IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DIABETES PREDICTION USING MACHINE LEARNING

Dr.Senthil Kumar M
Associate professor, Department of Computer Science and Engineering,
SRM Valliammai Engineering College, Kattankulathur, Chengalpattu district, Tamil Nadu, India.

Sundar Naveen Kumar C
Department of Computer Science and Engineering,
SRM Valliammai Engineering College, Kattankulathur,Chengalpattu district, Tamil Nadu, India.

Sivanarayanan M
Department of Computer Science and Engineering,
SRM Valliammai Engineering College, Kattankulathur,Chengalpattu district, Tamil Nadu, India.

Vijaya Kumar S
Department of Computer Science and Engineering,
SRM Valliammai Engineering College, Kattankulathur,Chengalpattu district, Tamil Nadu, India.

ABSTRACT - Diabetes is a disease caused by a high level of glucose in the human body. Diabetes should not be ignored if it is not treated, then Diabetes can cause serious problems for a person such as: heart problems, kidney problems, high blood pressure, eye damage and can affect other parts of the human body. Curing diabetes can be easier if it is predicted earlier. In order to achieve this project goal we will be able to predict Diabetes in the human body or patient with the highest accuracy by applying, Various Machine Learning Strategies. Machine learning is used to predict accurate values with the available set of patient informations so the accurate values of the diabetes based on the individual can be predicted accurately. In this activity we will use the Learning Machine Planning and compile data to predict diabetes. Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Neighboring K-Nearest (KNN), Gradient Boosting (GB) and Random Forest (RF). Accuracy is different for all models compared to other models. Project Work gives a more exact or precise model appearance that the model can precisely foresee diabetes. Our outcome shows that Random Forest has accomplished higher exactness contrasted with other AI systems.

Keywords — Machine learning; vector support machine; predicting diabetes.

I. INTRODUCTION

Diabetes is becoming a very common disease in India. According to WHO there were almost 31.7 million diabetic patients in India in the year 2000 and it may go to 79.4 million by 2030. Fig 1 contains data about diabetic patients according to WHO. There is a need to control this sickness in India. Machine learning algorithm are the mathematical techniques, which are very useful in analysing large amount of data and suggesting some actions on the basis of that data. These algorithms are useful in analysing a data set and predict the values for a new entry. Many researchers are applying ML algorithms for prediction and control of various diseases. The results of machine learning algorithms are found very well in the prediction of different diseases. There is a need to apply machine learning algorithm to explore their power for diabetes prediction so that necessary precautions can be taken to avoid diabetes.

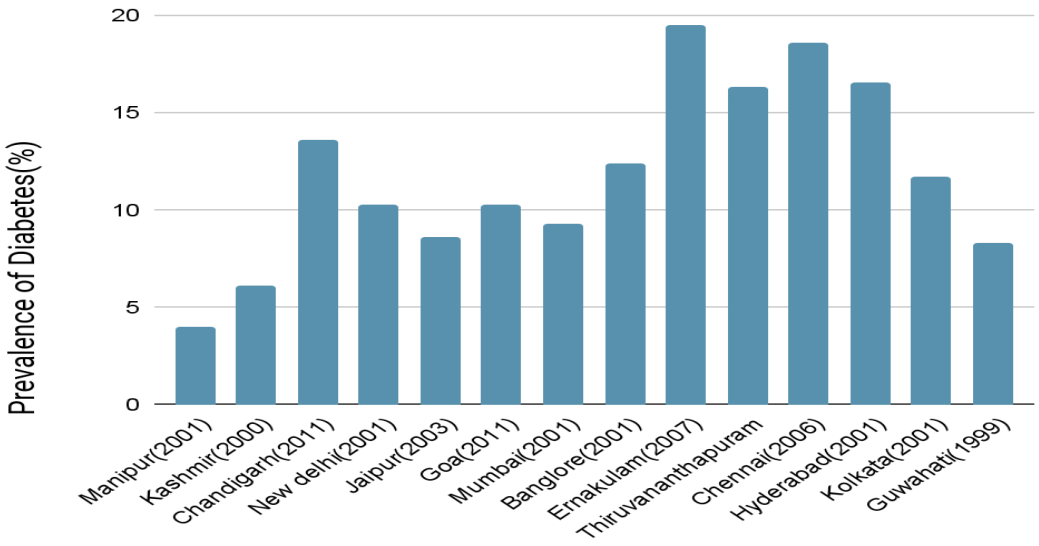


Fig 1. WHO reports on diabetes

II. RELATED WORKS

S. Chatterjee @al has demonstrated an analysis of unsupervised techniques to predict liver diseases in which they have clearly explained the possibilities of a person getting affected by the liver disease [5]. In addition to this, they have also monitored the increase and decrease in liver diseases. The authors J. Neelaveni and M. S. G. Devasana have discussed Alzheimer's disease prediction using Machine Learning algorithms in which they have elaborated in detail about various algorithms that are used in Machine Learning to predict the diseases that occur [1]. The authors Pouriyeh S. Vahid and G. Sannino have made a comprehensive study on the investigation and also made a comparison between various Machine Learning algorithm techniques that are used in predicting the heart diseases that occur [12].

M.Patil and Mishra are the authors who have proposed a model using Support Vector Machine for Lifestyle Disease Prediction which enhances the level of disease prediction using Machine Learning [14]. The authors S.Sampath and R.S.Raj have made a study on a comparison between support vector machines and Naïve Bayes Classifiers for predicting Diabetes using various techniques and algorithms [17]. Dr. M. Senthil Kumar has proposed a paper on Hybrid Fusion of Multimodal Medical Images for the Enhancement of Visual Quality in Medical Diagnosis[18].

III. MATERIALS AND METHODS

This paper has applied SVM machine learning algorithm and several other machine learning algorithms for diabetes prediction. The support vector algorithm is been implemented in python programming language and tested on a data set which is been acquired from kaggle . The whole model is created by using python programming language and for the interlinking of documents here Jupyter Notebook is mainly used. The dataset is widely divided into two parts the training part and the testing part. Then, the SVM model is trained accordingly with respect to the data. The outcome of this work is that the performance of SVM kernels on medical data set for disease prediction is proposed. The paper compares the performance of various machine learning algorithms The model is trained on different alogorithms available for Machine learning and its prediction accuracies are calculated by testing set. . The best Algorithm is chosen and used for diabetes prediction. Figure 2 shows the chart of the model which is proposed. The model proposed is implemented in Python programming language and tested with a data set of 768 patients. The data set is freely available on Kaggle with the name Pima Indians Diabetes Database for research. The data set is available in the form of a CSV comma separated values file which is best suited for python programming language. The prediction accuracy of the SVM ML algorithm depends on the selection of its model.

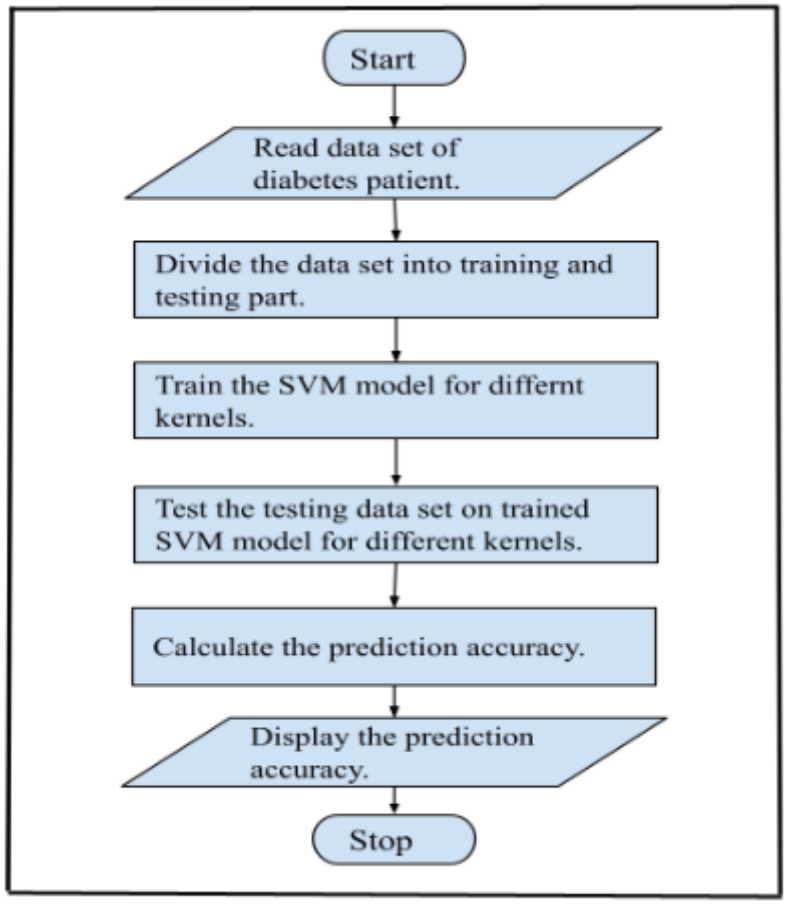
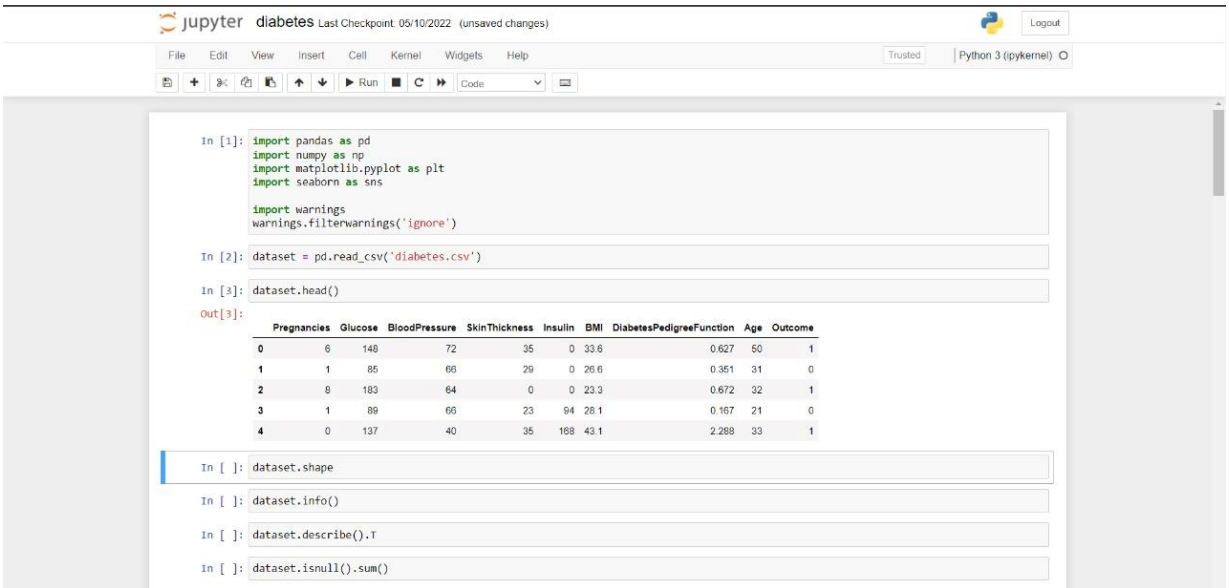
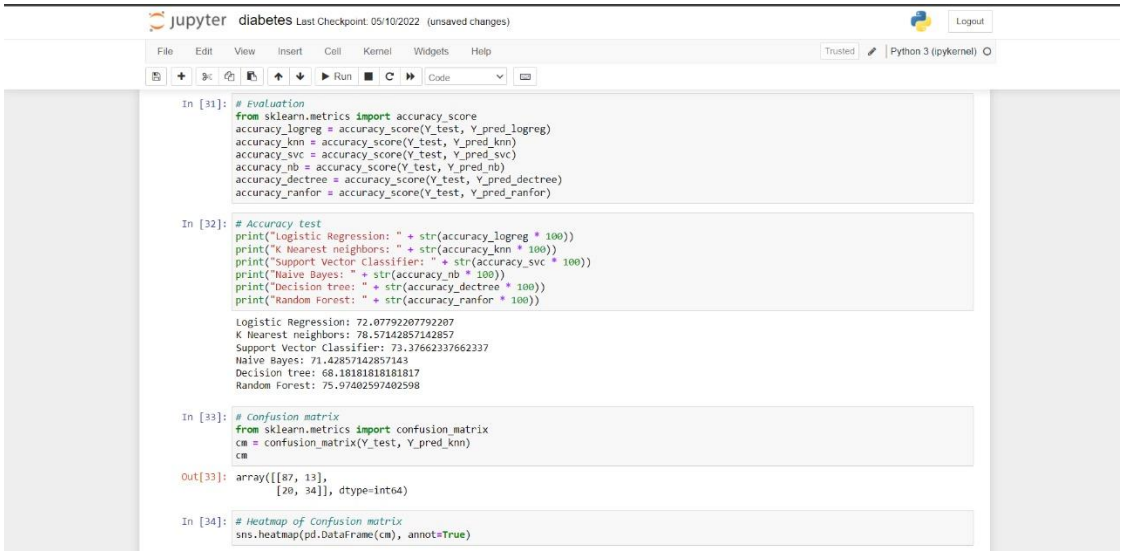
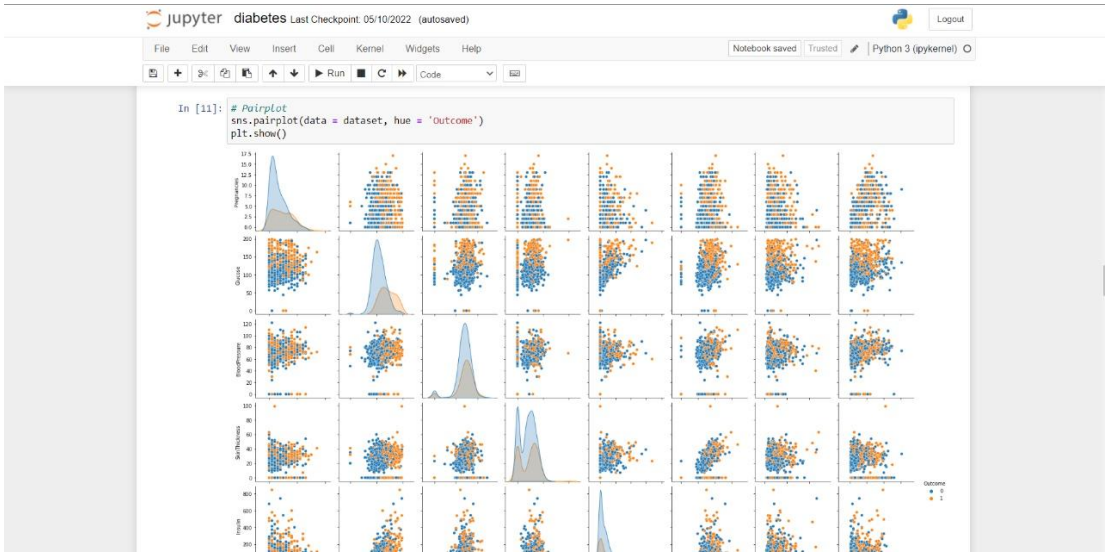


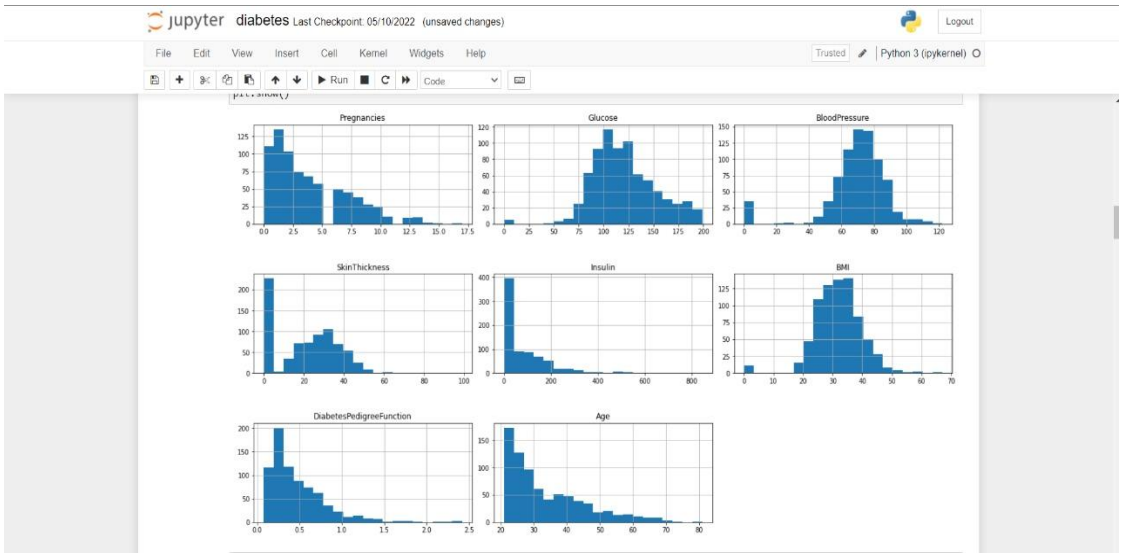
Fig2. Flow diagram of the project.

The Following Figures Representing the Proposed Project:









IV. DISCUSSION

In this undertaking, the goal is to foresee regardless of whether the individual has Diabetes in view of different highlights like Glucose level, Insulin, Age, BMI. We will utilize the Pima Indians dataset from the UCI Machine learning storehouse. We will foster this undertaking in six stages which follows information social event to show sending. Our plans for future work include developing an Android application for the suggested hypothetical diabetes monitoring system, including the proposed categorization and prediction algorithms, and deploying it. Genetic algorithms, in line with the proposed prediction method, may be investigated for better monitoring.

V. CONCLUSION

In this paper, four pieces of the SVM are utilized for forecast and their expectation precision is estimated. It is found that the RBF segment best performs for the diabetes gauge of Indian patients as its assumption precision is found best among the four sections. In future the RBF SVM bit can be tried in expectation of different sicknesses, for example, Cancer, Thyroid and so forth. Further the work can be reached out to look at the presentation of SVD on other execution measurements, for example, accuracy, review and F-measure and can measure up to other existing methods. In future other AI and Deep learning calculations can likewise be applied for diabetes forecast and their exhibition can measure up to SVD.

REFERENCES

1. J. Neelaveni and M. S. G. Devasana, "Alzheimer Disease Prediction using Machine Learning Algorithms," 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020, pp. 101-104.
2. V. Vats, L. Zhang, S. Chatterjee, S. Ahmed, E. Enziama and K. Tepe, "A Comparative Analysis of Unsupervised Machine Techniques for Liver Disease Prediction," 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Louisville, KY, USA, 2018.
3. Gavhane, G. Kokkula, I. Pandya and K. Devadkar, "Prediction of Heart Disease Using Machine Learning," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2018, pp. 1275-1278.
4. R. Atallah and A. Al-Mousa, "Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method," 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), Amman, Jordan, 2019, pp. 1-6.
5. S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" ,in IEEE Access, vol. 7, pp. 81542-81554, 2019.
6. M. A. Alim, S. Habib, Y. Farooq and A. Rafay, "Robust Heart Disease Prediction: A Novel Approach based on Significant Feature and Ensemble learning Model", 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2020, pp. 1-5.
7. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia and J. Gutierrez, "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease", 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, 2017, pp. 204-207.
8. M. Patil, V. B. Lobo, P. Puranik, A. Pawaskar, A. Pai and R. Mishra, "A Proposed Model for Lifestyle Disease Prediction Using Support Vector Machine", 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), angalore, 2018, pp. 1-6.
9. S. R. Alty, S. C. Millasseau, P. J. Chowienczyc and A. Jakobsson, "Cardiovascular disease prediction using support vector machines", 2003 46th Midwest Symposium on Circuits and Systems, Cairo, 2003, pp. 376-379 Vol. 1.
10. R. S. Raj, D. S. Sanjay, M. Kusuma and S. Sampath, "Comparison of Support Vector Machine and Naïve Bayes Classifiers for Predicting Diabetes", 2019 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE), Bangalore, India, 2019, pp. 41-45.
11. Dr. M. Senthil Kumar, “A Hybrid Fusion of Multimodal Medical Images for the Enhancement of Visual Quality in Medical Diagnosis”, Computer Aided Intervention and Diagnostics in Clinical and Medical Images, Volume & Issues-31, pp. 61-70, ISSN No: 2212-9391.