BioGecko

Vol 12 Issue 02 2023
ISSN NO: 2230-5807

MACHINE LEARNING-BASED DIABETES PREDICTION

Dr. Manisha Patil
Department of Computer Science, Indira College of Commerce and Science, Savitribai Phule Pune University, Pune, Maharashtra, India.
Email: manisha.patil@iccs.ac.in

Abstract
Diabetes is a chronic disease that affects millions of people worldwide. Early detection and prediction of diabetes can help in preventing its complications and improving the quality of life of patients. Machine learning techniques have shown great promise in predicting diabetes, utilizing various features such as age, gender, BMI, blood pressure, and glucose levels. This paper presents a comprehensive overview of the state-of-the-art machine-learning techniques used in the diabetes prediction model. Overall, machine learning-based approaches show great promise for accurate and timely prediction of diabetes, which could help in reducing the burden of this disease on society.

Keywords: Diabetes prediction, Machine learning, Classification

I. INTRODUCTION

This is the start of the body text of your paper. You can use headings like the one above to divide your paper into sub-topics. Use level 1 headings first, then level 2 headings if you need further divisions inside those, and so on. Do not use a level of heading unless there will be at least two headings of that level. You don't have to use any headings at all if it doesn't make sense to divide your paper in that way. Appropriate numbering is automatically applied to headings. You do not have to number them yourself, just make sure the right heading style is applied to each one. Level 1 and 2 headings (as well as the paper title) should be written with title case capitalization, while level 3 and 4 headings are written in sentence case.

Machine learning is indeed becoming increasingly important in various fields, including healthcare. The abundance of medical data available is a valuable resource that can be leveraged to improve patient outcomes and enable early disease detection. Predicting diabetes using blood data and other patient information is a good example of how machine learning can be used in healthcare.

Classification methods, such as decision trees, logistic regression, support vector machines, and neural networks, are commonly used in machine learning to predict the likelihood of a certain outcome. In this case, the outcome is whether a patient is likely to have diabetes or not. The algorithms learn from the training data, which consists of known cases of diabetes and non-diabetes patients, to build a model that can accurately predict the outcome for new, unseen data.

Handling complex data and tasks Machine learning algorithms are capable of handling complex tasks and data. They can process and analyze large volumes of data, identify patterns, and make predictions accurately and efficiently.

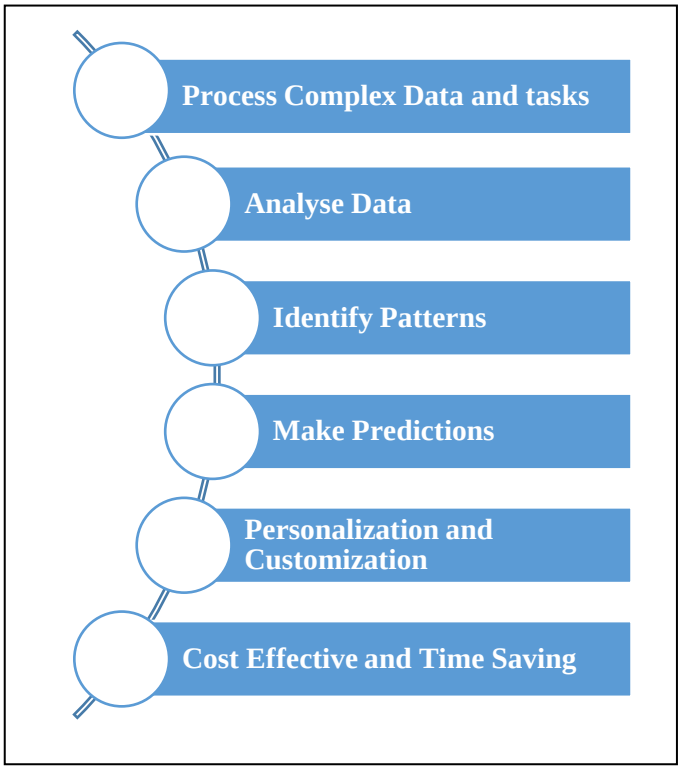


Fig 1 Feature of Machine Learning

• Personalization and customization Machine learning algorithms can be used to create personalized experiences for users.
• Cost-effective and time-saving Machine learning algorithms can automate various tasks, reducing the need for manual labor and saving time and money for organizations.
Overall, machine learning has numerous advantages and is a powerful tool that can be used to solve a wide range of problems in different fields. As technology continues to evolve, the potential applications of machine learning are only going to increase.
It is worth noting that machine learning is not a replacement for medical professionals. The predictions generated by the algorithms are intended to aid healthcare providers in making informed decisions and should not be used as a substitute for clinical judgment. Additionally, it is important to ensure that patient privacy is protected when using medical data for machine learning purposes.
Health Care centers diabetes report generation takes time so eager patients get relief if he/she would get to know if he is probable for diabetes. Many times, tests related to diabetes needs to be done regularly and the patient also needs to keep up with diet based on readings/data needed to predict diabetes. E.g., The patient today's BMI is 25.0 if he or she can check if this BMI grows to 30 or so will he/she be prone to diabetes Yes, a person will control his or her diet if NO he/she can at least be a relief for a week/day after fresh reading he or she would manage himself/herself.

II. LITERATURE REVIEW

Diabetes is a chronic disease that affects millions of people worldwide. Early detection and prediction of diabetes can help in preventing the onset of the disease and managing its progression. Machine learning algorithms have been widely used in predicting the onset of diabetes using various datasets. This literature review aims to provide an overview of the existing research in the field of diabetes prediction using machine learning techniques.

The study of, M. A. Alomari and A. Al-Mallah.[1], provides an overview of the existing literature on predicting type 2 diabetes using machine learning algorithms. The authors discuss various factors that contribute to the onset of diabetes and the datasets used in the prediction models. The article also highlights the performance of different machine learning algorithms, such as support vector machines, decision trees, and logistic regression, in predicting diabetes.
S. M. Naik et al.[2], the study focuses on the use of machine learning algorithms for predicting diabetes mellitus. The authors discuss various factors that influence the onset of diabetes, such as age, BMI, blood pressure, and glucose levels. The article also provides an overview of different machine learning algorithms, such as artificial neural networks, support vector machines, and decision trees, used for predicting diabetes mellitus.
H. Alharbi et al. [3], provide an overview of the use of machine learning algorithms for predicting diabetes. The authors discuss various datasets used in the prediction models, such as the Pima Indian diabetes dataset and the National Health and Nutrition Examination Survey dataset. The study also highlights the performance of different machine learning algorithms, such as k-nearest neighbors, logistic regression, and support vector machines, in predicting diabetes.
S. A. Alwafi et al.[4], this systematic literature review focuses on the use of machine learning algorithms for predicting type 2 diabetes mellitus. The authors discuss various datasets used in the prediction models, such as the diabetes dataset from the UCI Machine Learning Repository and the diabetes dataset from the University of California, San Francisco. The study also highlights the performance of different machine learning algorithms, such as random forests, artificial neural networks, and support vector machines, in predicting type 2 diabetes mellitus.
Machine learning algorithms have been widely used in predicting the onset of diabetes using various datasets. Different machine learning algorithms, such as support vector machines, decision trees, and artificial neural networks, have been used for predicting diabetes. The performance of these algorithms depends on the dataset used, the features selected, and the pre-processing techniques employed. Further research is needed to improve the accuracy of the prediction models and to develop personalized approaches for predicting diabetes.

III. RESEARCH METHODOLOGY

In this paper, the research objective is to develop a machine learning-based model and experimented tool designed to predict diabetes-prone patients.
The feature of this framework is discussed in this section.

Objective:

1. Predict if the patient is probably diabetic or not.
2. Making an efficient system that will solve the above-provided problem, provided that the users have the ease to interact with the interface

To achieve the above objectives, the system should have the following key features:

- User-friendly interface: The system should have a simple and intuitive interface that allows patients to easily input their data, including demographic information and medical test results.
- Data input: The system should allow patients to input data related to their medical history, lifestyle, and other relevant factors that could impact diabetes risk.
- Machine learning algorithms: The system should use machine learning algorithms, such as logistic regression or decision trees, to predict the probability of a patient having diabetes based on their input data.
- Accurate predictions: The algorithms used by the system should be trained on a large dataset of medical data to ensure accurate predictions.

- Interpretation of results: The system should provide clear and easy-to-understand feedback to patients, including their predicted diabetes risk and recommendations for further action if necessary.
- Privacy and security: The system should protect patient privacy and ensure the security of their data, in accordance with applicable laws and regulations.
- By incorporating these features into the system, patients will have a powerful tool at their disposal to help manage their diabetes risk and take control of their own health.

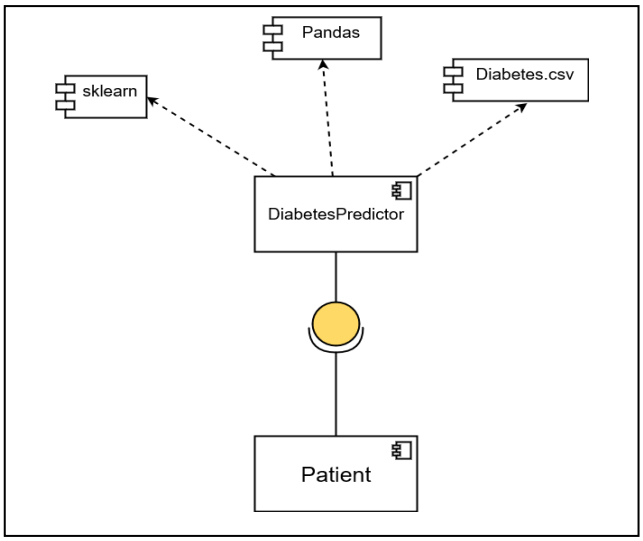
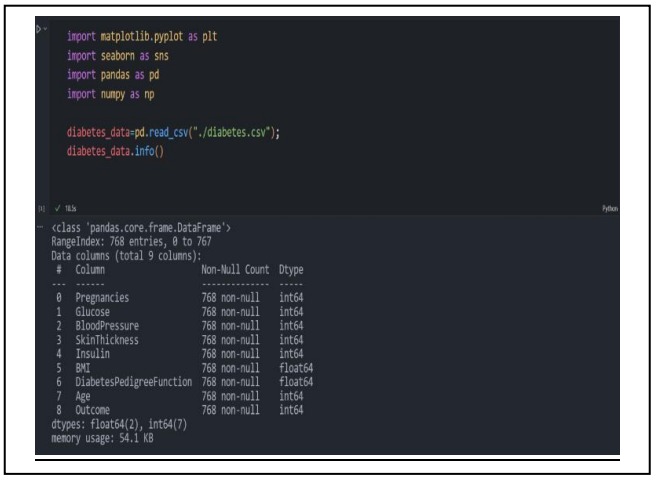


Fig 2: Components of the working experimented tool

IV. RESULTS AND ANALYSIS

The dataset uses seven input variables and one outcome variable. Various classification algorithms are used to train the model.



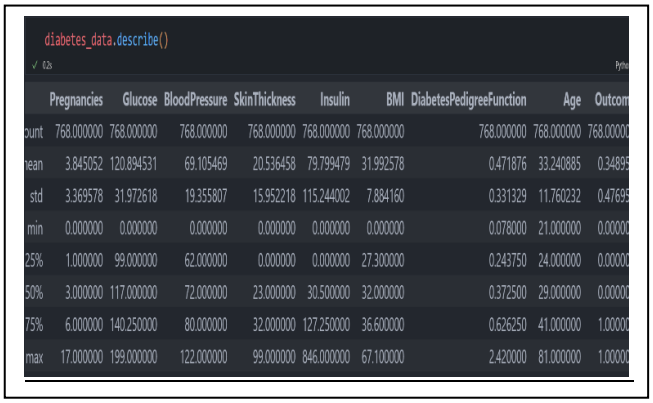
```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
import numpy as np

diabetes_data=pd.read_csv("./diabetes.csv");
diabetes_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Pregnancies               768 non-null    int64
 1   Glucose                   768 non-null    int64
 2   BloodPressure             768 non-null    int64
 3   SkinThickness             768 non-null    int64
 4   Insulin                   768 non-null    int64
 5   BMI                       768 non-null    float64
 6   DiabetesPedigreeFunction  768 non-null    float64
 7   Age                       768 non-null    int64
 8   Outcome                   768 non-null    int64
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

Fig 3: Features considered in the Dataset

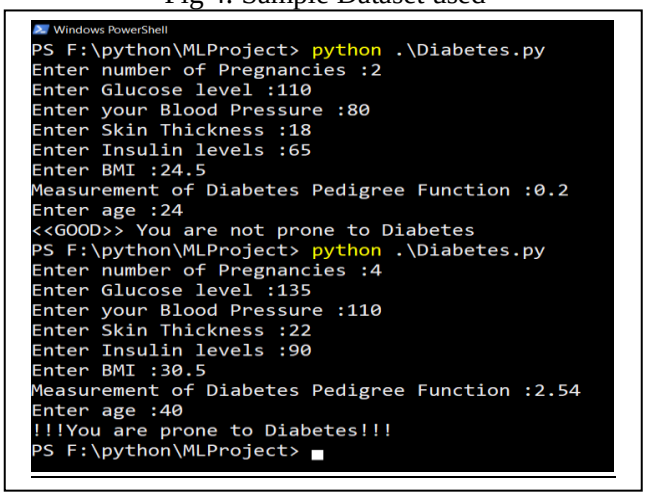
Here one can see the uploaded dataset with detail values observed with patient details as per the decided features.



diabetes_data.describe()

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcom
ount	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.00000
nean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	0.471876	33.240885	0.34895
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	0.331329	11.760232	0.47695
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000	21.000000	0.00000
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	0.243750	24.000000	0.00000
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	0.372500	29.000000	0.00000
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	0.626250	41.000000	1.00000
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000	81.000000	1.00000

Fig 4: Sample Dataset used



```
Windows PowerShell
PS F:\python\MLProject> python .\Diabetes.py
Enter number of Pregnancies :2
Enter Glucose level :110
Enter your Blood Pressure :80
Enter Skin Thickness :18
Enter Insulin levels :65
Enter BMI :24.5
Measurement of Diabetes Pedigree Function :0.2
Enter age :24
<<GOOD>> You are not prone to Diabetes
PS F:\python\MLProject> python .\Diabetes.py
Enter number of Pregnancies :4
Enter Glucose level :135
Enter your Blood Pressure :110
Enter Skin Thickness :22
Enter Insulin levels :90
Enter BMI :30.5
Measurement of Diabetes Pedigree Function :2.54
Enter age :40
!!!You are prone to Diabetes!!!
PS F:\python\MLProject>
```

Fig 5: Prediction result provided to the patient

Here patients can see the prediction result, which is helpful for the patient to improve their health issues.

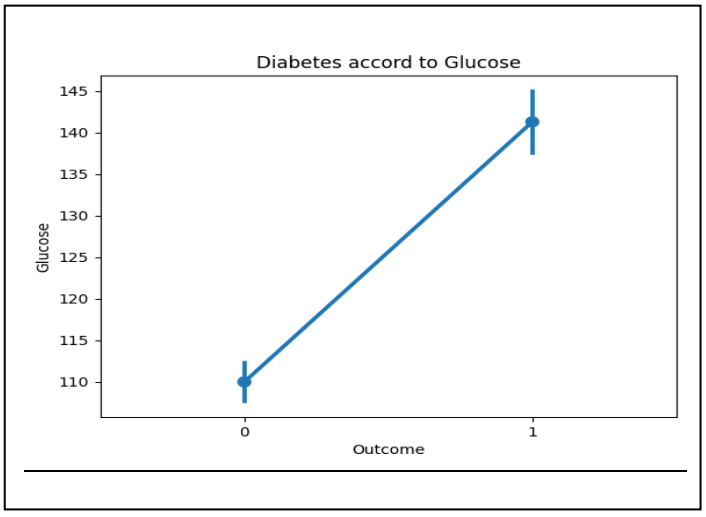


Fig 7 Outcome accord to Glucose

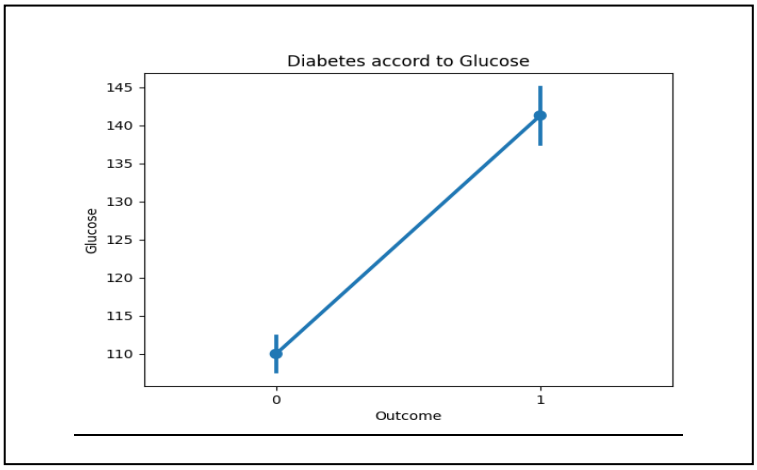


Fig 7 Outcome accord to Glucose

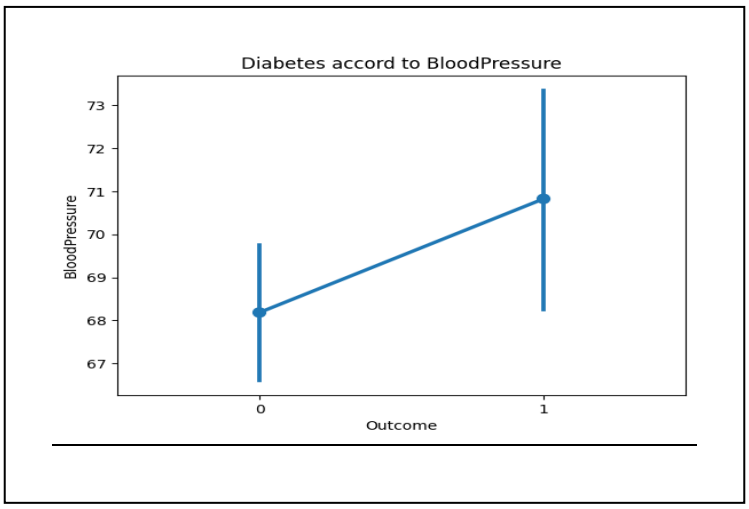


Fig 8 Outcome accord to blood pressure

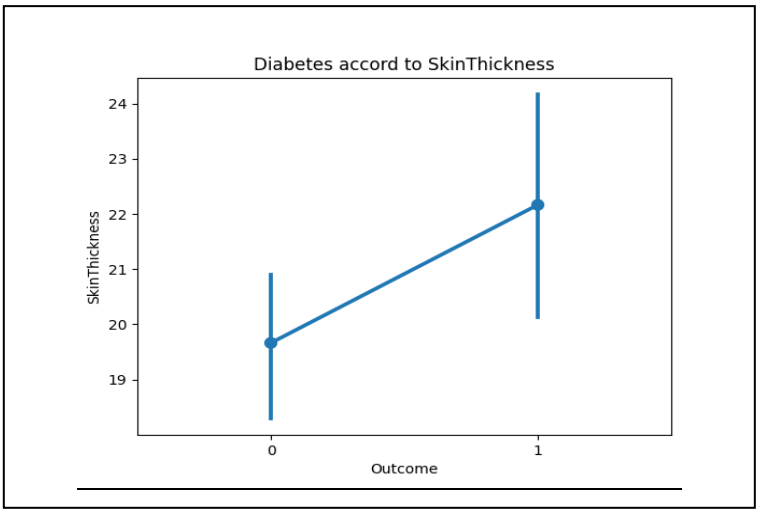


Fig 9 Outcome accord to SkinThickness

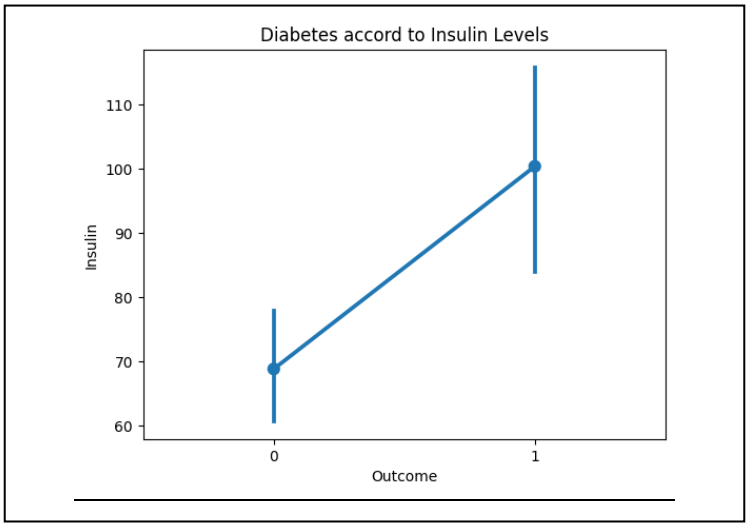


Fig 10 Outcome accord to Insulin Levels

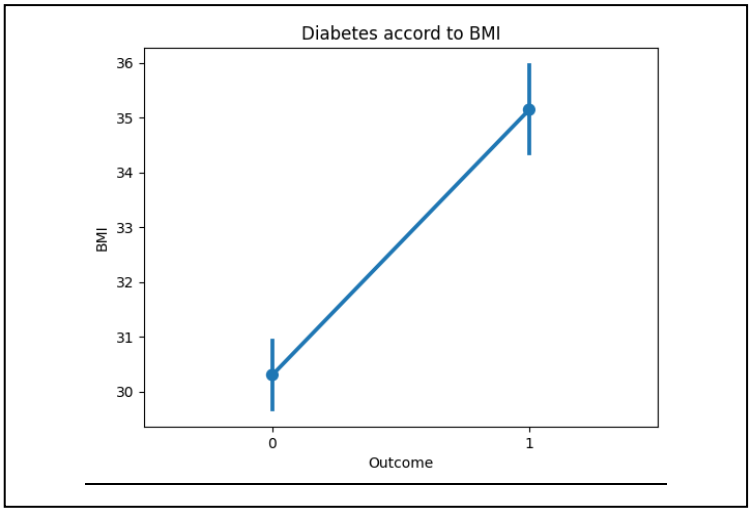


Fig 11 Outcome accord to BMI

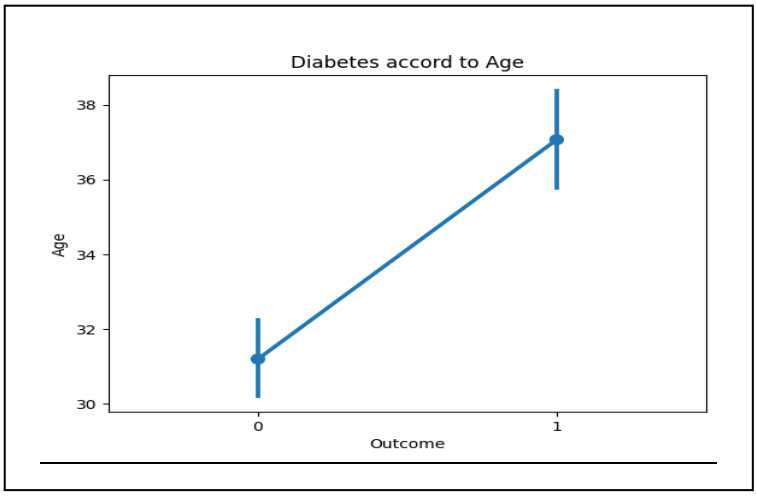


Fig 12 Outcome accord to Age

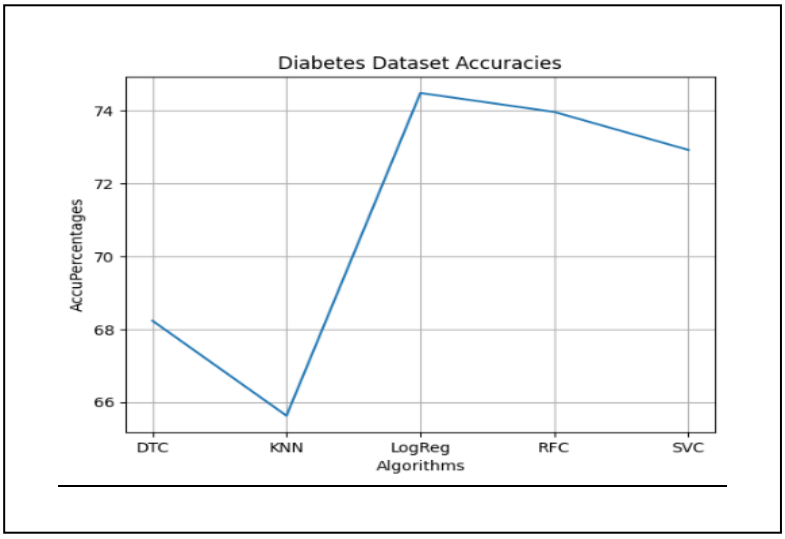


Fig 13 Accuracies with various ML training Models

The analysis is also depicted in the given graphs. Input parameters are considered during the analysis and the prediction of the graph from figure 6 to 12. Figure 13 depicts the performance accuracies of various ML models trained to predict diabetic-prone patients or not. Among the various ML models like Decision tree, KNN, Logistic regression, Random Forest, and Support Vector machine- the performance accuracy of the logistic regression model is highest for the dataset.

V CONCLUSION

This research paper depicts how machine learning works in general mostly the classification techniques. During this study, it is observed that every prediction is not true it serves as a probability at last human experience serves as the most accurate. Machines just merely help clients to progress through the result by eliminating half of the burden. Research work also shows that a higher magnitude of data points leads to positive results, while lower magnitude data points are more likely to get negative results. This fact in real life will not work every time.

REFERENCES

[1] D. V. Lindberg and H. K. H. Lee, “Optimization under constraints by applying an asymmetric entropy measure,” *J. Comput. Graph. Statist.*, vol. 24, no. 2, pp. 379–393, Jun. 2015, doi: 10.1080/10618600.2014.901225.

[2] B. Rieder, *Engines of Order: A Mechanology of Algorithmic Techniques*. Amsterdam, Netherlands: Amsterdam Univ. Press, 2020.

[3] I. Boglaev, “A numerical method for solving nonlinear integro-differential equations of Fredholm type,” *J. Comput. Math.*, vol. 34, no. 3, pp. 262–284, May 2016, doi: 10.4208/jcm.1512-m2015-0241.

[4] Pooja N. and R. N. Awale. "Prediction of Diabetes using Machine Learning Techniques: A Review." International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), vol. 6, no. 4, 2021, pp. 506-509.

[5] M. A. Soliman, et al. "A review of machine learning techniques for the prediction of diabetes." International Journal of Advanced Research in Computer Science, vol. 11, no. 5, 2020, pp. 10-16.

[6] A. M. Salem, et al. "An Integrated Machine Learning Approach for Diabetes Prediction." International Journal of Advanced Computer Science and Applications, vol. 12, no. 6, 2021, pp. 67-73.

[7] N. M. Salem, et al. "A Hybrid Machine Learning Approach for Diabetes Prediction." International Journal of Advanced Computer Science and Applications, vol. 12, no. 6, 2021, pp. 74-81.

[8] S. Mishra and S. K. Sharma. "Predictive Modeling of Diabetes Using Machine Learning Techniques: A Systematic Review." Journal of Medical Systems, vol. 44, no. 4, 2020, pp. 1-15.
[9] A. F. M. Fahad, et al. "A Comparative Study of Machine Learning Techniques for Diabetes Prediction." International Journal of Computer Applications, vol. 182, no. 45, 2018, pp. 16-22.
[10] A. Gupta and A. Pal. "Diabetes Prediction using Machine Learning Algorithms: A Review." International Journal of Engineering and Technology Innovation, vol. 11, no. 3, 2021, pp. 216-225.
[11] R. Venkataramanan and N. Deepa. "Predictive modeling of diabetes using machine learning techniques." Journal of Healthcare Engineering, vol. 2016, 2016, pp. 1-8.
[12] S. Sivakumaran, et al. "A Review on Diabetes Prediction Using Machine Learning Techniques." International Journal of Control and Automation, vol. 12, no. 2, 2019, pp. 119-128.
[13] M. M. Abd-Eldayem and A. S. El-Said. "Predicting diabetes using machine learning techniques: a comparative study." International Journal of Advanced Research in Computer and Communication Engineering, vol. 9, no. 7, 2020, pp. 2361-2366.