A SURVEY ON DIABETES USING MACHINE LEARNING

M. Koteswara Rao[1], R. Hari[2], N. Narendra Kumar[3], P. Manoj Kumar[4], B. Ashok[5]

[1,2,3,4]Students of ECE ,Chalapathi Institute of Engineering and Technology, Lam, Guntur.
[5]Assistant Professor of ECE, Chalapathi Institute of Engineering and Technology, Lam, Guntur.

ABSTRACT:
Diabetes Detection plays a crucial role in order to diagnose and give proper medication for any disease. Diabetes can be detected by considering some parameters like glucose, BMI, insulin, age etc. A careful detection of diabetes is very important for precise analysis of patient health condition. The proposed method for detecting diabetes is done by using one of the machine learning classifiers which is referred as Random Forest Classifier. Random forest is a classifier "which consists of a number of decision trees in it on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset". All the necessary parameters for detecting diabetes is taken as data set which is imported, analysed by using random forest algorithm. The main aim of the project which is diabetes detection can be build by using Numpy, Pandas and sklearn. The model is evaluated based on the validation parameters. Experimental results show that the model performs well on test data with better precision and accuracy. This project can help people to make a preliminary judgment about diabetes according to their daily physical examination data and it can serve as a reference for doctors.

KEYWORDS: Microcontroller, Heart rate, Body temperature, Remote monitoring. used in order to build and test the model. The dataset **KEYWORDS:** Automatic, Neural network, Traffic signs, Recognition.

1. INTRODUCTION
Diabetes is a disease whereby blood sugar (glucose) is not metabolized in the body. This increases the glucose in the blood to alarmingly high levels. This is known by the name hyperglycemia. In this condition, body is unable to produce sufficient insulin. The other possibility is that body cannot respond to the produced insulin. Diabetes is incurable, it has to be controlled. A diabetic person can develop severe complications like nerve damage, heart attack, kidney failure and stroke. According to statistics in 2017, an estimated 8.8% of global population has diabetes. This is likely to increase to 9.9% by year 2045. Hyperglycemia caused by diabetes, create abnormalities in the cardiovascular system Diabetes causes cardiovascular autonomic neuropathy (CAN) which completely upsets the nervous system and results in diminished variability in heart rate. A variety of machine learning techniques has been proposed for the automated detection of diabetes in a non-invasive way. Machine learning techniques, which can self-learn from data, have been increasingly employed for detecting diabetes now-a-days. In this project one of the machine learning algorithm namely Random Forest Algorithm is, -diastolic blood pressure (mmHg), which is considered in this model consists of several predictor variables and one target variable

2. REVIEW
Significant work has been reported on Pima Indian diabetes datasets (PID). These studies applied different methods to the given problem, and achieved high classification accuracies using the dataset taken from the University of California, Irvine (UCI) machine learning repository [10]. This database provides a well validated data resource to explore the prediction of diabetes. The eight variables in the

dataset include: number of times pregnant plasma glucose concentration at 2 hour in an oral glucose tolerance test triceps skin fold thickness (mm) 2-h serum insulin (IU/ml), body mass index (weight in kg/height in m), diabetes pedigree function, and age (years). While PID is one of the mostly used datasets for prediction of type 2 diabetes, some researchers prefer to investigate diagnosis using data from hospitals, and to incorporate their own parameters of interest. Kazemnejadet al. used the Tehran Lipid and Glucose Study dataset which consists of variables like age, body mass index, waist-to-hip This is proposed to achieve through machine learning and deep learning classification algorithm ratio, gender, history ofhyperlipidemia, and history of hypertension [11]. In another study conducted by Dey*et al*. on data of 530 patients from Sikkim Manipal Institute of Medical Sciences, risk factors such as random blood sugar test results, fasting blood sugar test results, post plasma blood sugar tests, age, sex, and occupation were taken into account [12].

The third National Health and Nutrition Examination Survey (NHANES III, http://www.cdc.gov/diabetes/) dataset resulted from a survey conducted on a US population. The eighteen variables identified as important for diabetes risk prediction include body mass index, height, weight, waist circumference, waist-to-hip ratio, age, sex, race/ethnicity, taking blood pressure medication, taking cholesterol medication, gestational diabetes, high blood pressure, high cholesterol, history of diabetes (any blood relative), history of diabetes (parent or sibling), history of diabetes (parent), history of diabetes (sibling), and exercise, For machine learning we are going to use SVM algorithm For deep learning we are going to use neural network algorithm The proposed system improves accuracy of prediction through deep learning techniques. The dataset which had used in this model consists of 769 patient diagnostic measurements. This dataset consists of totally 9 parameters in it, out of those 8 are predictor(independent) variables and one is the target(dependent) variable this target can also be referred as outcome. This dataset is extracted from Kaggle which is referred from PIMA INDIAN DIABETES DATABASE

3. **PROPOSED SOLUTION**

The proposed system study is classification of india. A single dataset can be provided as an all these PIMA dataset for diabetes as binary classification3 algorithms with minimal or input to no problem Modification. A common scalar can be used to normalize the Available: input provided to these 3 algorithms
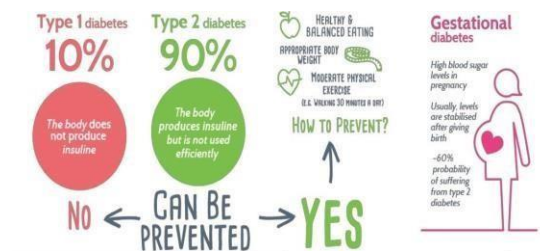
4. **CONCLUSION**

We developed a Prediction Engine which enables the user to check whether he/she has diabetes or heart disease. The user interacts with the Prediction Engine by filling a form which holds the parameter set provided as an input to the trained models. The Prediction engine provides an optimal performance

compared to other state of art approaches. The Prediction Engine makes use of three algorithms to predict the presence of a disease namely: Support Vector Machine (SVM), K- Nearest Neighbours (KNN) and Naïve Bayes.The reason to choose these three algorithms are: They are effective, if the training data is large.

5.REFERENCES

1. Priyanka Sonar, Prof. K. JayaMalini, "Diabetes Prediction Using Different Machine Learning Approaches", Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019) IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4.

2. Samrat Kumar Dey, Ashraf Hossain, Md. Mahbubur Rahman, "Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm",   2018 21st International Conference of Computer and Information Technology (ICCIT).

3. Deeraj Shetty, KishorRit, Sohail Shaikh,Nikita Patil, "Diabetes Disease Prediction Using Data Mining", 2017 International Conference   on Innovations   in Information,   Embedded and Communicatio Systems (ICIIECS)

4. ZhilbertTafa,   NerxhivanePervetica, Bertran   Karahoda,   "An Intelligent   System for Diabetes   Prediction", 4th Mediterranean Conference on Embedded Computing MECO – 2015