can be viewed as a maladaptation by maternal systems to pregnancy, which may include mechanisms such as insufficient insulin secretion, dysregulated hepatic glucose output, mitochondrial dysfunction and lipotoxicity [13]. It is most often diagnosed in the second or third trimester because of the increase in insulin-antagonist hormone levels that occurs currently [20]. Insulin is usually the first choice of diabetes medicine for gestational diabetes and it is not harm to the baby [14].The management of gestational diabetes typically involves a combination of lifestyle changes, monitoring blood sugar levels, and, in some cases, medication.Follow a regular meal schedule: Eating meals at regular intervals can help regulate blood sugar levels. Women with gestational diabetes should aim to eat three meals a day and include snacks as needed.

2.Related work

Yuvraj et al. [21] proposed a system for diabetes prediction using three different ML algorithms including Random Forest, Decision Tree, and the Naïve Bayes. He got highest accuracy i.e. 94% using Random Forest algorithm.

Deepti Sisodia et al. [22] the author used Pima Indians Diabetes Database and got 76.30% accuracy using Naïve Bayes machine learning algorithm.

Zou Q et al. [23] proposed a system based on techniques on random forest and got the highest accuracy 80.84% when all attributes are used.

Sambit Mohanty et al. [24] In classification, Author had taken the random forest algorithm. For hybrid approach, they chosen XGBoost algorithm. These two algorithms were implemented and compared in order to explore the prediction accuracy in diabetes for two different machine learning approaches and got the mean score 74.10% in XGBoost algorithm.

Swapna G. et al. [25] The classification system proposed can help the clinicians to diagnose diabetes using ECG signals with a very high accuracy of 95.7% and author use of deep learning techniques in diabetes detection with HRV as input data achieving an accuracy of 95.1%.

Naz. et al. [26] author achieved the accuracy by functional classifiers Artificial Neural Network (ANN), Naive Bayes (NB), Decision Tree (DT) and Deep Learning (DL) lies within the range of 90–98%. Among the four of them, DL provides the best results for diabetes onset with an accuracy rate of 98.07% on the PIMA dataset.

S. Wei. et al. [27] author use many Machine learning algorithms like DNN, SVM etc, and got 77.86% accuracy using 10-fold cross-validation.

Safdari et al. (2019) proposed a machine learning-based approach for predicting the risk of diabetic nephropathy (DN) in patients with diabetes. The authors used a combination of clinical and demographic data, along with laboratory test results, to train various machine learning models, including decision trees and artificial neural networks. The authors demonstrated that their approach achieved high accuracy in predicting the risk of DN in patients with diabetes.

Wang et al. (2022) proposed a machine learning-based approach for predicting the risk of diabetic peripheral neuropathy (DPN) in patients with diabetes. The authors used a combination of clinical and demographic data, along with nerve conduction study results, to train various machine learning models, including logistic regression and support vector machines. The authors demonstrated that their approach achieved high accuracy in predicting the risk of DPN in patients with diabetes and could be used to inform personalized management strategies for patients at high risk of developing DPN.

Liang et al. (2022) proposed a machine learning-based approach for predicting the risk of gestational diabetes mellitus (GDM) in pregnant women. The authors used a combination of demographic and clinical data, along with laboratory test results, to train various machine learning models, including decision trees and random forests. The authors found that their approach achieved high accuracy in predicting the risk of GDM and could be used to inform targeted screening and prevention strategies for pregnant women at high risk of developing GDM.

Liu et al. (2022) proposed a machine learning-based approach for predicting the risk of diabetic ketoacidosis (DKA) in patients with type 1 diabetes. The authors used a combination of clinical and demographic data, along with continuous glucose monitoring data, to train various machine learning models, including neural networks and decision trees. The authors demonstrated that their approach achieved high accuracy in predicting the risk of DKA and could be used to inform personalized insulin dosing and glucose monitoring strategies for patients with type 1 diabetes.

Vaishnavi M et al. (2021) This study reviewed various ML techniques such as support vector machine (SVM), artificial neural networks (ANN), k-nearest neighbor (KNN), and decision trees for diabetes prediction. The authors concluded that SVM and ANN are the most effective ML techniques for diabetes prediction.

3.World-wide Diabetes Status Study and Report Analysis

Diabetes is a chronic health condition that affects millions of people worldwide. According to the International Diabetes Federation (IDF), approximately 463 million adults (20-79 years old) were living with diabetes in 2019. This number is expected to rise to 700 million by 2045.

A study published in Nature Medicine in 2021 found that a diet high in saturated fat and low in fiber may increase the risk of developing type 2 diabetes by altering the gut microbiome. The study used a mouse model to demonstrate that a high-fat, low-fiber diet caused changes in the gut microbiome that led to insulin resistance and other metabolic problems.

In the Journal of Clinical Endocrinology & Metabolism in 2021 found that high levels of a protein called suPAR may be a biomarker for the development of type 2 diabetes. The study used data from a large, long-term study of women to demonstrate that higher suPAR levels were associated with an increased risk of developing diabetes.

The World Health Organization's (WHO) Global Diabetes Compact, launched in 2021, aims to reduce the global burden of diabetes through a range of actions, including promoting healthy diets and physical activity, improving access to diabetes care and medicines, and strengthening health systems to better prevent and manage diabetes.

Country having diabetes rate as per MedicalNewsToday Report-:

- China — over 1 billion
- India — over 1 billion
- U.S. — 338 million
- Indonesia — 275 million
- Pakistan — 235 million

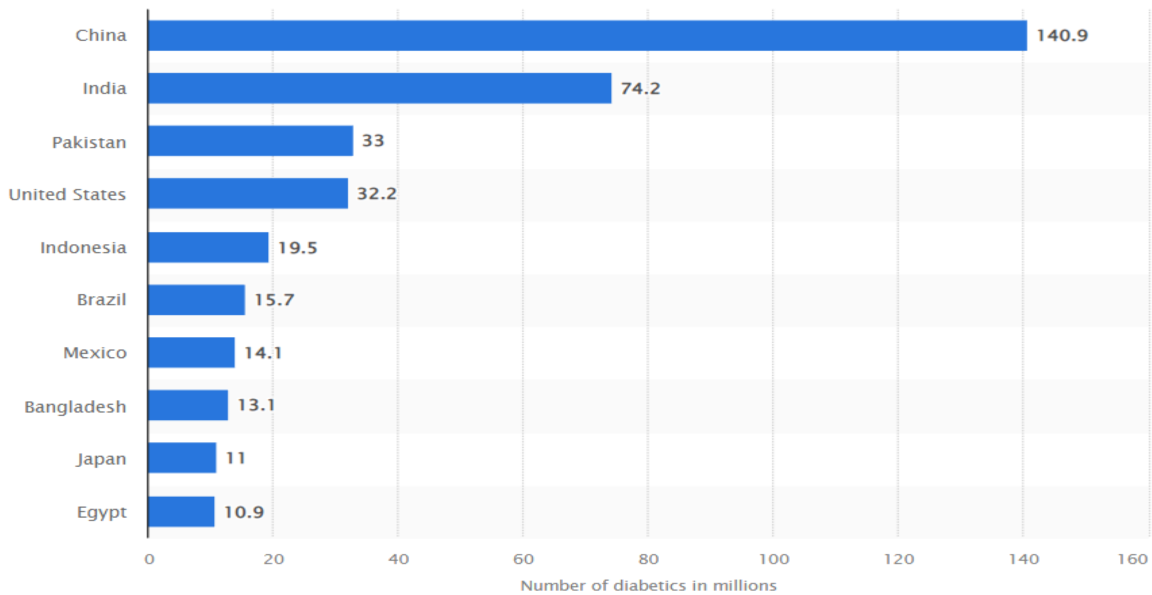


Fig.1-Countries with the highest number of diabetics worldwide in 2021

[based on Statista report]

3.a.The Burden of Diabetes in India: A Comprehensive Report, Prevention and Management Strategies

I. Study Reports

According to the International Diabetes Federation (IDF), India has the second-highest number of adults living with diabetes in the world, after China. In 2019, an estimated 77 million adults in India had diabetes, representing a prevalence of 8.9%.

The Indian government has taken several steps to address the growing burden of diabetes, including launching the National Programme for Prevention and Control of Cancer, Diabetes, Cardiovascular Diseases, and Stroke (NPCDCS) in 2010. The NPCDCS includes a range of interventions to prevent and manage diabetes, including promoting healthy lifestyles, providing screening and diagnosis services, and ensuring access to affordable and high-quality diabetes care.

II. Prevention and Management Strategies

Ayushman Bharat: Ayushman Bharat is a flagship healthcare scheme launched by the Indian government in 2018 to provide health coverage to thepoor and vulnerable sections of thesociety. Thescheme covers the cost of hospitalization for a range of diseases, including diabetes.

National Programme for Health Care of the Elderly (NPHCE): The NPHCE was launched by the Indian government in 2010 to provide comprehensive healthcare services to the elderly population. The programme includes screening and management of chronic diseases, including diabetes, in the elderly.

National Institute of Nutrition (NIN): The NIN is a research institute under the Indian Council of Medical Research (ICMR) that conducts research on nutrition and its role in the prevention and management of diseases, including diabetes. The institute also develops guidelines for the management of diabetes and conducts training programs for healthcare professionals.

Jan Aushadhi Scheme:Under this scheme, the government provides affordable generic drugs for diabetes and other chronic diseases through Jan Aushadhi stores located across the country.

Diabetes clinics: Diabetes clinics have been established in several hospitals and clinics across India to provide specialized care for patients with diabetes. These clinics are staffed by endocrinologists, diabetes educators, and dieticians who work together to develop individualized treatment plans for patients.

Patient education and support: Patient education and support are critical for successful diabetes management in India. Healthcare professionals provide patients with information on diabetes, its management, and the importance of self-care. They may also provide support through group counselling sessions or individual counselling.

4.Proposed System:

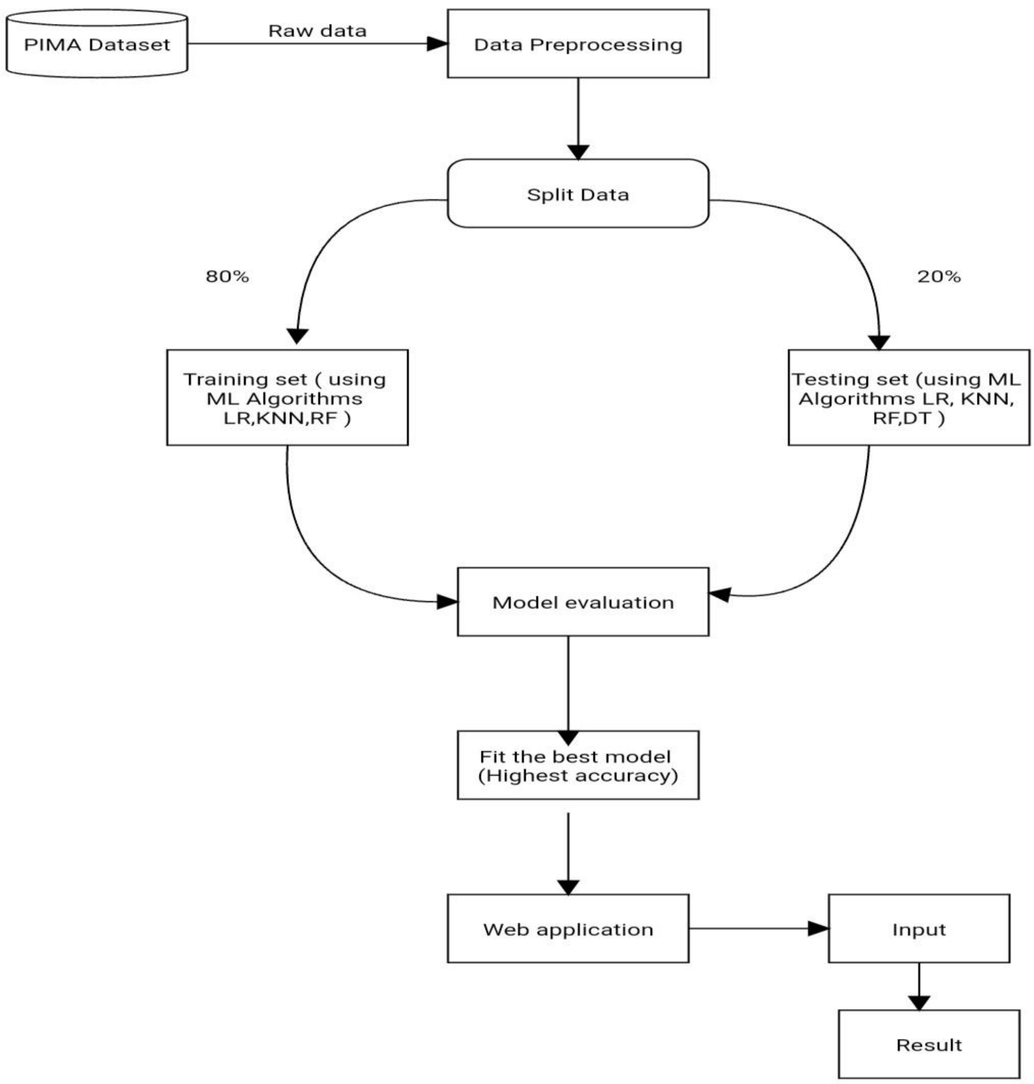


Fig.2 -: A Systematic Analysis of Flow Diagrams

5.Building Blocks of Algorithms:

Step-1: Collect the Diabetes dataset, import it and required libraries also.

Step-2: Data Pre-processing (Cleaning, normalization, missing value identification and feature extraction)

Step-3:Replace missing value with corresponding mean value.

Step-4: Split the dataset (80% as training set and 20% as test set).

Step-5: Classification model implementation.(Decision Tree, Random Forest, Linear Regression, Support Vector Machine)

Step-6: Select best fit model according to find highest Accuracy.

Step-7: Result and Analysis.

Data collection -: it includes data collection from a specific domain. In PIMA diabetes dataset having 8 attributes that are used to predict the diabetes.

Data Pre-processing -: It is a crucial step that helps to enhance the quality of data to promote the extraction of meaningful insights from the data.Information cleaning helped in recognizing the information required and expelling clamor and unessential information, especially captured factors not considered in this venture [28].

- Also known as data wrangling, cleaning, and feature engineering.

Missing value identification -: in the dataset which value will be missing replaced it with corresponding mean value.

Splitting dataset -: After data pre-processing, split the dataset into training set and testing set.

Training dataset (80%)	**Test dataset (20%)**

5.a.AComprehensive Exploration of Algorithms and Techniques

- **Logistic Regression -**:

  It is a supervised Machine Learning algorithm, used to predict the probability of a binary event.The logistic regression model estimates the coefficients of the input variables using maximum likelihood estimation, which finds the values of the coefficients that maximize the likelihood of the observed data. The model can be preventing overfitting.Logistic regression can be used for imbalanced class problems, where one class is much more prevalent than the other. Techniques such as oversampling, under sampling, or adjusting the class weights can be used to address the imbalance and improve the performance of the model.

  Examples -: Fraud detection, ‘spam’ or ‘no spam’ , patient is infected with a disease or not etc.
  Sigmoid function is trying to convert the Independent Variable into Expression of probability that ranges 0 and 1, w.r.t the dependent variable.

  $y(z) = \frac{1}{1+e^{-z}}$- (Sigmoid Function)

  The logistic regression model can be expressed as:

  $$h_\theta(x) = y(\theta^T x) = \frac{1}{1 + e^{-\theta Tx}}$$

  Where, x is independent variable (which will be transform)
  e is Euler’s constant (2.718)
  y is output

  - There are three main types of logistic regression.
    a. Binary
    b. Multinomial
    c. Ordinal

- **Decision Tree** –:

It is a supervised learning algorithm, approach that identifies way to split a dataset depending upon the conditions. Decision tree are used for both classification and Regression problem. A decision tree reaches its decision by performing a sequence of task.The process of building a decision tree involves recursively splitting the data based on the input variables, until a stopping criterion is met, such as reaching a maximum depth, a minimum number of samples in a node, or a minimum reduction in impurity.For classification trees, popular splitting criteria include Gini impurity, entropy, and classification error. For regression trees, popular splitting criteria include variance reduction, mean squared error, and mean absolute error.

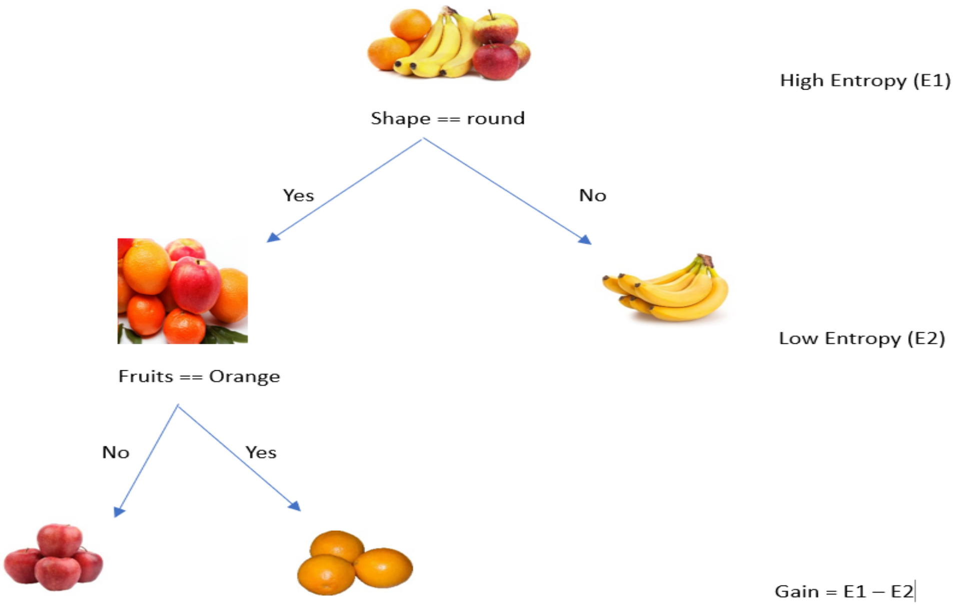


Fig.3 -: Decision TreeType equation here.

- Entropy : $H(S) = -P_{(+)}\log_2 P_{(+)} - P_{(-)}\log_2 P_{(-)}$

  Where, s = subset of training example

  $P_{(+)}$ / $P_{(+)}$ = positive & negative class in S

  Let us take an Example,

  - Impure set (2 Yes / 1 No)

    $H(s) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} \approx 0.92$

  - Pure set (2 Yes / 0 No)

  $H(s) = -\frac{2}{2}\log_2\frac{2}{2} - \frac{0}{2}\log_2\frac{0}{2} = 0$

  The equation for a decision tree is:

$$f(x) = \begin{cases} c_j & if\ x \in R_j \\ f_{left}(x) & if\ x_i \leq t_j \\ f\text{right}(x) & if\ x_i > t_j \end{cases}$$

Where,x is the input feature vector.

R_j is the region of the input corresponds to the j-th leaf node.
c_j is the predicted output value for the j-th leaf node.
t_j is the threshold value for the j-th decision node.
$f_{left}(x)$ is the function that predicts the output value for the left child node of the j-th decision node.
$f_{right}(x)$ is the function that predicts the output value for the rightchild node of the j-th decision node.

- **Support Vector Machine (SVM) -:**

It is a set of supervised machine learning methods used for classification, regression, and outliers' detection.The name "support vector" comes from the fact that SVMs find the hyperplane that maximizes the margin between the closest points from each class, which are called the support vectors. The goal of the SVM is to find best hyperplane with larger distance between two classes in N-Dimensional space, so that in future when new data point will come, we can easily put in the correct category.SVMs are less prone to overfitting than other machine learning algorithms, such as decision trees and neural networks.In classification tasks, they can handle both binary and multiclass problems. In regression tasks, they can handle both linear and nonlinear relationships between the input and output variables.SVMs use kernel functions to map the input variables to a higher-dimensional space, where the relationship between the input and output may be linear. The most common kernel functions are the linear kernel, polynomial kernel, radial basis function (RBF) kernel, and sigmoid kernel. The choice of kernel function depends on the problem at hand and the characteristics of the data.

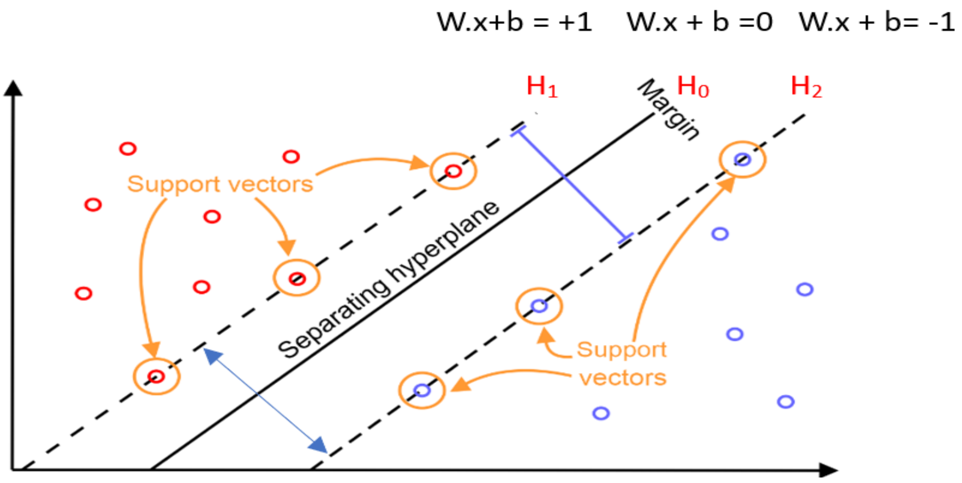


Fig.4 – Linear SVM

Form of equation defining the decision surface separating the classes is a hyperplane of the form:

$w^T x + b = 0$

wherew is the weight vector

x is input vector

b is bias

The distance between H_0 and H_1 is then:

$| W \bullet x+b|/||W||=1/||W||$

The total distance between H1 and H2 is : $2/||w||$

we can also write :

$w^T x + b \geq 0$ for di = +1 ---------- (1)
$w^T x + b < 0$ for di = –1---------- (2)
Margin of Separation (d): the separation between the hyperplane and the closest data point for a given weight vector w and bias b.
In fig.4 there are no datapoints between H1 and H2:

$x_i \bullet W+b \geq +1$ when $y_i = +1$---------------- (3)
$x_i \bullet W+b \leq -1$ when $y_i = -1$----------------- (4)

By combiningEquation (3) and (4), we can write:

$y_i (x_i \bullet W) \geq 1$

set of weights W (or W_i), one for each feature, whose linear combination predicts the value of y.
The equation for a linear SVM can be represented as:

$$f(x) = sign(w.x + b)$$

Where, x is the input feature vector.
w is the weight vector that defines the direction of the hyperplane.
b is the bias term that shifts the hyperplane away from the origin.
The equation for a Non-linear SVM can be represented as:

$$f(x) = sign(\sum_{i=1}^{N} \alpha_i y_i k(x, x_i) + b)$$

Where, x is the input feature vector.
N is the number of training samples.
y_i is the output label of the i-th training sample.
$K(x,x_i)$ is the kernel function that computes the similaritybetween the input feature vector and the i-th training sample in the feature space.

α_i is the Lagrange multiplier that determines the importance of the i-th training sample in the final decision boundary.
b is the bias term that shifts the decision boundary.

- **Random Forest -:**

Random forest is an ensemble learning method, which combines multiple individual models (decision trees) to improve the overall predictive accuracy. Ensemble learning is a technique where several models are trained and combined to achieve better performance than any individual model.Random forest, like other machine learning algorithms, must balance the trade-off between bias and variance.They can handle both categorical and continuous input variables, and can learn complex nonlinear relationships between the input and output.Random forest can be used for feature selection, which is the process of selecting a subset of input features that are most relevant for the prediction task.Random Forests have high accuracy compared to other machine learning algorithms.

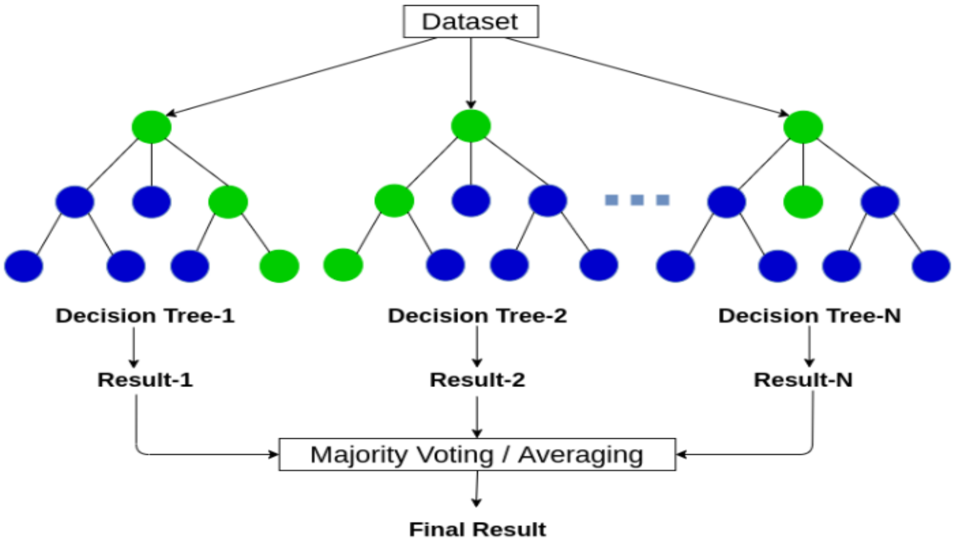


Fig.5 – Random Forest

Image source : Analytics Vidya

The Equation of Random Forest is :

$$f(x) = \frac{1}{M}\sum_{m=1}^{M} f_m(x)$$

Where, x is the input feature vector.

$f_m(x)$ is the prediction of the m-th decision tree.

M is the number of decision trees in the Random Forest.

The dataset

The PIMA diabetes dataset contains 768 instances as shown in fig (1), and contains 9 attributes that are listed below.

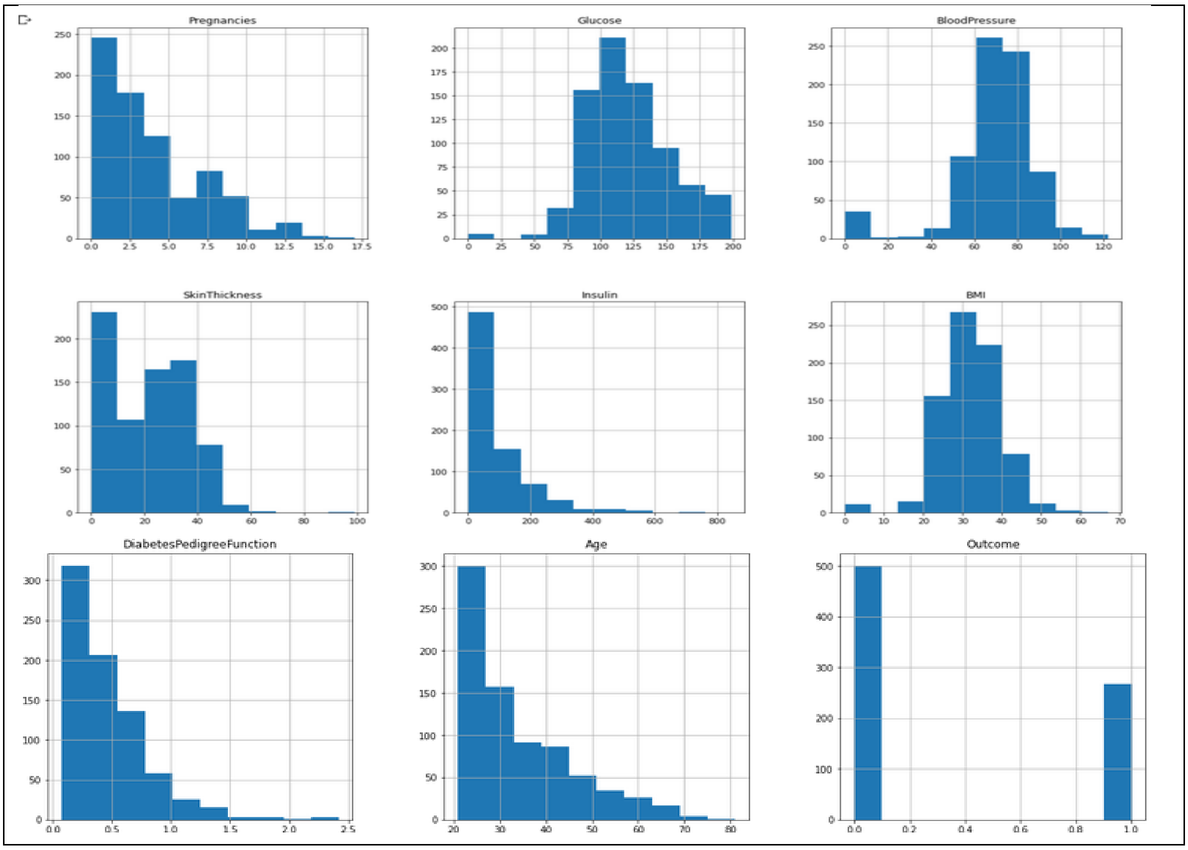
Table-**1**

```
[ ]
     Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI  \
0              6      148             72             35        0  33.6
1              1       85             66             29        0  26.6
2              8      183             64              0        0  23.3
3              1       89             66             23       94  28.1
4              0      137             40             35      168  43.1
..           ...      ...            ...            ...      ...   ...
763           10      101             76             48      180  32.9
764            2      122             70             27        0  36.8
765            5      121             72             23      112  26.2
766            1      126             60              0        0  30.1
767            1       93             70             31        0  30.4

     DiabetesPedigreeFunction  Age  Outcome
0                       0.627   50        1
1                       0.351   31        0
2                       0.672   32        1
3                       0.167   21        0
4                       2.288   33        1
..                        ...  ...      ...
763                     0.171   63        0
764                     0.340   27        0
765                     0.245   30        0
766                     0.349   47        1
767                     0.315   23        0

[768 rows x 9 columns]
```

<u>Data Visualization</u>

Plotted histogram by considering the Dataset.

Table – 2



Result -:

Using pipelining, we got Accuracy 79.87 for RandomForestClassifier

Table – 3: Pipelining Results

Algorithm	**Accuracy**
Logistic Regression	76.62
K-Nearest Neighbors	76.62
Decision Tree	73.37
RandomForestClassifier	79.87
Support Vector Machine	73.37

```
lr Test Accuracy :76.62337662337663
kn Test Accuracy :76.62337662337663
svc Test Accuracy :73.37662337662337
tree Test Accuracy :73.37662337662337
rf Test Accuracy :79.87012987012987
```

Fig-: Accuracy Of various Model

Conclusion -:

Machine learning models are helps to predict weather a person affected in diabetes or not.We evaluate our approach using a dataset of patients with and without diabetes and demonstrate its potential for accurate diabetes risk prediction. In this study, various machine learning algorithms are applied and Random Forest gives highest accuracy i.e., 80%. The PIMA Indian open-source dataset used in this work. Diabetes prediction is important for avoiding further complication of the disease. Machine Learning has great contribution for diabetes risk prediction.

References-:

[1]Rother KI. Diabetes treatment--bridging the divide. N Engl J Med. 2007 Apr 12;356(15):1499-501. Doi: 10.1056/NEJMp078030. PMID: 17429082; PMCID: PMC4152979.

[2]Åkerblom, Hans K., and Mikael Knip. "Putative environmental factors in type 1 diabetes." *Diabetes/metabolism reviews* 14.1 (1998): 31-68.

[3] Jennifer B. Marks; In Pursuit of Type 1 Diabetes Prevention. *Clin Diabetes* 1 October 2002; 20 (4): 168–169. https://doi.org/10.2337/diaclin.20.4.168

[4] K.M. Venkat Narayan; Glycaemic Control and Cardiovascular Disease in Patients with Type 1 Diabetes. *Clin Diabetes* 1 April 2006; 24 (2): 88–89. https://doi.org/10.2337/diaclin.24.2.88

[5,6] Michael Pignone; Rosiglitazone Reduced the Incidence of Diabetes in Patients with Impaired Glucose Tolerance. *Clin Diabetes* 1 January 2007; 25 (1): 23–24. https://doi.org/10.2337/diaclin.25.1.23

[7]Krolewski, Andrzej S., et al. "Epidemiologic approach to the etiology of type I diabetes mellitus and its complications." *New England Journal of Medicine* 317.22 (1987): 1390-1398.

[8]Rattan Juneja, Jerry P. Palmer, Type 1 1/2 Diabetes: Myth or Reality? Autoimmunity, 10.3109/08916939908995974, **29**, 1, (65-83), (2009).

[9]Larissa Avilés-Santa, Noel Maclaren, Philip Raksin, Immune-mediated disease, and secondary failure to oral therapy in type 2 diabetes mellitus, Journal of Diabetes, and its Complications, 10.1016/S1056-8727(02)00253-2, **18**, 1, (10-17), (2004).https://doi.org/10.1016/S1056-8727(02)00253-2

[10] Janet B. McGill," Re-examining Misconceptions About β-Blockers in Patients with Diabetes", *Clin Diabetes* 1 January 2009; 27 (1): 36–. https://doi.org/10.2337/diaclin.27.1.36

[11]Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo Jr, J. L., ... & National High Blood Pressure Education Program Coordinating Committee. (2003). Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. *hypertension*, *42*(6), 1206-1252.

[12]Kampmann U, Madsen LR, Skajaa GO, Iversen DS, Moeller N, Ovesen P. Gestational diabetes: A clinical update. World J Diabetes. 2015 Jul 25;6(8):1065-72. Doi: 10.4239/wjd.v6.i8.1065. PMID: 26240703; PMCID: PMC4515446.

[13] Moyce Gruber BL, Dolinsky VW. The Role of Adiponectin during Pregnancy and Gestational Diabetes. Life (Basel). 2023 Jan 21;13(2): 301.DOI: 10.3390/life13020301; PMCID: PMC9958871.

[14]"Gestational Diabetes". *NIDDK*. September 2014. Archived from the original on 16 August 2016. Retrieved 31 July 2016.

[15] Jena, P.K.; Khuntia, B.; Palai, C.; Nayak, M.; Mishra, T.K.; Mohanty, S.N. A Novel Approach for Diabetic Retinopathy Screening Using Asymmetric Deep Learning Features. *Big Data Cogn. Comput.* **2023**, *7*, 25. https://doi.org/10.3390/bdcc7010025

[16]O'Gara PT, Kushner FG, Ascheim DD, Casey DE, Chung MK, de Lemos JA, et al. (January 2013). "2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines".Circulation. 127 (4): e362–e425.doi:10.1161/CIR.0b013e3182742cf6. PMID 23247304

[17] Richard J. MacIsaac, George Jerums, Elif I. Ekinci. ARTICLE| VOLUME 25, ISSUE 2, P141-148, MARCH 2018.Doi: https://doi.org/10.1053/j.ackd.2017.11.003

[18]Hu EA, Pan A, Malik V, Sun Q. White rice consumption and risk of type 2 diabetes: meta-analysis and systematic review. BMJ. 2012 Mar 15;344: e1454. Doi: 10.1136/bmj.e1454.PMID: 22422870; PMCID: PMC3307808.

[19]Selvin E, Steffes MW, Zhu H, Matsushita K, Wagenknecht L, Pankow J, Coresh J, Brancati FL. Glycated haemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. N Engl J Med. 2010 Mar 4;362(9):800-11. Doi: 10.1056/NEJMoa0908359. PMID: 20200384; PMCID: PMC2872990.

[20]Soldavini J (November 2019). "Krause's Food & The Nutrition Care Process". *Journal of Nutrition Education and Behavior*. **51** (10): 1225. ISSN 1499-4046.doi:10.1016/j.jneb.2019.06.022

[21]Yuvaraj, N.; Sri Preethaa, K.R. Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster. *Clust. Comput.* **2019**, *22*, 1–9.

[22] Deepti Sisodia, Dilip Singh Sisodia, 'Prediction of Diabetes using Classification Algorithms', Volume 132, 2018, Pages 1578-1585.https://doi.org/10.1016/j.procs.2018.05.122

[23] Zou Q, Qu K, Luo Y, Yin D, Ju Y, Tang H. Predicting Diabetes Mellitus with Machine Learning Techniques. Front Genet. 2018 Nov 6;9:515. Doi: 10.3389/fgene.2018.00515. PMID: 30459809; PMCID: PMC6232260.

[24] Barik, S., Mohanty, S., Mohanty, S., Singh, D. (2021). Analysis of Prediction Accuracy of Diabetes Using Classifier and Hybrid Machine Learning Techniques. In: Mishra, D., Buyya, R., Mohapatra, P., Patnaik, S. (eds) Intelligent and Cloud Computing. Smart Innovation, Systems and Technologies, vol 153. Springer, Singapore. https://doi.org/10.1007/978-981-15-6202-0_41.

[25] Swapna G., Vinayakumar R., Soman K.P.' Diabetes detection using deep learning algorithms '.Volume 4, Issue 4, Pages 243-246, ISSN 2405-9595, https://doi.org/10.1016/j.icte.2018.10.005.

[26] Naz, H., Ahuja, S. Deep learning approach for diabetes prediction using PIMA Indian dataset. J Diabetes MetabDisord 19, 391–403 (2020). https://doi.org/10.1007/s40200-020-00520-5

[27] S. Wei, X. Zhao and C. Miao, "A comprehensive exploration to the machine learning techniques for diabetes identification," 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), Singapore, 2018, pp. 291-295, Doi: 10.1109/WF-IoT.2018.8355130.

[28] Danish Ather, Suman Madan, Manjushree Nayak, Rohit Tripathi, Ravi Kant, Sapna Singh Kshatri, Rituraj Jain, "Selection of Smart Manure Composition for Smart Farming Using Artificial Intelligence Technique", *Journal of Food Quality*, vol. 2022, Article ID 4351825, 7 pages, 2022. https://doi.org/10.1155/2022/4351825

[29]

GOYAL, C. (2021, may 31). *Bagging- 25 Questions to Test Your Skills on Random Forest Algorithm*. Retrieved from Analytics Vidya: https://www.analyticsvidhya.com/blog/2021/05/bagging-25-questions-to-test-your-skills-on-random-forest-algorithm/