IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A SURVEY ON DIABETES AND HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES

Tintu George, Dr. A. Hema

PhD Scholar, Associate Professor

Department of Computer Science, Department of Computer Application

Kongunadu Arts and Science College, Coimbatore, India, Kongunadu Arts and Science College, Coimbatore, India

Abstract: Diabetes is a group of metabolic diseases that is due to a high level of sugar in the blood for a long period of time. People with diabetes are also more likely to have many other conditions to raise the risk of heart disease. Diabetes and heart disease are the two most complex diseases that are globally affecting many peoples in society. According to WHO, the report says the diabetes patients list has been increasing day by day and due to diabetes 1.6 million deaths was caused in the year 2016 and by 2030 approximately 700 million peoples will be affected by diabetes worldwide. Therefore early prediction of diabetes is very necessary for that Data mining has played a main role in the field of diabetes and heart disease research. The main objective of this paper is to predict diabetes and heart disease which help doctors and many other medical practitioners to predict the disease before it occurs.

Index Terms **- Data mining, Diabetes, Heart disease prediction**

I. INTRODUCTION

Data Mining is the process of searching and extracting of knowledge from huge volumes of primary data to discover the patterns. Data Mining is used to discover knowledge from data and producing it in a structure that the humans can understand easily therefore Data mining is also called Knowledge Discovery in Data (KDD). KDD is a task of discovering the knowledge from the database. [3]. Patterns discovery, Predictions, Grouping, Actionable information are the few Data mining properties.

Data mining plays a vital role and it assists the researchers to extract knowledge from large data. Commonly used data mining techniques are association rules, clustering, classification, prediction, etc. also data mining techniques are used for many other applications in predicting the disease in the health industry. Prediction is a process that includes few variables or fields in the dataset to predict unknown values. This paper analyzes various data mining techniques for Diabetes and Heart disease predictions also to analyse the risk factor of heart disease based on diabetes data using ensemble learning.

1.1. DIABETES

The most common disease in day-to-day life is related to diabetes. Diabetes was one of the hazardous persistent diseases that lead to many other chronic diseases. Diabetes is defined as a group of metabolic confusions that are mainly caused by the high amount of glucose in the blood. World Health Organization says that diabetes patients occur in the world is most common in the developed countries also it shows a hike in coming decades [1]. Diabetes was common among adults mainly in the middle-aged it is due to changes in lifestyles also many children are also affected.

Two types of Diabetes they are type 1 and type 2. Type 1 diabetes is enforced for infusing insulin through medicines or injections. Type 2 diabetes creates insulin, but not effectively used by the body. Mostly 90-95% of adults are affected by type 2 diabetes. 12 million men and 11.5 women are affected with diabetes. Diabetes may also lead to many other diseases like heart disease.

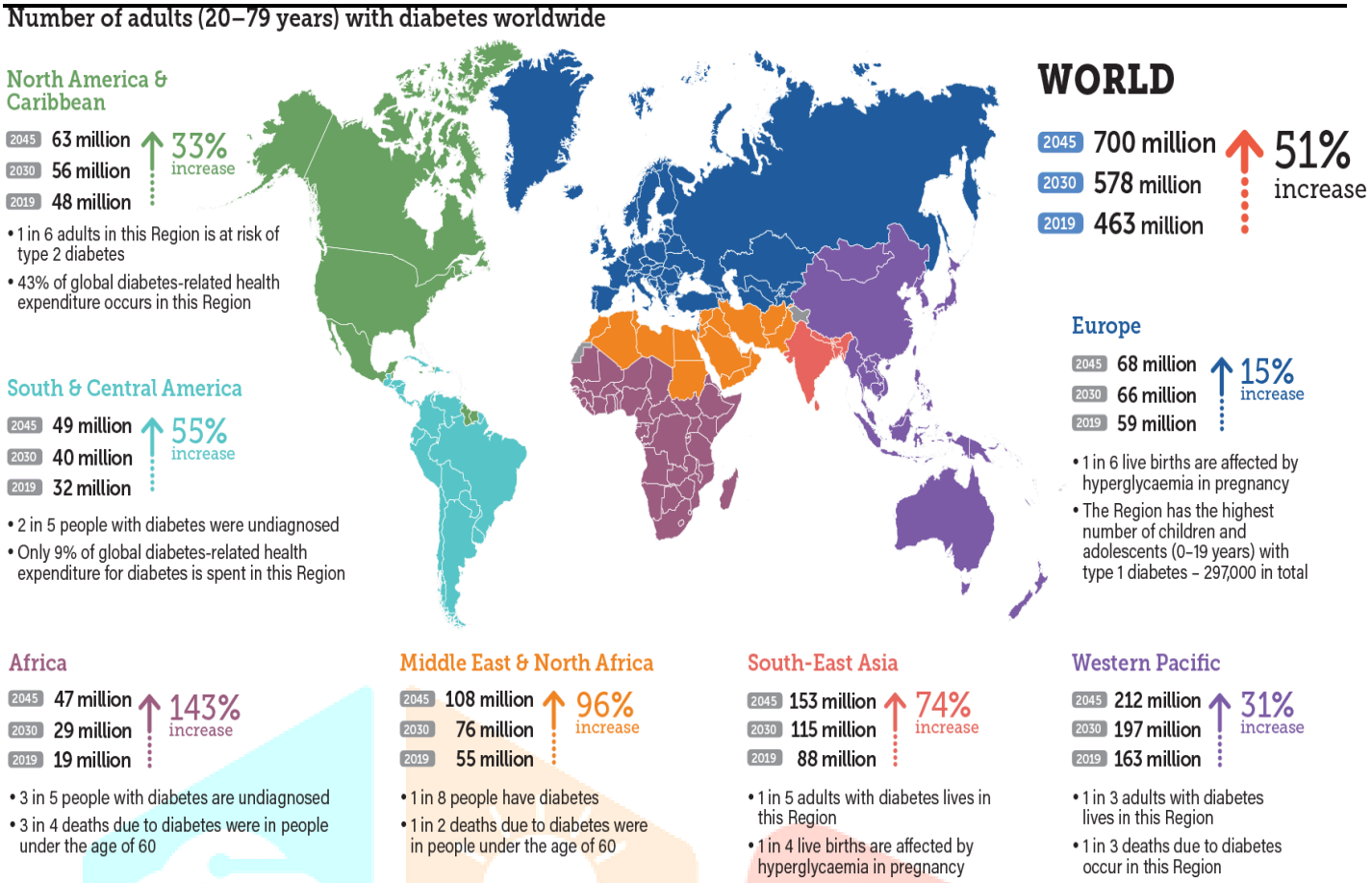

Figure 1: Global diabetes data report 2010-2045[17]

1.2. HEART DISEASE

Heart disease is the largest cause of death. Heart disease plays a vital role in the field of healthcare, mainly in cardiology. The operating system of the human body is the heart. If the heart is not functioning properly it will affect other human bodies[2]. The factor that affects heart disease is mainly diabetes, Family history, Smoking, age, High BP, Cholesterol, Improper diet, etc. Therefore, predicting diabetes and heart disease is needful to save human life[13].

II. LITERATURE SURVEY

N. Yuvar [7] has proposed work on health care systems to enhance the accuracy of diabetes prediction. The work was proposed for the implementation of machine learning algorithms in Hadoop clusters for diabetes prediction. . Hadoop cluster-based frameworks are used to supports the processing and storing of large datasets in a cloud environment. Random forest, decision tree, and naive Bayes algorithms are used. The results show that the Random forest produces high accuracy [80%] than decision tree and naive Bayes algorithm.

Rairikar et al, [5] proposed a work on the prediction of heart disease based on a genetic algorithm using a backpropagation algorithm. 13 attributes are used in the prediction of heart disease. The results show that it requires more time for the classification of heart diseases when compared with decision Tree and Naïve Bayes.

Suvarna, Sali, and Salmani [6] developed an algorithm for heart attack prediction. The authors combine data mining and modified PSO. The PSO algorithm was modified and new version was developed and it's called Constricted PSO. The proposed algorithm achieved 55% of classification accuracy which was better than normal PSO.

Amira,and Mona[8] proposed work for predicting diabetes using data mining techniques. The work was carried out by Clustering using the k- means algorithm. For the work, the authors have used a cluster model with 2733 instances within that 12 attributes are considered. The data was pre-processed to remove the noisy data, and missing values are replaced to increase processing time. The processed Patient dataset are utilized through the Business intelligence application to produce better results.

Talha Mahboob Alam, et.al [9] proposed work on diabetes mellitus prediction. This work had been represented various features among the association. PCA (Principal Component Analysis) algorithm was employed for the selection of features. The diabetes mellitus was carried out using 3 classification models. The models were identified as ANN, RF, and K-means. The results showed that the ANN has optimal performance than other classifiers with an optimal classification accuracy of 75.7%.

Ayman Mir,Sudhir N. Dhage[10] Proposed a work on predicting diabetes disease using WEKA tool and employed using Naive Bayes, Support Vector Machine, Random Forest and Simple CART algorithm. The result showed that SVM performed best in prediction with 0.7913 accuracy. Naive Bayes has 0.77 accuracy than Random Forest and Simple CART. Naive Bayes has less than SVM in training time. Simple CART has highest training time.

Samuel et al. [11] developed work on heart failure risk prediction using an integrated decision support system based on ANN and Fuzzy analytic hierarchy process (Fuzzy_AHP). The authors considered 13 attributes. (Fuzzy_AHP) technique was used to calculate the global weights based on their individual contribution. Fuzzy_AHP methods were examined by using online clinical dataset. The performance of this method shows 91.10% in accuracy.

Tuli et al. [13] has been proposed work on Ensemble Deep Learning for automatic heart disease prediction on intergraded IoT-based Fog computing environment. Fog provides service with IoT equipment and handles the cardiac patient's

data that was requested by users. Health-Fog can be programmed to give the best accuracy for various fog computing scenarios. For training and prediction, Deep learning methods with high precision are needed. Deep learning networks are utilized by new communication techniques and models to assembly a high-level accuracy with low latencies.

Table 1: Existing diabetes prediction using data mining techniques

S.NO	AUTHOR	METHODS	ALGORITHM	YEAR	ACCURACY
1	N. Yuvar [7]	Diabetes prediction	Random forest, decision tree, naive bayes	2017	80%
2	Talha Mahboob Alam, et.al [9]	Diabetes prediction	ANN, RF, and K-means	2019	75.7%
3	Ayman Mir et al [10]	Predicting diabetes	Naive Bayes, SVM, Random Forest and Simple CART.	2018	77%
4	Brisimi et al.[16]	Heart disease, Diabetes	*K*-LRT, a likelihood ratiotest-based method, and Joint Clustering and Classification (JCC)	2019	77.06%
5	Kamrul et al[17]	Diabetes prediction	ML classiers (k-nearest Neighbour, Decision Trees, Random Forest, AdaBoost, Naive Bayes, and XGBoost and (MLP)	2020	95%

Table 2: Existing heart disease prediction using data mining techniques

S.NO	AUTHOR	METHODS	ALGORITHM	YEAR	ACCURACY
1	Ali, Liaqat [15]	Prediction of heart disease	X^2 -statistical model and optimal DNN	2019	93.33%
2	Suvarna et al [6]	Heart attack prediction	Particle Swarm Optimization	2010	55%
3	Samuel et al. [11]	Heart failure risk prediction	ANN and Fuzzy Fuzzy_AHP	2017	91.10%
4	Tuli et al. [12]	Heart disease prediction	Ensemble Deep Learning	2020	89%
5	Vivekanandan et al[14]	Prediction of heart disease	Modified DE with Fuzzy AHP FFNN	2017	83%

III. REFERENCE

[1] Shaw, J. E., R. A. Sicree, and P. Z. Zimmet. (2010) "Global Estimates of the Prevalence of Diabetes For 2010 And 2030."
[2]. Sethilkumar Mohan, Chandrasegar Thirumalai, Gautam Shrivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," IEEE- Access – 2019.
[3]. Han, J., Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006.
[4]. Sethilkumar Mohan, Chandrasegar Thirumalai, Gautam Shrivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," IEEE- Access – 2019
[5] Rairikar A., Kulkarni V., Sabale V., Kale H. and Lamgunde A., Heart disease prediction using data mining techniques, Proceeding of International Conference on Intelligent Computing and Control(I2C2), 23-24 June 2017.
[6] Srinivas, K., "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques", IEEE Transaction on Computer Science and Education (ICCSE), 2010.
[7]. N. Yuvaraj ,K.R. SriPreethaa1 ―Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster‖, Springer – 7 December2017.
[8]. Amira Hassan Abed and Mona Nasr (2019), "Diabetes Disease Detection through Data Mining Techniques", Int. J. Advanced Networking and Applications 2019.
[9] Mahboob Alam, T., Iqbal, M., Ali, Y., Wahab, A., Ijaz, S., & Imtiaz Baig, T. et al. (2019). A model for early prediction of diabetes.
[10] Ayman Mir ,Sudhir N. Dhage‖ Diabetes Disease Prediction using Machine Learning on Big Data of Healthcare‖ (ICCUBEA)-2018.
[11] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, ``An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction," Expert Syst. Appl., vol. 68, pp. 163_172, Feb. 2017.

[12]. S. Tuli, N. Basumatary, S. S. Gill, M. Kahani, R. C. Arya, G. S.Wander, and R. Buyya, ``HealthFog: An ensemble deep learning based smart healthcare system for automatic diagnosis of heart diseases in integrated IoT and fog computing environments," Future Gener. Comput. Syst., , Mar. 2020.
[13] Senthilkumar Mohan, Chandrasegar Thirumalai, And Gautam Srivastava," Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" IEEE-ACESS, 2019.
[14] T. Vivekanandan and N. C. Sriman Narayana Iyengar, ``Optimal feature selection using a modi_ed differential evolution algorithm and its effectiveness for prediction of heart disease," Comput. Biol. Med., vol. 90, pp. 125_136, Nov. 2017.
[15] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, ``An automated diagnostic system for heart disease prediction based on_2 statistical model and optimally con_gured deep neural network," IEEE Access, vol. 7, pp. 34938_34945, 2019,
[16]. T. S. Brisimi, T. Xu, T. Wang, W. Dai, W. G. Adams, and I. C. Paschalidis, ``Predicting chronic disease hospitalizations from electronic health records: An interpretable classification approach," Proc.IEEE, 2018.
[17]. https://diabetesatlas.org/data/en/world/