

# Review of Predictive Analysis Techniques for Analysis Diabetes Risk

Sonali Vyas<sup>1</sup>, Rajeev Ranjan<sup>2</sup>, Navdeep Singh<sup>3</sup>, Arohan Mathur<sup>4</sup>

<sup>1</sup>Assistant Professor, AIIT, Amity University, Jaipur Rajasthan

<sup>2</sup>Research Scholar, AIIT, Amity University, Jaipur Rajasthan

<sup>3,4</sup>Student, AIIT, Amity University, Jaipur Rajasthan

<sup>1</sup>vyas.sonali86@gmail.com, <sup>2</sup>rajeev.ranjan.rajveer@gmail.com

<sup>3</sup>navdeepsykar@gmail.com, <sup>4</sup>arohanm@icloud.com

**Abstract**—Diabetes mellitus is one of the most commonplace persistent illnesses in nearly all nations, and keeps booming in numbers and importance, as economic development and urbanization result in changing lifestyles characterized by reduced physical hobby and expanded weight problems. In this paper we have reviewed various case related to diabetes mellitus based on different predictive analytics algorithm and we found that single algorithm is not sufficient for predictive analytics.

**Keywords**—Diabetes, predictive analytics, blood sugar, insulin

## I. INTRODUCTION

Diabetes may be prohibited by the assist of insulin infusions, a compact eating routine as well as systematic workout. Diabetes prompts significantly additional sickness. e.g. Visual impairment, circulatory strain, coronary sickness, along with kidney ailment as well as so on.

Diabetes are four types of variety

Type 1 Diabetes: - Insulin isn't delivering sufficient by pannacers then type-1 diabetes take place in body. It may arise any stage of life. e.g.: kids, youngsters [4].

Type 2 Diabetes: - Insulin doesn't adequate amount and it is not sufficient for body need, then type-2 diabetes arises. Because of parent's inheritance, seniority, corpulence expands the danger of getting write 2 diabetes. For the most part happens at 40 years old [1].

Gestational Diabetes: -It is the third principle shape, significantly happens with the pregnant ladies because of abundance glucose equal level in the body [2].

Pregestational Diabetes: It is another form of diabetes and it occurs when insulin-subordinate diabetes earlier getting to be pregnant [3].

### A. Predictive analysis

Predictive analysis is an analytics process which is analyze historical data cum new data to forecast behavior and trends activity for future prediction goal, it is closely related to advanced analytics. It includes statistical analytics techniques and machine learning algorithm and it create predictive model and it is place for future event happening. We use predictive analysis technique to detect fraud, reducing risk improve operation, diagnosis diseases improve marketing and much more. The crucial point of perceptive examination relies upon getting associations between sensible elements and the foreseen factors from past occasions and mishandling them to predict the dark outcome.

## I. EXISTING PREDICTIVE ANALYSIS TECHNIQUES

### A. Machine learning techniques

Machine learning technique is a technique which is deal in scientific field and it is learn from experience. The tenure ML is closely related to AI.

#### 1. Supervised learning

Machine learning technique which is learning particularly a purpose i.e. Objective purpose that is an appearance of a form unfolding the statistics. e.g.: -GA algorithm, KNN algorithm, SVM algorithm etc.

#### 2. Unsupervised learning

The machine learning technique which is trying on the way to determine the unseen arrangement of records among variables.

### B. Data mining techniques

Data Mining is an analytic action advised to analyze abstracts in seek of constant forms analytical associations amid attributes, and again to authenticate the allegation by applying the identified forms to novel subsections of records. The final ambition of abstracts withdrawal is anticipation - and projecting abstracts withdrawal is a lot of accepted blazon of abstracts mining and one that has a lot of absolute business applications: -Classification, Clustering, Association technique etc.

### C. Nature-inspired algorithm

Nature is a great teacher and guide for human from ancient era and it is important for humans, birds and animal too. Nature-inspired algorithm is algorithm which is inspired from nature, animal and bird specification and many algorithms inspired from nature for either search or optimization of any problem.

Nature-inspired algorithm are mainly two types: -Evolutionary algorithm and swarm algorithm. Evolutionary algorithm example is Genetic algorithm and swarm algorithm example are PSO algorithm, Ant algorithm etc. and it is played an important role in every field of human life.

**TABLE 1: Review of technique related to predictive analysis for diabetesYear**

Year	Author	Methodology	Dataset	Objective	Problem identification
2018	R.Haritha et al. [5]	Firefly and Cuckoo Search Algorithms	Pima Indian Diabetes	To develop an extraction approach can be applied for different types of data	Change no. of features used to train and used another nature-inspired algorithm also
2018	Tarik A. Rashid, Saman M. Abdullah [6]	Artificial Bee Colony, GA, NN algorithm	Pima Indian Diabetes	To performance enhance and resolve overfitting	To change the indexes of the weights for more optimal in classification
2018	Yinghui Zhang et al. [7]	A Feed forward NN	Pima dataset	This network model be trained several times to obtain an accuracy	Improving the training method and changing the activation function to better predict DM in early stages
2018	Mustafa S. Kadhm et al. [8]	K-means Clustering and Proposed Classification Approach	Pima Indian Diabetes	The proposed system focused on the features analysis and classification parts	More clusters are used to get the best separation points and most accurate results
2018	N.Mohana Sundaram [9]	Elman Neural Network	Pima Indian Diabetes	Improves the performance of accuracy classification	Use other function to improves the performance
2018	MinyechilAlehegn et al. [10]	Machine Learning Algorithm	UCI machine learning	The comparison of individual algorithm is done	It can be applying in large amount of data for future extension
2018	D. Christy Sujatha [11]	K-means algorithm	Pima Indian Data set	Predictive analytics using k means algorithm is developed	Implement other clustering algorithms and accuracy and sensitivity of all algorithms may be compared
2018	Shuja Mirza et al. [12]	SMOTE and Decision tree algorithm	Clinical dataset	Reduction in class imbalance can result in improvement of prediction rate	To reduction in class imbalance can result in improvement of prediction rate
2018	MansourehKhojandiJazi [13]	Radial Base Function, NN Method Combined with Colonial Competition Optimization Algorithm	Re-hospitalization of diabetic patients	To achieve better results	Use different algorithm like NN algorithm, Evolutionary algorithm, Genetic algorithm, SVM algorithm
2018	MaisYasen et al. [14]	Neural Networks using Dragonfly Algorithm	National Institute of	Optimization accuracy	For better result used another procedure and

			Diabetes and Digestive and Kidney Diseases		make appropriate comparison.
2018	Han Wu et al. [15]	Data mining	Pima Indian Diabetes Dataset	Targeted to type-2 DM through effective design of a model	Other sophisticated procedure implemented for the study of DM
2018	Ratna Nitin Patil et al. [16]	KNN and Naïve Bayes Genetic Algorithm	Pima dataset	Upgrading the Performance in Diabetes Detection	More efficient techniques for the classification of diabetes patients can be implemented
2018	R.Sengamuthu et al. [17]	Data mining techniques	Pima Indian Diabetes Dataset	Analysis to predict diabetes mellitus	Use more DM associated statics to prevent DM.
2017	Damodar Reddy Edla et al. [18]	Bat Optimization classification algorithm	PID dataset	Proposed Diabetes-Finder for effective diagnosis of Type-2 Diabetes	Used other algorithm for improved accuracy
2017	Dilip Kumar Choubey and Sanchita Paul [19]	GA and RBF NN algorithm	Pima Indian Dataset	To design a classification system for diabetes	This technique can as well be reused in other kinds of diseases
2017	Maham Jahangir et al. [20]	Enhanced Class Outlier with Automatic Multilayer Perceptron	UCI (Pima Indians)	To create a prediction framework for diabetes prediction	For superior diabetes classification design effective system
2017	Sushant Ramesh et al. [21]	Deep Learning Approach	Online Machine Learning UCI repository.	To Identify Diabetes	Improved result by increasing the size of the data set
2017	Seyyed Mohammad Hossein Dadgar and Mostafa Kaardaan [22]	Neural Network with Genetic Algorithm	Pima Indians Diabetes Data Set	Feature Selection to Predict Diabetes	Used other algorithm for improved accuracy
2017	Wenqian Chen et al. [23]	K-means and Decision Tree	Pima Indian Diabetes Data	Type 2 DM forecast model	Used large amount of medical statics for effective result
2017	Ratnanitinpatil et al. [24]	PCA, K-Means algorithm	Pima Indian Dataset	To predicting type-2 diabetes with accuracy	Further extended to deal datasets with multiple classes
2017	Dr.Chandan Banerjee et al. [25]	Evolutionary Algorithm	Pima Indian Diabetes dataset	Parameter Estimation for diabetes prediction	Used evolutionary search CV method for expand the work
2017	Harsha Sethi et al. [26]	ANN, KNN, Naïve Bayes, SVM algorithm	400 datasets from diverse section of the society	To accurate diagnosis of diabetes	Enhance number of verity in the database
2017	Ahmed Hamza Osman, Hani MoetqueAljahdali [27]	K-SVM algorithm	UCI Pima Indian dataset	Analyzed and mined diabetes dataset for diagnosis purpose	To integrate one of the optimization techniques for potential enhancement.
2017	Vaishali R et al. [28]	Genetic algorithm MOE fuzzy classification procedure	Pima dataset	Decrease data redundancy for enhance the speed of performance and accuracy for sorting	Analyzed and addressed missing data for feature selection

2017	IoannisKavakiotis et al. [29]	ML along with DM process	Electronic Health Records	DM identify through ML along with DM process	Use more medical statistics
2017	S.Selvakumar et al. [30]	Binary logistic regression, KNN Algorithm, Multilayer perceptron neural network	Multi-dimensional healthcare dataset	Implementation of algorithm for the diabetes data	Use other data set and work on accuracy too
2017	Ordoñez Barrios et al. [31]	Auto Classification technique	Virtual questionnaire hosted in Google Forms	To enhance the accuracy of the developed predictive model	Used large data set
2017	RamalingaswamyCheruku et al. [32]	Spider monkey algorithm	Pima Indians Diabetes dataset.	For better diabetes diagnosis use effective SM ruler Miner	Use other data set and work on accuracy too
2017	Shital Tambade et al. [33]	SVM algorithm	Pima Indian Diabetes data set	To classify the patient into diabetic positive or negative	Accuracy may be improved
2017	KaramathAteeq, and Dr. Gopinath Ganapathy [34]	MPSO-NN algorithm	Pima Indian Diabetes data set	For classifying the diabetic	To test the same algorithm with the increased number of instances
2016	Alby S. and B. L. Shivakumar [35]	General Regression Neural Networks (GRNN)	Pima Indian Dataset	Detect diabetes at a very early stage	To use a better technique with the same dataset and features for improved accuracy levels
2016	S. R. Priyanka Shetty, Sujata Joshi [36]	Data mining technique	Laboratories data	Design a tool for Diabetes Prediction and Monitoring	Work on accuracy
2016	Rabina, Prof. Meenakshi Sharma [37]	Bacterial Foraging Optimization Algorithm And ANN	UCI Diabetes diseases dataset	Innovative automatic model to diagnose the diabetes disease	To use Genetic Algorithm to improve optimization and use other SVM's kernel or another classifiers algorithm
2016	G. Thippa Reddy, NeeluKhare [38]	Firefly Algorithm, BAT Algorithm	Sree Diabetic Care Center	To develop prediction algorithm for diabetes	Applying other optimization techniques to increase the accuracy
2015	Omar S. Soliman, Eman Abo ElHamd [39]	BA With Chaotic Levy Flights	Pima data set	This is help for better as well as early treatment of patients	Proposed algorithm might be tested on other chronic diseases
2015	T. Santhanam and M.S Padmavathi [40]	K-means, GA	Pima Indian Diabetes	For classify the diabetes dataset used SVM method with reduction of dimensionality	Work on missing data for better goal

## II. CONCLUSIONS

This assessment paper focuses on several predictive analysis procedures and approaches and it is utilizing premature

estimate of a several cases of diabetes from patient record. The different approach analytics procedures are applied in health records field for foreseeing case of diabetes and to find out

effective ways to treat them in better manner. In future, comparative analysis will be conducted for analyzing the performance and application of each algorithm so that the best one can be applied for predictive analysis. Assessment of Diabetes Assumption displays that particular method to identify diabetes is not very sophisticated tactic for initial diabetes detection and it is not fully accurately predicting diseases. That's why we need a smart hybrid predictive analytics diabetes diagnostic system that can effectively work with accuracy and efficiency.

## REFERENCES

- [1] Nilesh Jagdish Vispute, Dinesh Kumar Sahu, Anil Rajput, "ASurvey on naive Bayes Algorithm for Diabetes Data Set Problems", International journal for research in Applied Science & Engineering Technology (IJRASET), Volume 3 issue XII, December 2015
- [2] Haldurai Lingaraj, Rajmohan Devadass, Vidya Gopi, Kaliraj Palanisamy, " Prediction of Diabetes Mellitus using Data Mining Techniques": A Review, Journal of Bioinformatics & Cheminformatics, February 19, 2015.
- [3] Isha Vashi, Prof. Shailendra Mishra, "A Comparative Study of Classification Algorithms for Disease Prediction in Health Care", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 9, September 2016.
- [4] Vrushali Balpande, Rakhi Wajgi, " Review on Prediction of Diabetes using Data Mining Technique", International Journal of Research and Scientific Innovation (IJRSI) | Volume IV, Issue IA, January 2017 | ISSN 2321-2705.
- [5] Haritha, R., Babu, D. S., & Sammulal, P. (2018). A Hybrid Approach for Prediction of Type-1 and Type-2 Diabetes using Firefly and Cuckoo Search Algorithms. International Journal of Applied Engineering Research, 13(2), 896-907.
- [6] Rashid, T. A., & Abdullah, S. M. (2018). A Hybrid of Artificial Bee Colony, Genetic Algorithm, and Neural Network for Diabetic Mellitus Diagnosing. Aro-THE Scientific JOURNAL of Koya University, 6(1), 55-64.
- [7] Zhang, Y., Lin, Z., Kang, Y., Ning, R., & Meng, Y. A Feed-Forward Neural Network Model For The Accurate Prediction Of Diabetes Mellitus.
- [8] Kadhm, M. S., Ghindawi, I. W., & Mhawi, D. E. (2018). An Accurate Diabetes Prediction System Based on K-means Clustering and Proposed Classification Approach. International Journal of Applied Engineering Research, 13(6), 4038-4041.
- [9] Sundaram, N. M. (2018). An Improved Elman Neural Network Classifier for classification of Medical Data for Diagnosis of Diabetes. International Journal of Engineering Science, 16317.
- [10] Alehegn, M., & Mulay, R. J. D. P. Analysis and Prediction of Diabetes Mellitus using Machine Learning Algorithm.
- [11] Sujatha, D. C., Kumar, D. M., & Chandrakumar Peter, M. (2018). Building Predictive Model for Diabetics Data using K Means Algorithm. Journal Homepage: <http://www.ijmra.us>, 8(7).
- [12] Mirza, S., Mittal, S., & Zaman, M. (2018). Decision Support Predictive model for prognosis of diabetes using SMOTE and Decision tree. International Journal of Applied Engineering Research, 13(11), 9277-9282.
- [13] Habibi, N., & Harouni, M. (2018). Estimation of Re-hospitalization Risk of Diabetic Patients based on Radial Base Function (RBF) Neural Network Method Combined with Colonial Competition Optimization Algorithm. Majlesi Journal of Electrical Engineering, 12(1), 109-116.
- [14] Yasen, M., Al-Madi, N., Obeid, N., Sumaya, P., & Abdullah II, K. Optimizing Neural Networks using Dragonfly Algorithm for Medical Prediction.
- [15] Wu, H., Yang, S., Huang, Z., He, J., & Wang, X. (2018). Type 2 diabetes mellitus prediction model based on data mining. Informatics in Medicine Unlocked, 10, 100-107.
- [16] Patil, R. N., & Tamane, S. C. (2018). Upgrading the Performance of KNN and Naïve Bayes in Diabetes Detection with Genetic Algorithm for Feature Selection.
- [17] Sengamuthu, M. R., Abirami, M. R., & Karthik, M. D. (2018). Various Data Mining Techniques Analysis to Predict Diabetes Mellitus.
- [18] Edla, D. R., & Cheruku, R. (2017). Diabetes-finder: a bat optimized classification system for type-2 diabetes. Procedia Computer Science, 115, 235-242.
- [19] Choubey, D. K., & Paul, S. (2017). GA\_RBF NN: a classification system for diabetes. International Journal of Biomedical Engineering and Technology, 23(1), 71-93.
- [20] Jahangir, M., Afzal, H., Ahmed, M., Khurshid, K., & Nawaz, R. (2017). ECO-AMLPL: A Decision Support System using an Enhanced Class Outlier with Automatic Multilayer Perceptron for Diabetes Prediction. arXiv preprint arXiv:1706.07679.
- [21] Ramesh, S., Caytiles, R. D., & Iyengar, N. C. S. (2017). A Deep Learning Approach to Identify Diabetes. Advanced Science and Technology Letters, 145, 44-49.
- [22] Dadgar, S. M. H., & Kaardaan, M. A Hybrid Method of Feature Selection and Neural Network with Genetic Algorithm to Predict Diabetes.
- [23] Chen, W., Chen, S., Zhang, H., & Wu, T. (2017, November). A hybrid prediction model for type 2 diabetes using K-means and decision tree. In Software Engineering and Service Science (ICSESS), 2017 8th IEEE International Conference on (pp. 386-390). IEEE.
- [24] Patil, R. N., & Tamane, S. C. A Novel Scheme for Predicting Type 2 Diabetes in Women: Using Kmeans with PCA AS Dimensionality Reduction.
- [25] Banerjee, C., Paul, S., & Ghoshal, M. (2017). An Evolutionary Algorithm based Parameter Estimation using Pima Indians Diabetes Dataset. International Journal on Recent and Innovation Trends in Computing and Communication, 5(6), 374-377.
- [26] Sethi, R., Goraya, A., & Sharma, V. (2017). Artificial Intelligence based Ensemble Model for Diagnosis of Diabetes. International Journal of Advanced Research in Computer Science, 8(5).
- [27] Osman, A. H., & Aljhdali, H. M. (2017). Diabetes disease diagnosis method based on feature extraction using K-SVM. Int J Adv Comput Sci Appl, 8(1).
- [28] Vaishali, R., Sasikala, R., Ramasubbareddy, S., Remya, S., & Nalluri, S. (2017, October). Genetic algorithm based feature selection and MOE Fuzzy classification algorithm on Pima Indians Diabetes dataset. In Computing Networking and Informatics (ICCN), 2017 International Conference on (pp. 1-5). IEEE.
- [29] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data

- mining methods in diabetes research. *Computational and structural biotechnology journal*, 15, 104-116.
- [30] Selvakumar, S., Kannan, K. S., &GothaiNachiyar, S. (2017). Prediction of Diabetes Diagnosis Using Classification Based Data Mining Techniques. *International Journal of Statistics and Systems*, 12(2), 183-188.
- [31] [31] Barrios, O., Alberto, D., Infantes, V., Raphael, E., Aguirre, A., & Alexander, J. (2017, August). Predictive modeling for presumptive diagnosis of type 2 diabetes mellitus based on symptomatic analysis. In *Electronics, Electrical Engineering and Computing (INTERCON)*, 2017 IEEE XXIV International Conference on (pp. 1-4). IEEE.
- [32] Cheruku, R., Edla, D. R., &Kuppili, V. (2017). SM-RuleMiner: Spider monkey based rule miner using novel fitness function for diabetes classification. *Computers in biology and medicine*, 81, 79-92.
- [33] Nidhi Chandra, Sunil Kumar Khatri, Subhranil Som, (2017) "Anti-Social Comment Classification based on kNN Algorithm", 6th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), IEEE Conference, indexed with SCOPUS Sep. 20-22, 2017, Amity University Uttar Pradesh, Noida, India.
- [34] Tambade, S., Somvanshi, M., Chavan, P., & Shinde, S. (2017). SVM based Diabetic Classification and Hospital Recommendation. *International Journal of Computer Applications*, 167(1).
- [35] Ateeq, K., & Ganapathy, G. (2017). The novel hybrid Modified Particle Swarm Optimization–Neural Network (MPSO-NN) Algorithm for classifying the Diabetes. *International Journal of Computational Intelligence Research*, 13(4), 595-614.
- [36] Alby, S., &Shivakumar, B. L. (2016). A Novel Approach for Prediction of Type 2 Diabetes. *International Journal of Advanced Research in Computer Science*, 7(4).
- [37] Shetty, S. P., & Joshi, S. (2016). A Tool for Diabetes Prediction and Monitoring Using Data Mining Technique. *IJ Information Technology and Computer Science*.
- [38] Shobha Tyagi, Subhranil Som, Qamar Parvez Rana (2017) "Trust based Dynamic Multicast Group Routing ensuring Reliability for Ubiquitous Environment in MANETs", *International Journal of Ambient Computing and Intelligence (IJACI)*, Volume 8, Issue 1, ESCI, Scopus Indexed, ISSN: 1941-6237, DOI: 10.4018/IJACI, Pages 70 – 97, January – March 2017. (<http://www.igi-global.com/journals/abstract-announcement/158348>)
- [39] Sharma, R. P. M., & SSCET, B. (2016). Diabetes Prediction by using Bacterial Foraging Optimization Algorithm And Artificial Neural Network. *methods*, 6(1).
- [40] Thippa Reddy, G., &Khare, N. (2016). FFBAT-optimized rule based fuzzy logic classifier fordabetes. In *International Journal of Engineering Research in Africa* (Vol. 24, pp. 137-152). Trans Tech Publications.
- [41] Soliman, O. S., &Elhamd, E. A. (2015). A chaotic levy flights bat algorithm for diagnosing diabetes mellitus. *International Journal of Computer Applications*, 111(1).
- [42] Santhanam, T., & Padmavathi, M. S. (2015). Application of K-means and genetic algorithms for dimension reduction by integrating SVM for diabetes diagnosis. *Procedia Computer Science*, 47, 76-83.