

Diabetes Risk Detection Review Using Machine Learning Techniques

Sanat Kumar Sahu

Department of Computer Science, Govt. Kaktiya.P.G. College Jagdalpur (C.G.), India

*Corresponding Author: sanat.kosa1@gmail.com

DOI: <https://doi.org/10.26438/ijcse/v9i8.8486> | Available online at: www.ijcseonline.org

Received: 17/Aug/2021, Accepted: 20/Aug/2021, Published: 31/Aug/2021

Abstract— Data Mining (DM) and Machine Learning (ML) are used as training algorithm for learning classification and feature selection technique (FST) from data. The DM and ML are contemporary concepts that are used to classify data with remarkable accuracy and efficiency. This paper contains a collection of research publications that utilized DM and ML techniques to diagnose diabetes. The survey's objective was to determine the study objective, diabetic type, data sets and technologies employed, as well as the results.

Keywords— Classification, data mining, diabetic disease, feature selection technique, machine learning

I. INTRODUCTION

The extended progress in advanced data and storage technology innovation has offered to raise of the rapidly growing amount of data and data storage in different types of media, for example, the World Wide Web, the information distribution centre and data-wide. While valuable information may be unclear in the data, it is not difficult to extract vast amount of data without groundbreaking hardware. In order to enrich such information, a new direction emerged in the 1980s called Data Mining with the help of the widespread current day computing gadgets [1]. With the aid of clinical studies and diagnostics, it is important to make the proper diagnosis of the patient for a best improvement. As a result, data identification is critical for health condition diagnosis. In the field of medical research and healthcare, there are many different types of computer-based decision support systems. These are computer-assisted decision-making technologies that assist doctors and other health practitioners in making accurate prescriptions.

II. RELATED WORK

The research review papers were selected on the foundation of the

- 1) Papers related to DM and ML Methods
- 2) Papers also related to Diabetes detection

Papers that were includes the parameters were picked and analyzed. The data was addressed in the article and provided in a tabular manner. The table 1 shows the detailed review of selected papers. The table includes

1. Serial Number,
2. Name of author and year,
3. Datasets
4. Algorithms
5. Outcomes,
6. Refernces

Table 1: Performances of existing work all ready done by different authors

| Sr . No . | Name of author and year | Dataset | Algorithms | Outcomes | Ref . |
|-----------|-------------------------|-------------------------------------------------------------|-----------------------------------------------------------------|--------------------------------|-------|
| 1 | Altamimi (2020) | diabetes in children detection | NB,RF, SVM,DT | SVM accuracy 76.4% | [2] |
| 2 | Kareem et al.(202) | Early Stage diabetes dataset | RF , MLP and RBFN | RBFN accuracy 98.8 % | [3] |
| 3 | Pethunachiyar (2020) | diabetes dataset | NB,LR,SVM and NN | SVM with kernel Polynomial 90% | [4] |
| 4 | Kaur et al. (2020) | diabetes | SVM | 75.3% accuracy | [5] |
| 5 | Zhang et al. (2020) | Tongue images of 296 diabetic subjects and 531 non diabetic | SVM,GA, PCA | SVM+PCA accuracy 78.77% | [6] |
| 6 | Perveen et al.(2016) | Canadian Primary Care Sentinel Surveillance Network | adaboost and bagging ensemble techniques using C4.5, Chi-square | ADABOOST perform better | [7] |
| 7 | Han et al. (2015) | Open dataset China Health and Nutrition Survey (CHNS) | SVM, RF , C4.5, NBTree, BP NN, SVM +C4.5, SVM +RF | SVM + RF has higher precision | [8] |
| 8 | Soliman et | Pima Indian | PSO LS- | Classificat | [9] |

| | | | | | |
|----|---------------------|------------------------------|-------------------|-----------------------------------------------------|------|
| | al.(2014) | Diabetes Dataset | SVM | ion accuracy of 97.833%. | |
| 9 | Kumari et al.(2013) | Pima Indian Diabetes Dataset | SVM | 78 | [10] |
| 10 | Uzer et al.(2013) | Pima Indian Diabetes Dataset | SVM+ABC Algorithm | Classification accuracy 79.29% for diabetes dataset | [11] |
| 11 | Giveki et al.(2012) | Pima Indian Diabetes Dataset | MI-MCS-FWSVM | 93.58% accuracy | [12] |

III. METHODOLOGY

The review work started with the study of background information of diabetic disease diagnosis using DM and ML techniques. In the present research work, we have reviewed the of following various DM techniques;

A. NEURAL NETWORK

The artificial neural network could be a mathematical model that we have a tendency to a while referred to as "neural network". This neural network form used the structure and process of the biological nervous network[13]. In diverse cases, artificial neural networks continuously amend once analyzing different types of parallel information processing unit. It is a tool for neural network modelling nonlinear application statistical data. We have a tendency to use this model to investigate complex data. The ANN method reviewed MLP and RBFN.

B. DECISION TREE

The decision tree is used in the form of a prediction model to go from observations about an item (representation in branches) to find out the target value of the item (representation in the leaves) by learning the decision tree. It is one of the predicted modelling approaches utilized in statistics, data processing and machine learning. Tree models [14] where the objective variable will take a unique set of values are called classification trees; in these tree structures, the leaves signify class labels and the branches represent these categories that lead those class labels. In decision trees, where intention variables are able to acquire continuous value (generally the genuine number) are referred regression trees. The Decision tree like RF, CART, DT have used the different authors.

C. STATISTICAL METHODS

Data analysis [15] is a tool to extracting the genuine result from the dissolution of the data. Data mining is the area of statistics that works in the direction of this goal. Whereas they will overlap, they are two very different procedures

that need different skills. Statistics represent the most piece of data mining which incorporates the whole method of data analysis. The statistical method like SVM and NB are used different authors in their works.

IV. CONCLUSION AND FUTURE SCOPE

For each of the datasets described above, the different classification algorithms, Decision Tree, Neural Network, statistical method used in various authors of their research work. The ability to use DM to deal with the avalanche of clinical data gathered from patients and created from diabetes monitoring and training is a great tool that may assist researchers and clinicians deliver better health care for people suffering by this modern-society disease. Through use of modified DM and ML, as well as a variety of optimization method, has shown encouraging results. In future alternative robust and computationally efficient model are going to develop and FST and Dimension Reduction Technique (DRT) is enforced in the alternative dataset.

REFERENCES

- [1] J. Han, M. Kamber, and J. Pei, Data mining: concepts and techniques, Third. Elsevier, 2012.
- [2] A. M. Altamimi, "Performance Analysis of Supervised Classifying Algorithms to Predict Diabetes in Children," J. Xi'an Univ. Archit. Technol., vol. XII, no. III, pp. 2010–2017, 2020.
- [3] A. Kareem, L. Shi, L. Wei, and Y. Tao, "A Comparative Analysis and Risk Prediction of Diabetes at Early Stage using Machine Learning Approach A Comparative Analysis and Risk Prediction of Diabetes at Early Stage using Machine Learning Approach," Int. J. Futur. Gener. Commun. Netw., vol. 13, no. 3, pp. 4151–4163, 2020.
- [4] G. A. Pethunachiyar, "Classification of diabetes patients using kernel based support vector machines," in 2020 International Conference on Computer Communication and Informatics, ICCCI 2020, 2020, pp. 22–25.
- [5] H. Kaur and G. Kaur, "Prediction of Diabetes Using Support Vector Machine," Int. J. Res. Eng. Appl. Manag., vol. 05, no. 02, pp. 470–473, 2019.
- [6] J. Zhang et al., "Diagnostic Method of Diabetes Based on Support Vector Machine and Tongue Images," Hindawi BioMed Res. Int. Res. Int., vol. 2017, 2017.
- [7] S. Perveen, M. Shahbaz, A. Guergachi, and K. Keshavjee, "Performance Analysis of Data Mining Classification Techniques to Predict Diabetes," in Procedia Computer Science, 2016, vol. 82, no. March, pp. 115–121.
- [8] L. Han, S. Luo, J. Yu, L. Pan, and S. Chen, "Rule extraction from support vector machines using ensemble learning approach: An application for diagnosis of diabetes," IEEE J. Biomed. Heal. Informatics, vol. 19, no. 2, pp. 728–734, 2015.
- [9] O. S.Soliman and E. AboElhamd, "Classification of Diabetes Mellitus using Modified Particle Swarm Optimization and Least Squares Support Vector Machine," Int. J. Comput. Trends Technol., vol. 8, no. 1, pp. 38–44, 2014.
- [10] A. Kumari and R. Chitra, "Classification Of Diabetes Disease Using Support Vector Machine," Int. J. Eng. Res. Appl., vol. 3, no. 2, pp. 1797–1801, 2013.
- [11] M. S. Uzer, N. Yilmaz, and O. Inan, "Feature selection method based on artificial bee colony algorithm and support vector machines for medical datasets classification," Sci. World J., vol. 2013, 2013.

- [12] D. Giveki, H. Salimi, G. Bahmanyar, and Y. Khademian, "Automatic Detection of Diabetes Diagnosis using Feature Weighted Support Vector Machines based on Mutual Information and Modified Cuckoo Search," *cornell*, **2012**.
- [13] M. Rambhajani, W. Deepanker, and N. Pathak, "A Survey on Implementation of Machine Learning Techniques for Dermatology Diseases Classification," *Int. J. Adv. Eng. Technol.*, **vol. 8, no. 2, pp. 194–202, 2015**.
- [14] M. He, D. Jianan, and Z. Sinian, "Kaggle Competition : Product Classification," *Kaggle Compet. Prod. Classif.*, **2015**.
- [15] M. Leshno, "Chapter 25 statistical methods for data mining," *no. Dm*, **2005**.

AUTHORS PROFILE

Dr. Sanat Kumar Sahu is working as an Assistant Professor in the Department of Computer Science, Govt. Kaktiya PG College, Jagdalpur (Bastar) Chhattisgarh, India. He has more than 11 years teaching Experience. His area of interest includes soft computing, machine learning, and data mining.. He has more than 20 research paper in national and international journals.

