

# Machine Learning Model for Prediction of Diabetes Mellitus

Manjiri Mahadev Mastoli, Urmila R. Pol, R.D.Patil

**Abstract:** Today world is extensively affected by endocrine disease Diabetes Mellitus which is commonly known as diabetes. There is a need for an effective model which can predict diabetes and its types at the early stages with accuracy. To improve the accuracy of prediction and to achieve better efficiency, a new Machine Learning based Model (MLM) is proposed. This Machine Learning Model (MLM) has ability to predict the diabetes and its categories as type 1, type 2 and Gestational diabetic with which the patient is suffering from. The proposed Machine Learning Model is innovative for diagnosis of diabetes is more accurate as compared to other existing approaches. This is a novel method from which one can combine power of an expert system with the machine learning environment.

**Keywords :** Artificial intelligent (AI), Machine Learning Model(MLM), Knowledge-base, Diabetes mellitus (DM), Decision Tree Algorithm..

## I. INTRODUCTION

Artificial intelligence will have an outstanding impact on healthcare domain. Data science and machine learning can either be used for analysis or prediction in healthcare. Healthcare is a revolutionary and promising industry for implementing the data science solutions for Medical Images Analysis, Drug Discovery, Genetics Research and Predictive Medicine. As the technology becomes more common, in case of healthcare domain, it is expected that artificial intelligence will help prognosis, diagnose a wide area of disease. This emulates the decision making ability of a human expert. According to the International Diabetes Federation (IDF) it is estimated that 415 million people are living with diabetes in the world and India is the second largest country with 69 million diabetes people. These numbers are the motivation for the this respective research, where there is need of computerized system which will predict the diabetes in the early stages. Diabetes is a disorder of carbohydrate metabolism. This is characterized by an impaired ability of the body to produce insulin or respond insulin and thereby maintain proper levels of glucose (sugar) in the blood. Diabetes broadly categorized into three main types as type 1 diabetes, type 2 diabetes, and gestational diabetes. In the case of type 1 diabetes, destruction of the islets of Langerhans of the pancreas happened because of autoimmune systems, this pancreas is source of insulin. From the total population of

diabetes people it accounts only 5-10%. Type 2 diabetes is very common and accounts 90-95% of the total diabetes population, it's strongly associated with obesity and is a result of resistance and deficiency of insulin. Gestational diabetes is a temporary condition in which blood glucose (sugar) levels increase during pregnancy state and it will return to normal conditions after delivery. The women, which are suffered from gestational diabetes has future risk of being diabetic[1].

In this research article authors have proposed a novel method for prediction of diabetes and its types. This predictive research will help an individual for knowing their own risk of being diabetic and the respective action should be done for avoiding such health care problem. The purpose of this study is to design a Machine Learning Model for diabetes diagnosis. Acquiring right symptoms of the patient is an important factor in the application of rules. These knowledge combinations determine whether a person is diabetes patient, along with its types such as type1 diabetes, type 2 diabetes and gestational diabetes. A Machine Learning Model was tested on 10 patients. It has achieved exact results as doctors. A Machine Learning Model that researcher have designed can be used effectively and efficiently to diagnoses for diabetes types. It helps patients in undeveloped countries where the number of doctors is not enough. This intelligent Machine Learning Model aims to reduce the dependence on doctors. It will help both doctors and patients to make more accurate and quicker decisions[2].

The lateral sections of this works are designed as follows II- Description of The Dataset, III- Designing The Knowledge Base for Machine Learning Model, IV- Designing The Knowledge Base For Machine Learning Model, V- Machine Learning Based Model VI. Results and Discussion VII. Conclusion and at last VII- References

## II. DESCRIPTION OF THE DATASET

This dataset contains details of symptoms and diabetes type associated with the values given to each symptom. Total 64 symptoms are considered which will categorize diabetes into three different types as Type-1 Diabetes, Type-2 Diabetes, and Gestational Diabetes. Therefore dataset contains total 65 columns in which 64 are symptoms and last one with prediction value as PredictionOfDiabetes with 3 different values as Type-1 Diabetes, Type-2 Diabetes, and Gestational Diabetes. The researcher has designed the dataset such a way that each row in the dataset represents one test case for each diabetes type. These can be rule for the model and set of rules are termed as knowledge base. Therefore the dataset used for this Machine Learning Model itself is knowledge base for Machine Learning Model.

Total 120 records are entered into the dataset or on other hand those are total rule for the Machine Learning Model.

Revised Manuscript Received on January 15, 2020

\* Correspondence Author

Manjiri Mahadev Mastoli\*, Research Scholar, Department of Computer Science Shivaji University Kolhapur, Maharashtra,  
Email: it.manjiri@gmail.com

Dr.Urmila R. Pol, Assistant Professor, Department of Computer Science, Shivaji University Kolhapur, Maharashtra, India

Mr.Rahul D.Patil, Quality Assurance Engineer, Menon Bearing Ltd, Kolhapur, Maharashtra

## Machine Learning Model for Prediction of Diabetes Mellitus

Designing the rules and selection of rule for Machine Learning Model is complicated task for which one can need expert knowledge which will enhance prediction accuracy.

### III. DESIGNING THE KNOWLEDGE BASE FOR MACHINE LEARNING MODEL

The table 1 contains list of major symptoms of diabetic patients. Those are classified on the basis of their types and table has been designed with the help of values. These values are zero and one in the case symptoms which will take binary values and the series of numbers when symptoms are varies from more than two values. In case of symptoms which are having two categories either True OR False assigned binary values 0 and 1 (table 2). For example FamilyHistory represented by the binary value 0 and 1. In the case of multivalued symptoms it will take series of number. Age with three categories Young\_Adult\_Old represented by the number with values 11-12-13, Obesity with three categories Low\_Normal\_Obese represented by the number with values 21-22-23, Hypertension with four categories Normal\_Elevatal\_High\_VeryHigh represented by the number with values 3-4-5-6, HDL\_Cholesterol with three categories Low\_Medium\_High represented by the number with values 1-2-3, Triglyceride with four categories Normal\_BoarderLine\_High\_VeryHigh represented by the number with values 2-3-4-5, This table 1 is prepared on the basis of primary data collection and secondary data collection such as physicians data, books, Internet, medical journals [4], diabetes patients etc. The Table 1 is considered as one of the three input for the system. In the same way with different combination of symptoms for each type of diabetes are considered while designing rule in CSV file.

#### Dataset as Knowledge base:

The table 1 contains list of major symptoms of diabetic patients. Those are classified on the basis of their types and table has been designed with the help of values. These values are zero and one in the case symptoms which will take binary values and the series of numbers when symptoms are varies from more than two values. Columns represented as A-Type1Diabetes, B- Type2Diabetes, C- Gestational Diabetes.

**Table 1: Dataset as Knowledge Base**

Sr. No.	Symptoms	A	B	C
1	Family_History	0	1	0
2	Age	11	12	12
3	Obesity	22	23	21
4	Previous_IFG_IGT	0	1	0
5	Hypertension	0	5	5
6	HDL_Cholesterol	0	3	0
7	Triglyceride	0	5	0
8	Inceased_Thirst	1	1	0
9	Increased_Urinate	1	1	0
10	Increased_Appetite	1	1	0
11	Weight_Variation	0	1	0
12	Impaired_Vision	0	1	0

13	Tiredness	1	1	0
14	Impatience	0	1	0
15	Infection	0	1	0
16	Itchy_Skin	0	0	0
17	Depression_Stress	0	0	0
18	Tingling_Sensation	0	0	0
19	Fruity_Breath_Odour	1	0	0
20	Bed_Wetting	1	0	0
21	Slow_Healing_Wound	0	1	0
22	FamilyHis_Pregnancy	0	0	1
23	Previous_Pregnancy	0	0	1
24	BabyOver_9Pd_PrePreg	0	0	1
25	Sleeplessness	1	1	0
26	Trembling	1	1	0
27	Sweating	1	1	0
28	Anxiety	1	1	0
29	Confusion	1	1	0
30	Weakness	1	0	0
31	Mood_Swings	1	0	0
32	Nausea	1	0	0
33	Vomiting	1	0	0
34	Dry_Skin	0	1	0
35	Aches&Pains	0	1	0
36	Recurresnt_fungal_infectn	0	1	0
37	Nightmares	1	1	0
38	Seizures	1	1	0
39	Sadness	1	1	0
40	Unconsciousness	1	1	0
41	Numbness	1	1	0
42	VaginalMycoticInfectn	1	1	1
43	Rapid_Heart_Beat	0	1	0
44	Recurring_Gum_Infe	0	0	0
45	Impotency	1	1	0
46	high blood Pressure	0	0	1
47	Sleep_Walking	1	1	0
48	Makeg_unusual_noises	1	1	0
49	Leg_Cramps	1	1	0
50	Slurred_Speech	1	1	0
51	Flushed_face	1	1	0
52	Pale_Skin	1	0	0
53	LossOfMenstruation	1	1	0
54	Stomach_Pain	1	1	0
55	Deep_Breathing	1	1	0
56	Areas_Darked_Skin	0	1	0
57	Difficult_Concentrating	1	1	0

58	Dehydration	1	1	0
59	LackOfCoordination	1	1	0
60	Hist_Heart_Stroke	0	0	0
61	Poly_Ovary_Syndrome	0	0	0
62	LowbloodSugar_NewbornBaby	0	0	1
63	WaistSize02cmM88cmF	0	0	0
64	WaistHipRatio.9M.85F	0	0	0

Following table 2 contains the number representation for the symptoms, which are used for enhancing analysis.

**Table 2: Number Representation for Symptoms**

Boolean Value Symptoms			Multi Value Symptoms		
Family History	True	1	Age	Young	1
	False	0		Adult	2
				Old	3
Obesity	Low	21	Triglyceride	Normal	2
	Normal	22		Boarder Line	3
	Obese	23		High	4
Hypertension	Normal	3	HDL_Cholesterol	Very High	5
	Elevatal	4		Low	1
	High	5		Medium	2
	Very High	6		High	3

#### IV. PROPOSED METHODOLOGY

The secondary data is collected from books, internet, medical journals and primary data collection done by providing questionnaires to physicians about symptoms affecting diabetes. The rules for this Machine Learning Model were designed after data analysis and expert suggestion. These rules are entered into CSV file which is dataset for the machine learning model has been used for further processing. Google Colaboratory known as Colab is a cloud service based on Jupyter Notebooks is used for execution of machine learning algorithm [3][4] for this Machine Learning Model. This dataset contains the knowledge of the Machine Learning Model in the form of rows with numbers which are symptoms of diabetes disease. This respective Machine Learning Model is predicting the type of diabetes which is patients suffering from. Here dataset with values which are rules and knowledge base of the Machine Learning Model. Some of the rules for including three categories of diabetes diagnosis that can be interpreted are as follows.

*Family\_History 1, Age 11, Obesity 22, Previous\_IFG\_IGT 0, Hypertension 0, HDL\_Cholesterol 0, Triglyceride 0, Inceased\_Thirst 1, Increased\_Urinate 1, Increased\_Appetite*

*1, Weight\_Variation 0, Impaired\_Vision 0, Tiredness 1, Impatience 0, Infection 0, Itchy\_Skin 0, Depression\_Stress 0, Tingling\_Sensation 0, Fruity\_Breath\_Odour 1, Bed\_Wetting 1, Slow\_Healing\_Wound 0, FamilyHis\_Pregnancy 0, Previous\_Pregnancy 0, BabyOver\_9Pd\_PrePreg 0, Sleeplessness 1, Trembling 1, Sweating 1, Anxiety 1, Confusion 1, Weakness 1, Mood\_Swings 1, Nausea 1, Vomiting 1, Dry\_Skin 0, Aches&Pains 0, Recurresnt\_fungal\_infectn 0, Nightmares 1, Seizures 1, Sadness 1, Unconsciousness 1, Numbness 1, VaginalMycoticInfectn 1, Rapid\_Heart\_Beatm 0, Recurring\_Gum\_Infe 0, Impotency 1, high blood Pressure 0, Sleep\_Walking 1, Makeg\_unusual\_noises 1, Leg\_Cramps 1, Slurred\_Speech 1, Flushed\_face 1, Pale\_Skin 1, LossOfMenstruation 1, Stomach\_Pain 1, Deep\_Breathing 1, Areas\_Darked\_Skin 0, Difficult\_Concentrating 1, Dehydration 1, LackOfCoordination 1, Hist\_Heart\_Stroke 0, Poly\_Ovary\_Syndrome 0, LowbloodSugar\_NewbornBaby 0, WaistSize02cmM88cmF 0, WaistHipRatio.9M.85F 0. Last Column as type 1 diabetes.*

*Family\_History 1, Age 12, Obesity 23, Previous\_IFG\_IGT 1, Hypertension 5, HDL\_Cholesterol 3, Triglyceride 5, Inceased\_Thirst 1, Increased\_Urinate 1, Increased\_Appetite 1, Weight\_Variation 1, Impaired\_Vision 1, Tiredness 1, Impatience 1, Infection 1, Itchy\_Skin 0, Depression\_Stress 0, Tingling\_Sensation 0, Fruity\_Breath\_Odour 0, Bed\_Wetting 0, Slow\_Healing\_Wound 1, FamilyHis\_Pregnancy 0, Previous\_Pregnancy 0, BabyOver\_9Pd\_PrePreg 0, Sleeplessness 1, Trembling 1, Sweating 1, Anxiety 1, Confusion 1, Weakness 0, Mood\_Swings 0, Nausea 0, Vomiting 0, Dry\_Skin 1, Aches&Pains 1, Recurresnt\_fungal\_infectn 1, Nightmares 1, Seizures 1, Sadness 1, Unconsciousness 1, Numbness 1, VaginalMycoticInfectn 1, Rapid\_Heart\_Beatm 1, Recurring\_Gum\_Infe 0, Impotency 1, high blood Pressure 0, Sleep\_Walking 1, Makeg\_unusual\_noises 1, Leg\_Cramps 1, Slurred\_Speech 1, Flushed\_face 1, Pale\_Skin 0, LossOfMenstruation 1, Stomach\_Pain 1, Deep\_Breathing 1, Areas\_Darked\_Skin 1, Difficult\_Concentrating 1, Dehydration 1, LackOfCoordination 1, Hist\_Heart\_Stroke 0, Poly\_Ovary\_Syndrome 0, LowbloodSugar\_NewbornBaby 0, WaistSize02cmM88cmF 0, WaistHipRatio.9M.85F 0. Last Column as type 2 diabetes.*

The CSV file with first 64 columns with value of symptoms for rules and prediction value as last column in the dataset as type of diabetes is saved in knowledge base of the Machine Learning Model. After executing the machine learning algorithm on this Machine learning based expert Sytem will Predict the diabetes type of entered symptoms on command line

#### V. MACHINE LEARNING BASED MODEL

Decision tree classifier is a systematic approach for multiclass classification. It poses a set of questions to the dataset. The decision tree classification algorithm can be visualized on a binary tree. On the root and each of the internal nodes, a question is posed and the data on that node is further split into separate records that have different characteristics.





# Machine Learning Model for Prediction of Diabetes Mellitus

making the classification diabetes types. This model predicts the diabetes type of a target symptoms by learning simple decision rules inferred from the data features. This is the generalized method from which one can convert expert systems to machine learning platform. Regular and good research is always in demand, which help for more accurate prediction in area of Healthcare.

## ACKNOWLEDGMENT

The authors thank to Kadsiddheshwar Swamiji and team of physician at Siddhagiri Hospital & Research Center, Kaneri, Kolhapur for their support and Dr.R.K.Kamat, coordinator, Department of Computer science, Shivaji university for his assistance and motivation. Department of Computer science, Shivaji university for their infrastructure support.

## REFERENCES

1. Joshi SR, Parikh RM. India - diabetes capital of the world: now heading towards hypertension. J Assoc Physicians India. 2007;55:323-4
2. Kumar A, Goel MK, Jain RB, Khanna P, Chaudhary V. India towards diabetes control: Key issues. Australas Med J. 2013;6(10):524-31.
3. Colaboratory: Frequently Asked Questions, Jun. 2018, [online] Available: <https://research.google.com/colaboratory/faq.html>.
4. Pessoa, Tiago & Medeiros, Raul & Nepomuceno, Thiago & Bian, Gui-Bin & Albuquerque, V.H.C. & Filho, Pedro Pedrosa. (2018). Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications. IEEE Access. PP. 1-1. 10.1109/ACCESS.2018.2874767
5. Oniiri Ernest E , Ndukwe Victor U , Igwe Nkechinyere , Olise Saviour C. "Simulation of an Expert System for Diabetes Diagnosis" International Journal of Advanced Computing, ISSN:2051-0845, Vol.48, Issue.1 1614-1621
6. Dilip Kumar Choubey, Sanchita Paul, Vinay Kumar Dhandhenia "Rule based diagnosis system for diabetes" Biomedical Research 2017; 28 (12): 5196-5209
7. Tawfik Saeed Zeki, Mohammad V. Malakooti, Yousef Ataeipoor, S.Talayeh Tabibi , " An Expert System for Diabetes Diagnosis", American Academic & Scholarly Research Journal Vol. 4, No. 5, Sept 2012
8. Anindito Yoga Pratama, Dewi Agushinta R., and Remi Senjaya , "Design Of Mobile Expert System For Diabetes Risk Diagnosis And Information", Journal of Information Systems, Volume 9, Issue 1, April 2013, 32-36
9. Hanslal Prajapati<sup>1</sup>, Anurag Jain<sup>2</sup>, Sanjay Kumar Pal<sup>3</sup> " An Enhance Expert System for Diagnosis of Diabetes using Fuzzy Rules over PIMA Dataset", International Journal of Advance Engineering and Research Development (IAERD) Volume 4, Issue 9, September-2017, e-ISSN: 2348 - 4470, print-ISSN: 2348-6406
10. Manjiri M. Mastoli, R.V.Kulkarni, Urmila R. Pol,(2019) "Fuzzy Expert System for Prediction of Indian General Election Results", International Journal of Computer Sciences and Engineering, Vol.-7, Issue-5, May 2019, E-ISSN: 2347-2693
11. Meysam Rahmani Katigari, Haleh Ayatollahi, Mojtaba Malek, and Mehran Kamkar Haghghi, "Fuzzy expert system for diagnosing diabetic neuropathy", World J Diabetes. 2017 Feb 15; 8(2): 80-88.
12. A.D.Dhivya , A.Felix , "A Fuzzy Rule based Expert System for T2DM Diagnosis" International Journal of Engineering & Technology, 7 (4.10) (2018) 432-435
13. Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. (2012) Reporting and Methods in Clinical Prediction Research: A Systematic Review. PLoS Med 9(5): e1001221. <https://doi.org/10.1371/journal.pmed.1001221>

## AUTHORS PROFILE



**Ms. Manjiri Mahadev Mastoli** pursued Master of Computer Application from CSIBER, Shivaji University, Kolhapur, Maharashtra. She is currently pursuing Ph.D. and currently working as research Scholar in Department of Computer Sciences, Shivaji

University, Kolhapur, Maharashtra. Her main research work focuses on Artificial Intelligence and Machine Learning, Data Science, She has 5 years of teaching experience and 5 years of Research Experience..



**Dr. Urmila Pol** pursued Bachelor of Science from Shivaji University and Master of Computer Application from CSIBER, Shivaji University, Kolhapur, Maharashtra. She is currently working as Assistant Professor in Department of Computer Sciences, Shivaji University, Kolhapur, Maharashtra. She is Member of International Association of Engineer. Currently, she is guiding four Ph.D and one M.Phil student. Her main research work focuses on Artificial Intelligence, Data Science, Big Data Analytics, LMS and Open Source Technologies, She has 20 years of teaching experience and 15 years of Research Experience.



**Mr. R.D. Patil** pursued Bachelor of Engineering Shivaji University, Kolhapur and Master of Business Administration from IGNOU. He is currently currently working as Quality Assurance Engineer at Menon Bearing Ltd, Kolhapur, Maharashtra. His main research work focuses on Operational Research, Artificial Intelligence and Machine Learning. He has 8 years of Quality Assurance experience and 4 years of Research Experience.