International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Prediction of Diabetes Using Machine Learning Algorithms

K.V. Haranadh[1], K. Chakri Dhanush[2]

Students of the Department of Electronics and Communication Engineering, R.V.R & J.C College of Engineering

---***---

Abstract - In the medical field, it is essential to predict diseases early to prevent them. Diabetes is one of the most dangerous diseases all over the world. In modern lifestyles, sugar and fat are typically present in our dietary habits, which have increased the risk of diabetes. To predict the disease, it is extremely important to understand its symptoms. Currently, machine-learning (ML) algorithms are valuable for disease detection. This article presents a model using a fused machine learning approach for diabetes prediction. The conceptual framework consists of two types of models: Support Vector Machine (SVM) and Artificial Neural Network (ANN) models. These models analyze the dataset to determine whether a diabetes diagnosis is positive or negative. The dataset used in this research is divided into training data and testing data with a ratio of 70:30 respectively. These three models gave us different accuracies and simple to use. A cloud storage system stores the fused models for future use. Based on the patient's real-time medical record, the fused model predicts whether the patient is diabetic or not. The proposed fused ML model has a prediction accuracy, which is higher than the previously published methods.

Keywords: Logistic Regression, SVM, ANN, ML techniques

1. INTRODUCTION

Diabetes is one of the most dangerous chronic diseases that could lead to others serious complicating diseases. Diabetes diseases are also called as diabetes mellitus, which describes a set of metabolic disease. Diabetes leads to many other kinds of diseases and that are- heart attack, blindness, kidney diseases and so on. Diabetes is also called as Diabetes Mellitus is a chronic disease and is considered as one of the deadliest diseases. Diabetes disease can be categorized as Type 1 or Type 2. If the pancreas does not create adequate amount of insulin in body, is called as Type 1. In Type 2, the body either cannot effectively use the insulin that it produces or an inadequate amount of insulin is released into the bloodstream Type 1 disease generally occurs in children and adolescents, but it can occur in older people. Type 2 diabetes is normally milder compare to people have type 2 diabetes. Type 1 diabetes can be cured by inserting insulin into the fatty tissue under the skin of patient. However, Type 2 diabetes can be cured by having a healthy diet, weight and exercising. Many of diseases can be prevented if diabetes can be diagnosed in the early stages. Early diagnosis and prediction of disease is possible due to recent technological development of IoT, Artificial Intelligence (AI) and Block chain in the current healthcare system. AI presented a paradigm shift in diabetes care from conservative management approaches to construct the targeted data-driven precision care. IoT offers connected environment to the smart healthcare system. ML and deep learning are AI based techniques. ML has a potential of improving efficiency and decrease the cost of treatment in the healthcare system. Various texts are available for diagnosis and prediction of diabetes based on data mining and ML. Data mining and ML methods are equally important to their specific objective. Data mining techniques are useful to extract rules and pattern from the vast amount of diabetes data set, while ML is significant to learn and automate the machine along with pattern recognition. Several ML techniques are used to form digital support in diabetes care. These include support vector machine (SVM), Logistic Regression (LR), neural network, Principal Component Analysis (PCA) based algorithm for better diabetes care. Various texts have been available for automatic diabetes detection, prediction and management via ML and AI.

In this paper, we will review the several ML techniques for diabetes detection and prediction. There are mainly two categories of learning i.e. supervised and unsupervised learning that made foremost impacts in the detection, prediction and treatment of diabetes. This literature survey firstly focused on key words associated to the supervised and secondly on unsupervised ML techniques mainly from 2018 to 2020. The remainder of the paper is arranged as follows. Section 2 and section 3 represent supervised and unsupervised ML techniques respectively to the analysis, diagnosis, classification and prediction of diabetes disease. Section 4 deliberated the findings of the review as a part of result and discussion. Lastly, section 5 concludes the paper. Supervised learning algorithms take direct feedback for the prediction. Supervised learning can be categorized in classification and regression methods. Are some popular algorithms of supervised learning? The basic objective of classification techniques is to detect and predict of the possibility of diabetes in patients with maximum accuracy. National Institute of Diabetes and Digestive Kidney Disease dataset and many techniques like Data transformation, Association rule mining is also used in. In this study, clustering techniques are used to predict diabetes with maximum accuracy. Artificial neural network outperformed with highest accuracy the classification algorithms were used to predict diabetes with maximum accuracy by applying various ML algorithms for instance SVM, NB Classifier, and DT. Experiments were implemented on PIDD (Pima Indian Diabetes Data Set) database and Naïve Bayes has gain highest accuracy i.e. proposed ensemble method that accuracy on PIDD dataset. The K-means clustering algorithm first employed to discover and delete outliers in diabetes data set and then classification algorithm SVM was applied. K-means algorithm was used to categorize the patients into Healthy and Diabetic clusters. In this work, healthcare dataset of pregnant women from healthcare is taken to build a predictive diabetics model. K-Means algorithm is found to be Further, due to having large dimension of diabetes dataset, it is significant to identify principal components or attributes that are participating in the detection and prediction of diabetes. Limited text is available that applied unsupervised learning for the prediction.

2.RELATED WORK

For this project, we have referenced numerous research papers and have gleaned valuable insights from them.

In [1] ‘‘An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier”

In [2] “Prediction of diabetes using machine learning algorithms in healthcare”

In [3] “Computer Vision and Machine Intelligence in Medical Image Analysis”

In [4] “An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier”

3. PROPOSED METHODOLOGY

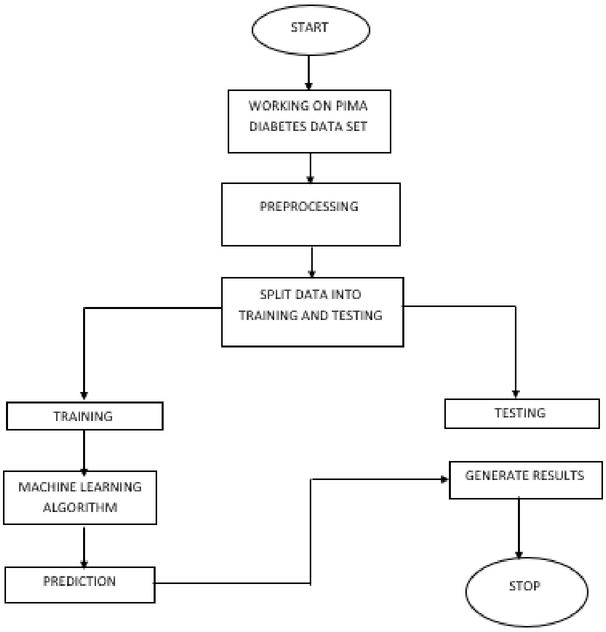

START
WORKING ON PIMA DIABETES DATA SET
PREPROCESSING
SPLIT DATA INTO TRAINING AND TESTING
TRAINING
TESTING
MACHINE LEARNING ALGORITHM
GENERATE RESULTS
PREDICTION
STOP

Fig1.Block Diagram

Proposed several machine learning models to classify attack are not, but none have adequately addressed this misdiagnosis problem. That can be used for this purpose are Stevens Multi Parameter Prediction of Diabetes Empowered with Fused Machine Learning. Also, similar studies that have proposed models for evaluation of such tumors mostly do not consider the heterogeneity and the size of the data Therefore, we propose a machine learning-based approach which combines a new technique of pre-processing the data for features transformation, SVM, Logistic regression ML algorithm give the best accuracy techniques to eliminate the bias and the deviation of instability and performing classifier tests based. Advantages of this model are having highest accuracy and reduces time complexity.

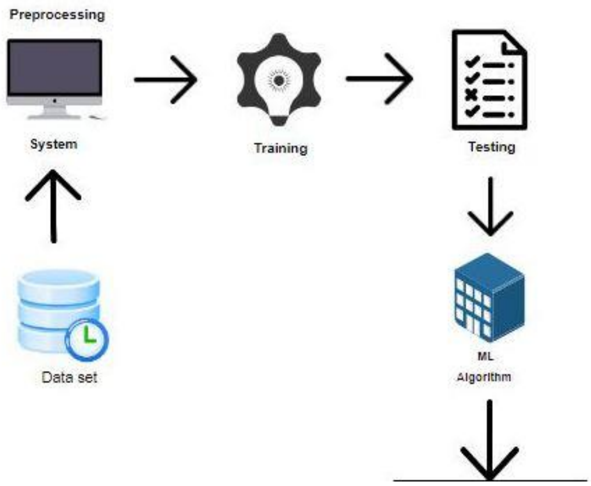


Fig2.Architecture

Our methodology consists of mainly three stages as shown in the above figure

1.Data acquisition and Pre-processing

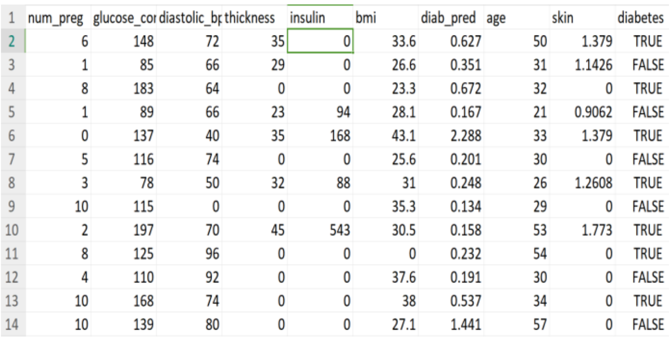
2.Training and Testing

3.Prediction of Diabetes

3.1 Data acquisition and Pre-processing

The dataset is PIMA Indians Diabetes (PID) dataset of 768 female diabetic patients from the Pima Indian population near Phoenix, Arizona.

This dataset consists of 268 diabetic patients (positive) and 500 non-diabetic patients (negative) with eight different attributes.



1	num_preg	glucose_cor	diastolic_bp	thickness	insulin	bmi	diab_pred	age	skin	diabetes
2	6	148	72	35	0	33.6	0.627	50	1.379	TRUE
3	1	85	66	29	0	26.6	0.351	31	1.1426	FALSE
4	8	183	64	0	0	23.3	0.672	32	0	TRUE
5	1	89	66	23	94	28.1	0.167	21	0.9062	FALSE
6	0	137	40	35	168	43.1	2.288	33	1.379	TRUE
7	5	116	74	0	0	25.6	0.201	30	0	FALSE
8	3	78	50	32	88	31	0.248	26	1.2608	TRUE
9	10	115	0	0	0	35.3	0.134	29	0	FALSE
10	2	197	70	45	543	30.5	0.158	53	1.773	TRUE
11	8	125	96	0	0	0	0.232	54	0	TRUE
12	4	110	92	0	0	37.6	0.191	30	0	FALSE
13	10	168	74	0	0	38	0.537	34	0	TRUE
14	10	139	80	0	0	27.1	1.441	57	0	FALSE

Fig3.PIMA dataset

Here Eight attributes are present to determine the target variable i.e. diabetes.

In this study, the PIMA Indians Diabetes dataset, which is taken from the Kaggle data repository and is frequently preferred for diabetes prediction, is used. The access link is https://data.world/data-society/pima-indians-diabetes-database . The dataset’s goal is to diagnose whether or not a patient has diabetes based on certain diagnostic metrics provided in the collection. All patients here, in particular, are PIMA Indian women over the age of 21.

The dataset includes the following measurements and ranges of clinical and physical characteristics. Pregnancies (number, [0–17]), glucose (value, [0–199]), blood pressure (mm Hg, [0–122]), skin thickness (mm, [0–99]), insulin (mu U/mL, [0–846]), BMI (kg/m2, [0–67.1]), diabetes pedigree function (PDF) (value, [0.078–2.42]), age (years, [21–81]), and outcome (Boolean- 0, 1). The data are entirely numerical and comprise a total of 8 features and 768 samples. Fig 3 shows a few samples from the dataset.

Then the data is pre-processed using python language. Preprocessing means removing null values and outliners of the collected data is liable to be influenced as reckless. Besides this, the data quality is important as it affects the prediction results and accuracy to a large extent. Therefore, Datasets need to be properly balanced and divided between testing and training data at a certain ratio, so that sampling can be done efficiently for better prediction outcomes. Sampling is a process of selecting a representative portion of data for extracting characteristics and parameters from large datasets consistently; therefore, it can contribute in a better manner concerning the training model of machine. For maintaining that consistency, we need to apply some sampling techniques (linear sampling, shuffled sampling, stratified sampling, and automatic sampling) on the dataset, that sampling techniques randomly splits the dataset into subsets and evaluates the prediction model. Those diverse sampling techniques perform dissimilar permutation and combination of a representative set of information from the collected data.

Finally, the dataset is ready to train and test using different machine learning algorithms.

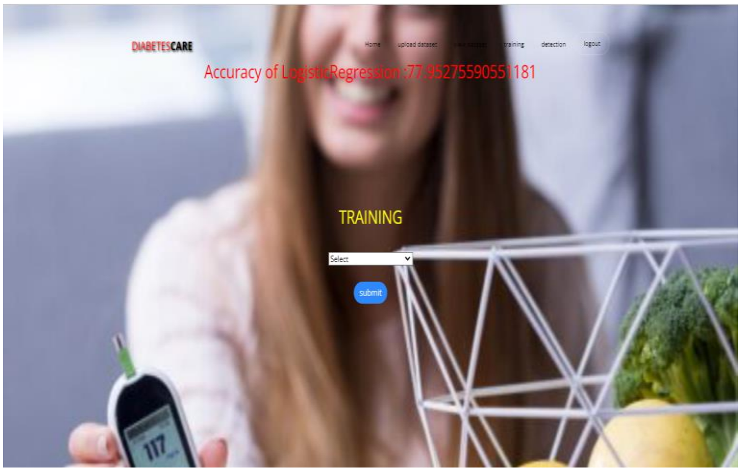
3.2 Training and Testing

The second phase of the proposed framework is reflected by the testing layer. The testing layer acquires dataset from medical database, and loads preprocessed training model from the cloud. A fused model is used to predict whether a diabetes diagnosis is positive or negative. Prediction accuracy is calculated by comparing the required output with the actual output.

The ANN model is trained with the preprocessed training dataset. We have divided the preprocessed data into training and test data with 70:30 ratio on the basis of class base split. For training the data we have used Bayesian regularization function with 5% is used for testing and 5% for validation, and the remaining 90% is used for training.

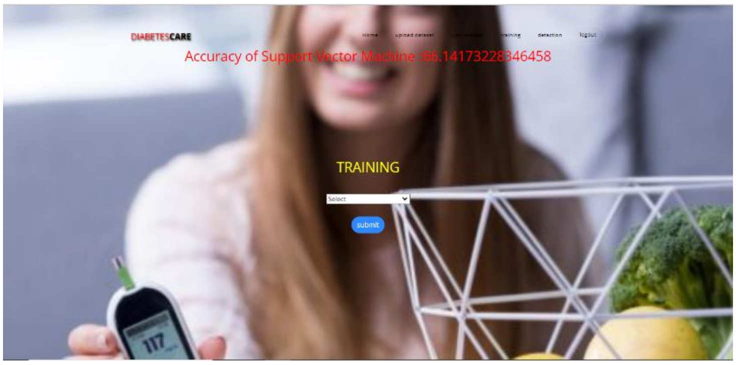
The data is trained using SVM, LOGISTIC REGRESSION and ANN algorithms.

We Acquired different accuracies for different algorithms as shown in fig 4,5 and 6.



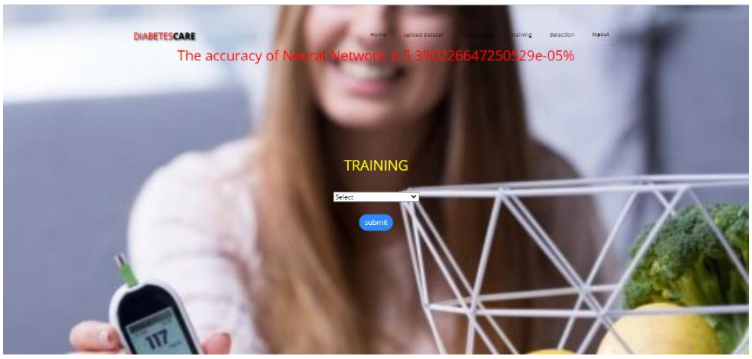
Accuracy of LogisticRegression :77.95275590551181

Fig4.Accuracy of Logistic Regression



Accuracy of Support Vector Machine :66.14173228346458

Fig5.Accuracy of SVM

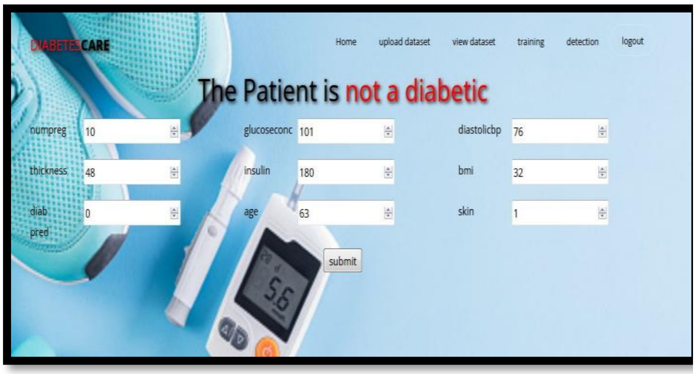


The accuracy of Neural Network is 5.390226647250529e-05%

Fig6.Accuracy of ANN

3.3 Prediction

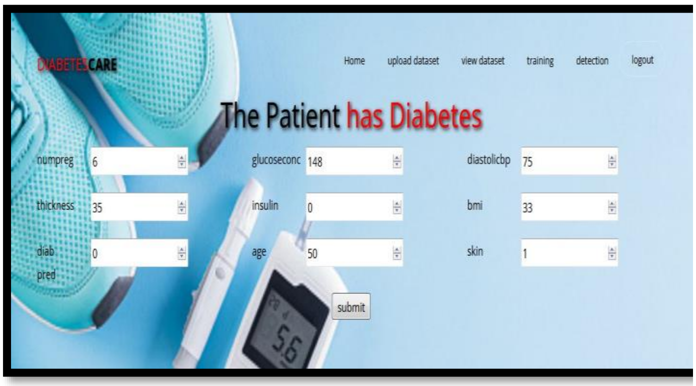
Then choose the algorithm to train which gives different accuracies. By entering the attribute values such as number of pregnancies, skin thickness, insulin and other attribute values as shown in fig 7,8. we get output result whether positive or negative.



The Patient is not a diabetic

numpreg 10
glucoseconc 101
diastolicbp 76
thickness 48
insulin 180
bmi 32
diab pred 0
age 63
skin 1

Fig7.Predicted as not diabetic



The Patient has Diabetes

numpreg 6
glucoseconc 148
diastolicbp 75
thickness 35
insulin 0
bmi 33
diab pred 0
age 50
skin 1

Fig8. Predicted as diabetic

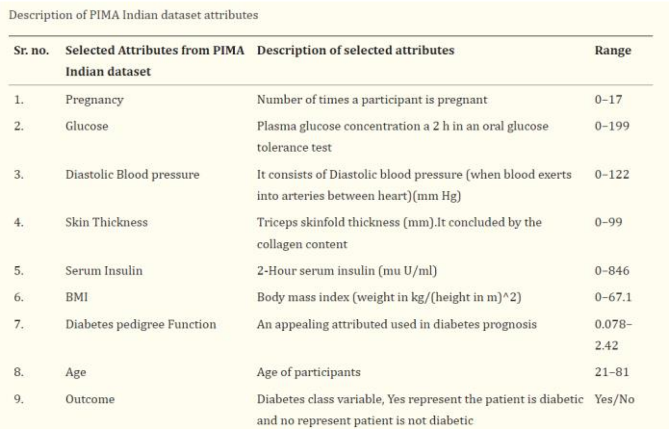
4. DATASET AND ALGORITHMS

In this section we discuss about all the datasets and ANN

4.1. Dataset

The dataset includes the following measurements and ranges of clinical and physical characteristics. Pregnancies (number, [0–17]), glucose (value, [0–199]), blood pressure (mm Hg, [0–122]), skin thickness (mm, [0–99]), insulin (mu U/mL, [0–846]), BMI (kg/m2, [0–67.1]), diabetes pedigree function (PDF) (value, [0.078–2.42]), age (years, [21–81]), and outcome (Boolean- 0, 1). The data are entirely numerical and comprise a total of 8 features and 768 samples. Fig 3 shows a few samples from the dataset.

Then the data is pre-processed using python language. Preprocessing means removing null values and outliers of the collected data is liable to be influenced as reckless. Besides this, the data quality is important as it affects the prediction results and accuracy to a large extent. Therefore, Datasets need to be properly balanced and divided between testing and training data at a certain ratio, so that sampling can be done efficiently for better prediction outcomes.



Description of PIMA Indian dataset attributes

Sr. no.	Selected Attributes from PIMA Indian dataset	Description of selected attributes	Range
1.	Pregnancy	Number of times a participant is pregnant	0–17
2.	Glucose	Plasma glucose concentration a 2 h in an oral glucose tolerance test	0–199
3.	Diastolic Blood pressure	It consists of Diastolic blood pressure (when blood exerts into arteries between heart)(mm Hg)	0–122
4.	Skin Thickness	Triceps skinfold thickness (mm).It concluded by the collagen content	0–99
5.	Serum Insulin	2-Hour serum insulin (mu U/ml)	0–846
6.	BMI	Body mass index (weight in kg/(height in m)^2)	0–67.1
7.	Diabetes pedigree Function	An appealing attributed used in diabetes prognosis	0.078–2.42
8.	Age	Age of participants	21–81
9.	Outcome	Diabetes class variable, Yes represent the patient is diabetic and no represent patient is not diabetic	Yes/No

Fig9.Description of PIMA dataset attributes

4.2. Algorithms

4.2.1. Logistic Regression:

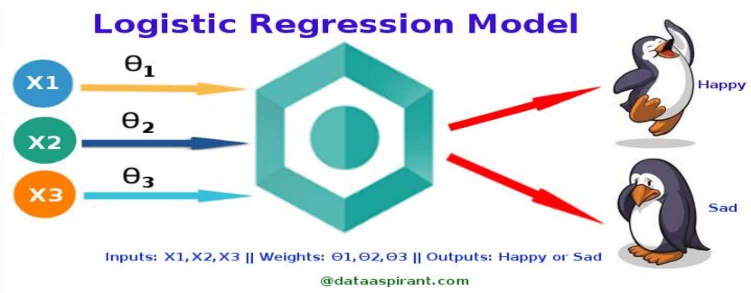

Logistic Regression Model
X1
X2
X3
ϴ1
ϴ2
ϴ3
Happy
Sad
Inputs: X1,X2,X3 || Weights: ϴ1,ϴ2,ϴ3 || Outputs: Happy or Sad
@dataaspirant.com

Fig10.Logistic Regression Model

Logistic Regression was used in the biological sciences in early twentieth century. It was then used in many social science applications. Logistic Regression is used when the dependent variable(target) is categorical. If we use linear regression for this problem, there is a need for setting up a threshold based on which classification can be done. Say if the actual class is malignant, predicted continuous value 0.4 and the threshold value is 0.5, the data point will be classified as not malignant which can lead to serious consequence in real time.

From this example, it can be inferred that linear regression is not suitable for classification problem. Linear regression is unbounded, and this brings logistic regression into picture. Their value strictly ranges from 0 to 1.

Where to use logistic regression

Logistic regression is used to solve classification problems, and the most common use case is binary logistic regression, where the outcome is binary (yes or no). In the real world, you can see logistic regression applied across multiple areas and fields.

The three types of logistic regression

1. **Binary logistic regression** - When we have two possible outcomes, like our original example of whether a person is likely to be infected with diabetes or not.
2. **Multinomial logistic regression** - When we have multiple outcomes, say if we build out our original example to predict whether someone may have the flu, an allergy, a cold
3. **Ordinal logistic regression** - When the outcome is ordered, like if we build out our original example to also help determine the severity of a diabetes, sorting it into mild, moderate, and severe cases.

4.2.2 Support Vector Machine:

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:

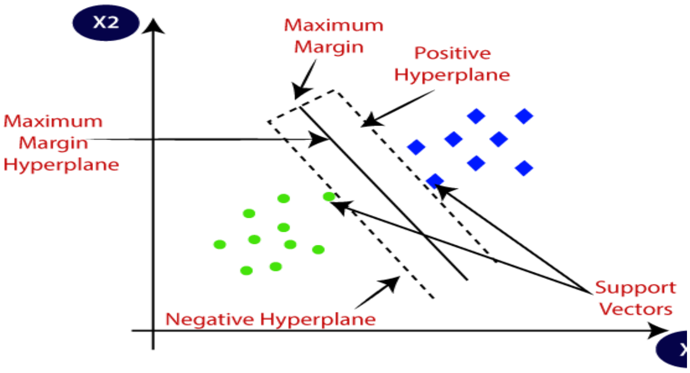


Fig11.Support vector machine

SVM can be of two types:

- **Linear SVM:** Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
- **Non-linear SVM:** Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

Hyperplane and Support Vectors in the SVM algorithm:

Hyperplane:

There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyperplane of SVM.

The dimensions of the hyperplane depend on the features present in the dataset, which means if there are 2 features (as shown in image), then hyperplane will be a straight line. And if there are 3 features, then hyperplane will be a 2-dimension plane.

We always create a hyperplane that has a maximum margin, which means the maximum distance between the data points.Now we used two hyperplanes,positive as upper hyperplane and negative as lower hyper plane . if data point is nearer to upper hyperplane then prediction will be positive and if data point is nearer to lower hyper plane then prediction is negative.

4.2.3.Neural Network

An artificial neural network (ANN) is the piece of a computing system designed to simulate the way the human brain analyzes and processes information. It is the foundation of artificial intelligence (AI) and solves problems that would prove impossible or difficult by human or statistical standards. ANNs have self-learning capabilities that enable them to produce better results as more data becomes available.

An ANN has hundreds or thousands of artificial neurons called processing units, which are interconnected by nodes. These processing units are made up of input and output units. The input units receive various forms and structures of information based on an internal weighting system, and the neural network attempts to learn about the information presented to produce one output report. Just like humans need rules and guidelines to come up with a result or output, ANNs also use a set of learning rules called backpropagation, an abbreviation for backward propagation of error, to perfect their output results.

An ANN initially goes through a training phase where it learns to recognize patterns in data, whether visually, aurally, or textually. During this supervised phase, the network compares its actual output produced with what it was meant to produce—the desired output. The difference between both outcomes is adjusted using backpropagation. This means that the network works backward, going from the output unit to the input units to adjust the weight of its connections between the units until the difference between the actual and desired outcome produces the lowest possible error.

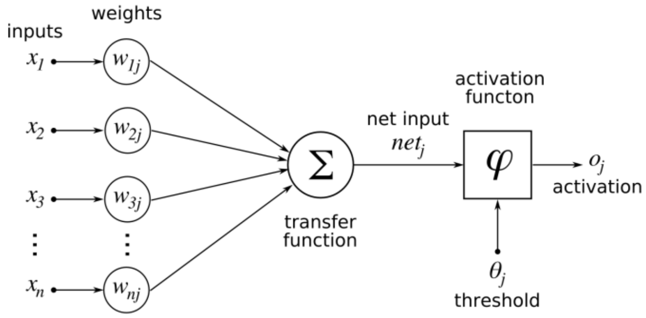
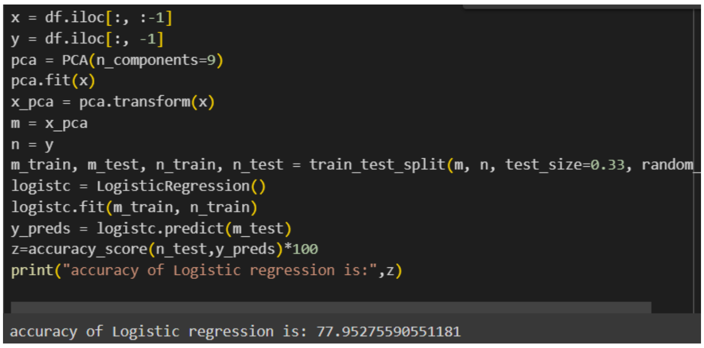


Fig12.Artificial Neural Network

By using ANN,input data is given and weights for individual attributes also given and with these a transfer function is obtained. Threshold values will be changed and based on the predictions weights will be changed.Finally,the model will be ready to predict using transfer function.This tranfer function predicts whether person has diabetes or not.

5. RESULTS AND DISCUSSIONS

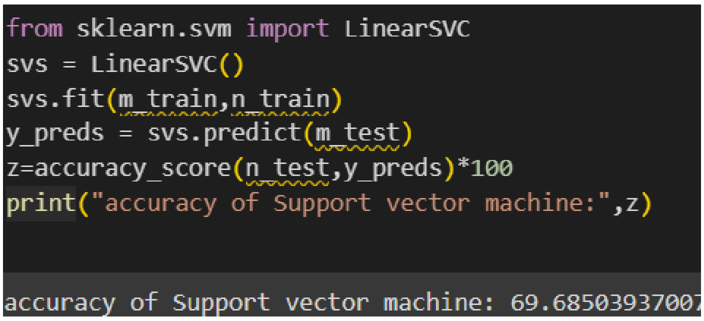
After testing our dataset with three different models we got three different accuracies as shown in the figure 13,14 and 15. Code run on google colab by importing sklearn. Three models are trained and tested; their accuracy is given below.



```
x = df.iloc[:, :-1]
y = df.iloc[:, -1]
pca = PCA(n_components=9)
pca.fit(x)
x_pca = pca.transform(x)
m = x_pca
n = y
m_train, m_test, n_train, n_test = train_test_split(m, n, test_size=0.33, random_
logistc = LogisticRegression()
logistc.fit(m_train, n_train)
y_preds = logistc.predict(m_test)
z=accuracy_score(n_test,y_preds)*100
print("accuracy of Logistic regression is:",z)

accuracy of Logistic regression is: 77.95275590551181
```

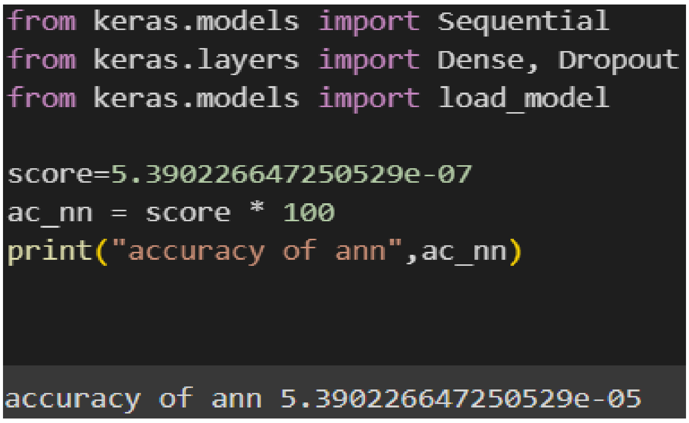
Fig13.Accuracy of Logistic regression



```
from sklearn.svm import LinearSVC
svs = LinearSVC()
svs.fit(m_train,n_train)
y_preds = svs.predict(m_test)
z=accuracy_score(n_test,y_preds)*100
print("accuracy of Support vector machine:",z)

accuracy of Support vector machine: 69.6850393700
```

Fig14.Accuracy of Support Vector Machine



```
from keras.models import Sequential
from keras.layers import Dense, Dropout
from keras.models import load_model

score=5.390226647250529e-07
ac_nn = score * 100
print("accuracy of ann",ac_nn)

accuracy of ann 5.390226647250529e-05
```

Fig15.Accuracy of ANN

For model 1, model 2 and model 3 we got an accuracy of 77.95% ,69.68% and 53.9% respectively. The difference between them is that we changed the model and with this dataset logistic regression gave us better accuracy than other two models. By taking more samples, we will definitely get very good accuracy.

6.CONCLUSION

Though different models had been used for the prediction of diabetes, the accuracy of the proposed models in disease prediction has always been the main concern of researchers. Therefore, a new model is required in order to achieve higher prediction accuracy in diabetes prediction. Prediction of Diabetes Empowered with Fused Machine Learning proposed a machine learning based diabetes decision support system by using decision level fusion. Two widely used machine learning techniques are integrated in the proposed model and ANN also used. The proposed decision system has achieved the accuracy of 77.14, which is higher than the other existing systems. Through this diagnosis model, we can save several lives. Moreover, the death ratio of diabetes can be controlled if the disease is diagnosed and preventative measures are taken in early-stage.

REFERENCES

1. F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra, Computer Vision and Machine Intelligence in Medical Image Analysis. London, U.K.: Springer, 2019.
2. World Health Organization (WHO). (2020). WHO Reveals Leading Causes of Death and Disability Worldwide: 2000–2019. Accessed: Oct. 22, 2021. [Online].Available: https://www.who.int/news/item/09-12-2020-who reveals-leading-causes-of-death-and-disability-worldwide-2000-2019
3. A. Frank and A. Asuncion. (2010). UCI Machine Learning Repository. Accessed: Oct. 22, 2021. [Online]. Available: http://archive.ics.uci.edu/ml
4. G. Pradhan, R. Pradhan, and B. Khandelwal, ‘‘A study on various machine learning algorithms used

for prediction of diabetes mellitus,'' in Soft Computing Techniques and Applications (Advances in Intelligent Systems and Computing), vol. 1248. London, U.K.: Springer, 2021, pp. 553–561, doi: 10.1007/978-981-15-7394-1_50.

5. S. Kumari, D. Kumar, and M. Mittal, ''An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier,'' Int. J. Cogn. Comput. Eng., vol. 2, pp. 40–46, Jun. 2021, doi: 10.1016/j.ijcce.2021.01.001.
6. M. A. Sarwar, N. Kamal, W. Hamid, and M. A. Shah, ''Prediction of diabetes using machine learning algorithms in healthcare,'' in Proc. 24th Int. Conf. Autom. Comput. (ICAC), Sep. 2018, pp. 6–7, doi: 10.23919/IConAC.2018.8748992.
7. S. K. Dey, A. Hossain, and M. M. Rahman, ''Implementation of a web application to predict diabetes disease: An approach using machine learning algorithm,'' in Proc. 21st Int. Conf. Comput. Inf. Technol. (ICCIT), Dec. 2018, pp. 21–23, doi: 10.1109/ICCITECHN.2018.8631968.
8. A. Mir and S. N. Dhage, ''Diabetes disease prediction using machine learning on big data of healthcare,'' in Proc. 4th Int. Conf. Comput. Commun. Control Autom. (ICCUBEA), Aug. 2018, pp. 1–6, doi: 10.1109/ICCUBEA.2018.8697439.
9. S. Saru and S. Subashree. Analysis and Prediction of Diabetes Using Machine Learning. Accessed: Oct. 22, 2022. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3368308
10. P. Sonar and K. JayaMalini, ''Diabetes prediction using different machine learning approaches,'' in Proc. 3rd Int. Conf. Comput. Methodologies Commun. (ICCMC), Mar. 2019, pp. 367–371, doi: 10.1109/ICCMC.2019.8819841.
11. S. Wei, X. Zhao, and C. Miao, ''A comprehensive exploration to the machine learning techniques for diabetes identification,'' in Proc. IEEE 4th World Forum Internet Things (WF-IoT), Feb. 2018, pp. 291–295, doi: 10.1109/WF-IoT.2018.8355130.
12. M. F. Faruque and I. H. Sarker, ''Performance analysis of machine learning techniques to predict diabetes mellitus,'' in Proc. Int. Conf. Electr., Comput. Commun. Eng. (ECCE), Feb. 2019, pp. 7–9, doi: 10.1109/ECACE.2019.8679365.
13. B. Jain, N. Ranawat, P. Chittora, P. Chakrabarti, and S. Poddar, ''A machine learning perspective: To analyze diabetes,'' Mater. Today: Proc., pp. 1–5, Feb. 2021, doi: 10.1016/J.MATPR.2020.12.445.
14. N. B. Padmavathi, ''Comparative study of kernel SVM and ANN classifiers for brain neoplasm classification,'' in Proc. Int. Conf. Intell. Comput., Instrum. Control Technol. (ICICICT), Jul. 2017, pp. 469–473, doi: 10.1109/ICICICT1.2017.8342608.
15. J. Liu, J. Feng, and X. Gao, ''Fault diagnosis of rod pumping wells based on support vector machine optimized by improved chicken swarm optimization,'' IEEE Access, vol. 7, pp. 171598–171608, 2019, doi: 10.1109/ACCESS.2019.2956221.