

A Study on Techniques for Diabetes Prediction model using Machine Learning

Naresh Kumar^{1*}, Dr. Mukesh Kumar²

¹ Research Scholar, Sunrise University, Alwar Rajasthan

² Associate Professor, Sunrise University, Alwar Rajasthan

Abstract - The aim of this study is to the using Artificial Intelligence and machine learning in healthcare is a recent research field. In this area of research, Artificial Intelligence and machine learning for healthcare such as Treatment efficiency, Healthcare management, Fraud and abuse detection, analyzing the behavior of patients, Customer relationship management, etc. is already done. This research is intended to propose an Artificial Intelligence and machine learning model for healthcare, finding the qualitative attributes of patients, preparation of dataset for mining and diagnosis of diabetes by Artificial Intelligence and machine learning to improve the treatment effectiveness which will save the time and enhance the same for faster treatment and analysis. The researcher has studied the existing system and its problem is being identified. The researcher has provided an appropriate Artificial Intelligence and machine learning model for diabetic healthcare. Medical field faces new difficulties like new diseases, cost, new therapeutics, fast decisions etc. Since medical decision making demands utmost accuracy of diagnosis, it is a tedious, demanding and challenging task for physicians. An automated system which helps in disease diagnosis, prognosis, which will benefit the medical. This has attracted researchers to design medical decision support systems with utmost accuracy. Numerous data science applications are found in the clinical related regions, for example, Medical gadget industry, Hospital Management and Pharmaceutical Manufacturing. Data science realistic in social insurance industry assumes a significant job in identification and analysis of the diseases. The individual examination will locate the most helpful and concealed information from the dataset, and structure the prescient model. This is the reason behind the use of data science. Famously data science has tremendous application regions in the human services area. Data science can either be utilized for analysis, for example, pattern-identification, testing of hypothesis, risk evaluation or forecast, for example, AI models that make appropriate expectations that are the probability of an occasion happening later on, in light of known information factors.

Keywords - Techniques, Diabetes Prediction Model, Machine Learning, Artificial Intelligence, Healthcare management

-----X-----

INTRODUCTION

The world has now moved into the era of Data Science, and tools that have traditionally been applied to another domain are now being considered in health care. These data come from a variety of sources, such as routine, systematic data input, reports, claims data, surveys, and data gleaned from biometric monitoring. Researchers must discover the finest data science strategy, deep learning and machine learning methodologies available to apply to these datasets given the enormous abundance of data that is being made available. Insulin resistance or insulin shortage can cause diabetes, which is defined by an excess of glucose in the bloodstream [ADF 2014]. It is by far one of the most significant global public health challenges of our day. The International Diabetes Federation (IDF) estimates

that by 2030, the number of individuals with diabetes will rise from 350 million to 550 million.[1]. Out of every hundred, eighty per cent of diabetes-related deaths occur in underdeveloped countries like India. Numerous studies have shown that diabetes is on the rise in both developing and developed countries around the world. This has necessitated the development of a scientifically sound and technologically advanced treatment for diabetes [2]. Diabetes mellitus also has a significant financial impact on the global healthcare system and the general public's global budget. Indirect and direct expenses associated to productivity loss, early death, and the negative impact of diabetes on GDP can be used to estimate the burden of the disease on society. According to recent cost estimates from a routine review, diabetes is

anticipated to cost the global economy almost US\$830 billion annually [3,4]. The International Diabetes Federation's analyses show that the global health care system spent more on diabetes treatment and diabetes from 2004 to 2014. The increase in the number of people with diabetes led to an increase in the amount of money spent on diabetes treatment per person [5]. Regardless of the current state of medicine, administrations are well-equipped to collect and preserve data through the use of a variety of monitoring and information gathering equipment. Specific tools are needed to obtain data, save and load it for information inquiry, and to operationally use it [6] for broad measures of information gathered in pharmaceutical services databases. The growing volume of data makes it extremely difficult to extract the relevant information for research. For this reason, the emerging multidisciplinary field of information revelation in databases, i.e. KDD, can be utilized by healthcare informatics to meet this need. In order to aid in the analysis of information as well as the discovery of regularities embedded within the information, this technique includes statistics, machine learning, AI, and pattern recognition [7]. Data science principles from AI, ML, DM, and DL are the focus of this study. Several methods are provided for removing clinical data for analysis, forecasting, checking, and understanding administration of Diabetes Mellitus.

HEALTHCARE AI AND DM APPLICATIONS OVERVIEW:

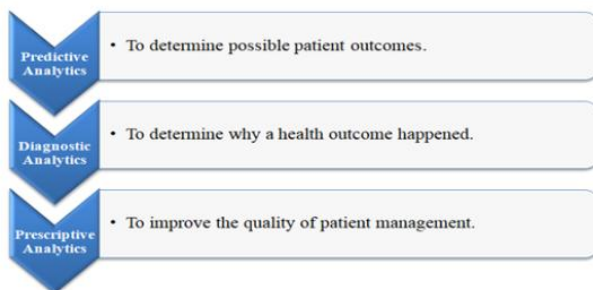


Figure 1: Healthcare AI and DM Applications

Many new researches proposes that the majority of data science analytics researches use cases as well as evolving uses for clinical data mining categories into following types

- **Predictive:** In predictive analytics the companies, business and healthcare professionals apply artificial intelligence, machine learning, data mining to look at quiet records so as to decide conceivable patient results, for example, the chance of an exacerbating or improving wellbeing condition, or odds of acquiring an ailment in a person's and group of the patient.
- **Diagnostic:** Diagnostic Analytics is characterized as a type of cutting edge examination which analyzes information or

substance to decide why a wellbeing result or result occurred.

- **Prescriptive:** In Prescriptive Analytics when research firms create AI calculations to perform total investigations of patient information to improve the nature of patient administration, for example, taking care of patient belongings and planning the progression of errands employments, such as requesting test tests, among of clinical work force.

KNOWLEDGE DISCOVERY IN DATABASES (KDD)

Knowledge Discovery in Databases (KDD or DM) is a popular term for data mining, which is used to extract important information from large databases and information distribution centers. Just like logical sides, mining applications are used in business [8]. DM is defined as a method for locating examples and drifts in databases that can be used to develop predictive models [9]. Otherwise, it is described as the process of determining and investigating data, and then developing models based on that data to uncover previously unknown, obscure occurrences [5, 10]. There are many people who believe that mining databases and KDDs is the same thing as discovering new information through the process of information mining. During the period of information disclosure, information mining might be a significant advancement. It is depicted in the picture above as an iterative sequence: Data cleaning, Data coordination, Data selection, Data change, Data mining, Pattern assessment, and Knowledge introduction.

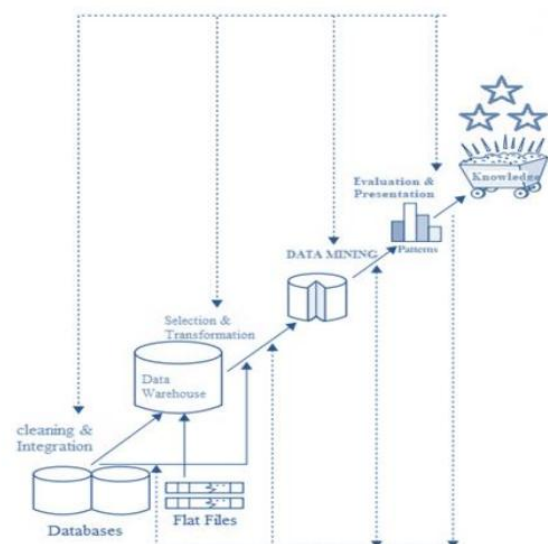


Figure 2: Data missing as a step in the process of knowledge discovery

Data mining is becoming increasingly popular in the healthcare sector, where the enormous volumes of information generated by clinical interactions are far too diverse and big to be

managed and analyzed using traditional ways. Data mining can uncover patterns, trends, and designs in massive datasets of varying types, allowing for better decision-making [11]. Data mining applications will provide numerous benefits and aids to the healthcare industry. Data science's impact on healthcare, its spread, and its characteristics are examined in the following study. This study focuses on the application of data science in the healthcare industry.

DIABETES PREDICTION MODEL

AI is going to have a huge impact on the healthcare industry. Data science and machine learning are useful in the healthcare industry for analysis and prediction. Data science applications for medical imaging analysis, drug discovery, genetics research, and predictive medicine are becoming increasingly popular in the healthcare industry. For example, in the healthcare industry, artificial intelligence is predicted to aid in prognosis and diagnosis for a wide range of diseases. This technology has advanced to the point where it is as capable as a human person in terms of making judgments. The International Diabetes Federation (IDF) estimated that 415 million individuals worldwide have hyperglycemia, with India in second place with 69 million people suffering from the condition. Type 1 diabetes, type 2 diabetes, and gestational diabetes are the three basic kinds of diabetes. Due to autoimmune processes, this pancreas is the primary supplier of insulin in the body. Only five to ten percent of diabetics have type 1 diabetes, which is the most common form of the disease. Type 2 diabetes is the most common kind of diabetes, and it is strongly linked to obesity. This is a result of both confrontation and a shortage of insulin, which affects a large percentage of the diabetes population (between 90% and 95%). A long-term ailment that causes blood sugar to rise when pregnant, gestational diabetes is a short-term condition that will return to normal after the baby is born. The women who have gestational diabetes are at risk of developing diabetes in the future. A computerized method that can anticipate diabetes in its earliest stages has been the driving force behind this research. Diabetes is caused by a malfunction in the body's ability to process carbohydrates. Reduction in the human body's ability to secrete insulin hormone or respond to secreted insulin, thereby maintaining the correct sugar levels, is characterized as this Type 1 diabetes, type 2 diabetes, and gestational diabetes are the three basic kinds of diabetes.

Using this machine learning model (MLM), it is possible to predict type 1, type 2, and gestational diabetes. Diabetic patients can be identified using this novel machine learning model. The accuracy is judged in light of the most recent approaches that are at our disposal. To combine the power of an expert system with machine learning, this is a revolutionary approach. In this chapter, a new strategy for predicting diabetes and its various forms

has been proposed by a researcher. A person's chance of developing diabetes can be predicted by this research, which will enable them take the necessary steps to avoid health care difficulties. One of the primary goals of this research is to develop a machine learning model for the early detection of diabetes. It's critical to know exactly what the user's or diabetic's symptoms are in order to apply and implement regulations. To identify whether a person has diabetes and the many types of diabetes they have, these knowledge combinations are used. There were 150 patients who participated in the testing of this algorithm. It has produced the same outcomes as doctors have done. ' Researchers have developed a machine learning model that may be used to accurately and quickly identify different forms of diabetes. Those in less developed countries, where there aren't enough doctors to go around, can benefit from this technology. The goal of this sophisticated Machine Learning Model is to lessen the need for doctors. Both doctors and patients will benefit from the improved accuracy and speed of decision-making.

DESCRIPTION OF THE DATASET

Details on the symptoms and diabetes type associated with each symptom are included in this data set. There are a total of 64 symptoms that are taken into account when classifying diabetes as Type 1, Type 2, or gestational. Since there are three different types of diabetes (Type 1 DIABETES, Type 2 DIABETES, and gestational diabetes), the dataset comprises a total of 65 columns, 64 of which are symptoms and the final one with a prediction value called Prediction Of Diabetes. Each entry in the dataset represents a test case for a different form of diabetes, as the researcher created the dataset to do. A knowledge base is a collection of rules that can be used in a machine learning model. As a result, the data set employed in this Machine Learning Model serves as the expert system's knowledge base. There are 220 records in the dataset or, alternatively, 220 rules in the expert system. Machine learning model rule design and rule selection is a difficult undertaking that may necessitate specialized expertise in order to improve prediction accuracy.

DESIGNING THE KNOWLEDGE BASE FOR MACHINE LEARNING MODEL

Table 1 lists the most common signs and symptoms of diabetes. Those are categorized by category, and a database based on values has been created. In the case of binary symptoms, these values are zero and one, and in the event of more than two values, a series of numbers. The binary values 0 and 1 are allocated to symptoms that are either True or False (table 4.2). For example, the binary values 0 and 1 are used to

represent FamilyHistory. It will take a series of numbers if the symptoms are multivalued. a three-tiered age scale The number 11-12-13 represents the numerical value of Young Adult Old. The numbers 21-22-23 symbolize the three levels of obesity: Low Normal Obese. Hypertension can be divided into four types: "Normal," "Elevated," "High," and "Very High." HDL Cholesterol is broken down into three groups: low, medium, and high. The numbers 1-2-3, which denote low, medium, and high, are used. The triglyceride has four classifications: "Normal," "BoarderLine," "High," and "VeryHigh." Primary and secondary data, including doctors' records, records from the Internet, and research articles on diabetes, were used to compile this table 1. Table 1 is one of the system's three inputs and may be found below. When formulating rules for a CSV file, distinct combinations of symptoms for each kind of diabetes are taken into account.

DATASET AS KNOWLEDGE BASE

Table-1

Sr. No.	Symptoms	Type1Diabetes	Type2Diabetes	Gestational Diabetes
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	Recurrsnt_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

Table 2: Number Representation for Symptoms

Boolean Value Symptoms			Multi Value Symptoms		
1. FamilyHistory	True	1	1. Age	Young	11
	False	0		Adult	12
				Old	13
2. Obesity	Low	21	3. Triglyceride	Normal	2
	Normal	22		BoarderLine	3
	Obese	23		High	4
4. Hypertension	Normal	3		VeryHigh	5
	Elevatal	4	5. HDL_Cholesterol	Low	1
	High	5		Medium	2
	VeryHigh	6		High	3

CONCLUSION

The review of literature and present study suggest that the artificial intelligence and machine learning methods have virtually endless applications in the healthcare industry. Today AI and ML are helping to simplify administrative processes in hospitals, personalize medical treatment and treat infectious

diseases. Data science research has remarkable potential to improve the diagnosis of diabetes mellitus and also a prediction of diabetes types which is helpful for medical practitioners and patients. Developing a machine learning based model for diabetes diagnosis is a time saving process. Applications of AI and Machine Learning techniques in the logical research are hopefully increasing in the upcoming years. The machine Learning model is an innovative technique from which researcher combines the power of the machine learning technique with the AI. The machine learning model uses simple decision tree classification algorithm that is supervised learning process which is non-parametric and applicable for forecasting the values of diabetes mellitus and grouping diabetes types. This machine learning model predicts the diabetes mellitus and diabetes type of a selected sign or symptoms by learning simple entered rules conditional from the datasets. This is universal method from which one can convert expert systems to machine learning platform.

REFERENCES

- Whiting, David & Guariguata, Leonor & Weil, Clara & Shaw, Jonathan. (2016). IDF Diabetes Atlas: Global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes research and clinical practice. 94. 311-21. 10.1016/j.diabres.2011.10.029.
- Sherwani, S. I., Khan, H. A., Ekhzaimy, A., Masood, A. & Sakharkar, M. K. Significance of hba1c test in diagnosis and prognosis of diabetic patients. Biomarker Insights11, BMI-S38440 (2016).
- NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since: a pooled analysis of 751 population-based studies with 4*4 million participants. Lancet 2016; published online April 7. [http://dx.doi.org/10.1016/S0140-6736\(16\)00618-8](http://dx.doi.org/10.1016/S0140-6736(16)00618-8).
- Seuring T, Archangelidi O, Suhrcke M. The economic costs of type 2 diabetes: A global systematic review. PharmacoEconomics. 2015; 33(8): 811–31.
- IDF Diabetes Atlas, 6th ed. Brussels, International Diabetes Federation; 2018.
- Nada Lavrac , "Selected techniques for data mining in medicine" , Artificial Intelligence in Medicine 16 (2019) 3–23
- Frawley W, Piatetsky-Shapiro G, Matheus C. Knowledge discovery in databases: an overview. In Piatetsky-Shapiro G, Frawley W, editors. Knowledge discovery in databases. Menlo Park, CA: The AAAI Press, 2016.
- HianChyeKoh and Gerald Tan,—Data Mining Applications in Healthcare, journal of Healthcare Information Management – Vol 19, No 2.
- Kincade, K. (2018). Data mining: digging for healthcare gold. Insurance & Technology, 23(2), IM2-IM7. [10]Milley, A. (2000). Healthcare and data mining. Health Management Technology, 21(8), 44- 47
- Christy, T. (2017). Analytical tools help health firms fight fraud. Insurance & Technology, 22(3), 22-26
- Jiawei Han, Micheline Kamber, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers is an imprint of Elsevier., 500 Sansome Street, Suite 400, San Francisco, CA 94111, ISBN 13: 978-1-55860-901-3
- S.Yamini , Dr.V.Khanaa , Dr.Krishna Mohantha - A State of the Art Review on Various Data Mining Techniques, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 5, Issue 3, March 2016
- U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI Mag., pp. 37–54, 2016.
- J.-J. Yang, J. Li, J. Mulder, Y. Wang, S. Chen, H. Wu, Q. Wang, and H. Pan, "Emerging information technologies for enhanced healthcare," Comput. Ind., vol. 69, pp. 3–11, 2015
- N. Wickramasinghe, S. K. Sharma, and J. N. D. Gupta, "Knowledge Management in Healthcare," vol. 63, pp. 5–18, 2015
- Shortliffe, EH.,Perrault, LE., (Eds.). Medical informatics: Computer applications in health care and biomedicine (2nd Edition). New York: Springer, 2020.
- Denis Rothman, "Artificial Intelligence by Example", Ingram short title (2018),1788990544,50-250
- Nick Bostrom,"Superintelligence: Paths, Dangers, Strategies", Oxford University Press, 2014, ISBN 0199678111, 9780199678112
- Shai Shalev-Shwartz, Shai Ben-David "Understanding Machine Learning", Cambridge University Press, United States of America, ISBN 978-1-107-05713-5
- Y. LeCun, Y. Bengio, G. Hinton, Deep learning Nature, 521 (7553) (2015), pp. 436-444
- Joshi SR, Parikh RM. India - diabetes capital of the world: now heading towards hypertension. J Assoc Physicians India. 2017;55:323–4
- Kumar A, Goel MK, Jain RB, Khanna P, Chaudhary V. India towards diabetes control: Key issues. Australas Med J. 2018;6(10):524–31.

Corresponding Author

Naresh Kumar*

Research Scholar, Sunrise University, Alwar
Rajasthan