International journal of Inventive Research in Science and Technology ISSN: 2583-5637 (Online)

Volume 4 Issue 11 November 2025

PREDICTION OF HEALTHCARE EFFECTIVE DIABETES DISEASE USING SUPERVISED AIML ALGORITHM

RAJSHREE SINGH JAT
M.Tech Schollar,,MIT,Bhopal
jatpalak59@gmail.com

Prof. Preeti Mishra
Asst. Professor MIT,Bhopal
pritypreet85@gmail.com

ABSTRACT :- One of the major threats to human health today is Diabetes Mellitus (DM). Diabetes is a metabolic disease where a person suffers from increased sugar levels because either the pancreas does not produce sufficient insulin for the body or the cells do not respond to the insulin. Persistent Diabetes leads to malfunction, injury and failure of organs such as kidneys, eyes, nerves, blood vessels and heart.

The clinical data includes several tests needed to diagnose Diabetes mellitus, depending upon the healthcare personal experience. Hence, it is essential to predict the symptoms of Diabetes at its beginning stage to prevent its growth with appropriate medical diagnosis. These symptoms need to be wisely classified for early detection of Diabetes. Hence, the design of a classifier is essential for detecting Diabetes disease with optimal cost and better performance.

In this research, the main objective is to classify the data as diabetic or non-Diabetic and improve the classification accuracy. It presents an automatic prediction system for Diabetes mellitus through machine learning techniques. It considers several limitations of traditional classifiers and provides a significant relationship between patients' symptoms with diabetes diseases and the blood sugar rate. Machine learning offers reliable and excellent support for predicting a DM with the correct case of training and testing. Diagnosis of Diabetes mellitus requires good support of machine learning classifiers to detect diabetes disease in its early stage since it cannot be cured at later stages and subsequently bring more complications to a person's health system.

In this thesis, the Gradient boosting machine learning technique is implemented to train the Diagnosis of Diabetes and to classify the diabetes patients in two class values. Positive diabetes patients are defined by class '0' value, and negative diabetes patients are defined by class' 1'. The total Diagnosis diabetes dataset is 1145. All datasets applied to the Gradient boosting machine learning technique are divided into two groups. The first group consists of 815 datasets belonging to non-Diabetes, and the second group includes 330 datasets of Diabetes.

Keywords: Machine Learning, Supervised Learning, Diabetes Mellitus, Health System, Gradient Boosting, Python, Colab Platform, Recall, Accuracy

1.Introduction

1.1 Introduction

In several countries including United States, China, UK healthcare is a thriving domain that has industrialized as a leading sector related to economic growth and employment as well as functional overheads. From the study of the global spending on healthcare, it is anticipated that the general healthcare expenditure will reach double in the following 5 years [1, 2]. In India, public health spending is expected to increase to ₹486000 crores by 2022 from ₹267000 crores in the year 2018 (Ministry of Health & Welfare 2017). The overall financial outlay to health and wellbeing is expected to rise 45% from the financial year 2018 to 2022. Non value added and unproductive activities including diagnostic and medication errors, improper usage of antibiotics, readmissions, and deception create a substantial amount of spending for care. Approximately, 5.2 million Indians demise annually due to clinical faults and the consequences of malicious activities [3].

A suitable Clinical Decision Support System (CDSS) can handle these problems and accelerate the transformation towards a value based clinical system to deliver passable care and high services [4]. Digital therapeutics has been recognized to be a renowned intervention for treatment in disease control, and both medical professionals and patients are benefitting from CDSS. At present, the healthcare industry is adopting information and communication technologies for its organizational venture to deliver high quality care services in both commercial and technical aspects [5].

DM has developed as a long term deadly disease across the world and especially in developing nations. Several worldwide and nationwide epidemiological studies have observed the increasing number of diabetics across the globe. Recently, a study exposed that around 463 million individuals between the ages of 20 and 79 years are diabetic patients and it is projected that 552 million people are expected to get affected by the year 2030 [6]. Besides, DM is considered the 5th leading cause of disease oriented demises and is claimed to cause one death in every 6 seconds globally [7]. Unfortunately, 46.5% of DM patients have not been identified.

Worldwide around 422 million people suffers from diabetes. Diabetes mellitus can be understood to dissimilar types, first category being Type 1 diabetes and the other one being Type 2 diabetes [5, 6]. For case of about all the diabetes cases about 5 to 10% of cases have only the Diabetes "Type1". The Diabetes "Type 2" is about remaining 90% from all the aroused cases. The type 1 diabetes occurs to often kids or adults or during their adolescence days which is caused by the partial dis functioning of the Pancreas. It will not show any symptoms at the beginning stage because the Pancreas will be functioning partially [7]. This type 1 diabetes gets critical only when 80 90% of Insulin creating Pancreatic cells devastated. In insulin dependent diabetes which is resultant of very known chronic Hyper glycemia and the body is unable to regulate its own sugar, level which will lead to the shooting of sugar levels in the blood. This type 2 diabetes happens mostly to grown up people and affect more heavy and obese adults. When conducting the quantitative research, the diagnosis of the diabetes mellitus is the difficult part because the terms like the A1C, WBC (white blood cell) count, fibrinogen and parameters such as the hematological indices are ineffective because of some shortcomings [8].

The only way for the diabetic patient to live with this disease is to preserve the blood glucose level as normal as possible without extreme higher or lower levels [8], and this is achieved when the patient undergoes a proper medication which may include consuming oral drugs or some form of insulin, exercise, and nutrition. Furthermore, managing DM is also a perplexing, expensive, and difficult task for medical experts. There are huge vital data to store about the patients and diseases that support the doctors in making optimum clinical verdicts to improve the life expectancy of the patients. This makes the conventional machine learning based diagnosis approaches stop or degrades the training procedures as the algorithm becomes susceptible to over fitting due to the huge irrelevant and redundant attributes in a high dimensional database.

1.2 Diabetes Mellitus Occurs

Digestion is a process which requires food to be broken down into various forms of energy or sources of nutrients. Carbohydrates being taken by our body get converted into glucose by our body. Now, we have glucose present in the body which needs insulin to reach to its final destination i.e. the cells within the body so that which helps in building tissues and muscles. Pancreas which are present behind our stomach are responsible for making insulin. When the insulin is released into the blood, glucose enters into the blood by the body cells. If someone, has diabetes then this may be possible due to the various reasons listed below:

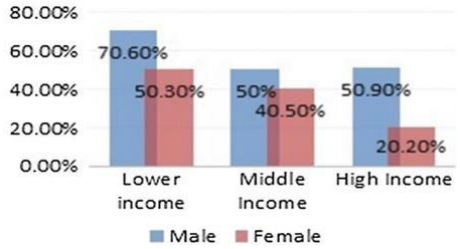

Figure 1.1: Comparison of death rate of diabetes among different classes of people

Pancreases are unable to form the enough amount of insulin required. When the pancreas forms the required amount of insulin, but the beta cells present in the blood does not retaliate to the production thus leaving it in the blood. Thus, Glucose remains unabsorbed thus allowing it to rise beyond levels [10].

1.3 ML Technique: Future for Diabetes Care

There has been a colossal impact in the healthcare sector with the proliferation of cutting edge technologies including AI, DM, ML etc. These techniques have demonstrated their effectiveness in disease prediction from massive amounts of healthcare data [11]. They are used to perform prognostic modeling, that is, the exploitation of data and statistics to calculate imminent results from past experiences.

Efficient ML approaches have been employed to develop algorithms in order to support frameworks in diagnosing diabetes in early stage or its subsequent complications. ML enables a constant and problem free system for monitoring biomarkers and symptoms. Several ML approaches have been proposed, including unsupervised, supervised, and reinforcement learning approaches. This is obviously practical because ML techniques are driven by data. Since the huge volume of big disease related data fed into the diagnosis system, ML can save substantial human effort. In the ML approach, models are trained on this data and deliver the more accurate output according to the input data. Some might analyze features of clinical scans, whereas others use blood test data collected from patients. Because there are several signs of the disease, the parameters differ accordingly. With several established techniques, investigators have investigated different algorithms and tweaked many hyper parameters to achieve outcomes that seem most appropriate for real time applications.

2. LITERATURE REVIEW

Isfafuzzaman Tasin et al. [1], in diagnosis or detection of temperature variations in the human skin are obtained from good resolution thermal images captured by the highly sensitive thermal imaging system. Historically, the first thermal image of human was acquired with a British prototype system; the "Pyroscan" were made in London for research in rheumatology, chronic fatigue, and pain management between 1959 and 1961. The thermal imaging

system has been utilized for many clinical applications in hospitals since the late 20th century. The thermal imaging system has been used for mass fever detection localization of swelling in dentistry biometric authentication of facial signature and palm vein. Recently, it has also been utilized to record the presence and position of the fetus and other physiological factors associated with pregnancy. This technique is effectively utilized for screening of temperature variations in minimally symptomatic patients, during the outbreak of the coronavirus disease –19.

Olisah et al. [2], is characterized in the form of intensity differences on a thermal image. The difference in the thermal pattern of pixel intensities results in texture variations in an image. These texture variations can be extracted in the form of statistical features and used for classification. As the temperature distribution of DM patients is different from healthy subjects, asymmetry amongst the features may be analyzed to gather significant information about the pathology. The human examination of plantar thermal images is susceptible to faults owing to carelessness, color blindness, fatigue, and repetitive tasks. The utility of the textural and higher order statistical features in finding the temperature distribution changes in plantar thermal images has been determined. These features have been utilized to differentiate diabetic feet from healthy ones.

Deberneh et al. [3], in this work, many feature extraction approaches are explored to design and develop efficient machine learning frameworks intended for detecting diabetic feet in plantar thermal images. In extension to the broadly utilized asymmetry method using textural and temperature features, the subsequent feature extraction approaches are also analyzed, for plantar thermal images. When the performance of several extracted features could not be used seemingly, machine learning methods have been utilized for training and automated classification. Many CNN models are employed to automatically segment, classify diabetic foot ulcers from color (RGB) images. The most significant features are also extracted from different layers of the pre trained CNN models using transfer learning. The process of analysing and collecting various databases and extracting important chunk of data which is used for the effective analysis is known as the Data mining technique. In the diabetic field, predictive analytics method is being used for diagnosis, prediction, self management, and prevention being done. In the recent trend, the research analysis helps in the diabetes prediction that plays a vigorous role in the case of High mortality and high morbidity, disease compilation and prevention.

Nikos Fazakis et al. [4], the dataset contained nine variables, eight of which contained patient information. The ninth variable was the class that predicted whether patients would develop diabetes or not. Outliers and missing values make up the dataset. We have removed the outliers from the dataset using our proposed method. The mean filter method was used to impute missing values that were present in the dataset, preserving the dataset's consistency. Weka was used in each and every one of the experiments. It is used for “Data Mining” and classification purpose. Classification algorithms may be applied to a dataset. The other applications of weka are visualizing, regression, purpose of data mining methods. Full form of Weka is Waikato Environment for knowledge Analysis.

Naveen Kishore G et al. [5], One of the most fatal and long-lasting diseases that causes an increase in glucose is diabetes. If diabetes is left undiagnosed and untreated, many problems arise. However, the advancement of AI approaches addresses this primary problem. Estimates of exactness are made based on correctly and incorrectly ordered cases. With a precision of 76.30 percent, the obtained results demonstrate that Naive Bayes prevails over relatively different calculations. R.O.C. bends are used in a legitimate and methodical way to verify these results. For the diagnosis of the diabetes these indices are necessary for research papers these indices are used. In-order to increase the A1C several treatment methods are used such as the ingestion of liquor, narcotics and salicylates. A1C can be increased by the process of vitamin-c ingestion but when analysed by the method of chromatography the count of A1C remains the same the increased output of A1C count can be obtained only through the electrophoresis analysis.

Chatrati et al. [6], there are a couple of simulated intelligence methodologies that are used to carry out judicious examination over colossal data in various fields. Perceptive assessment in clinical benefits is a troublesome endeavor regardless can assist experts with making gigantic data taught ideal decisions about open minded's prosperity. Perceptive examination in clinical consideration, man-made intelligence estimations are used in this assessment work. Six distinct AI calculations are applied to a dataset of the patient's clinical record for the purpose of analysis. The applied calculations' execution and precision are examined and analyzed. The best calculation for predicting diabetes is determined by correlating the various AI methods used in this study. One of the leading causes of disability and death is type II diabetes. These are incomplete and noisy data. A comprehensive estimation of missing data is presented in this paper. Correlation-based estimation models and k-NN were two of the five approaches to imputation of missing values that were compared.

Hasan et al. [7], have proposes the affirmation of Indian correspondence through marking movements using a mind blowing man-made awareness gadget, convolutional brain associations (CNN). The catch technique used in this work

is selfie mode continuous gesture-based communication video, allowing a conference-weakened individual to freely use the S.L.R. versatile application. CNN preparation is carried out using three distinct example estimates, each containing distinct subject and survey point arrangements. The prepared CNN is tested using the remaining two examples. With our selfie-based communication data, distinct CNN models were planned and tested for improved acknowledgment precision. In contrast to other classifier models written about the same dataset, our acknowledgment rate was 92.88%.

2.1 Problem Formulation

In the present era, diabetic disease results in increase within the death rate in most of the country. Diabetes disease is becoming a common disease occurring to humans due to an inadequate production of insulin and also due to high amount of sweetness in the body fluids. Before generating the clinical examination, several symptoms are adopted in cause of diabetic disease.

To effectively diagnose a health issue, proper care must be taken while considering the relevant parameters such as the patient's daily routine, eating habits, medical history, etc. For efficient use of machine learning in predicting and diagnosing the health problem, the above mentioned parameters are necessary as an input variable for successfully formulating an algorithm. The problem formulation is carried out as follows:

- A most effective tool is needed for early stage prediction and diagnosis of disease such as Diabetes Mellitus, Cancer, Heart problems etc., with the highest accuracy and lowest misclassification.
- In medical science, disease data diagnosis involves many medical tests that is needed to diagnose a specific disease. The effective diagnosis depends on the health care personal experience since a less experience person can misdiagnose a health problem.
- The number of false positives is quite high in some specific cases, which can be further reduced using the Machine learning algorithm.
- The Classifier to be designed should be efficient, convenient, and, most importantly, must be capable of classifying and predicting the Diabetes patient with the highest accuracy and minimal misclassification.

3. MACHINE LEARNING

ML is perhaps the most discussed subject in the realm of innovation. Along with the guarantees and advantages, it is also generally related to discussions and discussions. Individuals who don't know about the nature and benefits of ML or have accepted their data from dishonest sources regularly peer down on ML and are frightened of it also. In any case, every one of the unusual and strange things that have come out about ML is most likely legends and bogus worries.

3.1 Machine Learning Approach

The word ML and AI are driving the figures by a considerable degree. This consistent ascent in the notoriety of AI is a direct result of its rising use in our everyday lives. It is, these days, being utilized in different gadgets and machines just as devices. Along these lines, to get rid of such fantasies, let us view the short history of ML [15].

Their usefulness will increment by an enormous degree, becoming a necessary resource for humankind. ML can be utilized in practical fields of epistemology. It is inspected that the acoustics signals have been gained from the rotating machines, and they have been utilized with the wavelets for the helpful determination of the outcomes.

3.1.1 Differences Between Traditional Programming and Machine Learning

Conventional Programming: The information is taken care of by the P.C., and a program is run. This program then presents yield at that point, utilizing the provided information. ML Pre tackled information and the subsequent yield are taken care of to the P.C. These two information sources are utilized to make a program. This program then, at that point can do the work of customary programming [16, 17].

3.1.2 Elements of ML As ML is a muddled and tangled field, it's difficult to comprehend its nuts and bolts. It is likewise a steadily developing field. Subsequently, it is feasible to see new improvements in the space consistently. For example, it is accepted that consistently in excess of a couple hundred, new calculations are fostered everywhere. This brings the quantity of in general ML calculations to a total that is bigger than 10,000. Despite the fact that a ton of assortment is found in the calculations of ML, every one of them depends on three fundamental ideas that are as per the following. He utilized improved eigen vector calculations for identifying various kinds of deficiencies in machine components [18]. The vibration signals were utilized to address them in the structure eigen vectors comparing to each condition so the arrangement is performed. The outcome showed that the eigen vector portrayal of the vibration signal was smarter to utilize. Notwithstanding, addressing the vibration signal as eigen vector requests solid subject information.

3.2 Different Types of Machine Learning Technique

3.2.1 Logistic Regression

Calculated relapse [15] works with likelihood and chips away at discrete arrangement of classes. Different classes are given to the observations. Linear vs logistic regression is represent in figure 3.1. By utilizing this classifier, we can foresee regardless of whether an individual has specific sickness.

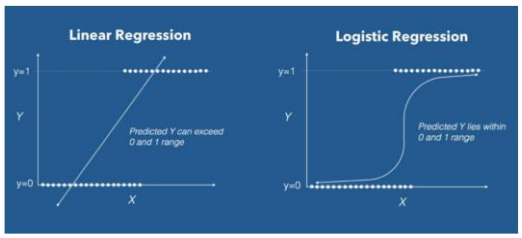

Figure 3.1: Linear Regression VS Logistic Regression
Graph Image: Data Camp

3.2.2 Decision Tree Classifier

A decision tree classifier is a simple but powerful classification model. It is a supervised learning model having a flowchart like structure that creates a training model from the dataset is shown in figure 3.2. A decision tree is a tool for the detection of patterns, relationships, and knowledge from data. Decision trees give a straightforward visualization of data. These are easy to interpret for decision-making in real-time applications. There are different issues related to training with large data sets and the construction of decision trees. To handle have these problems different by different researchers. These approaches related to decision trees are summarized in this chapter. The emphasis is given on the approaches to create accurate and optimized decision trees. A decision tree is a supervised classifier having a tree structure made up of a set of nodes. These nodes are categorized into internal nodes and leaves. Internal nodes are also called decision nodes because attributes are tested at these nodes to take a decision and partition the data set. Leaves are used to represent predicted class or decision class. In a decision tree construction, splitting criteria is used for partitioning attributes.

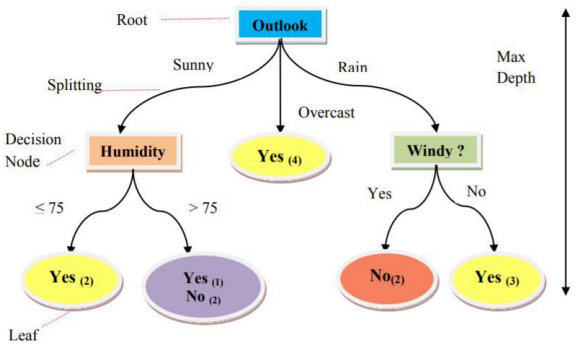

Figure 3.2: Simple Decision Tree

The greedy approach selects a globally optimal solution at each stage. The selection made by a greedy approach may depend on choices made so far in the construction of a tree. These choices do not depend on future choices or all the solutions of the sub problem. In each iteration, it makes one greedy choice after another. This leads to reducing each specified problem into a smaller one [16].

4. PROPOSED METHODOLOGY

4.1 Introduction

The primary goal of this work is to categories data as diabetes or non-diabetic and to enhance classification accuracy. It offers an automated diabetes mellitus prediction system based on machine learning methods. With the proper scenario of training and validation, machine learning offers a dependable and great assistance for DM prediction. Diabetes mellitus detection requires significant assistance from machine learning classifiers to identify the illness at a preliminary phase, as it cannot be treated, posing significant complications for our health system. This study aided in the development of a classification system for DM prediction.

In latest technologies, monitoring of health is a highly important job. Diabetes Mellitus is amongst the most problematic illness both in developing & developed countries [24]. The science and medicine diagnosis of illness data consists of a variety of medical tests that are required to identify a specific condition and the diagnosis will rely on the doctor expertise, if a less expertise physician may detect an issue wrongly. So, a doctor has to examine a number of problems and variables that makes the physician work tough for identifying the illness. In 2004, it was projected approximately 3.4 million individuals suffered with high blood sugar [25], a condition that results in diabetes mellitus. Diabetes is gradually becoming one amongst leading causes of chronic infections, with prevalence expected to almost double form 1.5 million over 3.5 million for next two decades if no further measures are taken to combat the disease. Furthermore, some analysis has been conducted on illnesses that often result in misdiagnosis as well as an elevated mortality rate of approximately 98,000 individuals each year as a result of this medical mistake. Early diagnosis of diabetes mellitus is critical, especially given that between 50% as well as 80% of patients with diabetes remain unaware of the disease.

4.2 Proposed Methodology

4.2.1 XG Boosting Classifier

XG Boosting is an effective machine learning method that may be used for a variety of classification problems, including the classification of breast cancer. With the use of the ensemble learning technique known as gradient boosting, a powerful predictive model is produced by combining the predictions of several weak learners, often

decision trees. It's a preferred option because it frequently produces high accuracy and can manage intricate data interactions.

Assemble a collection of characteristics (attributes) that are associated to breast cancer cases. These characteristics could include things like patient age, tumour kind, and size. Whether a case is benign (non cancerous) or malignant (cancerous), the target variable should show this XG Boosting. Make a training set and a testing/validation set from breast cancer dataset. The Gradient Boosting model is trained on the training set, and its performance is assessed on the testing/validation set. Give details on hyperparameters such learning rate, number of boosting iterations, and maximum depth of each tree. Tuning hyperparameters is crucial for improving model performance. Depending on the precise objectives of breast cancer prediction analysis, evaluate the model's performance on the testing/validation set using the appropriate metrics, such as accuracy, precision, recall, F1 score, or ROC AUC. Although Gradient Boosting models are not the easiest to grasp, learning about the significance of particular features.

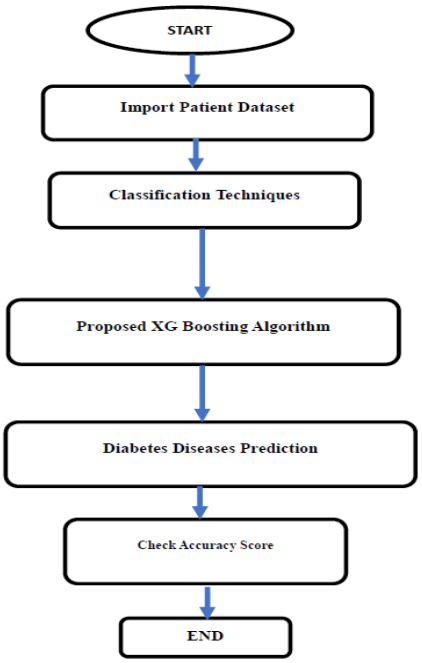

Figure 4.1: Flow chart of Proposed Algorithm

4.2.3 Writing the Program in Python The precise syntax, semantics, and tokens of Python are contained in the Python Standard Library. I/O and a few other essential modules are among the built in modules that give access to fundamental system functions. Most of the Python Libraries are written in the C programming language. There are more than 200 essential modules in the Python standard library.

5. SIMULATION RESULTS

5.1 Simulation

Simulation Tool

Python tool with colab platform is used for this project. All of them are free and open source. Python is a general programming language and is broadly utilized in a wide range of disciplines like general programming, web improvement, programming advancement, information investigation [40]. Python is utilized for this venture since it is entirely adaptable and simple to utilize and furthermore documentation and local area support is exceptionally huge. Colab is a hosted Jupyter Notebook service that requires no setup to use and provides free access to computing resources.

Data Set

The investigation is based on the U.C.I. machine learning store's diabetic Pima Indian dataset, which contains 1145 information occurrences and 9 traits. The data set is downloading in dataworld.com.

Dataset Description

The investigation is based on the U.C.I. machine learning store's diabetic Pima Indian dataset [12], which contains 1145 information occurrences and 9 traits. All of the women in this dataset are Pima Indians and are at least 21 years old. Their age is indicated by either a "0" or a "1," with a "0" indicating a negative test for diabetes and a "1" indicating a positive test. Table 5.1 shows the number of features, classes and patterns and table 5.2 represent the confusion matrix of prima Indians diabetes dataset.

Table 5.1: Description of benchmark dataset for diabetic for pima Indians

Datasets	No. of features	No. of classes	No. of patterns
Pima India Diabetic Dataset	8	2	1145

Table 5.2: Confusion Matrix

Datasets	No. of classes	No. of patterns
Actual class	TP	FN
	FP	TN

Table 5.3: Description of Diabetic Data Set

Data Set	No. of Attributes	Feature Set
Diabetic	9	No. of times of pregnant Plasma glucose concentration Diastolic blood pressure Triceps skin fold thickness Serum insulin Body mass index Diabetes pedigree function Age of patient Class' 0' or '1'

Parameter Details:- The different significant parameter used for Gradient Boosting are center, spread and weight. The different attributes used for diabetic data set are described in table 5.3.

Evalution metrics: By and large, the evaluation of an order issue depends on data called as a disarray framework, with the number of testing tests effectively grouped and inaccurately arranged represented as takes after.

So, the accuracy can be measured according to Eq. 5.1

$$Accurancy = \frac{TN + TP}{TN + TP + FN + FP}$$

5.2 Result Analysis

Machine Learning is a thought that agrees over the machine to take data from instances and former knowledge, and learn from historic data to make predictions based on the learning of the past data and that too without being programmed by any programmer i.e. we can use previous data for future predictions. In this case, instead of programmer writing the code, what a naïve user can do is feeding data to the generic algorithm, and the logic is build based on trained data by the algorithm/ machine. For e.g. When we shop online, while looking for a product, we have noticed that similar products are recommended to us to what we were looking for and we also notice the following quotation “the person who purchased this product also purchased this” type of combination of products. This recommendation is done using machine learning. Many a times we get a phone call from the bank or the finance company asking us to take a loan or purchase an insurance policy.

Figure 5.1 shows the histogram of attributes and the range of dataset attributes and code used to create it.

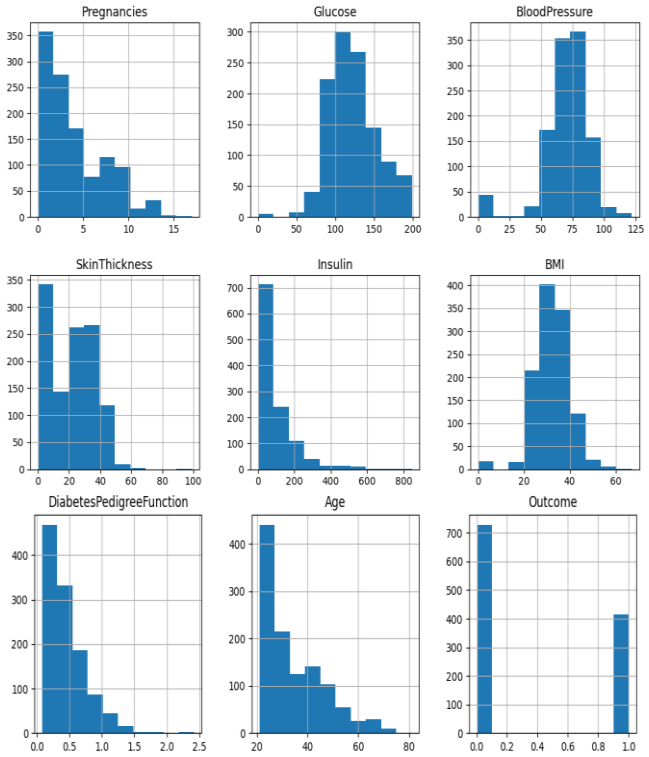

Figure 5.1: Histogram of Dataset

Figures 5.2 and 5.3 show the status of Diabetes health, ranging from healthy to severely unhealthy. Blue bar represents Diabetes disease, and the red bar represents not Diabetes disease.

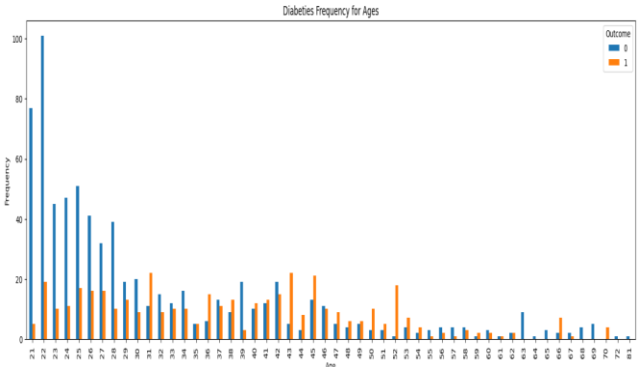

Figure 5.2: Bar Plot of the Number of Diabetes Frequency for Ages

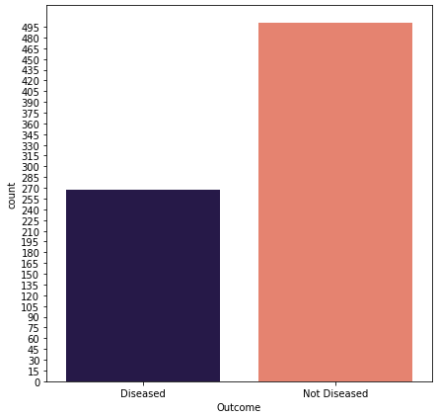

Figure 5.3: Bar Plot According to Outcomes

Machine Learning algorithms are trained by using a training data set for model creation. When some new input data from the attributes is feed to the algorithm of machine learning, prediction is done on the basis of the selected model is shown in figure 5.4. Then the predictions are measured for accuracy. If the accuracy of the input data is acceptable, deployment of the Machine Learning algorithm is done on input data. If the accuracy of input data is not acceptable, the algorithms with some new data are trained again and again with an arbitrary training data set.

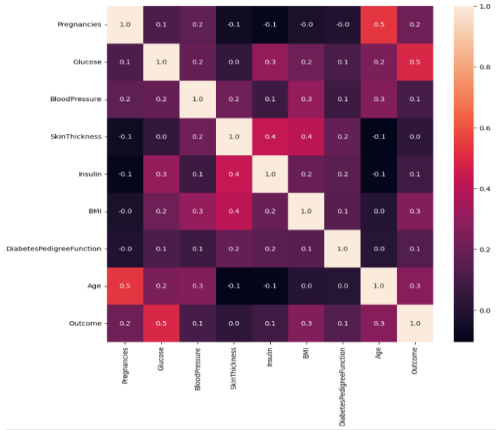

Figure 5.4: Bar Plot According to Diabetes Pedigree Function

In proposed demonstrating at least two related however unique scientific models are utilized and produce their outcomes are joined into a solitary score. Subsequent to playing out the AI approach for testing and preparing we find that exactness of the Inclination supporting is much proficient when contrasted with different calculations. Precision ought to be determined fully backed by disarray framework of every calculation as displayed in Figure 5.5, here number of counts of T.P., TN, F.P., F.N. are given and utilizing the condition of precision, esteem has been determined and it is presume that proposed calculation is best among them with 92.18% exactness and the correlation is displayed in Table 5.4.

Table 5.4: Assessment of Different Classification

Sr. No.	Algorithm	Accuracy
1	LR	74.82%
2	GNB	73.42%
3	RFC	83.56%
4	K-NN	71.32%
5	DT	80.76%
6	SVM	82.51%
7	Proposed Algorithm	91.23%

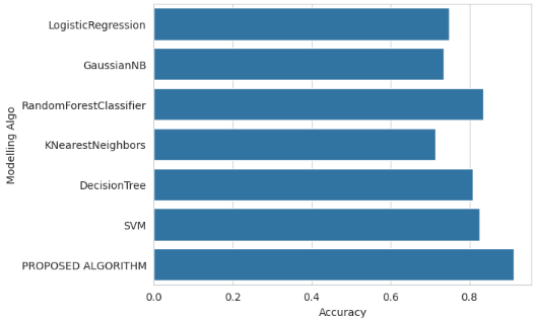

Figure 5.5: Bar Graph of the Different Classification Methods

Table 5.5 represents the accuracy for different ML classifier with Previous Isfafuzzaman Tasin et al. [1]. GB classifier is best accuracy compared to Previous Isfafuzzaman Tasin et al. [1]. Bar Graph of the previous and proposed Algorithm for Accuracy in Diabetes Dataset is representing in fig. 5.6.

Table 5.5: Comparison Result for Accuracy

Techniques	Previous Isfafuzzaman Tasin et al. [1]	Proposed Algorithm
DT	72%	80.76%
K-NN	73%	71.32%
LR	75%	72.82%
RF	76%	83.56%
SVM	78%	82.51%
NB	79%	73.42%
GB	81%	91.23%

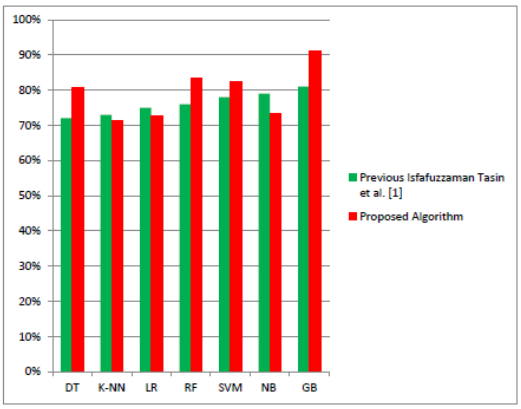

Figure 5.6: Bar Graph of the Previous and Proposed Algorithm for Accuracy

6. CONCLUSION AND FUTURE WORK

6.1 Conclusion

One significant health challenge worldwide is Diabetes detection early and proactively. The present study aims to establish a suitable prediction model that depends on a Machine Learning scheme for predicting Diabetes. Diabetes is a metabolic condition brought on by elevated levels of glucose in the blood. Patients with diabetes don't

have enough insulin in their bodies to control their sugar levels. Additionally, diabetes contributes to a number of other dangerous diseases. Due to the fact that diabetes is the root cause of numerous diseases in the human body, it is essential to detect this serious condition as soon as possible.

There have been a number of approaches taken in the past to make it easier for doctors to diagnose diabetes. However, there is a classification issue with it: a selection of an excessive number of samples, but it does not improve classification accuracy. The speed of the algorithm is better in a few situations, but the accuracy of the data classification is lower. A selected classification algorithm is used to test the diabetic data set. We applied a gradient-boosting machine learning algorithm to the Diabetes dataset's attributes in order to get the best results.

The diabetes dataset of 1145 Pima Indians: The test uses 330 diabetic and 815 non-diabetic participants. An ensemble of gradient boosting was used in the proposed algorithm to achieve an accuracy of 91.23%. As can be seen, the majority vote-based model employs NB, DT, and SVM classifiers, and its accuracy for the diabetes disease dataset is 73.42%, 80.76%, and 82.51%, respectively. Subsequently, the Inclination helping calculation gives the best exactness to Diabetes finding than the past calculation.

6.2 Future Scope

Generally speaking, we guarantee that this exploration has provided another guidance to evaluate individual wellbeing records with AI procedures. We propose two potential headings for future examination.

- Only supervised learning methods and predefined health datasets are used in the proposed approach. This structure will deal with electronic wellbeing records utilizing solo learning approaches in future.
- The proposed work may be expanded to include unstructured and semi-structured information in the future, but it currently only addresses structured data.

REFERENCES

[1] Isfafuzzaman Tasin, Tansin Ullah Nabil, Sanjida Islam, Riasat Khan, “Diabetes prediction using machine learning and explainable AI techniques”, Healthcare Technology Letters, pp. 01-10, Wiley 2022.

[2] Olisah, C.C., Smith, L., Smith, M., “Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective”, Comput. Methods Programs Biomed., Vol. 20, pp. 1–12, 2022.

[3] Deberneh, H.M., Kim, I., “Prediction of type 2 diabetes based on machine learning algorithm”, Int. J. Environ. Res. Public Health, Vol. 18, pp. 1–14, 2021.

[4] Nikos Fazakis, Otilia Kocsis, Elias Dritsas, Sotiris Alexiou, Nikos Fakotakis, and Konstantinos Moustakas, “Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction”, IEEE Access 2021.

[5] Naveen Kishore G, V.Rajesh, A.Vamsi Akki Reddy, K.Sumedh, T.Rajesh Sai Reddy, “Prediction of Diabetes using Machine Learning Classification Algorithms”, International Journal of Scientific & Technology Research, Vol. 9, No. 01, 2020.

[6] Chatrati, S.P., Hossain, G., Goyal, A., “Smart home health monitoring system for predicting type 2 diabetes and hypertension”, J. King Saud Univ. Comput. Inf. Sci., Vol. 34, No. 3, pp. 862–870, 2020.

[7] Hasan, M.K., Alam, M.A., Das, D., Hossain, E., Hasan, M., “Diabetes prediction using ensembling of different machine learning classifiers”, IEEE Access, Vol. 8, pp. 76516–76531, 2020.

[8] Cervantes, J., García-Lamont, F., Rodríguez, L., Lopez-Chau, A., “A comprehensive survey on support vector machine classification: Applications, challenges and trends”, Neurocomputing, Vol. 408, pp. 189–215, 2020.

[9] Pranto, B., “Evaluating machine learning methods for predicting diabetes among female patients in Bangladesh”, Information Vol. 11, pp. 1–20, 2020.

[10] Mohan, N., Jain, V., “Performance analysis of support vector machine in diabetes prediction”, In: International Conference on Electronics, Communication and Aerospace Technology, pp. 1–3, 2020.

[11] Muhammad Azeem Sarwar, 2Nasir Kamal, 3Wajeeha Hamid,4Munam Ali Shah, “Prediction of Diabetes Using Machine Learning Algorithms in Healthcare”, 24th International Conference on Automation & Computing, Newcastle University, Newcastle upon Tyne, UK, 6-7 September 2019.

[12] Rao G.A., Syamala K., Kishore P.V.V., Sastry A.S.C.S. ., “Deep convolutional neural networks for sign language recognition”, International Journal of Engineering and Technology(UAE) ,Vol: 7, Issue 5, pp: 62-70, 2018.

[13] Quan Zou, Kaiyang Qu, Yamei Luo, Dehui Yin, Ying Ju and Hua Tang, “Predicting Diabetes Mellitus With Machine Learning Techniques”, Springer, 2018.

[14] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, “Machine learning on big data: Opportunities and challenges,” Neuro computing, vol. 237, pp. 350–361, May 2017.

[15] J. B. Heaton, N. G. Polson, and J. H. Witte, “Deep learning for finance: deep portfolios,” Appl. Stoch. Model. Bus. Ind., vol. 33, no. 1, pp. 3–12, Jan. 2017.

[16] Reddy S.S., Suman M., Prakash K.N. ., “Micro aneurysms detection using artificial neural networks”, 2018, Lecture Notes in Electrical Engineering ,Vol: 434 ,Issue 3, pp: 409 to 417.

[17] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., Chouvarda, I., “Machine Learning and Data Mining Methods in Diabetes Research”, Computational and Structural Biotechnology Journal 15, 104–116, 2017.

[18] F Mercaldo V Nardone and A Santone "Diabetes Mellitus Affected Patients Classification and Diagnosis through Machine Learning Techniques" Procedia Computer Science vol. 112 pp. 2519-2528 2017.

[19] I Kavakiotis O Tsave A Salifoglou N Maglaveras I Vlahavas and I Chouvarda "Machine learning and data mining methods in diabetes research" Computational and structural biotechnology journal 2017.

[20] J Siryani B Tanju and TJ Eveleigh "A Machine Learning Decision-Support System Improves the Internet of Things Smart Meter Operations" IEEE Internet of Things Journal vol. 4 no. 4 pp. 1056-1066 2017.