## Original article

# Application of Artificial Intelligence in prediction type 2 diabetes mellitus: A Systematic Review

### [1]Mr. Jagdish D. Powar, [2]Dr. Rajesh Dase, [3]Dr. Deepak Bhosle

[1]PhD student in Biostatistics, Department of Community Medicine, MGM, Medical College & Hospital, Aurangabad, Maharashtra, India-431003.
[2] Associate Professor (Biostatistics), Department of Community Medicine, MGM, Medical College & Hospital, Aurangabad, Maharashtra, India-431003.
[3] Prof and Head, Department of Pharmacology, MGM, Medical College & Hospital, Aurangabad, Maharashtra, India-431003.

Corresponding author:  Dr. Rajesh Dase, Associate Professor (Biostatistics), Department of Community Medicine, MGM, Medical College & Hospital, Aurangabad, Maharashtra, India-431003

**Abstract**

Artificial Intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The healthcare industry produces large amount of healthcare data daily that can be used to extract information for predicting disease or disease diagnosis. Diabetes is chronic disease occurs due to insulin imbalance secretion. This present study was conducted with primary objective to review and evaluate the literature available on development of diabetes risk prediction models using Artificial Intelligence. In proposed review article, published original articles based on Artificial Intelligence in prediction diabetes mellitus were collected. The collected articles were scrutinized and selected articles of diagnosis of diseases or prediction of disease were included in the study. A total of 34 articles were included in the study. Different Artificial Intelligence algorithms and models were used in prediction of diabetes like Support vector machine, Decision tree, Naïve Baye, k-Nearest Neighbour, Random Forest, Logistic regression and Artificial Neural Network are the most useful prediction models for diabetes. Diagnosis of diseases at early stage is helpful for prevention and control the chronic diseases; such prediction of diseases is possible by the use of Artificial Intelligence. Artificial Intelligence can be used in the diagnosis of diabetes with higher prediction accuracy. Artificial Intelligence is helpful in prediction of diabetes at early stage by considering most common risk factors of Diabetes Mellitus.

**Key words:** Artificial Intelligence, Prediction of diabetes, Artificial Neural network, Prediction Model, Diagnosis of Diabetes.

## Introduction

Artificial intelligence (AI) is a new technology that simulates and extends human intelligence in machines that are programmed to mimic human actions[1]. Artificial intelligence is the ability of computer to complete the tasks such tasks usually done by humans with reasonable accuracy which eliminate human errors. Artificial Intelligence (AI) in healthcare is an overarching term used to describe the use of machine-learning algorithms and software to mimic human cognition in the analysis, presentation, and comprehension of complex medical and health care data. AI applications provide results with greater accuracy, efficiency and ease of use.

The healthcare industry produces large amount of healthcare data daily that can be used to extract information for prediction of diseases. AI can assist doctors, nurses, and other healthcare workers in their daily work. AI in healthcare can enhance preventive care and quality of life, produce more accurate diagnoses and treatment plans, and lead to better patient outcomes overall. AI can also predict and track the spread of infectious diseases by

36

analyzing data from a government, healthcare, and other sources. Applications of Artificial Neural Network (ANN) in health care include clinical Prediction, prediction of cancer, prediction of length of stay, image analysis and interpretation (e.g., automated electrocardiographic (ECG) interpretation used to diagnose myocardial infarction, and drug development[2].

Diabetes is chronic disease occurs due to insulin imbalance secretion. The prevalence of diabetes mellitus is rising rapidly throughout the world, and India is no exception. The International Diabetes Federation reported that 537 million adults (20–79 years old) are living with diabetes, or 1 in 10. This number is predicted to rise to 643 million by 2030 and 783 million by 2045[3]. Diabetes is responsible for 6.7 million deaths in 2021—1 in every 5 seconds[3]. The prevalence of diabetes has significantly risen in recent decades in India, which varies in rural and urban parts of the country. Type 2 diabetes is the most common type of diabetes in India. The burden of diabetes is rising rapidly in India and throughout the world. The prevalence of diabetes in India has risen from 7.1% in 2009 to 8.9% in 2019[4]. India is referred to as the "Diabetes Capital of the World" due to its increased prevalence of diabetes[5]. Undiagnosed diabetes is the major cause of chronic conditions. Some studies have reported that the prevalence of undiagnosed diabetes is more than 50% in India, whereas more than 75% of people around the world live with undiagnosed diabetes[3,4]. In India, an estimated 77 million people above the age of 18 are suffering from diabetes (type 2), and nearly 25 million are prediabetics (at a higher risk of developing diabetes in the near future)[6]. According to reports of World Health Organization (WHO), about 422 million people worldwide have diabetes, the majority living in low-and middle-income countries, and 1.5 million deaths are directly attributed to diabetes each year[7]. WHO reported that 3% increase in mortality rate due to diabetes between year 2000 to 2019[7].

In recent days various methods have been used for Prediction of diseases such as decision trees, the naïve Bayesian classifier, the support vector machine algorithm, random forest, multiple linear regression, the ANN algorithm, boosting and bagging algorithms, logistic regression and k-nearest neighbor etc. Risk prediction of disease at early stage of development will be helpful in prevention and management of chronic disease such as diabetes, cancer, heart disease, chronic kidney disease etc. Hence, this study was undertaken with the primary objective of reviewing and evaluating the literature available on the development of diabetes risk prediction models using Artificial Intelligence in India.

Methods

In the present research work, published original articles based on Artificial Intelligence in prediction of type 2 diabetes mellitus were collected. The articles were systematically searched on PubMed, Medline, Google scholar, Science Direct databases. The search keywords used on these databases as "Artificial Intelligence in Prediction of Diabetes", "Risk prediction models for diabetes", "Diabetes prediction using ANN" and "Prediction model for Diabetes". The collected articles were scrutinized and the relevant articles were included in the study.

Inclusion Criteria

The articles which satisfy the following criteria were included in the study: 1) articles published in peer-reviewed journals, 2) age of participants more than 18 years or older of both the gender, 3) use artificial intelligence in prediction of type 2 diabetes.

Article selection

The articles which consider development of prediction models for Prediction of type 2 diabetes. The summarized results of included articles were tabulated and presented. A total of 700 abstracts and original articles were scanned throughout the world and 34 articles were included in the final study.

Figure 1:
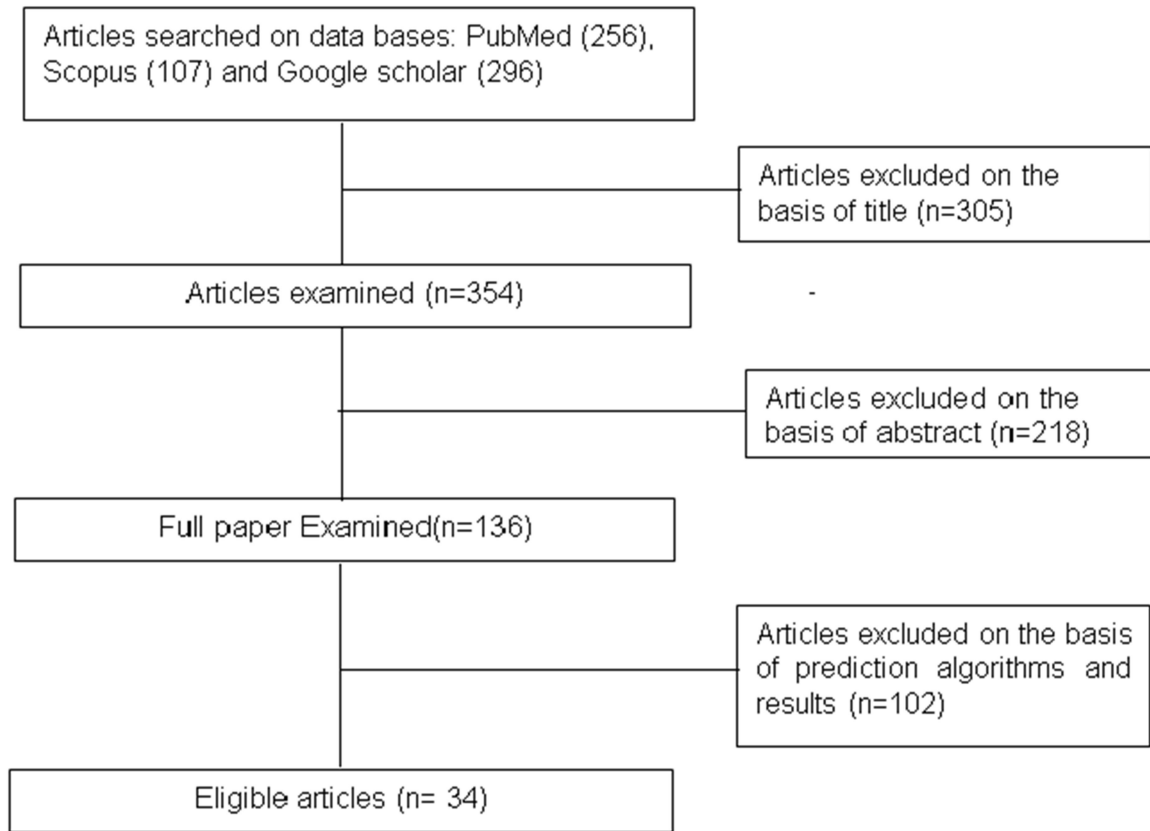


From figure-1, In this literature we found total 659 articles on PubMed, Scopus, Google Scholar and Science direct. Then 305 articles were excluded from the study based on the title of the research. A total of 354 abstracts were examined and 218 articles were excluded based on abstracts. A total of 136 full articles were examined and 102 articles were excluded based on methodology, prediction algorithms and statistical results and finally 34 Indian articles which consider development of prediction algorithms for type 2 diabetes were reviewed in the present research work.

**Results**

Figure 2: Year-wise publication on use artificial intelligence in prediction of type-2 diabetes
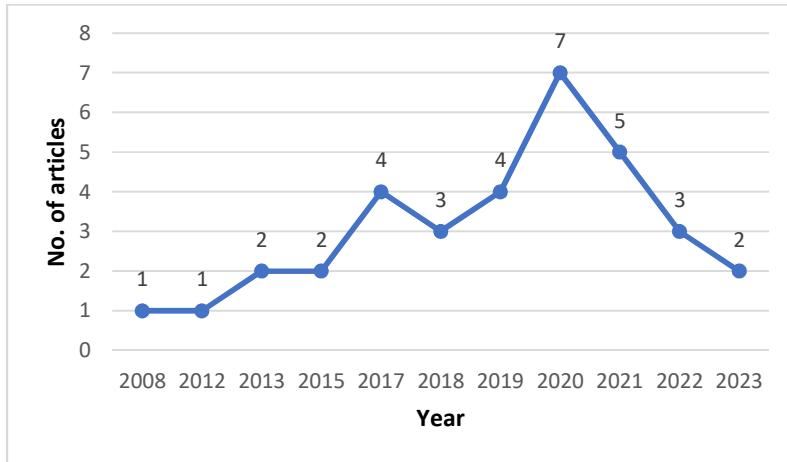


Figure-2, We found total 34 research articles in Indian literature which consider development of risk prediction models. The first article on prediction of diabetes was published in year 2008 by Rajeeb Dey et al., they used ANN model for prediction and recent publications were done in 2023. The maximum 7 articles were found in year 2020 and 2 articles were published recently in 2023 on diabetes prediction using artificial intelligence.

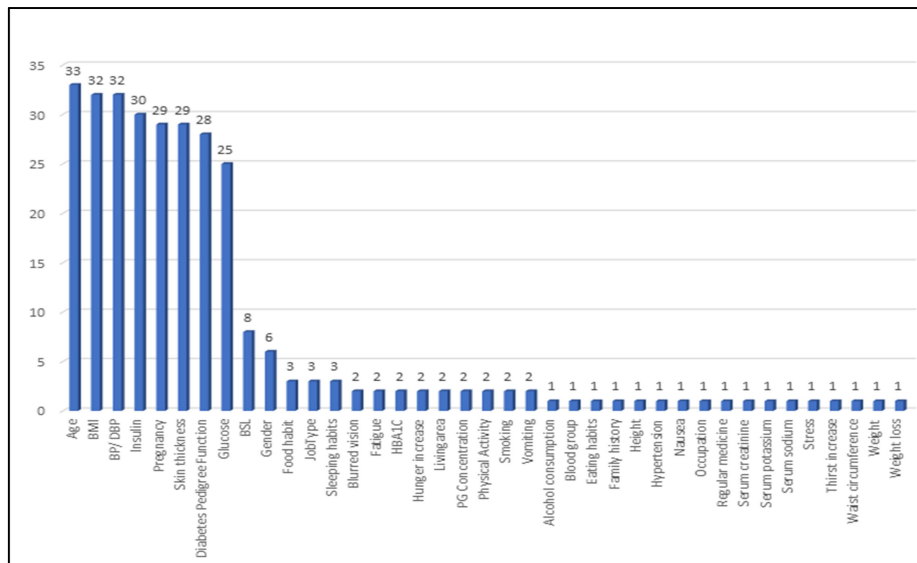Figure 3: Risk parameter used in prediction of diabetes



Figure-3 : The prediction models based on Pima Indian Diabetes Database used 8 risk parameters as no. of Pregnancy, BMI, Insulin level, Age, Blood pressure, Skin thickness, Glucose, Diabetes pedigree function. One researcher used 7 risk parameters from Pima Indian Diabetes Database as predictor in development of risk models based. Blood pressure (Systolic or diastolic) and Age used by all researcher in their prediction model. BMI and Insulin used by 32 and 30 researchers respectively in their prediction model. Skin fold thickness used as predictor by 27 authors in their research models. Pima Indian Diabetes database consist of diabetes pedigree function which was used by 28researchers in their prediction models. Pima Indian Diabetes Database consists of 768 female patients' data of which 26 researcher used no. of pregnancies as risk in the model development. Gender, Plasma

39

Glucose, HBA1C, hours of sleep, Serum creatinine, serum potassium, smoking, job type as other risk predictors in the prediction model development.

Figure 4: Most commonly used AI prediction model in prediction of diabetes
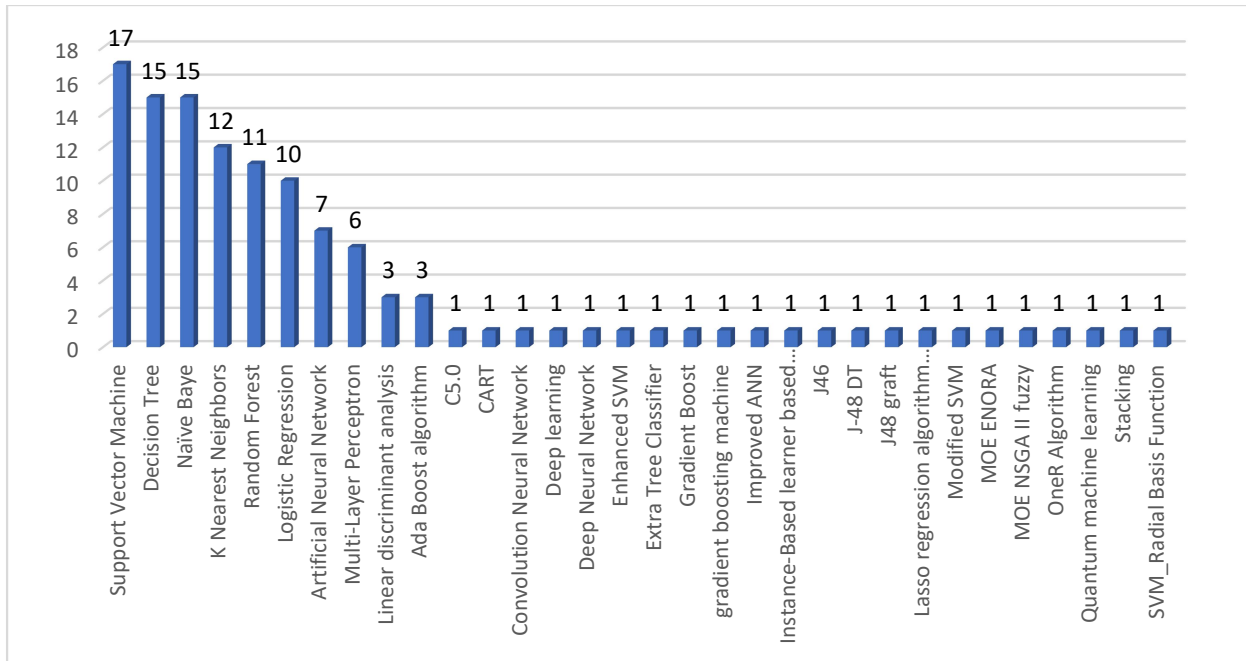


Support vector machine, Decision tree, Naïve Baye, k-Nearest Neighbour, Random Forest, Logistic regression and ANN are the most common used prediction models for diabetes.

Table 1: Details of eligible articles

| Sr.no. | Study (First Author) | Year | Data Source | Sample size | No. of attribute | Highest Accuracy | Highest Accuracy Model |
|---|---|---|---|---|---|---|---|
| 1 | Rajeeb Dey[8] | 2008 | Manipal Institute of Medical Sciences Hospital, Sikkim | 530 | 6 | 93% | ANN |
| 2 | Sonu Kumari[9] | 2012 | Survey | 100 | 13 | 93% | ANN |
| 3 | Ms. Divya[10] | 2013 | PIDD | 768 | 8 | Sen=88% | ANN |
| 4 | V. Anuja Kumari[11] | 2013 | PIDD | 460 | 8 | 78% | SVM |
| 5 | Ayush Anand[12] | 2015 | Self-Collected | 180 | 6 | 75% | CART |
| 6 | Veena Vijayan V[13] | 2015 | PIDD, Self-Generated for Validation | 768,200 | 8, 6 | 81% | Adaboost |
| 7 | S. Selvakumar[14] | 2017 | PIDD | 100 | 7 | 80% | KNN |
| 8 | P. Suresh Kumar[15] | 2017 | Diagnosis Lab located in Warangal, India | 650 | 14 | 99% | RF |

| 9 | N. Yuvaraj[16] | 2017 | National Institute of Diabetes | 75664 | 7 | 94% | RF |
| 10 | Vaishali R[17] | 2017 | PIDD | 768 | 8 | 83% | MOE NSGA II fuzzy |
| 11 | Deepti Sisodia[18] | 2018 | PIDD | 768 | 8 | 76% | NB |
| 12 | Suyash Srivastava[19] | 2018 | PIDD | 768 | 8 | 92% | ANN |
| 13 | Debadri Dutta[20] | 2018 | PIDD | 262 | 7 | 84% | RF |
| 14 | N. Sneha[21] | 2019 | archive.ics.uci.edu secondary data | 2500 | 15 | 78% | SVM |
| 15 | Aishwarya Mujumdar[22] | 2019 | Data Collected, PIDD | 800 | 10 | 96% | LR |
| 16 | Jatin N Bagrecha[23] | 2019 | PIDD | 768 | 8 | 84% | CNN |
| 17 | P. Prabhu[24] | 2019 | PIDD | 768 | 8 | 81% | DBN |
| 18 | Huma Naz[25] | 2020 | PIDD | 768 | 8 | 99% | DL |
| 19 | Neha Prerna Tiggaa[26] | 2020 | Self-collected/ generated, PIDD | 952 Self, 768(PIDD) | 16 | 93% | RF |
| 20 | Mitushi Soni[27] | 2020 | PIDD | 768 | 8 | 77% | RF |
| 21 | Abdulhakim Salum Hassan[28] | 2020 | PIDD | 768 | 8 | 90% | SVM |
| 22 | Jitranjan Sahoo[29] | 2020 | PIDD | 768 | 8 | 79% | LR |
| 23 | Naveen Kishore G[30] | 2020 | PIDD | 768 | 8 | 75% | RF |
| 24 | S. Thenappan[31] | 2020 | PIDD | 768 | 8 | 97% | SVM |
| 25 | P. Nagaraj[32] | 2021 | PIDD | 768 | 8 | 98% | Enhanced SVM, DNN |
| 26 | Namrata Nerkar[33] | 2021 | PIDD | 768 | 8 | 85% | ANN |
| 27 | R. Usharani[34] | 2021 | PIDD | 768 | 8 | 79% | LR |
| 28 | Satish Kumar Kalagotla[35] | 2021 | PIDD | 639 | 8 | 80% | Stacking |
| 29 | Radhanath Patra[36] | 2021 | PIDD | 768 | 8 | 83% | KNN |
| 30 | T. Madhubala[37] | 2022 | PIDD | 768 | 8 | 79% | ANN |
| 31 | SalliahShafi Bhat[38] | 2022 | Self-collected/ generated | 403 | 11 | 98% | RF |

| 32 | Himanshu Gupta[39] | 2022 | PIDD | 768 | 8 | 95% | DL |
| 33 | Aakash Shewani[40] | 2023 | PIDD | 1547 | 8 | 99% | ANN |
| 34 | Yogendra Singh[41] | 2023 | PIDD | 768 | 8 | 93% | lasso regression with ANN |

PIDD- Pima Indian Diabetes Database, ANN- Artificial Neural Network, SVM- Support Vector Machine, KNN- k-Nearest Neighbour, RF- Random Forest, LR- Logistic Regression, NB- Naïve Bayes Classifier, CART-Classification and Regression Trees, DL- Deep Learning, DNN- Deep Neural Network, DBN-Deep Belief Networks, CNN- Convolution Neural Network

Table-1, Author Aakash Shewani developed highest accuracy model of ANN based on PIDD. The maximum 16 predictor variables were used by Neha Prerna Tiggaa in their prediction model. Maximum sample size was 75664 and minimum sample size was 100 used by researcher in the model development.

**Discussion**

Total 34 original research articles were found in literature survey on prediction of Diabetes using AI in the Indian diabetic patients. The PIMA Indian Diabetes consists of female patient data with sample size n=768 of which 500 patients were non-diabetic and 268 were diabetic. There were 9 attributes in the in the Pima Indian Diabetes Database[6]. N. Yuvaraj[16] et al. published article based on prediction of diabetes, they have collected secondary data from National Institute of Diabetes of sample size n-75664 with 7 attributes. They used Support Vector Machine, Naïve Bayes, Decision tree, Random Forest, Neural Network models for prediction of diabetes and they observed highest accuracy for Random Forest with 94%. Rajeeb Dey[7] et al. collected 530 patient data from Hospital of Manipal Institute of Medical Sciences, Sikkim and used 6 attributes in model development of prediction of type 2 diabetes mellitus. They used ANN model for prediction purpose and found the accuracy of prediction was 93%. Aishwarya Mujumdar[22] et al. collected the data of 800 patients with 10 attributes and they used Random Forest Classifier, Decision Tree Classifier, Extra Tree Classifier, Ada Boost algorithm, Perceptron, Linear Discriminant Analysis algorithm, Logistic regression, K-Nearest Neighbour, Gaussian NB, Bagging algorithm, Gradient Boost Classifier algorithms in model. They found logistic regression algorithm is the highest accuracy algorithm with accuracy of 96%. Neha Prerna Tiggaa[26] et al. developed prediction model based on Pima Indian Diabetes database as well as self-collected database. They used 16 predictors in model and found highest accuracy for Random Forest with accuracy 93%. Sonu Kumari[9] et al. published paper in 2012 on 100 patient's dataset and they used 13 predictor variables in model as Gender, Age in years, Weight in kilograms, Height in feet's, Weight loss, Thirst increase, Hunger increase, Appetite increase, Nausea, Fatigue, Vomiting, Bladder-Vaginal infections, Blurred vision in their model. They have applied ANN and found 92.80% prediction accuracy of ANN algorithm. Ayush Anand[12] et al developed prediction model based on dataset of 180 patients, they used 6 risk predictors as Eating habits, sleeping habits, physical activity, BMI, waist circumference in their model of Classification and Regression Tree and their model achieved the accuracy of 75%. Total 25 researcher developed prediction models based PIDD and the Highest accuracy was observed 99% for ANN Model. Aakash Shewani[39] et al. used PIDD for model with sample size n=1547 and used simulation technique to increase sample size. Debadri Dutta[20] et al. used only 7 predictors and consider 262 patient's data from PIDD in their model and excluded no. of pregnancies. They found random forest as highest accuracy model. Maximum risk predictor variables were considered by Neha Prerna Tiggaa[26] et al. and they consider 16 variables in prediction model. Minimum 6 risk predictors were considered by Ayush Anand[12] et al. and Rajeeb Dey[7] et al. Author V. Anuja Kumari[11] used 460 patient's data in the prediction model and excluded missing values and

42

outliers from PIDD and found accuracy of 78% from Support Vector Machine classifier. Some researcher who used PIDD replaced the missing values with mean values in the respective predictor. As we found PIDD consists of missing values in Glucose level, Blood pressure Skinfold thickness, Insulin and BMI. Maximum missing values were for Insulin and Skinfold thickness with 374 and 227 respectively in PIDD.

**Conclusion**

Prediction of diseases at early stage is helpful for prevention and control the chronic diseases, such prediction of diseases is possible by the use of Artificial Intelligence. Artificial Intelligence can be used in the Prediction of diabetes with higher prediction accuracy by considering most common risk factors of Diabetes Mellitus.

From last one-decade Artificial Intelligence is used in the prediction of diabetes. The different models and algorithms were used for prediction purpose of diabetes by various researcher. Most responsible factors for diabetes were found as age, BMI, hypertension, parental history, and skin fold thickness. We hope that this systematic review of prediction of diabetes mellitus will also provide guidance to research in this new emerging sector of prediction of diabetes.

**Reference**

1. Dase RK, Pawar DD. Artificial Neural Network an Advanced Tool for Disease Diagnosis: A Review. International Journal of Current Medical and Applied Sciences 2020; 27(1):14-18.
2. International Diabetes Federation, Diabetes around the world in 2021. Last accessed on 18 June 2023. https://diabetesatlas.org/#:~:text=Diabetes%20around%20the%20world%20in%202021%3A,%20and%20783%20million%20by%202045
3. Pradeepa R, Mohan V. Epidemiology of type 2 diabetes in India. Indian Journal of Ophthalmology 2021;69: 2932-8.
4. TIMES OF INDIA.COM. Why India is diabetes capital of the world. Times Of India. 2022, November 14. https://timesofindia.indiatimes.com/india/why-india-is-diabetes-capital-of-the-world/articleshow/95509990.cms
5. Beagley J, Guariguata L, Weil C, Motala AA. Global estimates of undiagnosed diabetes in adults. Diabetes Research and Clinical Practice 2014; 103:150-60.
6. Pima Indians Diabetes Database. Data. World. Available online at: https://data.world/data-society/pima-indians-diabetes-database (retrieved 03 Jun 2023).
7. Diabetes Key facts World Health Organisation 5 April 2023. Last accessed on 01 Sep 2023 https://www.who.int/news-room/fact-sheets/detail/diabetes
8. Dey R, Bajpai V, Gandhi G, Dey B. Application of artificial neural network (ANN) technique for diagnosing diabetes mellitus. Region 10 and the Third international Conference on Industrial and Information Systems. IEEE 2008:1-4.
9. Kumari S, Singh A. A data mining approach for the diagnosis of diabetes mellitus. 7th International Conference on Intelligent Systems and Control (ISCO) 2013; 373–375.
10. Ms. Divya, Chhabra RR, Kaur S, Ghosh S. Diabetes Detection Using Artificial Neural Networks & Back-Propagation Algorithm. International Journal of Scientific & Technology Research 2013; 2 (1):9-11.
11. V Anuja Kumari, R Chitra. Classification Of Diabetes Disease Using Support Vector Machine. IJERA 2013; 3(2): 1797-1801.
12. Vijayan VV, Anjali C. Prediction and diagnosis of diabetes mellitus — A machine learning approach. IEEE Recent Advances in Intelligent Computational Systems 2015;122–127.
13. Anand A, Shakti D. Prediction of diabetes based on personal lifestyle indicators. 1st International Conference on Next Generation Computing Technologies 2015: 673-676.
14. Selvakumar S, Kannan KS and Nachiyar GS. Prediction of Diabetes Diagnosis Using Classification Based Data Mining Techniques. International Journal of Statistics and Systems 2017; 12(2):183-88.

15. Kumar PS, Pranavi S. Performance analysis of machine learning algorithms on diabetes dataset using big data analytics. International Conference on Infocom Technologies and Unmanned Systems 2017: 508-513.

16. Yuvaraj N, Preethaa KR. Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster. Cluster Comput 2019; 22(S1) :1–9.

17. R. Vaishali, R. Sasikala, S. Ramasubbareddy, S. Remya, S. Nalluri. Genetic algorithm-based feature selection and MOE Fuzzy classification algorithm on Pima Indians Diabetes dataset. International Conference on Computing Networking and Informatics (ICCNI), 2017:1-5.

18. Sisodia D, Sisodia DS. Prediction of Diabetes using Classification Algorithms. Procedia Computer Science 2018; 132:1578–85.

19. Srivastava S, Sharma L, Sharma V, Kumar A et al. Prediction of Diabetes Using Artificial Neural Network Approach. Engineering Vibration, Communication and Information Processing Springer Nature Singapore Pte Ltd;2019: 679–87.

20. Dutta D, Paul D and Ghosh P. Analysing Feature Importances for Diabetes Prediction using Machine Learning. IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference 2018: 924-928.

21. N. Sneha, Gangil T. Analysis of diabetes mellitus for early prediction using optimal features selection. J Big Data 2019; 6(13) :1-19.

22. Mujumdar A, Vaidehi V. Diabetes Prediction using Machine Learning Algorithms. Procedia Computer Science 2019; 165: 292–9.

23. Bagrecha JN, Chaithra GS, Jeevitha S. Diabetes Disease Prediction using Neural Network.International Journal for Research in Applied Science & Engineering Technology 2019;7(4): 3888–93.

24. Prabhu P, Selvabharathi S. Deep belief neural network model for prediction of diabetes mellitus.3rd International Conference on Imaging, Signal Processing and Communication 2019: 138-142.

25. Naz H, Ahuja S.Deep learning approach for diabetes prediction using PIMAIndian dataset. Journal of Diabetes & Metabolic Disorders 2020; 19: 391–403.

26. Tigga NP, Garg S. Prediction of type 2 diabetes using machine learning classification methods. Procedia Computer Science 2020;167: 706–716.

27. Soni M, Varma S. Diabetes Prediction using Machine Learning Techniques, IJERT 2020; 09(09): 921-25.

28. Hassan ASI Malaserene, A Anny Leema. Diabetes Mellitus Prediction using Classification Techniques. IJITEE 2020; 9(5): 2080-2084.

29. Sahoo J, Dash M, Patil A. Diabetes Prediction Using Machine Learning Classification Algorithms. International Research Journal of Engineering and Technology 2020; 07(08): 1664-69.

30. Kishore N, Rajesh V, Vamsi A, Reddy A, Sumedh K, Rajesh T. Prediction of diabetes using machine learning classification algorithms. International Journal of Scientific & Technology Research 2020; 9(01): 1805-08.

31. Thenappan S, Valan Rajkumar M, Manoharan PS. Predicting diabetes mellitus using modified support vector machine with cloud security. IETE Journal of Research 2022; 68(6): 3940–50.

32. Nagaraj P, Deepalakshmi P. Diabetes prediction using enhanced SVM and Deep Neural Network learning techniques: An algorithmic approach for early screening of diabetes. International Journal of Healthcare Information Systems and Informatics 2021; 16(4): 1–20.

33. Nerkar N, Inamdar V, Kajrolkar L, Barve R. Diabetes Prediction using Neural Network. International Research Journal of Engineering and Technology 2021; 08(02): 330-333.

34. Usharani R, Shanthini A. Neuropathic complications: Type II diabetes mellitus and other risky parameters using machine learning algorithms. Journal of Ambient Intelligence and Humanized Computing 2021:1-23.

35. Kalagotla SK, Gangashetty SV, Giridhar K. A novel stacking technique for prediction of diabetes. Computer Biology and Medicine 2021; 135:1-11.

36. Patra R, Khuntia B. Analysis and prediction of Pima Indian Diabetes Dataset using SDKNN classifier technique. ICRIET 2020; 1070(1): 012059.

37. Madhubala T, Umagandhi, Sathiamurthi. Diabetes Prediction using Improved Artificial Neural Network using Multilayer Perceptron. International Journal of Electrical and Electronics Engineering 2022; 9(12):167–79.

44

38. Bhat SS, Selvam V, Ansari GA, Ansari MD, Rahman MH. Prevalence and early prediction of diabetes using Machine Learning in North Kashmir: A case study of district Bandipora. Computational Intelligence and Neuroscience 2022:1-12.

39. Gupta H, Varshney H, Sharma TK, Pachauri N, Verma OP. Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction. Complex & Intelligent Systems 2022; 8(4): 3073–87.

40. Shewani A, Kaur N. Diabetes Classification and Prediction using Machine Learning Approach. IJCRT 2023;11(2): a682-87.

41. Singh Y, Tiwari M. A novel hybrid approach for detection of type-2 diabetes in women using lasso regression and artificial neural network. I. J. Intelligent Systems and Applications 2022;14(4):11–20.