

FORECASTING DIABETES

V.Kandasamy^{1*}, Sathish Kumar. G^{2a}, Sanjay Preeth. D^{3a}, Sanjay. V^{4a}

**Assistant Professor, Department of Information Technology, Panimalar Engineering College*

^aStudent, B.Tech Information Technology, Panimalar Engineering College

-----***-----

Abstract

In Today's world the Diabetes has become one of the most frequent diseases of all age generation. It is a chronic disease which has not yet initiate a cure for it. According to the IDF (International Diabetes Federation) just about 382 million people are living, furthermore endure from this disease. Diabetes Mellitus, alias diabetes which is on the whole caused due to the disparity and swell in the blood glucose level. It a consequential disease that the imbalance and not taking precaution measure for this disease may act on various organ like the human kidney, eye, heart, nerves, etc. and sometimes even silent death. Data science methods have the possible to benefit other scientific fields by discarding new light on familiar questions. scientific field in data science commerce with the ways in which machines learn from experience. The main point of this project is to enlarge a system which can perform early prediction of diabetes for a patient with a higher accuracy by combining the results of different machine learning techniques.

Keywords: Diabetes, Data science, Machine learning

1. INTRODUCTION

In the contemporary world, the Diabetes is the vigorous disease amid the people and even betwixt the youth. In some case even after unquestionable months the new born babies are also found to be Diabetic. Sugar has come one of the fundamental human needs in their routine. Foods like bread, fruits, chocolates, dairy products etc. hold at least a skosh of sugar. When we eat those nourishment our body smash them into the bloodstream and as a result Diabetes develops. Diabetes Mellitus means a high sugar level in the blood which is very often junk as urine. A Diabetic Patient has the habit of urinating frequently. This is essentially due to the evacuation of the excess glucose level in the blood. DM can be divided into several distinct types. However, there are two major clinical types, 1) type 1 diabetes (T1D) 2) type 2 diabetes (T2D), according to the etiopathology of the disorder. T2D emerge to be the most common construct of diabetes (90% of all diabetic patients), mainly personified by insulin resistance. The main causes of T2D include lifestyle, physical activity, dietary habits and heredity, whereas T1D is thought to be overdue to auto immunological destruction of the Langerhans islets hosting pancreatic- β cells. Gestational diabetes is the one that turnup in pregnant women who suddenly advance a high blood sugar. In two thirds of the cases, it will renege during subsequent pregnancies. There is a vast chance that type 1 or type 2 diabetes will occur after a pregnancy affected by gestational diabetes. Machine learning is the logical field merchantly with the ways in which machines learn from encounter. For numerous researchers, the term "machine learning" is impalpable of the term "artificial intelligence", given that the plausibility of learning is the most characteristic of an entity called cleverly within the broadest sense of the word. With the rise of Machine Learning passage, we have the capacity to discover an scheme to this precipitate, we have created a framework exploiting information mining which has the capacity to await whether the understanding has diabetes or not. Information mining has the capacity to disengage covered up information from a colossal sum of diabetes-related information. Since of that, it encompasses a citable part in diabetes investigate, presently more than ever. The point of this cross- examine about is to create a framework which can foresee the diabetic chance level of a quiet with a better accuracy. This project has centred on creating a system based on some classification strategies specifically, Support Vector Machine, Logistic regression, Decision algorithms, Random Forest

2. WHY A DIABETIC SHOULD HAVE A THIS SYSTEM

In the leading edge of years number of Diabetic patients have been incessantly compounding and there is no unending cure till now. Each one of them can only control the adverse effect of this disease, they should unceasingly keep trail of the sugar level. Even more the elderly people cannot travel recurrently to check their sugar level. Using this system everyone can regularly check their sugar level and as a fence they can also take precautionary measures if they are going to be diabetic. The main is to find out the Diabetes at it early stage.

3. METHODOLOGY

In this we might learn more or less the different classifiers utilized in machine learning to anticipate diabetes. We should to clarify our presumed technique to progress the exactness. Certain unlike strategies were employed in this paper. The distinctive strategies utilized are depicted beneath. The yield is the exactness measurements of the machine learning models. Then, the show can be utilized in expectation.

We have accustomed the Scikit-learn library to enforce our study. The graphical representation of our proposed approach is exhibited in FIGURE 1.

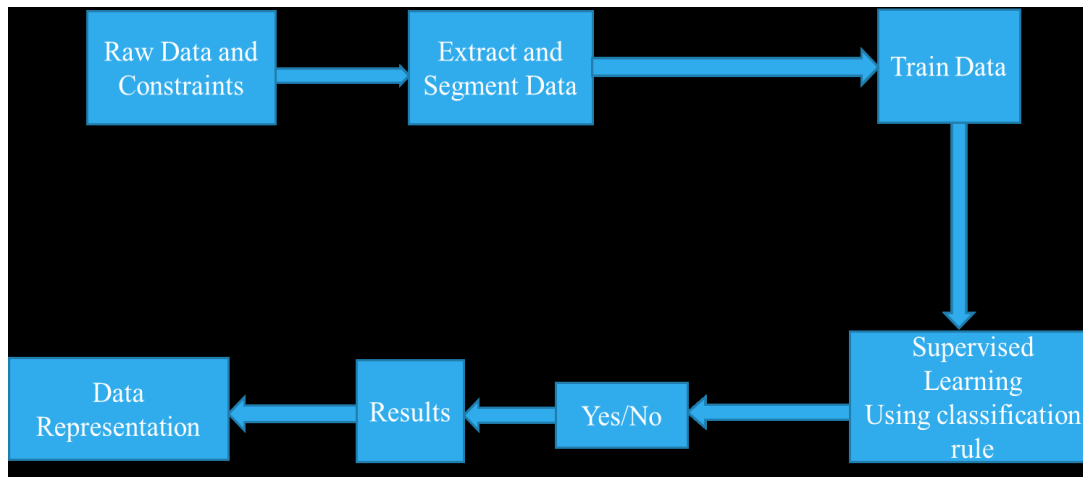


Figure 1: Architecture Diagram

4. DATA INPUT

We have used some data from database. The data were escalated using direct questionnaires and were decisive by a doctor. In this step, we input the data in our proposed system.

4.1 Data Pre-processing

In this section, we have proselyted the unremitting emphasizes like as Polyuria, Sudden weight loss, Polydipsia, Weakness, Gender, Genital thrush, Polyphagia, Visual blurring, Partial paresis, Itching, Irritability, Muscle stiffness, Delayed healing, Obesity, Alopecia, and Class into their numeric counterpart.

4.2 Data Splitting

We separated our dataset into two sections in this section: a training dataset and a research dataset. We used 75 percent of the data for training and 25 percent for testing.

4.3 Applying Machine Learning Algorithms

In this step, we have subdued eleven different classes namely: Logistic Regression (LR), Gaussian Process (GP), Adaptive Boosting (AdaBoost), Decision Tree (DT), Random Forest (RF), with our example data. After training these classifiers, each of them has been used to vaticinate the presence of diabetes in the patients for the testing dataset.

4.4 Best Model

In this analysis the result rides on some part of this research. However, which algorithm gives the best true positive, false positive and also gives the best exactness are the best algorithms in this analysis.

5. PROPOSED MODEL

This study queries the example data from database. The proposed approach outline is tracing in the FIGURE 2. A comparative analysis was achieved on four ML algorithms (Decision tree, Logistic Classifier, Random Forest).

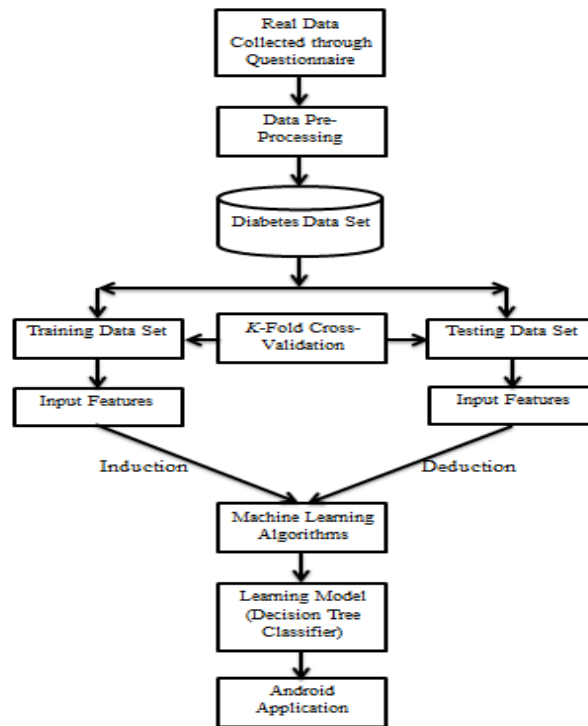


Figure 2: Data flow diagram

The proposed system is mainly heady on development of machine learning model and also it helpful in the medical sector to identify the diseases. This offer system is an mechanization to presage the diabetes using old patient’s data.

5.1 System Design

Designing of system is the process in which it is used to outline the interface, modules and data for a system to prescribe the impportunity to quench . System design is seen as the application of the system theory. The main thing of the design a system is to elaborate the system architecture by giving the data and information that is obligatory for the implementation of a system.

Machine Learning Algorithms: Four Machine learning algorithms (Decision tree, Logistic Classifier, Random Forest , Ada boost) are queried for the division of diabetes dataset into three classes: *High, Medium* and *Low*.

5.2 Decision Tree

This classifier generates a decision tree based on which, it endowed the class values to each data point. Here, we can vary the maximum number of features to be considered while manufacturing the model. The delicacy on the training set is 100% and the test set accuracy is also good.

5.3 Feature Importance in Decision Trees

Feature importance rates how necessary each feature is for the decision a tree makes. It is a number between 0 and 1 for each feature, where 0 means “not used at all” and 1 means “perfectly predicts the target”. Feature “Glucose” is by far the most important feature.

5.4 Logistic Regression

The application of the Logistic regression model has comprised eminently in countless spaces such as the organic sciences. The Logistic regression calculation is avail oneself of when the objective is to classify information items into categories. As a directive in logistic regression the target variable is binary, which advocate that it as it were contains information classified as 1 or 0, which in our case refers to a quiet that's positive or negative for diabetes. The reason of our logistic regression calculation is to perceive the best fit that's effusively sensible to portray the relationship between our target variable and the indicator factors.

5.5 Random Forest

The precision level is more particularly when compared to other calculations. The proposed show gives the most excellent comes about for diabetic expectation and the result appeared that the expectation framework is able of anticipate the diabetes disease successfully, skilled and most critically, immediately.

5.6. Flowchart

The flowchart for the system, its working and analysis is given below:

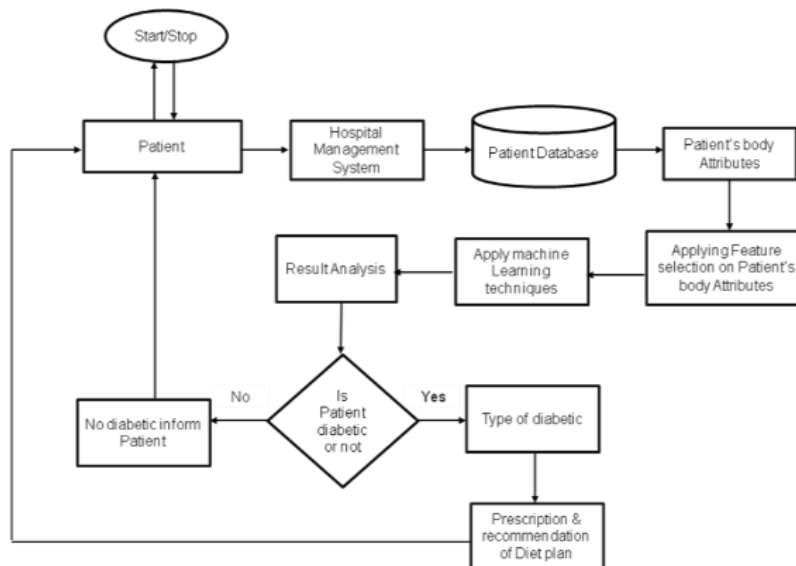


Figure 3: Flow chart diagram

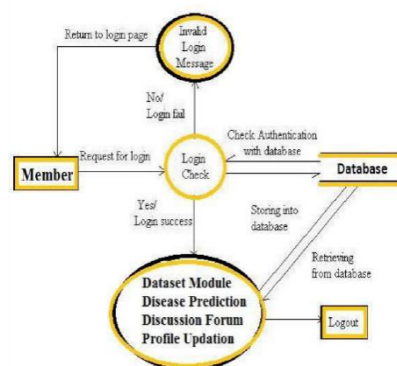


Figure 4: Module data flow diagram

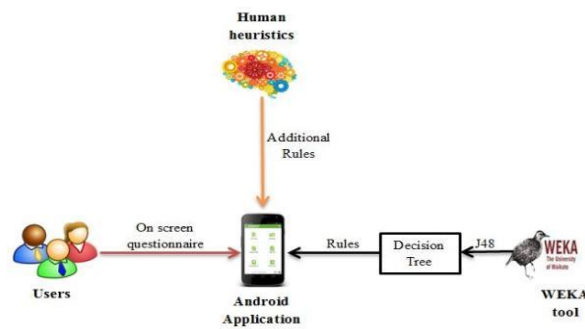


Figure 5: Working model of the proposed system

6. CONCLUSION

Machine learning has the unbelievable capacity to transfigure the diabetes chance forecast with the abetment of furtherance computational strategies and accessibility of rangy sum of epidemiological and hereditary diabetes chance dataset. Detection of diabetes in its early stages is the key for treatment. This work has explicit a machine learning avenue to anticipating diabetes levels. The procedure may excessively offer support analysts to accelerate a precise and viable device that will reach at the table of clinicians to assist them make way better choice around the disease status. This project mainly aims to propose a fruitful technique for earlier detection of the diabetes disease.

REFERENCE

- [1] Wikipedia contributors. (2020, October 26). Diabetes. In Wikipedia, Retrieved 06:20, October 31, 2020, from <https://en.wikipedia.org/w/index.php?title=Diabetes&oldid=985520269>, World Health Organization, (2020, October 28). Diabetes. Retrieved 06:20, October 31, 2020, from <https://www.who.int/health-topics/diabetes>
- [2] M. Sumithra and Dr. S. Malathi, "A Novel Distributed Matching Global and Local Fuzzy Clustering (DMGLFC) FOR 3D Brain Image Segmentation for Tumor Detection", IETE Journal of Research, doi.org/10.1080/03772063.2022.2027284, 2021
- [3] B.Buvanswari and T.Kalpalatha Reddy, "A Review of EEG Based Human Facial Expression Recognition Systems in Cognitive Sciences" International Conference on Eenergy, Communication,Data analytics and SoftComputing(ICECDs),CFP17M55-PRJ:978-1-5386-1886-8",August 2017.
- [4] Chethana, C., Subbiah Swaminathan, S. Sharanyaa, E. Sathish, R. Prathipa, and Anuradha Thakare. "Application Of Reverse Engineering in the Process of Utilization of Human Brain in Artificial Intelligence." Journal of Optoelectronics Laser 41, no. 3 (2022): 89-93.
- [5] M. Sumithra and Dr. S. Malathi, " Modified Global Flower Pollination Algorithm-based image fusion for medical diagnosis using computed tomography and magnetic resonance imaging", International Journal of Imaging Systems and Technology, Vol. 31, Issue No.1, pp. 223-235, 2021
- [6] K. Sridharan , and Dr. M. Chitra "SBPE: A paradigm Approach for proficient Information Retrieval , Jokull Journal" , Vol 63, No. 7;Jul 2013
- [7] Kaur, H., & Kumari, V. (2020). Predictive modelling and analytics for diabetes using a machine learning approach. Applied computing and informatics.
- [8] Sharanyaa, S., P. N. Renjith, and K. Ramesh. "Classification of Parkinson's disease using speech attributes with parametric and nonparametric machine learning techniques." 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS). IEEE, 2020.
- [9] M. Sumithra and Dr. S. Malathi, "3D Densealex NET Model with Back Propagation for Brain Tumor Segmentation", International Journal OfCurent Research and Review, Vol. 13, Issue 12, 2021.

- [10] B.Buvaswari and Dr.T. Kalpalatha Reddy, "EEG signal classification using soft computing techniques for brain disease diagnosis", Journal of International Pharmaceutical Research, ISSN : 1674-0440, Vol.46, No.1, Pp.525-528, 2019.
- [11] Sharanyaa, S., P. N. Renjith, and K. Ramesh. "An Exploration on Feature Extraction and Classification Techniques for Dysphonic Speech Disorder in Parkinson's Disease." In Inventive Communication and Computational Technologies, pp. 33-48. Springer, Singapore, 2022.
- [12] K. Sridharan, and Dr. M. Chitra "Web Based Agent And Assertion Passive Grading For Information Retrieval", ARPN Journal of Engineering and Applied Sciences, VOL. 10, NO. 16, September 2015 pp:7043-7048
- [13] Islam, M. F., Ferdousi, R., Rahman, S., & Bushra, H. Y. (2020). Likelihood prediction of diabetes at early stage using data mining techniques. In Computer Vision and Machine Intelligence in Medical Image Analysis (pp. 113-125). Springer, Singapore.
- [14] M. Sumithra and Dr. S. Malathi, "Segmentation Of Different Modalities Using Fuzzy K-Means And Wavelet ROI", International Journal Of Scientific & Technology Research, Vol. 8, Issue 11, pp. 996-1002, November 2019.
- [15] Sharanyaa, S., S. Lavanya, M. R. Chandhini, R. Bharathi, and K. Madhulekha. "Hybrid Machine Learning Techniques for Heart Disease Prediction." International Journal of Advanced Engineering Research and Science 7, no. 3 (2020).
- [16] M. Sumithra and S. Malathi, "A Survey of Brain Tumor Segmentation Methods with Different Image Modalities", International Journal of Computer Science Trends and Technology (IJCTST) – Vol. 5 Issue 2, Mar – Apr 2017
- [17] B.Buvaswari and Dr.T. Kalpalatha Reddy, "High Performance Hybrid Cognitive Framework for Bio-Facial Signal Fusion Processing for the Disease Diagnosis", Measurement, ISSN: 0263-2241, Vol. 140, Pp.89-99, 2019.
- [18] Sharanyaa, S., and M. Shubin Aldo. "Explore places you travel using Android." In 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 4796-4799. IEEE, 2016.
- [19] Emon, M. U., M. S., Meghla, T. I., Rahman, M. M., Al Mamun, M. S., Kaiser, M. S. (2020, November). Performance Analysis of Machine Learning Approaches in Stroke Prediction. In 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1464-1469). IEEE.
- [20] M. Sumithra and Dr. S. Malathi, "A Brief Survey on Multi Modalities Fusion", Lecture Notes on Data Engineering and Communications Technologies, Springer, 35, pp. 1031-1041, 2020.
- [21] Sharanyaa, S., and Madhumitha RP. "Eyeball Cursor Movement Detection Using Deep Learning." RP, Madhumitha and Rani. B, Yamuna, Eyeball Cursor Movement Detection Using Deep Learning (July 12, 2021) (2021).
- [22] M. Sumithra and S. Malathi, "A survey on Medical Image Segmentation Methods with Different Modalities", International Journal of Engineering Research and Technology (IJERT) – Vol. 6 Issue 2, Mar 2018.
- [23] B.Buvaswari and Dr.T. Kalpalatha Reddy, "ELSA- A Novel Technique to Predict Parkinson's Disease in Bio-Facial", International Journal of Advanced Trends in Computer Science and Engineering, ISSN 2278-3091, Vol.8, No.1, Pp. 12-17, 2019
- [24] K. Sridharan, and Dr. M. Chitra, "Proficient Information Retrieval Using Trust Based Search On Expert And Knowledge Users Query Formulation System, Australian Journal of Basic and Applied Sciences, 9(23) July 2015, Pages: 755-765.
- [25] Md Faisal Faruque, Iqbal H. Sarker, "Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus" IEEE 2019.
- [26] Sharanyaa, S., and K. Sangeetha. "Blocking adult account in osn's using iterative social based classifier algorithm." International Journal of Scientific Engineering and Science 2, no. 1 (2018): 33-36.
- [27] B.Buvaswari and Dr.T. Kalpalatha Reddy, "ACPT- An Intelligent Methodology for Disease Diagnosis", Journal of Advanced Research in Dynamical and Control Systems, ISSN : 0974-5572, Vol.11, No.4, Pp.2187-2194, 2019.

[28] Sumithra, M., Shruthi, S., Ram, S., Swathi, S., Deepika, T., "MRI image classification of brain tumor using deep neural network and deployment using web framework", *Advances in Parallel Computing*, 2021, 38, pp. 614–617.

[29] K. Sridharan , and Dr. M. Chitra "RSSE: A Paradigm for Proficient Information Retrieval using Semantic Web" , *Life Science Journal* 2013;10(7s), pp: 418-425

[30] Sharanyaa, S., S. Vijayalakshmi, M. Therasa, U. Kumaran, and R. Deepika. "DCNET: A Novel Implementation of Gastric Cancer Detection System through Deep Learning Convolution Networks." In *2022 International Conference on Advanced Computing Technologies and Applications (ICACTA)*, pp. 1-5. IEEE, 2022.