Prediction of Diabetes using Classification Algorithms

Aadesh Mallya[1] Pranav Kangane[2] Shivam Gulve[3] Aayush Gawane[4]
[1,2,3,4]Department of Computer Engineering
[1,2,3,4]Mukesh Patel School of Technology and Engineering, Mumbai, India

Abstract— Diabetes is considered to be one of the deadliest and chronic diseases that causes a rise in blood sugar. Many difficulties might occur if diabetes stays untreated and unidentified by a specialist. Some of the complications faced due to diabetes are excretory organ injury, typically resulting in chemical analysis, eye damage that may end in visual impairment, or associate degree enhanced risk for cardiopathy or stroke. Usually, the identifying process is tedious which consists of visiting a patient to a diagnostic center and then consulting a doctor. However, the rise in machine learning approaches solves this critical problem. The objective of this paper is to develop a system which can early predict the likelihood of diabetes in patients with maximum accuracy. Hence, machine learning classification algorithms are used in this experiment to detect diabetes at an early stage. Experiments are performed on Pima Diabetes dataset. The performances of various classification algorithms are evaluated on different performance measures such as precision, accuracy, and f-score. Results obtained show Logistic Regression outperforms with the highest accuracy of 82.46% comparatively other algorithms.

Keywords: Machine Learning, SVM, Logistic Regression, KNN, Decision Trees

I. INTRODUCTION

Presently, In the medical field, classification strategies are broadly used for classifying data into different classes. Diabetes is considered to be one of the deadliest diseases in the world. It affects the ability of the body in producing the hormone insulin, which in turn makes the metabolism of carbohydrates abnormal and raises the levels of glucose in the blood. A person generally suffers from high blood sugar in Diabetes. Intensifying their thirst, hunger and frequent urination are some of the symptoms caused due to high blood sugar [7].

Diabetes is also a creator of various sorts of diseases like heart failure, blindness, etc. It is examined as a vital serious health matter during which the measure of sugar substance cannot be controlled. Diabetes is not only affected by various factors like height, weight, hereditary factors and insulin but the major reason considered is sugar concentration among all factors. The early identification is the only remedy to prevent complications [6].

A diabetic patient is characterized by several symptoms which is more or less common for diabetics all around the world. While a pathological test for diabetes can give us a conclusive result as to whether a patient is diabetes positive or not, there are many who are unable to undergo such a test owing to financial restrictions, busy schedule etc. This experiment aims to make test for diabetes predictions easier i.e., based on some common knowledge data, without the need to undergo a test. In our experiment, we aim to test the efficiency of several different machine learning algorithms, in accurately predicting whether a patient with certain health attributes, is susceptible to being diabetes positive or not.

While there are many complex algorithms, namely, XGBoost, Naïve Bayes, etc., we will be, in our paper, focusing on the traditional classification algorithms namely, KNN, decision tree, support vector machines and logistic regression and compare their efficiency. While it is impossible to definitely predict the possibility of being diabetic, based on common knowledge health attributes, our model was able to achieve almost an 83% accuracy with minimum overfitting.

The remaining article is structured as follows. We describe the Literature Survey in Section II. Methodology and machine learning algorithms are described in Section III. Performance comparison is described in Section IV. The conclusion is discussed in Section V and Future Work in Section VI. Finally, Section VII concludes the references.

II. LITERATURE SURVEY

Various researches have been led in the diabetes area by using data mining as an incredible analytics tool to extricate information from huge accessible data. In this research, numerous prediction models were developed and implemented by various researchers using machine learning models and data mining techniques. Using a variety of classification algorithms such as Naive Bayes and Decision Trees, Aishwarya Iyer (2015) was able to discover hidden patterns in the diabetes dataset. Various performance measures were used to measure the effectiveness of the algorithm [1]. K. Rajesh (2012) used the C4.5 decision tree algorithm to discover hidden patterns in the dataset to classify effectively [2]. A model utilizing Random Forest was proposed by Mani Butwall (2015) to estimate diabetes conduct [3]. Sanakal and Jayakumari planned a model for demonstrating if the patient is determined to have diabetes. The model utilized 9 features for diagnosing and was based on the Fuzzy C-means clustering algorithm. In the wake of executing the model, FCM demonstrated 94.3% precision and positive prescient 88.57% [4]. Namayanja et al made a clustering model to explore the estimations in blood glucose and portions of normal. The model was created dependent on genuine diabetic patients' data. The model is meant to recognize, for example, targeted care for diabetic patients [5].

III. METHODOLOGY

The Methodology represents a description of the framework that is undertaken. It consists of various milestones that need to be achieved in order to fulfil the objective. We have undertaken different data mining and machine learning concepts. The following outline (Fig. 1) represents step-wise tasks that needs to be completed.

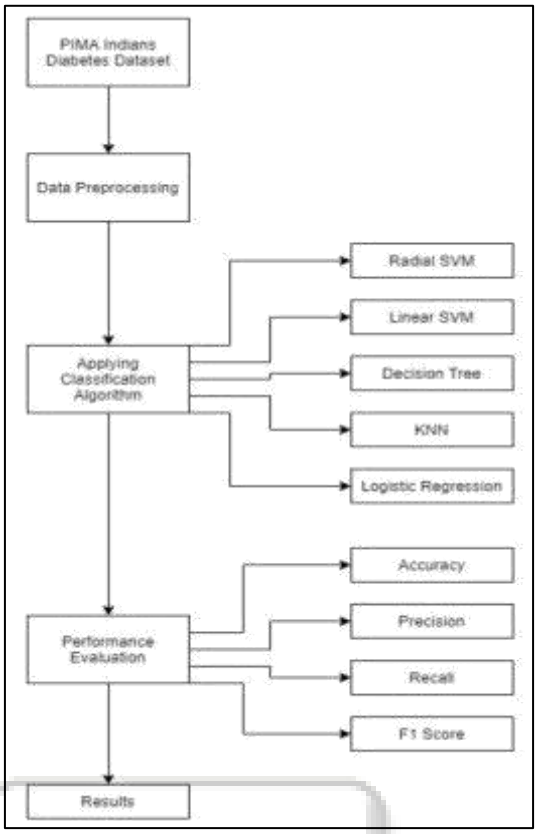

Fig. 1: Proposed Methodology

A. *Data Collection*

The proposed approach is assessed on Diabetes Dataset in particular Pima Indians Diabetes Dataset. This dataset is initially from the National Institute of Diabetes and Digestive and Kidney Diseases. The goal of the dataset is to anticipate whether a patient has diabetes, using various diagnostic measurements. This dataset includes clinical detail of 768 occasions which are females at any rate 21 years of age of Pima Indian legacy. The dataset contains 8 independent factors used to decide the final outcome. The dataset comprises of the following attributes:

- Pregnancies: Number of times pregnant
- Glucose: Plasma glucose concentration in an oral glucose test
- Blood Pressure: Diastolic blood pressure (mm Hg)
- Skin Thickness: Triceps skin fold thickness (mm)
- Insulin: 2-Hour serum insulin (mu U/ml)
- BMI: Body Mass Index
- Diabetes Pedigree Function
- Age
- Outcome: ’0’ treated as tested negative for diabetes and value of another class ’1’ is treated as tested positive for diabetes

B. *Data Preprocessing*

It is a process of transforming the raw, complex data into systematic understandable knowledge. It involves the process of finding out missing and redundant data in the dataset. In this dataset it was found that no attribute contains any null value. Since, all the attributes were distinct due to which no redundant data was present.

C. *Data Anlaysis*

Before applying any model to our dataset, we need to find out the characteristics of our dataset. Thus, we need to analyze our dataset and study the different parameters and relationships between these parameters. We can also find out the outliers present in our dataset. Outliers occur due to some kind of experimental errors and they need to be excluded from the dataset.

A correlation number gives the degree of association between two variables. The correlation number exists between +1 to -1. A positive number represents a positive correlation between two variables and vice versa. However, if the correlation number is 0, it shows that there is no correlation between two variables and they are independent of each other. The correlation Matrix gives an in-depth idea about the correlation among various parameters. A correlation matrix for 8 parameters in the following Fig. 2 is plotted.

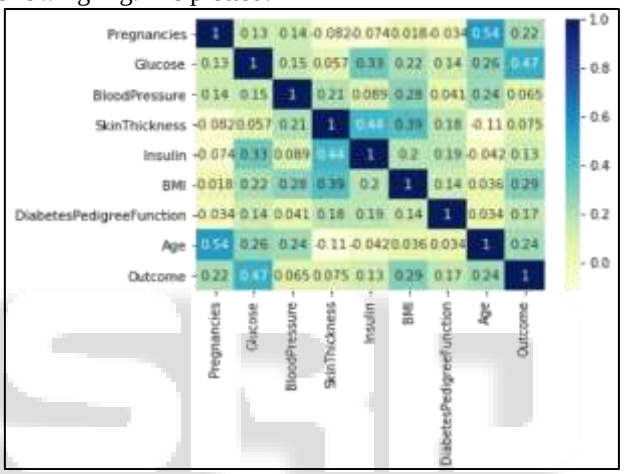

Fig. 2: Correlation Matrix

From the correlation matrix, it is observed that pregnancy and age have the highest correlation of 0.54 followed by glucose and outcome with a correlation of 0.47.

Fig.3 shows distribution of diabetic people age-wise whereas Fig. 4 shows distribution of non-diabetic people age wise. It is observed that 27.43 % of people in the range 51 – 60 have diabetes followed by 26.98% of people in the range of 41- 50. The second pie chart depicts that 27.04% of people in the range 21-30 are non-diabetic.

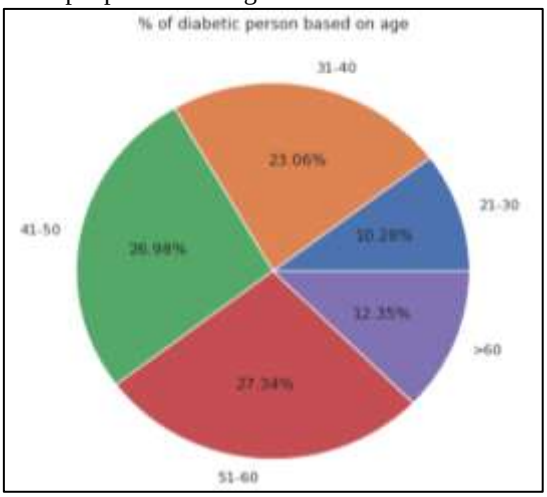

Fig. 3: Diabetic Person % based on age

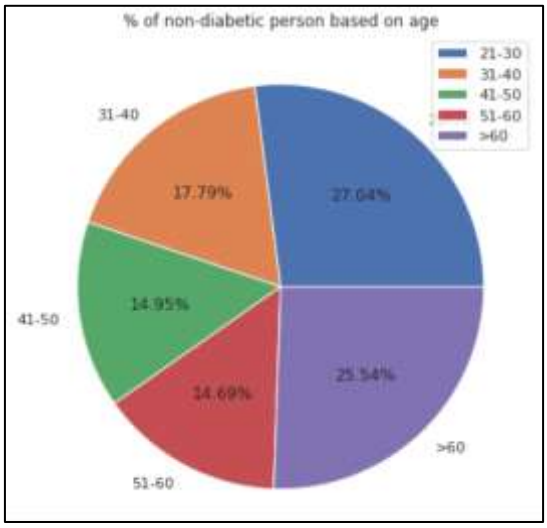

Fig. 4: Non- Diabetic Person % based on age

D. Algorithms Applied

Once the data is clean and we have gained insights about the dataset, we can apply an appropriate machine learning model that fits our dataset. We have selected five algorithms to predict the dependent variable in our dataset. The algorithms that we have selected are basically Radial SVM, Linear SVM, Decision Tree, Logistic Regression and KNN. These algorithms were implemented with the help of python's SciKit-learn Library.

1) Decision Tree

Decision Tree is a type of supervised learning algorithm used for both classification and regression tasks. It comprises a tree-like structure, where every internal node denotes a test on feature. The outcome of the test is denoted by every branch and the leaf node or terminal represents the class label [8]. For a given record, the value of the attributes is tested against the decision tree. For a particular tuple, a path is traced by the algorithm from the root node to the leaf node that holds the class prediction. Information gain is used to determine which feature to split on at each step while building the tree. Split with the highest information gain is taken as first split and the process continues until all children nodes are pure or gain is 0 [14]. Pure means, in a selected sample of dataset all data belongs to the same class.

a) Benefits:

- Handles categorical and numerical attributes
- It is resistant to outliers
- Easy to comprehend and utilize

b) Drawbacks:

- Prone to overfitting
- If some classes dominate then it creates biased trees.
- Need to be careful with parameter tuning.

c) Applications:

- Business: Utilized in the perception of probabilistic business models, utilized in customer relationship management [8].
- E-Commerce: Utilized in e-commerce to generate an online catalogue which is fundamental for the achievement of an e-commerce web site [8].
- Medicine: Decision trees are generally helpful in the diagnostics of different illnesses. And furthermore, utilized for Heart sound determination [8].
- Industry: It can be used in fault detection and non-destructive tests areas.

2) Logistic Regression

Logistic regression is a supervised learning classification algorithm. Binary logistic regression is a specialized type of logical regression used to predict binary responses such as pass/fail, yes/no. A binomial logistic regression consists of a dichotomous dependent variable and it predicts in which of the two categories, the observation falls into depending on several independent and continuous attributes. The outcome can only be in two categories like success/failure, True/False, etc.

a) Benefits:

- Logistic regression provides a good accuracy for many simple data sets and when the dataset is linearly separable, it performs really well.
- Logistic regression is easier to implement, interpret and it is very efficient to train.

b) Drawbacks:

- The most common limitation of Logistic Regression is the assumption of linearity between the independent variables and the dependent variable.
- Non-linear problems can't be solved with logistic regression because they have linear decision surfaces and the data which is linearly separable is usually not found in real-world scenarios.

c) Applications:

- Medicine: Medical researchers can make use to logistic regression to know how exercise and weight impact the probability of having a heart attack.
- Email Spam: It can be used to classify whether the email is spam or not and then putting it in either Inbox or Spam folder.
- Sentiment Analysis: Analysing the sentiment of an individual by using their review, tweets or social media is the objective of this use case.

3) Random Forests

Random forest is a supervised learning algorithm which is used for both classification and regression. A forest comprises trees. The algorithm creates a decision tree on arbitrarily selected data samples, gets the prediction from every tree and selects the most optimal solution by means of voting. Random forests are also useful in good feature selection. The algorithm calculates the relevance score of every attribute in the training phase. The score is then scaled down so that sum of all scores in 1. This score is then utilized to select the most important feature and drop the least important ones for model building.

a) Benefits:

- Considered as a highly accurate and robust method because of the number of decision trees participating in the process.
- The model takes the average of all predictions and cancels out the biases. Thus, it does not suffer from an overfitting problem.
- It can handle missing values.

b) Drawbacks:
- The prediction is slow due to the presences of multiple decision trees. For every input, all trees have to make prediction and then vote which makes the process time consuming.
- Interpretation of the path is difficult as compared to decision tree.

c) Applications:
- Banking Sector: The banking sector comprises of loyal and fraud customers and to classify them, Random forest analysis is used.
- Medicines: Medicines need a complex combination of specific chemicals. It is used in detection and prediction of drug sensitivity of a medicine. Also used in identifying medicines best combination.
- Stock Market: It is used in stock market analysis and for forecasting the behavior of stock market.

4) KNN

The K-Nearest Neighbour is a machine Learning algorithm that is based on the supervised Learning technique. This algorithm can be used for Regression as well as classification but it is mostly used for the classification problems. It assumes the similarity between the new data and available data and then it puts the new data into the category that is most similar to the available categories. KNN algorithm does not learn from the training set immediately as it stores the dataset and at the time of classification it performs an action on the dataset.

a) Benefits:
- K-NN just tags the new data entry-based learning from historical data as it does not explicitly build any model. Any new data entry will be tagged with the majority class in the nearest neighbour.
- The K-NN algorithm can be used both for classification and regression problems.
- There are no assumptions to be met to implement the K-NN algorithm whereas Linear regression has lots of assumptions that are to be met by data before it can be implemented.

b) Drawbacks:
- The K-NN algorithm functions admirably with a small number of input variables but as the number of variables grows it struggles to predict the output of new data points.
- One of the biggest issues with the K-NN algorithm is to choose the optimal number of neighbours to consider while classifying any new data entry.
- K-NN cannot comprehend any missing values.

c) Applications:
- Forecasting Stock market: it is used to predict the price of a stock on the basis of company performance measures and economic data.
- Medicine: Predict whether a patient who is hospitalized due to a heart attack will have a second heart attack based on demographic, diet and clinical measurements for that patient.

5) SVM

The SVMs are a powerful set of algorithms based on different kernel functions (we will be explaining this in detail later) mapping techniques to generalize classification on small dataset. The SVMs function on the assumption that a machine can make mistakes while classifying an object into different classes i.e., it takes the errors made by the model into consideration before trying to fit a line that best separates the classes [9].

For example – consider two classes of objects, apples, and oranges. The SVM will try to find the apple that looks the most like an orange, and an orange that looks the most like an apple (cases where the machine is most likely to misjudge) and then use them as support for maximum possible error in judgement. For example – consider two classes of objects, apples and oranges. The SVM will try to find the apple that looks the most like an orange, and an orange that looks the most like an apple (cases where the machine is most likely to misjudge) and then use them as support for maximum possible error in judgement. To explain this let's consider an example as shown in Figure [3].

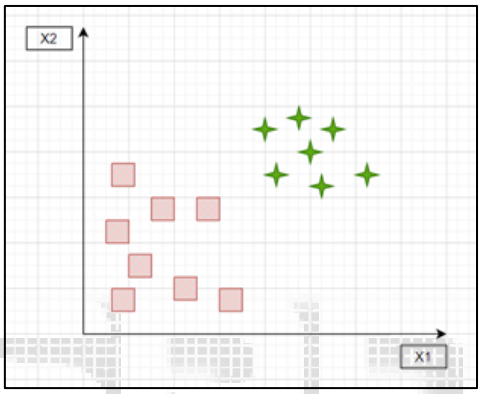

Fig. 5: Sample Graph demonstrating two classes of data

Clearly there are two classes into which every object can be classified. However, it becomes confusing for the machine to draw a line that can separate the two classes as there are multiple possibilities.

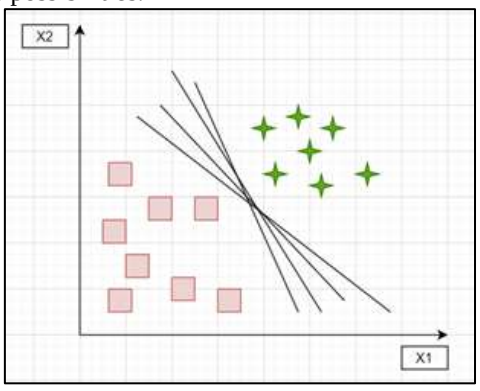

Fig. 6: Possible Regression on data

To handle this issue, the SVM chooses two points of opposite classes that are closest to each other and uses them as maximum permissible error value (epsilon function). The SVM would look something like this –

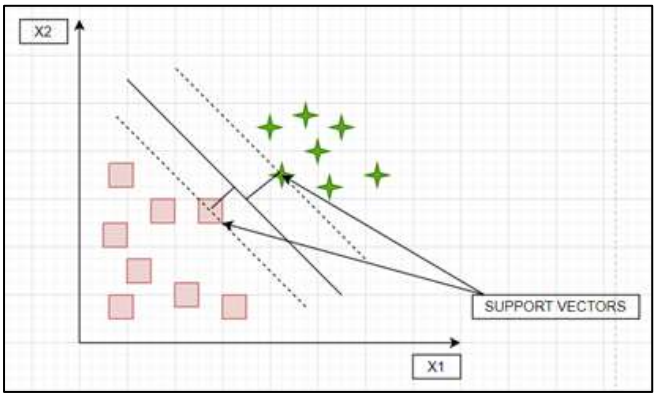

Fig. 7: SVM Classification on dataset

The two points are called supporting points when in reality they are vectors. This is why because in a multi-dimensional space when you have more than just two variables (you can have three five 10 or 100 variables) each point is actually no longer a point because you can't visualize it on a two- dimensional plane or even a three-dimensional space and therefore each of those points that we see here is considered is actually a vector in a multi-dimensional space. Hence, the more general term for points that we see here are vectors. So generally speaking, they are all vectors just in this particular example and we have two dimensions then we can call them points but in reality, there are pictures and Hence the name, support vectors machine.

Classification mechanism of Support Vector Machines can be applicable in three cases:

– linearly separable data
– linearly non- separable data
– non-linear data

a) Benefits:

Highly effective in multi-dimensional spaces.

Can even work in cases where the number of attributes or input spaces is higher than the number of records/samples. For example – text classification.

Highly memory efficient and lower space complexity owing to using support vector point in the expanded space, only during the training period.

b) Drawbacks:

Choosing regularization functions and kernel functions become crucial when the dataset is large, as overfitting can occur leading to very complex SVM models, which can increase the time complexity and inaccurate classifications.

SVMs use expensive five-fold cross validation to calculate the validation accuracy as they do not provide estimations.

1) Linear SVM

Linear kernel function is used when the dataset consists of data that is linearly separable, that is, it can be separated using a single line. It is one of the most common kernels. It is mostly used when there are a large number of features but a relatively low number of records in a particular dataset. One such example is that of text classification, as every single alphabet is a new input feature. Hence, linear kernel-based support vectors machines are most commonly used for text-classification.

An example of a dataset where linear kernel can be used is –

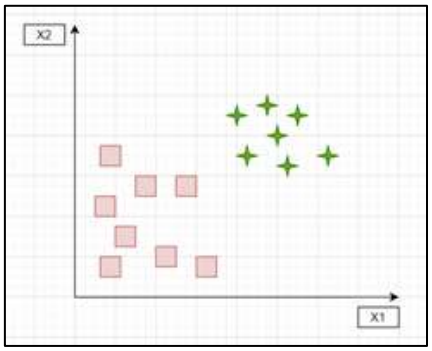

Fig. 8: Sample Graph demonstrating two classes of data

A linearly separable dataset can be applied with a linear kernel function for a support vector machine. The linear kernel function is described as –

$$K(x, x_i) = x \cdot x^T$$

a) Benefits:

– Linear kernel based SVM as easily modelled and easily trained.
– Training of an SWM with a linear function, only the requires the optimization of the C regularization parameter. On other hand, training of SVM of other kernels – polynomial kernel, RBF etc. require the optimization of the y parameter that requires a grid search to be performed, which only ends up increasing the execution time.

2) Radial SVM

The Radial Basis Function (RBF) kernels are the most widely used form of kernelization largely because of its similarity to the gaussian distribution function. The RBF kernel function is used to linearly non-separable and non-linear dataset. It uses two points to compute the similarity between the two classes of dataset.

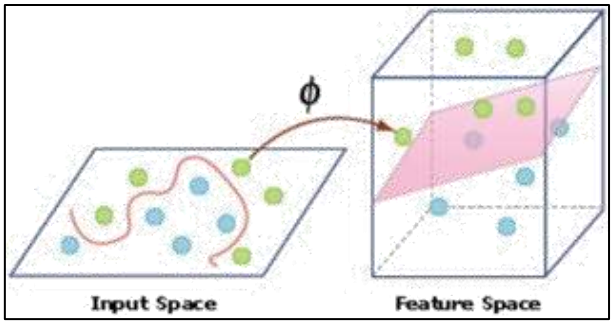

Fig. 9: RBF Kernel in SVM

The radial basis function kernel can be mathematically represented as follows:

$$K(X_1, X_2) = exp(-\frac{||X_1 - X_2||^2}{2\sigma^2})$$

where,

– ‘σ’ is the variance and our hyperparameter
– $||X_1 - X_2||$ is the Euclidean (L_2-norm) Distance between two points X_1 and X_2

The RBF kernel takes two input parameters – gamma and C, both of which are related to two factors -

– the decision region i.e., how the dataset in spread in a multi-dimensional plane
– the penalty for wrongly classifying a data point i.e., the epsilon function.

Unlike the linear kernel the RBF requires a grid search to be done to find the optimal value of C and especially, the gamma parameter which is highly sensitive to minute changes. Grid search area for gamma is taken under logarithmic scales. On the other hand, grid search for the C parameter, requires trying the powers of 10 (0.0001, 0.001,0.01,0.1,1,10,100 etc.)

RBF Kernel has striking similarity to the K-Nearest Neighbours Algorithm. It combines the best of both the algorithms. It takes the advantages of the k-nearest neighbour and combines them with the low space complexity of its own. The low space complexity is due to the fact that the algorithm requires the support vectors in the input space to be stored only during the training epochs and not along the entire dataset, including the validation.

a) Drawbacks:

– If the dataset consists of linearly separable data, then RBF, which is used for non-linear data, can lead to inefficient solution. Linear kernels in such a case can maximize the margins better compared to RBF and give us less denser solutions.
– Owing to the expanded feature space, RBF is inefficient compared to the linear kernel when the training dataset consists of a large number of data records.
– Linear SVM is can generate simpler SVM models compared to kernelized SVMs owing to overfitting in the kernels which leads to the generation of complex SVM models.

b) Applications:

– Face Detection
– Text and Hypertext Categorization
– Classification of Images
– Bioinformatics
– Handwriting Recognition
– Geo and Environmental Sciences

IV. PERFORMANCE

A. *Accuracy Measures*

1) Accuracy (A): It determines the accuracy of the algorithm in predicting instances.
2) Precision (P): Classifier's correctness is measure by Precision.
3) Recall (R): It measures classifier's completeness or sensitivity
4) F-measure: It is the weighted average of precision and recall.

B. *Performance Comparison*

	Precision score	Recall Score	F1 Score	Accuracy Score
SVM rbf kernel	0.510638	0.727273	0.600000	0.792208
SVM linear kernel	0.617021	0.743590	0.674419	0.818182
Logistic Regression	0.617021	0.763158	0.682353	0.824675
Decision Tree	0.765957	0.654545	0.705882	0.805195
K Nearest Neighbours	0.553191	0.702703	0.619048	0.792208

Table 1: Performance With all Attributes

For the second approach, we decided to optimize the model by removing useless additional attributes that were not contributing to the decision making. For this, we were required to know which input features contributed to the decision making and which did not. To do this, we used the random forest algorithm.

The random forest algorithm, after being fitted with the data provides a feature importance attribute that can be utilized to understand the importance of each input feature in the decision-making process of the algorithm. The scores highlight the input features that most heavily affect the target variable and those, which are least relevant to the target variable.

After fitting the model with the data, we used the feature importance_ function to map the scores of all input features and obtained the following result –

```
Glucose                    0.242098
BMI                        0.172574
Age                        0.135220
DiabetesPedigreeFunction   0.128324
BloodPressure              0.092903
Pregnancies                0.086774
SkinThickness              0.073109
Insulin                    0.068999
dtype: float64
```

We then decided to delete input features with feature importance score less than 0.1 and only keep those higher than the same in the dataset, and use that data to train the model. The aim was to ease the modelling process, make it faster and more accurate.
We obtained the following result –

	Precision score	Recall Score	F1 Score	Accuracy Score
Radial SVM	0.548387	0.708333	0.618182	0.781250
Linear SVM	0.548387	0.708333	0.618182	0.781250
Logistic Regression	0.580645	0.692308	0.631579	0.781250
Decision Tree	0.500000	0.563636	0.529915	0.713542
KNN	0.629032	0.650000	0.639344	0.770833

Table 2: Performance with selected attributes

V. CONCLUSION

In this research paper, we focused on one of the most important real-world medical problems which is the detection of diabetes at its early stage. Various machine learning algorithms were used in the prognosis of diabetes and various performance measures were used to evaluate them. Experimental results show that logistic regression was able to achieve the highest accuracy of 82.46 % while the least accuracy of 79.22% was achieved by KNN and Radial SVM. The experimental results after implementing the second approach were quite peculiar. We tried to ease the modelling process and make it faster but it took a toll on the accuracy of the model. While deleting seemingly insignificant input features made the modelling process significantly faster, it was found that the accuracy score of all the classification algorithms was reduced after feature selection. The accuracy score of decision tree was the worst

affected by attribute selection based on feature importance, as the number of attributes for the tree generation were significantly reduced, leading to inaccurate classification of data points. This shows that in order to predict diabetes all features play a significant role in determining the correct outcome. Even the most insignificant detail might play a major role in indicating whether a patient might be diabetes positive or not.

VI. FUTURE WORK

Considering that about 422 million people all over the world suffer from diabetes and with a majority of population living in low and middle-income countries, many people are unaware or simply, cannot afford a test for diabetes. A high blood sugar cannot definitively indicate that a person is suffering from diabetes as there are several factors that go into consideration. While a pathology test can conclusively indicate the result, our experiment aims to provide a reliable yet close proximation of chances whether a patient has diabetes or not.

In our future work, we aim to design an android/iOS native application that can be installed on any phone.

The application would take basic health attribute inputs from the user. These include –

- Pregnancies (for female users) – this is common knowledge data that does not require any test.
- Glucose levels – blood sugar levels can be tested using glucose sticks on devices such as AccuCheck etc. which is easily and economically viable for people of all income household.
- BMI – common knowledge data which can even be checked in any gym or doctor's clinic or can even be calculated at home. No specific test needed.
- Age

These few common knowledge input features can be utilized by the application which will provide input to the machine learning model we have designed and return the result to the user with an approximate percentage of his chances of being diabetes positive. A sequence diagram of the same is shown in Figure 8.

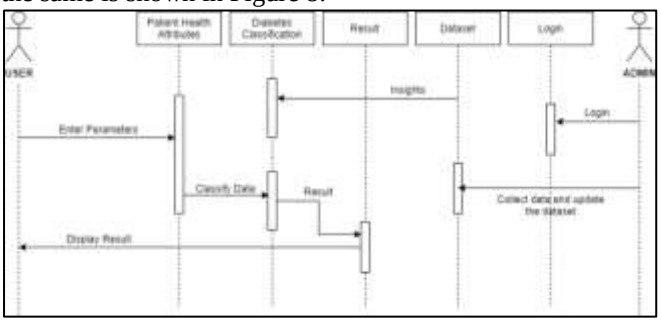

Fig. 10: Sequential Diagram

To avoid a false positive case, the patient however might have to get a medical test done ultimately. However, in our experiment, we have successfully designed a machine learning model that is accurate for 82.5% of the cases. We hope to utilize this for the betterment of the society and help people timely diagnose this deadly disease.

REFERENCES

[1] Aiswarya Iyer, S. Jeyalatha and Ronak Sumbaly," Diagnosis of Diabetes Using Classification Mining Techniques", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.1, January 2015.

[2] K. Rajesh and V. Sangeetha, "Application of Data Mining Methods and Techniques for Diabetes Diagnosis", International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 3, September 2012.

[3] Mani Butwall and Shraddha Kumar," A Data Mining Approach for the Diagnosis of Diabetes Mellitus using Random Forest Classifier", International Journal of Computer Applications, Volume 120 - Number 8,2015.

[4] Sanakal, Ravi, and Smt T. Jayakumari. "Prognosis of Diabetes Using Data mining Approach-Fuzzy C Means Clustering and Support Vector Machine."International Journal of Computer Trends and Technology 11.2 (2014): 94-8.

[5] Namayanja, J., & Janeja, V. P. (2012). An assessment of patient behavior over time periods: A case study of managing type 2 diabetes through blood glucose readings and insulin doses. Journal of Medical Systems

[6] Veena Vijayan; C. Anjali. Prediction and diagnosis of diabetes mellitus - A machine learning approach. 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS).

[7] K. Vijiya Kumar, B. Lavanya, I. Nirmala, S. Sofia Caroline. Random Forest Algorithm for the Prediction of Diabetes. 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)

[8] Sharma, Himani & Kumar, Sunil. (2016). A Survey on Decision Tree Algorithms of Classification in Data Mining. International Journal of Science and Research (IJSR). 5.

[9] G. Daqi and Z. Tao, "Support vector machine classifiers using RBF kernels with clustering-based centers and widths," *2007 International Joint Conference on Neural Networks*, Orlando, FL, 2007, pp. 2971-2976, doi: 10.1109/IJCNN.2007.4371433.

[10] Y. Liu and K. K. Parhi, "Computing RBF Kernel for SVM Classification Using Stochastic Logic," *2016 IEEE International Workshop on Signal Processing Systems (SiPS)*, Dallas, TX, 2016, pp. 327-332, doi: 10.1109/SiPS.2016.64.

[11] A. Patle and D. S. Chouhan, "SVM kernel functions for classification," *2013 International Conference on Advances in Technology and Engineering (ICATE)*, Mumbai, 2013, pp. 1-9, doi: 10.1109/ICAdTE.2013.6524743.

[12] I. S. Al-Mejibli, D. H. Abd, J. K. Alwan and A. J. Rabash, "Performance Evaluation of Kernels in Support Vector Machine," *2018 1st Annual International Conference on Information and Sciences (AiCIS)*, Fallujah, Iraq, 2018, pp. 96-101, doi: 10.1109/AiCIS.2018.00029.

[13] S. Ghosh, A. Dasgupta and A. Swetapadma, "A Study on Support Vector Machine based Linear and Non-

Linear Pattern Classification," *2019 International Conference on Intelligent Sustainable Systems (ICISS)*, Palladam, Tamilnadu, India, 2019, pp. 24-28, doi: 10.1109/ISS1.2019.8908018.

[14] Deepti Sisodia, Dilip Singh Sisodia, "Prediction of Diabetes using Classification Algorithm", Procedia Computer Science,2018

