Journal of Artificial Intelligence, Machine Learning and Neural Network
Vol: 03, No. 01, Jan-June 2023
ISSN: 2799-1172, DOI: 10.55529/jaimlnn.33.27.34

Research Paper

Diabetes prediction with the help of machine learning

Akash Ashok Choutele*
*MSc Data Analytics, De Montfort University, United Kingdom.

Article Info

Article History:
Received: 08 February 2023
Revised: 20 April 2023
Accepted: 27 April 2023
Published: 13 June 2023

Keywords:
Algorithms
Diabetes
Insulin
Machine Learning
Prediction Model

ABSTRACT

"The first WHO Global Report on Diabetes was launched on World Health Day 7th April 2016 and was dedicated to Diabetes" (Roglic G. WHO Global report on diabetes, 2016). Diabetes was recognized as a serious medical condition even in ancient times, but it does not appear that clinicians or other physicians dealt with it frequently. Increases in the prevalence of this disease over the past few decades have had a chilling effect on human health and development. This research is all about diabetic patients and how diabetes can be detected early using data science and machine learning algorithms. Based on diagnostic measurements, the purpose of the research is to make a diagnostic determination as to whether a patient suffers from diabetes.

Corresponding Author:
Akash Ashok Choutele
MSc Data Analytics, De Montfort University, United Kingdom.
Email: P2689151@my365.dmu.ac.uk

Copyright © 2023 The Author(s). This is an open access article distributed under the Creative Commons Attribution License, (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

When your blood glucose, which is sometimes referred to as blood sugar, is too high, an individual is at risk of developing an illness known as diabetes [1]. When there is an excessive amount of glucose in the blood, this can, over time, lead to a variety of health issues, including kidney disease, nerve damage, eye difficulties, and heart disease. You may take measures to either avoid diabetes or manage it if you already have it.

By the year 2030 [2], [3] the United Kingdom Diabetes organization anticipates that there will be 5.5 million individuals living with diabetes in the country. Type 2 diabetes accounts for approximately 90 percent of all cases of diabetes. Type 1 diabetes affects around 8 percent of persons who have diabetes. Roughly 2 percent of diabetic patients have one of the less common forms of the disease.

Most diabetic patients discover they are diabetic when it is too late to be completely curable; as a result, they must live with it for the rest of their lives while adhering to a rigorous lifestyle; this is one of the world's greatest diabetes-related challenges. Diabetes [4] imposes so many limitations on human existence [5] that it is impossible to live life to the fullest and produces stress, which is another enemy of health. The idea of our research is to combine data science and machine learning algorithms to develop a prediction model that can diagnose diabetes accurately at an early stage and prevent individuals from developing diabetes.

1.1 Background

The past twenty years have been a remarkable time in terms of the number of medical advancements and discoveries that have been made. It was previously believed that specific components of the digestive system were the ones that were deficient in diabetics; however, it was later revealed that the pancreas [6] was the organ that was responsible for producing the chemical that was absent in diabetics. The researchers made some success in discovering the identity of this molecule, which they subsequently determined to be insulin. Frederick Banting and Charles Best are credited with being the individuals who discovered this hormone for the first time [7].

There are four types of diabetes namely type 1 diabetes [8], type 2 diabetes [9], gestational diabetes [10], and diabetes insipid us [11]. Type 1 diabetes is a chronic disease in which the pancreas stops or severely reduces its insulin production. A spike in urination frequency is accompanied by several negative health outcomes, including dehydration, hunger, rapid weight loss, and extreme exhaustion. In those with type 2 diabetes, elevated blood sugar results from insufficient insulin production. In pregnant women, diabetes is called gestational diabetes. It may not exhibit any symptoms at first. Symptoms might include increased urination, perspiration, and thirst as the condition deteriorates. This condition, called diabetes insipid us, is brought on by internal fluid imbalance. This causes dehydration and frequent urination [12].

Because of advances in medical research, there has been a significant increase in the number of people who can be successfully treated for diabetes. However, there is no cure for this illness; the symptoms can be managed with medication, but the condition itself cannot be cured entirely. Diabetes has its own unique set of dangers [2], [3]. A person can die from complications related to it if they fail in controlling it at an early stage with the assistance of medicines. The number of people who pass away as a direct result of this illness has been reduced as a direct result of advances in medical technology; nonetheless, the primary concern has been the growing diabetes population each year.

1.2 Proposed Work

The reduction of the ever-increasing number of diabetes patients across the United Kingdom and the rest of the world would be the principal goal of the work that is being suggested. The suggested technology will, in the long run, make it possible to eliminate stress from human existence and help avoid diabetes.

1.3 Aim

The research is majorly focused on how a machine-learning model is capable of both preventing people from developing diabetes as well as predicting who will get diabetes [13]. A person's glucose levels, insulin levels, blood sugar levels, and body mass index (BMI) are all risk factors for developing diabetes. The best model can help to prevent the increasing number of diabetic patients and prevent people from becoming victims of this disease [14]. The performance goals are measured based on the accuracy of the model. The proposed model will use algorithms such as [15] regression, decision trees, and neural

networks. By taking into consideration all the factors that contribute to diabetes, the model will be able to do so.

1.4 Objective

- To ensure that the findings will be accurate and will not be based on any random guesses, it is intended to use methods of machine learning to replicate human intelligence. This will ensure that the conclusions will not be based on any assumption. This will be of benefit to a great number of people in warding off diabetes.
- Mainly focused on developing a model using algorithms in which the major processing burden is to partition the data into training and testing sets and then execute operations on the test set to acquire high-level accuracy.
- To develop a new system that helps to predict symptoms of diabetes to prevent diabetes from happening.
- To save time and cost for people who might have diabetes at the same time maintaining the precision of results.

1.5 Rationale

- **Data Quality:** The quality of the information that is given in a machine learning model is directly related to the level of accuracy it achieves. The accuracy of the model's predictions will suffer if the data are characterized by high levels of noise or bias. Our development entails maintaining a high level of data integrity, which ultimately leads to a precision of a high grade.
- **Limited Generalizability:** The usefulness of machine learning models is directly proportional to their capacity to generalize data that has not been seen before. If the model is trained on a certain group, then it is possible that it will not be able to effectively forecast the risk of diabetes for another community. However, in our model, we are going to extend our model in such a manner that we will only examine the elements that lead to the development of diabetes. This will ensure that our model is accurate even when applied to data that has not yet been gathered.

2. METHODOLOGY

- **Data Collection:** This is the initial phase, and this will involve gathering information on both those with diabetes and those who do not have the condition. This data could include demographic information, medical history, lifestyle characteristics, and any other variables that are pertinent to the discussion.
- **Cleaning the Data:** For the data to be suitable for use in the training of a machine learning model, it must first be cleaned and pre-processed to get rid of any invalid or missing data.
- **Feature Selection:** The data are reviewed to determine which characteristics are most essential in accurately predicting the likelihood of developing diabetes. This step is referred to as feature selection. These characteristics, which are collectively referred to as features, are sent into the machine learning model as input.
- **Training the Model:** After the data has been pre-processed, several different methods (including decision trees, random forests, and support vector machines) are utilized in order to train the machine learning model on the data.
- **Evaluation:** After the model has been trained, it is tested on a different set of data to see how well it predicted the data. Besides accuracy, recall, and the F1 score, there are several other ways to reach this goal.
- **Deployment of the Model:** If the model performs well enough on the evaluation dataset, it could be used in a clinical setting to predict how likely it is that a certain patient will get diabetes.
- **Maintenance of the Model:** To make sure that the model stays accurate over time, it should be checked regularly and changed as needed.

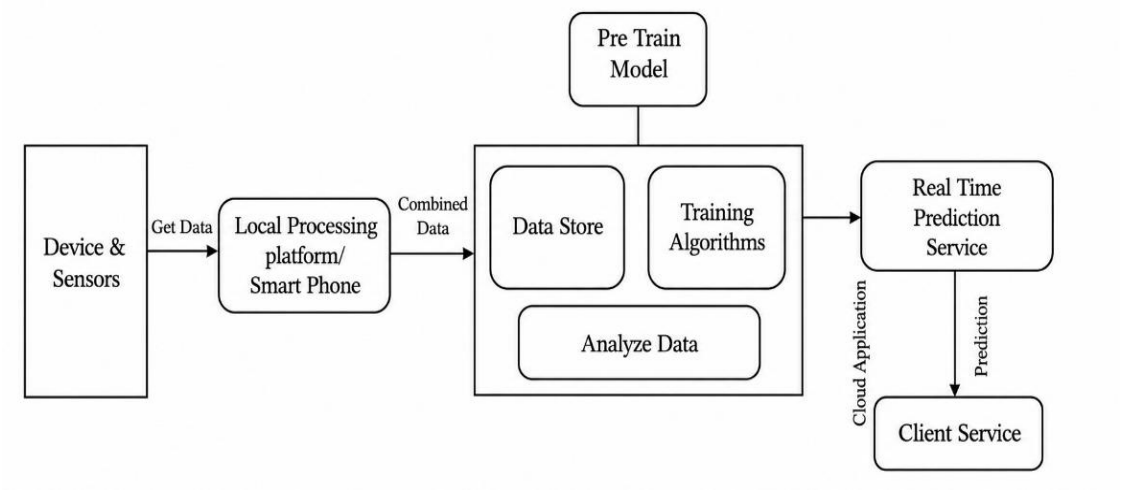


Figure 1. Framework of Diabetes Prediction System

2.1 Work Package

2.1.1 Work Package 1: Study of the Relevant Literature and the Fundamental Insights Gained

For the researcher to become more familiar with the technologies that are going to be included in this model, he is going to do some further study on those technologies (machine learning, artificial intelligence, programming languages, and their libraries). In addition to this, the researcher can do their own investigation into the development of the project at their own discretion. Institutional advisers who are employed in educational establishments at a higher level can participate and give guidelines if they so desire.

2.1.2 Work Package 2: Obtaining Appropriate Samples of Data

It is essential to have a clear understanding of the goals we have set for our model and what we hope to accomplish with it. This will help us figure out what kind of information we need to gather and how much of it we will need. The data of normal people has been taken into consideration, but the purity of the data is the most important aspect.

2.1.3 Work Package 3: Development

The relevant data sources should be successfully identified. Develop a way to organize the data so that it can be used to teach a machine learning model. Cleaning and pre-processing the data to remove any errors or inconsistencies is important, as is formatting the data so that it is easy to use for training and testing the model.

2.1.4 Work Package 4: Implementation

When examining the effectiveness of the model on data that is unknown, it is necessary to keep the training set distinct from the test set. The training set is what's used to teach the model, while the testing set is what's utilized to judge how well it does its job.

2.1.5 Work Package 5: Training and Evaluation of the Model

After all the data has been collected and all the necessary steps have been taken, it is used in several ways to train a machine-learning model. The model can be made more accurate by using the testing set to see how well it works and then making any changes that are needed.

2.2 Professional, Legal and Ethical Issues

- **Bias in the Data:** The accuracy of machine learning models is directly proportional to the quality of the data they are trained on. If the data that is used to train the model has a bias, then it is probable that the model will also have a bias. It is essential to give due consideration to the possible sources of bias in the data and to take measures to eliminate or reduce those sources.

- **Bias in Algorithms:** Machine learning algorithms may also be prejudiced, which can lead to erroneous or unjust findings. This can be prevented by avoiding certain practices. It is essential to conduct a thorough analysis of the performance of the model to check that it does not include any biases and to take measures to compensate for any biases that may be discovered.
- **Transparency:** It is necessary to be upfront about the methods and data used to construct the model, as well as the limits of the model and its possible influence on individuals. It is also important to be transparent about the constraints of the model itself.

2.3 Relevance to Beneficiaries

Using machine learning to make a diabetes prediction model can be helpful for people who may be at risk of getting diabetes because it can help them take steps to prevent or manage the disease. For example, if the prediction model can accurately detect people who are likely to get diabetes, those people may be able to take steps to lower their chance of developing the disease, such as making changes to their lifestyle (like improving their diet and getting more exercise) or seeing a doctor. Also, if the prediction model can accurately predict which people are most likely to get diabetes, it may be helpful for healthcare providers to use the model to target interventions for those people. This could make diabetes prevention and management efforts more effective.

2.4 Management Plan

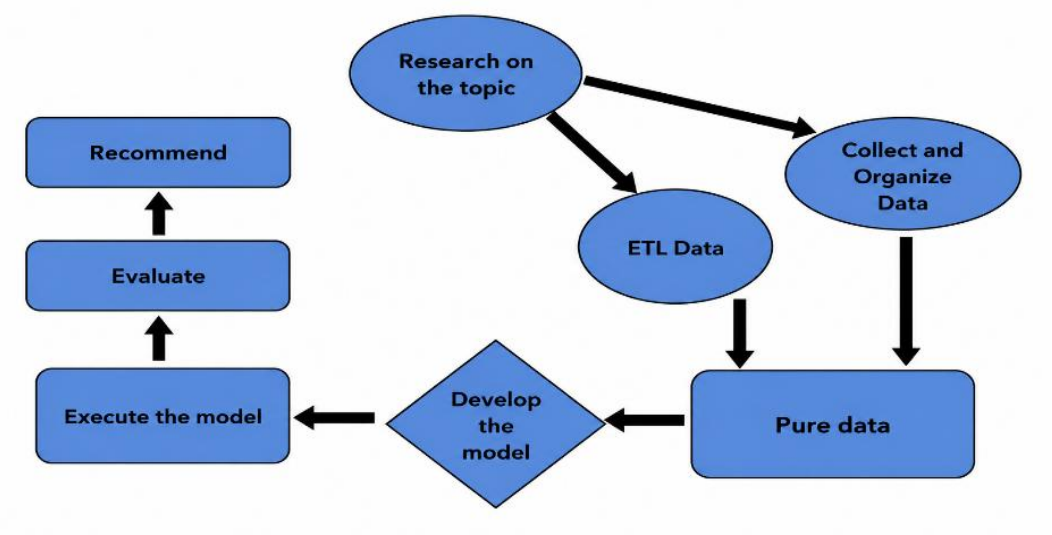


Figure 2. Management Plan

2.5 Justification of Resources

2.5.1 Personnel

The responsibilities of a PhD researcher include expanding on the model, gathering information about the current system, and attempting to make improvements in all these areas. In addition to this, the researcher is responsible for ensuring that the appointed supervisor is kept up to date with the current study.

2.5.2 Access to Resources

Since the researcher will be in charge of doing all the research according to the instructions given by the supervisor, the researcher will have to be a part of a higher institution in the end. When it comes to hiring new people for the research project, there won't be any extra costs. The only thing that will be required is for the researcher to make periodic visits to the institution on a regular basis for the project to continue its progress.

2.5.3 Hardware and Software

When we talk about building an algorithm for machine learning, the researcher won't need any more hardware. But the system should have at least the hardware that is needed to build an algorithm for machine learning. SAS Enterprise Miner is going to be the piece of software that is implemented during the process of the model's creation. To use the app, you need a valid SAS license, which you can get from the institution.

2.6 Gantt Char

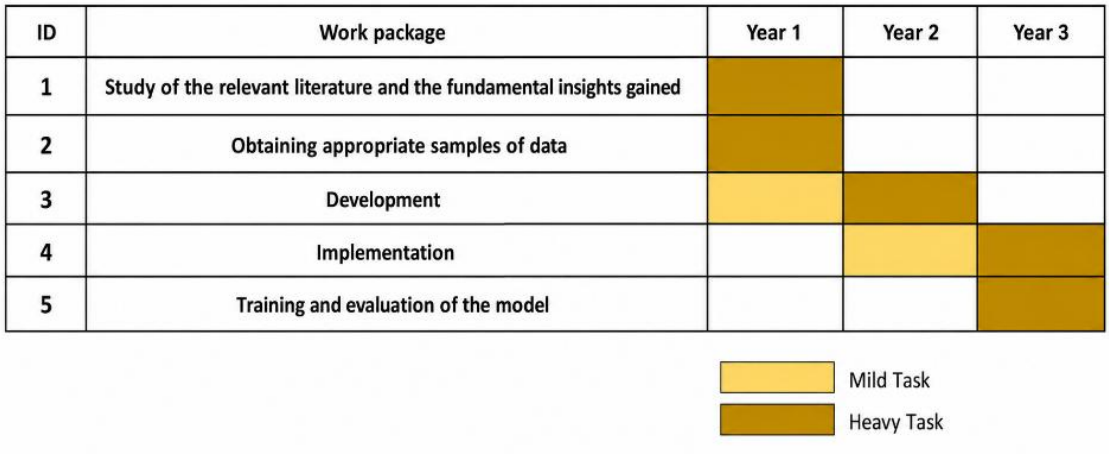


Figure 3. Gantt Char

Acknowledgments

The authors have no specific acknowledgments to make for this research.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Akash Ashok Choutele	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	

C: **C**onceptualization
M: **M**ethodology
So: **So**ftware
Va: **Va**lidation
Fo: **Fo**rmal analysis
I: **I**nvestigation
R: **R**esources
D: **D**ata Curation
O: Writing- **O**riginal Draft
E: Writing- Review& **E**diting
Vi: **Vi**sualization
Su: **Su**pervision
P: **P**roject administration
Fu: **Fu**nding acquisition

Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Informed Consent

All participants were informed about the purpose of the study, and their voluntary consent was obtained prior to data collection.

Ethical Approval

Not Applicable.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] G. R. Al, 'Who global report on diabetes: A summary', International Journal of Noncommunicable Diseases, 2016

[2] J. Chandra, Ed., Role of apoptosis in pancreatic beta-cell death in diabetes. 2001. doi.org/10.2337/diabetes.50.2007.S44

[3] J. Chandra, Ed., Apoptosis in pancreatic beta-cell death in diabetes. 2001. doi.org/10.2337/diabetes.50.2007.S44

[4] Anon., 2014. Diabetes UK. [Online] [Accessed 20 04 2023]

[5] Juan Oliva, A. F.-B. a. A. H., 2012. Health-related quality of life in diabetic people with different vascular risk. BMC public health, 12(1), pp. 1-8. doi.org/10.1186/1471-2458-12-812

[6] L. St-Onge, 'Pancreas development and diabetes', Current opinion in genetics & development, vol. 9, pp. 295-300, 1999. doi.org/10.1016/S0959-437X(99)80044-6

[7] M. Ahmed, 'History of diabetes mellitus', Saudi medical journal, vol. 23, no. 4. doi.org/10.15537/1658-3175.1664

[8] G. S. Eisenbarth, 'Type i diabetes mellitus', New England journal of medicine, vol. 314, no. 21, pp. 1360-1986, 1986 doi.org/10.1056/NEJM198605223142106

[9] E. F. L. G. R. R. H. W. H. H. Ralph A Defronzo, 'Type 2 diabetes mellitus', Nature reviews Disease primers, vol. 1, no. 1, pp. 1–22, 2015.

[10] P. C. C. Z. G. D. E. R. M. A. P. D. H David Mcintyre, 'Gestational diabetes mellitus', Nature reviews Disease primers, vol. 5, no. 1, pp. 1-19, 2019. doi.org/10.1038/s41572-019-0104-1

[11] M. Maghnie, 'Diabetes insipidus', Hormone Research in Paediatrics, pp. 42-54, 2003. doi.org/10.1159/000067844

[12] Wareham, N. G. F. a. N. J., 2010. Epidemiology of diabetes. Medicine, 38(11), pp. 602-606. doi.org/10.1016/j.mpmed.2010.08.007

[13] M. A. Jahan, 'Prediction of onset diabetes using machine learning techniques', International Journal of Computer Applications, vol. 180, no. 5, pp. 7-11, 2017. doi.org/10.5120/ijca2017916020

[14] A. C. A. D. Gupta, Recent developments in machine learning and data analytics. 2019.

[15] [Online]. Available: https://www.ijncd.org/article.asp?issn=2468-8827;year=2016;volume=1;issue=1;spage=3;epage=8;aulast=Roglic;type=3[Accessed. [Accessed: 12-May-2026].

How to Cite: Akash Ashok Choutele, (2023). Diabetes prediction with the help of machine learning. Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN), 3(1), 138–144. https://doi.org/10.55529/jaimlnn.33.27.34

BIOGRAPHY OF AUTHOR

	**Akash Ashok Choutele,** is currently pursuing an MSc in Data Analytics at De Montfort University. His academic interests include data science, machine learning, artificial intelligence, big data analytics, and predictive modelling. He is passionate about applying analytical techniques to solve real-world problems and developing intelligent data-driven solutions. His research focuses on emerging technologies, data processing methodologies, and advanced analytics applications across various domains. Email: P2689151@my365.dmu.ac.uk