CrossRef DOI of original article:

Design of Framework to Reduce the Risk of Diabetic using Machine Learning

dr.annandakumar

Received: 1 January 1970 Accepted: 1 January 1970 Published: 1 January 1970

Abstract

Diabetes is a serious disease that spreads rapidly around the world and is first surprising. Its danger factors are therefore quite high. A different technique to identify diabetes at an early stage is to gradually decline in general health. This approach will reveal the health of the body's organs before symptoms manifest. Using machine learning approaches, a framework for the diabetes prediction system is built in this research paper's first stages.

Index terms— diabetic prediction, support vector machine, accuracy.

1 I. Introduction

he risk factors for developing diabetes are numerous and growing rapidly. Diabetes is an unexpected, severe disease that is expanding throughout the world. 90 million of India estimated 1391.99 million people have diabetes, according to a November 14, 2019 article in The Hindu. Additionally, the ninth edition of the Atlas, published by the International Diabetic Foundation (IDF), forecast that there will be 134 million diabetics worldwide by 2045. This study article discusses how to create and develop a framework to use machine learning approaches to prevent future complications as well as to identify type-1, type-2, gestational, and pre-diabetic. A clinical dataset model that uses data gathered from digital surveys and hospitals to predict type-1, type-2, and prediabetic diabetes. The main purpose of study is to enhance people's lives and make them more relaxed. The Python Data Analysis Library (PANDA), or data wrangling, was used to make the csv, tsv, or sql database useable based on the clinical data set that was obtained. Machine learning (algorithms) approaches are used on the processed data. Slight variations with various datasets can aid to lead the accuracy as different ML algorithms are used to various data sets to determine accuracy. A framework that predicts diabetes will be created using training and testing that is based on maximal accuracy. We are talking about how to create a framework model using clinical data that has been gathered from surveys and hospital path labs that may predict if a patient will acquire type-1, type-2, or pre-diabetes as well as future risk factors. The risk factors for developing diabetes are numerous and growing rapidly. Diabetes is an unexpected, severe disease that is expanding throughout the world. 90 million of India's estimated 1391.99 million people have diabetes, according to a November 14, 2019 article in The Hindu. Additionally, the ninth edition of the Atlas, published by the International Diabetic Foundation (IDF), forecast that there will be 134 million diabetics worldwide by 2045. This study article discusses how to create and develop a framework to use machine learning approaches to prevent future complications as well as to identify type-1, type-2, gestational, and pre-diabetic. A clinical dataset model that uses data gathered from digital surveys and hospitals to predict type-1, type-2, and prediabetic diabetes. The main purpose of study is to enhance people's lives and make them more relaxed. The Python Data Analysis Library (PANDA), or data wrangling, was used to make the csv, tsv, or sql database useable based on the clinical data set that was obtained. Machine Learning Algorithms approaches are used on the processed data. Slight variations with various datasets can aid to lead the accuracy as different ML algorithms are used to various data sets to determine accuracy. A framework that predicts diabetes will be created using training and testing that is based on maximal accuracy. We are talking about how to create a framework model using clinical data that has been gathered from surveys and hospital path labs that may predict if a patient will acquire type-1, type-2, or pre-diabetes as well as future risk factors.

the plan, hyper planes are employed as decision boundaries. Support Vector Machine is design to made strong by the assistance of cost function and gradient in determining the margin between plane and data sets. When

the presence of diabetes is predicted and confirmed based on clinical data and symptoms, this research technique provides a complete solution by registering patients. Step-by-step solutions are provided by this system, including testing, consultation with a doctor, and the prescription of medicine. For this presentation, only SVM will be used however, there will be a sequence of all algorithms which will be used to design ge With various libraries like Matplotlib, Pyplot, and others, we use Python programming for backend coding. Python Panda is used for data analytics, and CSS is used for the front end. The author has the right to retain the idea. This study is ongoing, thus it must not be replicated or copied.

This essay is structured as follows: Section 1 presents the past research on diabetes done by various researchers. In part 3, problem identification is completed. Goal of the issue statement in Section 4. Section 5 provides a suggested technique for employing several ML algorithms to handle diabetes challenges. The accuracy rate, error rate, and other metrics are computed, and the SVM's results for predicting diabetes are provided in section 6.

2 II. Existing Work a) Literature of Diabetic Complication

First, we looked at a number of articles and talks on current hot topics in the healthcare machine learning. If diabetes is not treated promptly, a number of long-term complications have been described by the Mayo Clinic USA. The risk of problem increases with the duration of diabetes and the degree to which your blood sugar is under control. Diabetes problems might eventually become incapacitating or even fatal. Among the potential issues [1].

i. Cardiovascular Disease Diabetes may raises the risk of a amount of cardiovascular issues, such as coronary artery disease with chest discomfort (angina), heart attacks, strokes, and arterial constriction (atherosclerosis). Diabetes increase your risk of initial heart disease or stroke [5].

3 ii. Nerve Damage (Neuropathy)

The walls of the tiny blood arteries (capillaries) may feed nerves that can get damaged by too much sugar, mainly in your legs. The tingling, numbness, burning, or pain that may result from this typically starts at the tips of the toes or fingers and progressively moves upper. If uncared for, the afflict limbs can go fully numb. Problems with nausea, vomiting, diarrhoea, or constipation can outcome from damage to the nerves that control digestion. It could cause erectile dysfunction in males.

4 iii. Kidney Damage (Nephropathy)

The kidneys' millions of glomeruli, or groups of small blood vessels, filter waste from your blood. This sensitive filtration mechanism can be harmed by diabetes. A kidney transplant or dialysis may be necessary if there is severe damage that results in kidney failure or irreversible end-stage kidney disease.

5 iv. Eye Damage (Retinopathy)

Diabetes can cause diabetic retinopathy, which can damage the retina's blood vessels and result in blindness. Diabetes also raises the risk of glaucoma and cataracts, two devastating eye diseases.

6 v. Foot Damage

Various foot issues are more likely to occur when there is nerve injury in the feet or insufficient blood supply to the feet. Blisters and injuries that go untreated can get seriously infected and heal badly. An eventual toe, foot, or limb amputation may be necessary due to these illnesses.

7 vi. Skin Conditions

Diabetes patients are more likely to experience hearing issues.

8 vii. Hearing Impairment

Hearing problems are more common in people with diabetes.

9 viii. Alzheimer's Disease

Alzheimer's disease and other forms of dementia may be more likely in people with type 2 diabetes. The danger seems to increase as your blood sugar control declines. There are suggestions about how these conditions could be related, but none have been shown.

10 ix. Depression

Diabetes patients, both type 1 and type 2, frequently experience depressive symptoms. Diabetes control may be impacted by depression (mayoclinic.org, USA).

11 b) Previous Studies on Diabetic using Machine Learning

If the case study of the long-term consequence had not been examined utilising machine learning, our list of references would be lacking. In one debate during the 2019 IEEE International Conference on Deep Learning and Machine Learning in Turkey, it was suggested to use a generalised additive model to predict severe Retinopathy of Prematurity. Without the medical literature pertaining to our case study, severe retinopathy of prematurity, our list of references would be lacking (RoP). Retinopathy in infants, particularly those who weigh less than 1500 grammes at birth, can result in blindness if disease progresses to stages 4 and beyond. Thus, routine examinations by nurses, neonatologists, and ophthalmologists guarantee that therapy is administered if RoP exceeds a certain threshold, such as a diagnosis of severe RoP [2].

12 41

Year 2022 ( ) J Another international conference, ICITACEE, conducted in Indonesia, discussed the application of machine learning techniques such the Gray Level Occurence matrix, Support Vector Machine, and K Nearest Neighbor to predict diabetes using iridology.

The procedure starts with data gathering in the form of iris image acquisition and blood glucose level monitoring, after which image processing is carried out. Pre-image processing, image processing, and classification are the main three steps that the iris image processing process goes through. Image augmentation, image localisation, and image normalisation are all parts of the pre-image processing step. While feature extraction and ROI segmentation make up the image processing step. We also included references to four outstanding studies for a more omprehensive overview of diabetes and its long-term complications.

In-depth explanations of renal damage caused by uncontrolled diabetes were provided by Linta Antony, who also provided an algorithm.

For more general discussion on diabetic and its long-term complication we have also referred four excellent papers.

Linta Antony explained in detailed the about kidney disease because of diabetic unmanaged and has given an algorithm. BEGIN

13 III. Problem Identification

With 1,393 million people, India has the secondlargest population in the world and may surpass China as the most populous country by the middle of the decade due to people's busy lives and lack of time to care for or assess their own health. According to IDF USA, India has one in six diabetics, making it the country with the second-highest number of diabetics worldwide with an estimated 77 million and a startling 134 million in 25 years [23].

Because of the nation's population, economy, and demographics, major illnesses frequently progress into a variety of problems, and people tend to wait until something goes wrong before seeking care. As this diabetic causes rapid harm to families, societies, and countries, it may have an impact on how those entities develop, either directly or indirectly. This is a major issue that affects the younger generation as well, therefore it has to be identified quickly, treated effectively, and maintained with userfriendly systems (ML). If discovered at a critical time, identifying long-term complications is another significant issue [23].

14 IV. Objective

My long-term goals with this research study are to make it easy to identify long-term complications and determine if a person has diabetes if symptoms are recognised. I'll achieve this objective by achieving the following goals:

15 ? Determine if the person with diabetic of with what

type, based on feed dataset through system.

? Determine a trained dataset based on feed to act upon. ? Assess healthy habit procedure.

? Compare whether or not diabetic managed based on periodic feed dataset. ? Determine long term complications of diabetic through system. ? Assess diagnostic medical procedure.

16 V. Proposed Methodology

To understand a person with diabetes and the sort of machine learning algorithms and ANN that play a crucial part in fines, much study is necessary. Since accuracy is determined by how well the selected actions mirror the proper ones, machine learning is the process of teaching computers to adjust or adapt their actions (whether these actions are generating predictions or commanding a robot) in order to become more accurate.

Imagine playing Scrabble (or another game) against a machine. CASE STUDY: In the beginning, you might win every time, but after several games, it starts to defeat you, until eventually you never do. Either you're becoming worse, or the Scrabble-winning machine is getting smarter. This is a sort of generalisation because once it figures out how to defeat you, it may continue to apply the same tactics against other players rather than starting over with every new opponent [24]. set of examples with the right answers (targets), which is supplied. This is also known as imitation learning.

17 a) Types of Machine

ii. Unsupervised Learning The algorithm instead looks for commonalities between the inputs such that inputs that share anything are grouped together, rather than providing correct answers. Density estimation is a statistical method used in unsupervised learning.

18 iii. Reinforcement Learning

Between supervised and uncontrolled learning, this falls. Although the algorithm is informed when the solution is incorrect, it is not informed how to make it right. It must investigate and test out several options before figuring out how to provide the correct response. Because the monitor only assigns a score and makes no suggestions for improvement, reinforcement learning is sometimes referred to as learning with a critic.

19 iv. Evolutionary Learning

It is possible to think of biological evolution as a learning process whereby organisms make adjustments to increase their chances of reproduction and survival. We'll examine how this may be computer-modeled using the concept of fitness, which is equivalent to a rating for how effective the present solution is.

The majority of learning takes place in supervised learning, which will be the subject of the following several chapters. We'll examine what it is and the sorts of issues it may be used to tackle before we get started [24].

20 b) ML Algorithms

Logistic regression is a type of supervised learning techniques which deals the details of dependent variable and independent variable can be understood using sigmoid function. Actually logistic regression not used for regression problems rather is a type of machine learning classification problem where the dependent variable is dichotomous (0/1, -1/1, true/false) and independent variable can binominal, ordinal, interval or ratio-level [9]. The sigmoid/logistic function is given as?? = 1 1+?? ???

Where, y is the output that the result of weighted sum of input variables x. If output is greater than 0.5, the result is 1 else the result is 0 [5] [7].

Where, y is the result which is the result of weighted sum of input variables x. If the result is greather than 0.5, the result is 1 else the result is 0.

ii. K-Nearest Neighbor Classifier K-Nearest Neighbor (KNN) techniques are used to solve problems related to regression as well as classification, conversely it is being used to solve classification problems in business. Its main advantage is simplicity of translation and less computation time. In figure 2, (KNN example) the points (3.5, 6) and (4.5, 5.5) will be allocated in any one of the clusters. (Neha Prerna Tiggaa et. Al., 2019). The K nearest neighbor uses Euclidean distance function to calculate distances with existing data points and any new data point. Thus, (3.5, 6) will belong to the green cluster, whereas, (4.5, 5.5) will belong to the red cluster [6][13].

. In machine learning techniques, SVM is a supervised classifier that may be applied to both classification and regression. It is mostly used to address classification-related issues. SVM attempts to categorise data points in a multidimensional space using the proper hyperplane. A decision boundary for classifying data points is a hyperplane. With the widest possible gap between the classes and the hyperplane, the hyperplane classifies the data points. Support vector machine classification is shown in Figure 3

21 Decision Tree Classification

Method A Decision Tree is works on the principle of decision making. It can be described in form of tree and provides high accuracy and stability [6].

22 c) The Machine Learning Process

It briefly examines the process by which machine learning algorithms can be selected, applied, and evaluated for the problem.

23 i. Data Collection and Preparation

Target data is also necessary for supervised learning, which may need consulting subject-matter experts and making large time commitments. The quantity of data must also be taken into account. Machine learning methods require large quantities of data, ideally without too much noise, but as dataset sizes grow, so do computational costs, and it is typically hard to estimate the exact point at which there is enough data but not too much computer overhead.

24 ii. Feature Selection

We examined potential traits that may be advantageous for coin identification. It entails determining the characteristics that are most beneficial for the issue at hand. In the case of the coins example above, human common sense was used to select some potentially beneficial qualities and to ignore others; this generally needs previous knowledge of the issue and the data.

Algorithm Selection This book should be able to prepare you for selecting an acceptable algorithm (or algorithms) given the dataset since understanding the fundamental concepts of each algorithm and examples of how to apply them is exactly what is needed for this. Selection of Parameters and Models There are several algorithms that contain parameters that need to be manually specified or that need testing to get the right values.

Given the dataset, method, and parameters, training should consist of nothing more complicated than the application of computer resources to create a model of the data in order to forecast the results on fresh data.

Evaluation A system must be tested and assessed for correctness before it can be put into use, as shown in Fig. 4. demonstrates the process of diagnosing diabetes using data on which it was not trained.

Architecture diagram for diabetes prediction model. This model has five different modules.

25 VI. Evaluation

There are total 952 people, which include 372 women and 580 men, are chosen for study who are 18 years of age or older. A questionnaire that was selfprepared based on the required dataset for generating the model that might help to predict diabetes was given to the participants and is provided in Table 1, as base but based on this dataset further questionnaire may be prepared to calculate future prediction. The same tests were carried out on a different database called the PIMA Indian Diabetes database in order to confirm the model's validity. A sample dataset obtained using a questionnaire is shown in Figure 5. In this study, total 952 participants are selected aged 18 and above, out of which 580 are males and 372 are females. The participants were asked to answer a questionnaire shown in Table1 which was self-prepared based on the constraints that could lead to diabetes. In order to verify the validity of model same experiments were performed on another database called PIMA Indian Diabetes database shown in Table1. Figure 5 shows sample dataset collected through questionnaire [6].

26 VII. Expected Outcome of the Proposed Work

Following the use of several machine learning algorithms on the dataset, the accuracy results are as follows. The maximum accuracy is achieved with logistic regression (96%).

The following measure specified in the equation may be computed using the confusion matrices that were collected. True Negative (TN), False Positive (FP), False Negative (FN), and True Positive were the results of these matrices (TP). Because there are more nondiabetic cases than diabetic ones in both datasets, the TN is greater than the TP. As a result, all strategies provide worthwhile outcomes. The following measurements have been computed using the provided formulae in order to determine the precise accuracy of each approach. Only support vector machine algorithm has been analysed on PIMA dataset further more will be analysed on primary dataset with different algorithm.

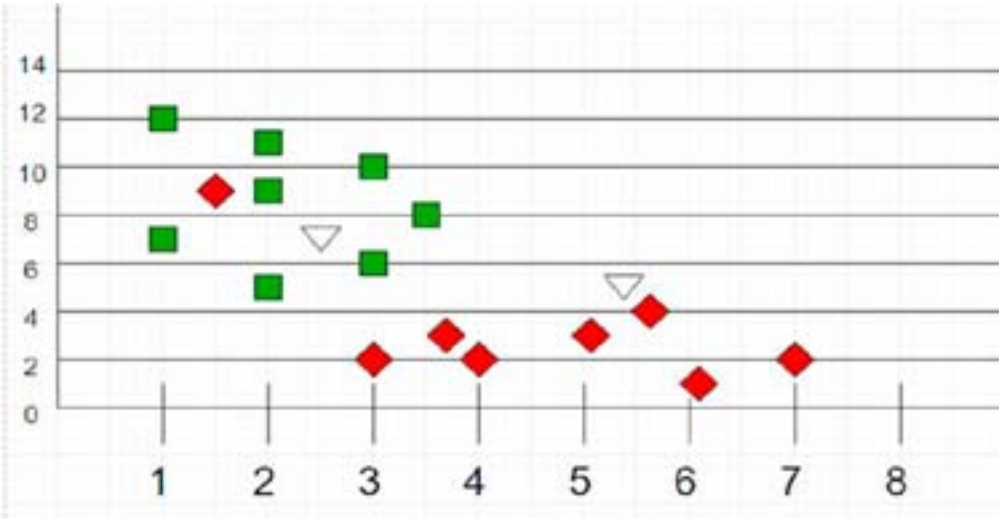
27 Accuracy Rate=

28 VIII. Conclusion

Because diabetes is a severe condition that can impair any part of the body if not detected early or treated promptly, the goal of this effort is to raise awareness of it and provide a framework to minimize risk of developing it. In the literature, only the PIMA dataset has been analyzed, whereas in this paper, the primary dataset has been collected and worked on machine learning technology that has involved to predict diabetic at any stage with certain parameters. Here, only the support vector machine algorithm has been used on PIMA as well as the primary dataset are used at the early phase of research to know the person is diabetic or undiabetic with 99% accuracy, and further many ML algorithms on PIMA and research survey primary datase.

1

Figure 1: Figure 1 :



2

Figure 2: Figure 2 :

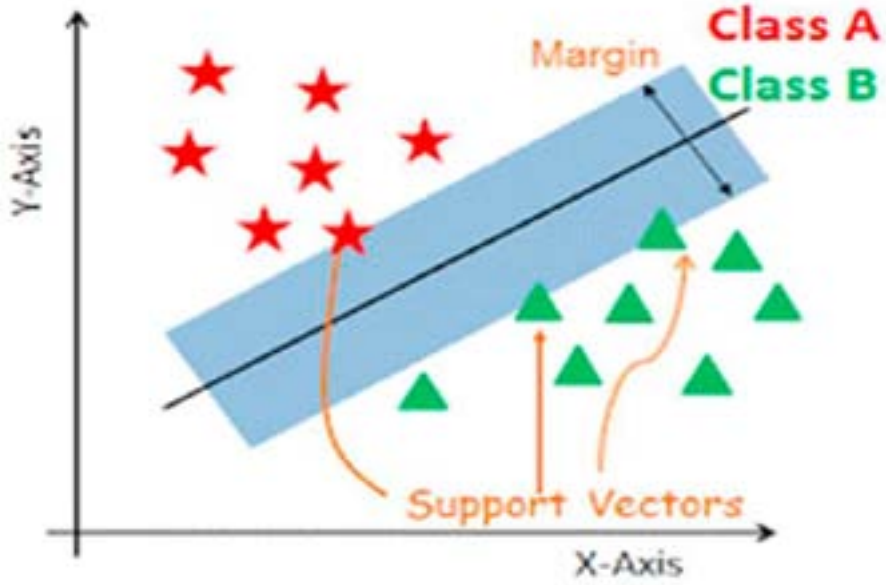
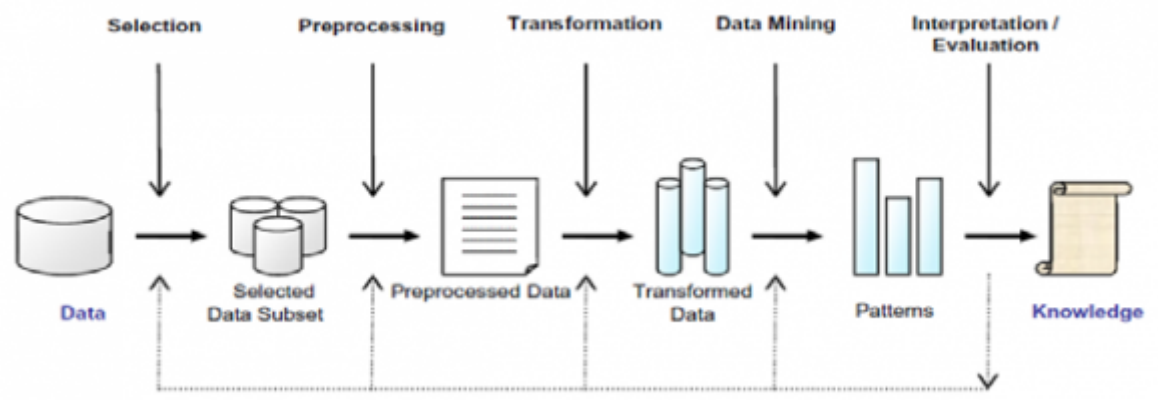


Figure 3:



3

Figure 4: Figure 3 :

1

Algorithm 1: Proposed Method Pseudo-Code

Figure 5: Table 1 :

2

Figure 6: Table 2 :

[Mayo Foundation for Medical Education and mResearch] , *Mayo Foundation for Medical Education and mResearch* MFMER.

[Antony ()] *A Comprehensive Unsupervised Framework for Chronic Kidney Disease Prediction*, Linta Antony . 10.1109/ACCESS.2021.3109168. 2021. (Digital Object Identifier)

[Rghioui ()] *A Smart Architecture for Diabetic Patient Monitoring Using Machine Learning Algorithms*, Amine Rghioui . 10.3390/healthcare8030348. 2020. (Healthcare 2020, 8, pp 348)

[Aminah (2019)] Ratna Aminah . *Diabetes Prediction System Based on Iridology Using Machine Learning" 6th Int. Conf. on Information Tech., Computer, and Electrical Engineering (ICITACEE)*, (Semarang, Indonesia) 2019. Sep 26-27, 2019. 978-1-7281-2689-0/19.

[Pal (2017)] 'Application of Machine Learning Algorithms on Diabetic Retinopathy'. Ridam Pal . *2nd IEEE International Conference On Recent Trends In Electronics Information & Communication Technology*, (India) 2017. 2017. May 19-20, 2017. 2017. IEEE. p. .

[Huang] 'Application of Machine Learning Techniques to Web-Based Intelligent Learning Diagnosis System'. Chenn-Jung Huang . HIS'04) 0-7695-2291-2/04. *Proceedings of the Fourth International Conference on Hybrid Intelligent Systems*, (the Fourth International Conference on Hybrid Intelligent Systems)

[Huang ()] 'Application of Machine Learning Techniques to Web-Based Intelligent Learning Diagnosis System'. Chenn-Jung Huang . *Fourth International Conference on Hybrid Intelligent Systems*, 0-7695-2291-2/04.

[Nabila Shahnaz Khan (2017)] 'Diabetes Predicting mHealth Application Using Machine Learning'. Nabila Shahnaz Khan . *2017 IEEE International WIE Conference on Electrical and Computer Engineering*, (Dehradun, India) 2017. 18-19 December 2017. WIECON-ECE

[Sonar ()] 'Diabetes Prediction Using Different Machine Learning Approaches'. Priyanka Sonar . *Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019) IEEE Xplore Part Number: CFP19K25-ART*, (the Third International Conference on Computing Methodologies and Communication (ICCMC 2019) IEEE Xplore Part Number: CFP19K25-ART) 2019. ISBN p. .

[Md and Hasan (2020)] *Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers*, Md , Hasan . 10.1109/ACCESS.2020.2989857. 2020. Received April 6, 2020, accepted April 18, 2020. April 23, 2020. (date of current version May 7, 2020. Digital Object Identifier)

[Mujumdar ()] 'Diabetic prediction using machine learning algorithm'. Aishwarya Mujumdar . *International conference on recent trends in advanced computing*, 2019. Scince Direct, 2019. 2019. Elsevier. 165 p. .

[Pranto] *Evaluating Machine Learning Methods for Predicting Diabetes among Female Patients in Bangladesh*, Badiuzzaman Pranto . 10.3390/info11080374. (Information 2020, 11, pp 374)

[Karatekin ()] 'Interpretable Machine Learning in Healthcare through Generalized Additive Model with Pairwise Interactions (GA2M): Predicting Severe Retinopathy of Prematurity*'. Tamer Karatekin . *Iternational conference on deep learning* 2019. 2019, 978-1-7281-2914-3/19.

[Hafiz Farooq and Al ()] 'Investigating Health-Related Features and Their Impact on the Prediction of Diabetes Using Machine Learning'. Ahmad Et Hafiz Farooq , Al . *Appl. Sci* 2021. 2021 p. 1173.

[Kavakiotis ()] 'Machine Learning and Data Mining Methods in Diabetes Research'. Ioannis Kavakiotis . *Computational and Structural Biotechnology Journal* 2017. 2017. 15 p. . (journal homepage: www.elsevier.com/locate/csbj)

[Ali ()] 'Machine Learning Techniques for Classification of Diabetes and Cardiovascular Diseases'. Berina Ali MECO . 978-1-5090-6742-8/17. *6th Mediterranean Conference On Embedded Computing*, (bar, montenegro) 2017. JUNE 2017. p. .

[Marsland and Book] Stephen Marsland , Book . *Machine Learning an algorithmic perspective*, CRC Press. (second edition)

[Ms and Sowjanya ()] *MobDBTest: A machine learning based system for predicting diabetes risk using mobile devices*, Ms , Sowjanya . 978-1-4799- 8047-5/15/$31.00_c. 2015. 2015. IEEE.

[Gensheng ()] 'Multi-Output Support Vector Machine Regression and Its Online Learning'. Hu Gensheng . 978-0-7695-3336-0/08. *2008 International Conference on Computer Science and Software Engineering*, 2008.

[Saroj Kumar Gupta ()] 'Noise Load Adaptive Filter Using Neural Network'. Saroj Kumar Gupta . 978-0-7695-3958-4/10 © 2010 IEEE. *2010 International Conference on Data Storage and Data Engineering*, 2010.

[Lejla ()] 'Predicting Diabetes in Healthy Population through Machine Learning'. Alic Lejla . *32 nd International Symposium on Computer-Based Medical Systems (CBMS)*, 2019. 2019. IEEE.

[Rajsavi and Anand (2018)] 'Predicting Mortality in Diabetic ICU Patients Using Machine Learning and Severity Indices'. S Rajsavi , Anand . PMC5961793. *AMIA Jt Summits Transl Sci Proc* 2018. 2018. 2018. 2018 May 18. p. . (Published online)

[Veena Vijayan (2015)] 'Prediction and Diagnosis of Diabetes Mellitus -A Machine Learning Approach'. V Veena Vijayan . *2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) | 10-12*, (Trivandrum) 2015. December 2015.

[Muhammad Azeem (2018)] 'Prediction of Diabetes Using Machine Learning Algorithms in Healthcare'. Sarwar Muhammad Azeem . *Proceedings of the 24th International Conference on Automation & Computing*, (the 24th International Conference on Automation & ComputingNewcastle upon Tyne, UK) 2018. September 2018. p. . Newcastle University

[Neha Prerna Tiggaa ()] 'Prediction of Type 2 Diabetes using Machine Learning Classification Methods'. Neha Prerna Tiggaa . *Internati-onal Conference on Computational Intelligence and Data Science (ICCIDS 2019)*, 2019. 2020. 167 p. .

[Bassam Farran (2019)] *Use of Non-invasive Parameters and Machine-Learning Algorithms for Predicting Future Risk of Type 2 Diabetes: A Retrospective Cohort Study of Health Data From Kuwait*, Bassam Farran . 10.3389/fendo.2019.00624. 2019. September 2019.