

Diabetes Prediction using Machine Learning Technique

Parag Nar
K.J. Somaiya College of Engineering
Vidyavihar, Mumbai- 400077

Bhakti Palkar, PhD
K.J. Somaiya College of Engineering,
Vidyavihar, Mumbai-400077

ABSTRACT

Diabetes Mellitus or diabetes is the disease which is also called as the silent killer of human body which is caused when human body's efficiency of producing or supplying or responding to the hormone insulin is impaired. The target of this research was to design an efficient predictive model which will have a very high selectivity and sensitivity to better identify patients who might be at danger of getting the disease. Which can help is early diagnosis of the disease and its treatment

This disease is detected depending on the patient's laboratory results. The main aim was to develop an application that will attempt to predict or detect if the person is suffering with diabetes or not. Accuracy of the system was aimed to be above 75% which was achieved and the system showed an accuracy of 80.51%

The final aim of the project was to also understand the concepts of Support Vector Machines and its usefulness into the medical field.

Keywords

Support vector machine, Machine learning, Diabetes prediction

1. INTRODUCTION

Diabetes mellitus is amongst the harmful diseases within the world. It causes various types of other disorders like heart failure, eyes related disorders like cataract or blindness, urinary organ diseases etc. In these cases the patient collect their reports from the diagnostic centre after consultation. Diabetes is a serious disease that is caused due to increasing level of sugar in blood due to body's inability to process the hormone insulin. The diagnosis of the disease can be easily done by the medical expert but if left untreated it can cause serious threat or big consequence to life. By using efficient machine learning technique will help to spot the trend and predict the whether the patient is diabetic or if the person is diabetic this ML system will help in preventing and treating the disease at its earlier stage itself..

2. LITERATURE SURVEY

2.1 Associated Machine Learning Techniques based on Diabetes based Predictions [1]

In this paper there has been a survey accomplished at the numerous exceptional techniques of diabetes prediction using different Machine learning techniques. There have been many exceptional explores accomplished on diabetes datasets. The used dataset is that the Pima Indian diabetes dataset with 768 records. Feature choice and clustering performed a totally crucial position for this version at the pima diabetes dataset. Accuracy of the subsequent algorithms are mentioned in this paper:

- Naïve Bayes Classifier
- MLP trained by Levenberg-Marquardt algorithm
- Hybrid ANN and FNN
- Ant Colony Optimization with Fuzzy Logic
- LDA-ANFIS system
- Genetic Programming
- General Regression Neural Network (GRNN)
- Multiple methods with bagging & boosting
- Impact of Small world network topology on ANN
- Convolutional Neural Network classification (CNN)
- Least Square Support Vector Machine Generalized Discriminant Analysis

Table 1: Shows the accuracy comparison on the pima Indian diabetes dataset [1]

Ref	Algorithms used	Best Algorithm	Accuracy %
[4]	Probabilistic NN, MLP-LM	MLP-LM	82.37
[5]	LS-SVM, GDA-LS-SVM	GDA-LS-SVM	82.05
[6]	ANN-FNN hybrid	ANN-FNN hybrid	84.24
[7]	RBF, Naïve Bayes, CART, ACO with Fuzzy Logic	ACO with Fuzzy Logic	79.48
[8]	LDA-ANFIS	LDA-ANFIS	84.61
[9]	GP-KNN, GP-SVM	GP-SVM	87
[10]	MLP, GRNN, RBF	GRNN	80.21
[11]	Feed Forward ANN, Small world network on ANN	Small world network on ANN	91.66
[12]	ID3, J48, Fast Rotational Forest, CNN	CNN	0.82 ROC
[13]	SOM-PCA-NN	SOM-PCA-NN	92.28
[14]	NB	Naïve Bayes	66.66

2.2 Diabetes prediction using different machine learning approaches: [2]

Every decision support system is known by how its accuracy is. The aim of this paper was to create a decision network for the prediction and diagnosis of diabetes with a very good precision. This paper was documented on research based of prediction of diabetes using various Artificial intelligence & Machine learning techniques as below

- Decision Tree
- Support vector machine
- Naïve bayes
- Artificial neural network

Table .2 Output Parameters of different algorithm [2]

Technique	Class	Precision	Recall
Decision Tree Classifier	0	0.78	0.71
	1	0.45	0.54
Support Vector Classifier	0	0.70	1.00
	1	0.00	0.00
Gaussian Naïve Bayes	0	0.82	0.79
	1	0.56	0.60
Artificial Neural Network	0	0.70	1.00
	1	0.00	0.00

Table .3 Output Parameters of different algorithm [2]

Technique	Class	F1 Score	Support	Accuracy
Decision Tree Classifier	0	0.74	187	74
	1	0.49	82	49
Support Vector Classifier	0	0.82	187	82
	1	0.00	82	0
Gaussian Naïve Bayes	0	0.80	187	80
	1	0.58	82	58
Artificial Neural Network	0	0.82	187	82
	1	0.00	82	0

2.3 A Decision Support System for Diabetes Prediction using machine learning and deep learning techniques : [4]

Decision Support System for Diabetes Prediction has been implemented in this research paper . Different Machine learning and deep learning algorithms had been studied with their comparison. Classifiers which include the Support

Vector Machine (SVM) and the Random Forest (RF) had been studied in the conventional Machine learning method. And in Deep learning method that they'd used a Fully Convolutional neural network for detection of diabetes. The Pima diabetes dataset of 768 samples every which had eight features. Out of those 768 samples 500 samples had been classified as non-diabetic and 268 had been classified as diabetic sufferers which was used

Table 4: Accuracy of different approaches [4]

Approach	Accuracy
SVM	65.38%
RF	83.67%
DL	76.81%

By the experimental results they found at the the Random Forest Classifier worked well with this dataset compared to the Support vector machine and Deep learning approach. For the diabetes prediction this study has performed proper comparative analysis between the machine learning algorithm like the SVM , RF and the deep learning approach.

3. DESIGN METHODOLOGY

Support Vector Machine (SVM) is that the foremost ordinarily used classification machine learning algorithm for disease prediction. it's widely accustomed predict diabetes, different sorts of cancer ,heart disorders etc. it is a supervised learning technique that's used for locating patterns for classification of data . SVMs were first introduced by Vapnik in 1960s for classification of knowledge. Earlier the classification is performed by drawing hyperplanes. In two class classification, this hyperplane is equidistant from both the classes. the knowledge instances which are accustomed define this hyperplane are stated as support vector. There is a margin that's described in SVM that's that the space among hyperplane and therefore the closest support vector. There need to be suitable separation through this hyperplane, the space of margin need to be as big as viable due to the fact that larger distance offers much less error. If the margin is near then it is greater chances to noise .The different very important gain of the SVM is that it may additionally resolve the trouble of excessive dimensionality with very a good ease, i.e., the problem that arises whilst there may be an outsized quantity of enter variables relative to the amount of reachable observations. This is due to the fact the SVM method is data-pushed and it's far viable with out theoretical framework, it need to do right type even within the instances where in pattern sizes are small. SVM makes use of precise kernel capabilities to assemble a hyper plane it really is used for type. These kernels are the thought for locating the proper hyper-plane among the many possible hyper-planes within the given situation that divides a selected vector. The 4 basic kernel functions are as follows:

- Sigmoid Kernel
- Polynomial Kernel
- Linear Kernel
- Radial Basis Function Kernel

Any of these kernel functions are used for designing the classifier according to the requirements of the matter. These kernels have the following values associates with it the Kernel

Function Type ,values of its parameters and values of kernel function Regularisation Parameters These values calculate meticulously irrespective of the choice of the kernel because they largely influence the results of the classifier and its accuracy. Any Error or miscalculations in these values can have a heavy toll on the output and also the end-results.

The designed system makes use of Machine Learning SVM algorithm to predict the risk of diabetes :

The System is trained on the following medical parameters using the Support Vector Machines Algorithm for classification:

Dataset is in csv file having various parameters like –

1. Plasma Glucose
2. Diastolic BP
3. Triceps Skin
4. Serum Insulin
5. BMI
6. Diabetes Pedigree
7. Age
8. Class (This has values 0 and 1; where 0: No diabetes Detected) and 1: Diabetes Detected)

This dataset is trained with SVM algorithm. The system has simple user interface which is designed using Tkinter which is Python's UI rendering engine

4. IMPLEMENTATION AND RESULTS

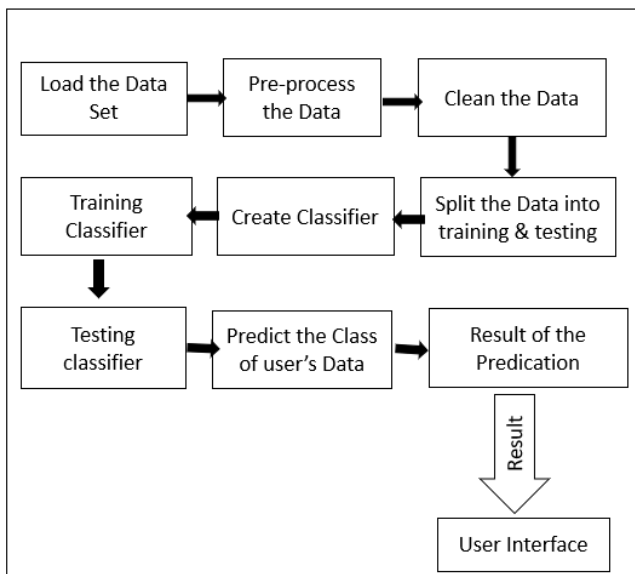


Fig .1 Flow diagram of the proposed system

```

Python 3.8.2 Shell
File Edit Shell Debug Options Window Help
Python 3.8.2 (tags/v3.8.2:7b3ab59, Feb 25 2020, 22:45:29) [MSC v.1916 32 bit (Intel)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
===== RESTART: C:\Users\acer\Desktop\special topic seminar\parag.py =====
*****Diabetes Prediction*****
Accuracy = 80.51948051948052
>>> |
  
```

Fig .2 Accuracy of Designed system

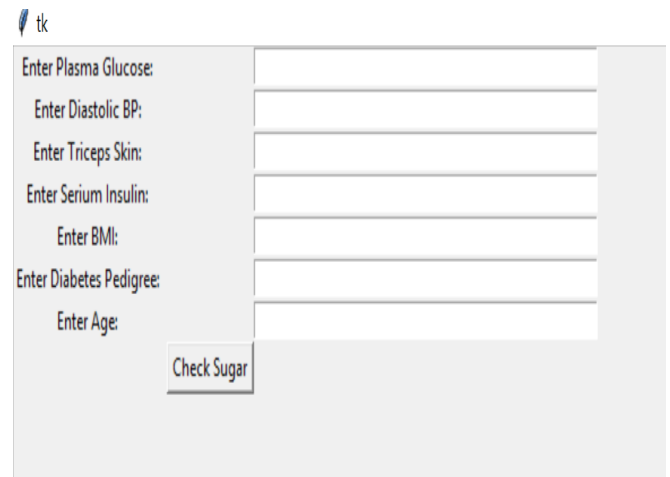


Fig .3 UI of Designed system

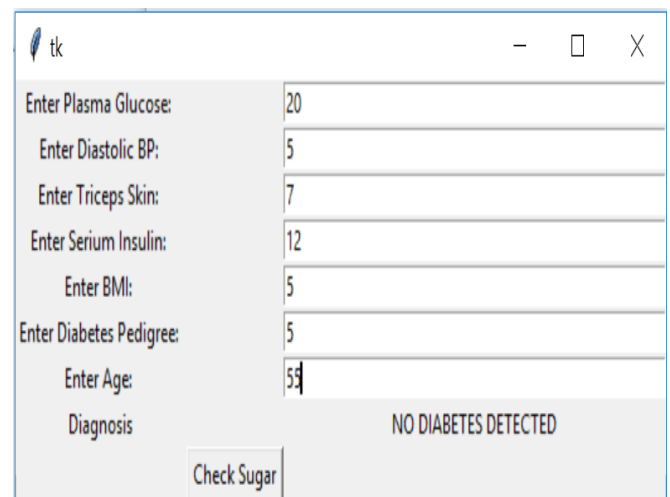


Fig .4 UI of designed system with result

5. CONCLUSION

In this research work various literature review was done about diabetes prediction & detection using various machine learning and deep learning approaches. As per the findings and literature survey from above journal papers it was found out that SVM algorithm was used in common for majority of the papers and it shows different accuracy with different diabetes datasets. In this research work the system was implemented using SVM algorithm on 7 parameter dataset and it was found out that it works perfectly with this dataset and it also gave a good accuracy of about 80.51% .User interface of the system was designed using tkinter package in python . This system could also be integrated in medical

healthcare system for disease detection

6. ACKNOWLEDGEMENT

Authors would like to thank Principal, K. J. Somaiya College of Engineering for providing all the necessary facilities required for this research.

7. REFERENCES

- [1] Shoh Mathew Jacob , Dr. Kumudharaimond , deepakanmani “Associated Machine Learning Techniques based On Diabetes Based Predictions”, IEEE International Conference on Intelligent Computing, IEEE, 2019.
- [2] Priyanka Sonar;K.JayaMalini , “Diabetes Prediction Using Different Machine Learning Approaches”, 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC),IEEE,2019
- [3] AmaniYahyaoui;AkhtarJamil;JawadRasheed;MirsatYesil tepe, "A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques",2019 1st International Informatics and Software Engineering Conference (UBMYK),IEEE,2019.
- [4] Hasan Temurtas , Nejat Yumusak , FeyzullahTemurtas , “A comparative study on diabetes disease diagnosis using neural networks”, Expert Systems with Applications 36, pg 8610–8615, 2009
- [5] Kemal Polat, Salih Gunes, Ahmet Arslan, “A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine”, Expert Systems with Applications 34, 2008
- [6] HumarKahramanli, NovruzAllahverdi, “Design of a hybrid system for the diabetes and heart diseases” , Expert Systems with Applications 35, pg 82–89, 2008
- [7] Mostafa FathiGanji, Mohammad SanieeAbadeh,“Using fuzzy Ant Colony Optimization for Diagnosis of Diabetes Disease”, Proceedings of ICEE , May 11-13, 2010
- [8] EsinDogantekin , Akif Dogantekin , DeryaAvci , LeventAvci , “An intelligent diagnosis system for diabetes on Linear Discriminant Analysis and Adaptive Network Based Fuzzy Inference System: LDA-ANFIS”, Digital Signal Processing 20, pg1248–1255, 2010
- [9] Muhammad Waqar Aslam, Zhechen Zhu, Asoke Kumar Nandi, “Feature generation using genetic programming with comparative partner selection for diabetes classification”, Expert Systems with Applications 40, pg 5402–5412, 2013
- [10] Kamer Kayaer, Tulay Yildirim, “Medical Diagnosis on Pima Indian Diabetes Using General Regression Neural Networks”, Proceedings of the International Conference on Artificial Neural Networks and Neural Information Processing, 2014
- [11] OkanErkaymaz ,Mahmut Ozer, “Impact of small world network topology on the conventional artificial neural network for the diagnosis of diabetes”, Chaos, Solitons and Fractals 83, pg 178–185, 2016
- [12] Tharani.Setal., “Classification using Convolutional Neural Network for Heart and Diabetics Datasets”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 12, December, 2016
- [13] MehrbakhshNilashi, Othman Ibrahim, Mohammad Dalvi, Hossein Ahmadi, Leila Shahmoradi,“Accuracy Improvement for Diabetes Disease Classification: A Case on a Public Medical Dataset”, Fuzzy Inf. Eng., pg 345-357, 2017
- [14] Thulasi K S et al., “Classification of Diabetic Patients Records Using Naïve Bayes Classifier”, 2nd IEEE International Conference on Recent Trends in Electronics Information & Communication Technology (RTEICT), May 19-20, 2017
- [15] P. Prabhu;S.Selvabharathi," Deep Belief Neural Network Model for Prediction of Diabetes Mellitus",2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC),IEEE,2019
- [16] Nongyao Nai-arun, Rungruttikarn Moungrmai,“Comparison of Classifiers for the Risk of Diabetes Prediction”, Procedia Computer Science 69, pg 132 –142, 2015
- [17] Ioannis Kavakiotis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, Ioanna Chouvarda, “Machine Learning and Data Mining Methods in Diabetes Research”, Computational and Structural Biotechnology Journal 15, pg 104–116, 2017.
- [18] Deepti Sisodiaa, Dilip Singh Sisodiab, “Prediction of Diabetes using Classification Algorithms, International Conference on Computational Intelligence and Data Science”, Procedia Computer Science 132, pg 1578–1585, 2018.