An Overview of Diabetes Prediction Analysis Methodologies and Techniques

Asst. Prof. Priyanka D. Nighot

(Assistant Professor)

Sonopant Dandekar Shikshan

Mandli, Palghar, Maharashtra, India

Abstract:

Diabetes mellitus is among the most prevalent chronic diseases globally and continues to rise due to urbanization, economic growth, and lifestyle changes. This paper analyses various diabetes mellitus instances using predictive analytics algorithms, highlighting the necessity for diverse approaches in predictive analytics.

Keywords:

Diabetes, Decision Tree, Random Forest, Naïve Bayes, SVM, and KNN

1. Introduction

Diabetes is a condition that arises from improper production or utilization of the insulin hormone by the body. This makes it possible for the blood to become overly sugary, or glucose. Insulin is a hormone that controls blood sugar. The World Health Organization reports that diabetes caused over 2.2 million deaths in 2012 and 1.6 million deaths in 2016. Nerves, kidneys, and other body organs may be harmed by uncontrolled diabetes. Diabetes symptoms include increased thirst, increased hunger, abrupt weight loss, and frequent urination. Thus, early forecasting is necessary to prevent loss of life and financial ruin. There are essentially three types of diabetes: Type 1 Diabetes, in which the body is unable to produce insulin. Type 2 Diabetes: insufficient body function to utilise insulin.

2. Current Methods of Predictive Analysis

1. Methods of Machine Learning

A growing method that enables machines to efficiently learn from existing records is called machine learning. Machine learning employs a variety of strategies to build mathematical models and provide predictions based on statistical information or expertise. These days, it is employed for a wide range of activities, such as email filtering, recommendation systems, speech recognition, image identification, auto-tagging for Facebook, and many more.

2. Data Mining Techniques

One of the most useful techniques for helping developers, academics, and individuals extract useful information from massive data sets was data mining. Another name for information discovery in databases is data mining. The knowledge discovery process includes data cleansing, data integration, data collecting, data transformation, data mining, pattern analysis, and knowledge presentation.

Basic table outlining some methodologies and techniques commonly used in the analysis of diabetes prediction:

Methodology/Technique	Detailed
The Logistic Regression	The Regression Logistic statistical technique for jobs involving binary classification, like determining if a patient has diabetes or not from a set of input features.
Decision Trees	These are tree-like structures that are effective for both classification and regression problems. Each node in a decision tree represents a feature, each branch a decision based on that feature, and each leaf node a predicted outcome.
Random Forest	Unknown Forest An ensemble learning technique that produces a large number of decision trees during training and outputs the mean prediction for regression or the mode of the classes for classification, providing improved robustness and accuracy.
Support Vector Machines (SVM)	Vector Support Systems (SVM) a regression analysis and classification model based on supervised learning. It maximizes the margin between classes by identifying the hyperplane that divides a dataset into the best possible classes.
Neural Networks	A computational model that resembles the structure and functions of the human brain by connecting nodes, or neurons. The output of the model is intended for use in complex pattern recognition tasks.
k-Nearest Neighbors (k-NN)	Narrowest NeighborsThe majority vote of the k nearest neighbors inside the feature space determines the result of this non-parametric regression and classification method.
Naive Bayes Classifier	A probabilistic classifier based on Bayes' theorem, assuming independence among features, commonly used in text classification but can be applied to various prediction tasks, including diabetes prediction.

Conclusion:

The different review papers place more emphasis on the methods and processes of predictive analysis for estimating and forecasting diabetes. Numerous methods are applied to diabetes datasets in order to more accurately forecast the treatment of diabetes. The process of comparing and contrasting several approaches makes it easier to determine which is best for future prediction. Therefore, hybrid methods for diabetes illness diagnosis and prediction are more efficient and produce more accurate results. Summary of key findings and insights. Recommendations for future research directions in diabetes prediction.

Bibliography:

1) Rawshani A, Rawshani A, Franzén S, Eliasson B, Svensson AM, Miftaraj M, McGuire DK, Sattar N, Rosengren A, Gudbjörnsdottir S. is the bibliography. Cardiovascular illness and mortality in people with type 1 and type 2 diabetes. Journal of Medicine in New England. 2017 Dec 7;376(15): 1407–18.

2) Mahajan A, Florez JC, McCarthy MI, Udler MS. Precision medicine and genetic risk scores for diabetes diagnosis. Reviews of endocrines. Apr 1, 2019; 40(2): 150-92.

3) Moons KG, Navis G, Bakker SJ, Beulens JW, Abbasi A, Peelen LM, Corpeleijn E, van der Schouw YT, Stolk RP, Spijkerman AM, van der AD. Type 2 diabetes prediction models: a comprehensive review of the literature and an independent, external validation investigation. Sep 18, 2012; Bmj; 345:e5900.

4) Zhang L, Sun X, Wang Y, Qi C, Zheng T. Fasting blood glucose levels are useful in predicting the emergence of diabetes in non-diabetic people who have had an ischemic stroke. 2015 Oct 1;36(10):1827–1833; Neurological Sciences.

5) Huang, Hu, X, Cai, Chen, P, Mai, Tang, and Huang Y. A meta-analysis examining the relationships between cardiovascular and all-cause mortality and prediabetes. Medical journal annals. 2014 Nov 3;46(8):684–92.

6) Herman WH, Zimmet P. Type 2 diabetes: a global epidemic that needs immediate attention and response. Diabetes Care. 2012;35(5):943–944. Dec. 1.

7) Sowers JR, Frye RL, Arora S, and Smakowski A. Type 2 diabetes and hypertension: a recipe for atherosclerosis? Diabetes Care, May 1, 2002; 25 (5): 809–14.

8) Lipska KJ, Lee SJ, Ross JS, Shah ND, Miao Y, and Steinman MA. Possible overtreatment of diabetic mellitus in elderly individuals who have strict glycemic management. 2015 Feb 10;313(5):534–45 in Jama.

9) Ali MK, Sattar N, and Gregg EW. Complications from diabetes are evolving. Diabetes & Endocrinology, The Lancet, 4(2), 99-100, Feb. 1, 2016.

10) Pivovarov R, Albers DJ, Sepulveda JL, Elhadad N. Laboratory test bias identification and mitigation. Biomedical Informatics Journal. 2014 Oct 1;51:24–34.

11) Vlahavas, I., Maglaveras, N., Salifoglou, A., Tsave, O., Kavakiotis, I., & Chouvarda, I. (2017). Data mining and machine learning techniques in diabetes research. 15: 104–116 in Computational and Structural Biotechnology Journal.

12) Dhar, S., Srivastava, G., and A. D. Dwivedi (2019). a comparison of machine learning methods for the diagnosis of diabetes. (pp. 1–5) in the 2019 International Conference on Automation, Computational, and Technology Management (ICACTM). IEEE.

13) Jain, A., and R. Gupta (2019). A comparison of machine learning techniques empirically for the prediction of diabetes. (pp. 827-832) in 2019 International Conference on Communication and Electronics Systems (ICCES). IEEE.

14) Yu et al. (2010); Liu et al. (2010); Valdez et al. (2010); Gwinn et al. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. BMC medical decision-making and informatics, 10(1), 16.

15) Bendale, Anthony, and T. Agrawal (2015). Assessment of machine learning algorithms' performance in predicting diabetes. Computational Intelligence and Communication Networks (CICN), International Conference, 2015, pp. 428–433. IEEE.