Prediction of Lifestyle Diseases Such as Diabetes Using Supervised Machine Learning Approach

Animesh Tayal[1], Amruta K. Chimote[2], S.R. Tandan[3]

[1,2]Research Scholar, Dr. C.V. Raman University, Bilapsur (C.G.) India.

[3]Associate Prof., Dept of CSE, Dr. C.V. Raman University, Bilapsur (C.G.) India.

ABSTRACT

Lifestyle diseases *are defined as diseases linked with the way people live their life. This is commonly caused by alcohol, drug and smoking abuse as well as lack of physical activity and unhealthy eating. Diseases that impact on our lifestyle are heart disease, stroke, obesity and type II diabetes. Habits that detract people from activity and push them towards a sedentary routine can cause a number of health issues that can lead to chronic non-communicable diseases that can have near life-threatening consequences. We are preparing model for prediction of these diseases which will help user to change daily habits and get healthy lifestyle. We are generating primary data for Proposed system as per suggested by medical practitioner, our model will analyzed the data and predict whether person is prone to these diseases or not using supervised machine learning technique. Supervised machine learning help to classify the available data and finds the relation between them which is necessary to predict future consequences.*

Keywords: Lifestyle diseases, Supervised Classification, Machine learning.

I INTRODUCTION

Diabetes diseases commonly stated by health professionals or doctors as diabetes mellitus (DM), which describes a set of metabolic diseases in which the person has blood sugar, either insulin production inefficient, or because of the body cell do not return correctly to insulin, or by both reasons. The day is now to prevent and diagnose diabetes in the early stages.

According to the WHO (world health organization) report in Nov 14, 2016 in the world diabetes day "*Eye on diabetes*" reported 422 million adults are with diabetes, 1.6 million deaths, as the report indicates it is not difficult to guess how much diabetes is very serious and chronic.

Diabetes diseases damage different parts of the human body from those parts some of them are: eyes, kidney, heart, and nerves. *William's text book of endocrinology was* predictable that in 2013 more than 382 million populations in the world or all over the world were with diabetes or had diabetes. There are so many people's are died every year by diabetes disease (DD) both in poor and rich countries in the world.

According to the centers for disease control and prevention (CDCP) they give information for the duration of 9 ensuing years that is between 2001 and 2009 type II diabetes increased 23% in the United States (US). There are different countries, organization, and different health sectors worry about this chronic disease control and prevent before the person death.

Diabetes. Most in the current time diabetes is grouped into two types of diabetes, type I and Type II diabetes. Type I diabetes this type of diabetes in heath language or in doctors' language this type of diabetes also called Insulin dependent diabetes illness. Here the human body does not produce enough insulin. 10 % of diabetes caused by this type of diabetes.

Type II diabetes this type of diabetes. According to CDA (Canadian Diabetes Association) during 10 years, between 2010 and 2020, expected to increase from 2.5 million to 3.7 million. Therefore, as the above mentioned Diabetes diseases needs early prevention and diagnosis to safe human life from early death .By considering how much this disseises is very series and leading one in the world. Moloud [2] Algorithms which are used in machine learning have various powers in both classification and predicting.

This study follows different machine learning algorithms to predict diabetes disease at an early stage. Such as, Logical regression, SVM to predict this chronic disease at an early stage for safe human life.

II RELATED WORK

(a) Describe and explain different classification Algorithms using different parameters such as Glucose, Blood Pressure, Skin Thickness, insulin, BMI, Diabetes Pedigree, and age. The researches were not included pregnancy parameter to predict diabetes disease (DD). In this research, the researchers were using only small sample data for prediction of Diabetes. The algorithms were used by this paper were five different algorithms GMM, ANN, SVM, EM, and Logistic regression. Finally. The researchers conclude that ANN (Artificial Neural Network) was providing High accuracy for prediction of Diabetes.

(b) Machine learning algorithms are very important to predict different medical data sets including diabetes diseases dataset(DDD).in this study they use support vector machines(SVM) ,Logistic Regression ,and Naïve Bayes using 10 fold cross validation to predict different/varies medical datasets including diabetes dataset(DD) .the researchers' was compare the accuracy and the performance of the algorithm based on their result and the researchers conclude that SVM(support Vector Machine) algorithm provides best accuracy than the other algorithm which are mentioned on the above . The researchers were use those machine learning algorithm on a small sample of data in this study factors for accuracy were identified such factors are Data origin, Kind, and dimensionality.

(c) CART (classification and Regression Tree) was used for generating fuzzy rule. Clustering algorithm also was used (principal component Analysis (PCA) and Expectation maximization (EM) for pre-processing and noise removing before applying the rule. Different medical dataset (MD) was used such as breast cancer, Heart, and Diabetes Develop decision support for different diseases including diabetes. The result was CART (Classification and Regression tree) with noise removal can provide effective and better in health/diseases prediction and it is possible to safe human life from early death.

(d) This study was the new approach that used KNN algorithm by removing the outlier/OOB(out of bag) using DISKR(decrease the size of the training set for K-nearest neighbor .and also in this study the storage space was minimized. There for ,the space complexity is become less and efficient .after removing a parameters or instances which have less effect or factor the researchers got better accuracy.

(e) Feature selection is one of the most important steps to increase the accuracy. Hoeffding Tree(HT) ,multi-layer perceptron (MP),Jrip,BayeNet,RF(random forest),and Decision Tree machine learning Algorithms were used for prediction .From different feature selection algorithm in this study they were use best first and greedy stepwise feature selection algorithm for feature selection purpose . The researchers conclude that Hoeffding Tree (HT) provides high accuracy.

(f) In this study the researchers concentrate on different datasets including Diabetes Dataset (DD).The researcher were investigate and construct the models that are universally good and capability for varies/different medical datasets (MDs).the classification algorithm did not evaluate using Cross validation evaluation method .ANN,KNN, Navie Bayes,J48,ZeroR,Cv Parameter selection, filtered classifier ,and simple cart were some of the algorithm used in this study. From those algorithms Naïve Bayes provide better accuracy in diabetes dataset (DD) in this study. The two algorithms KNN and ANN provide high accuracy in other datasets on this study.

(g) By using CPCSSN(Canadian primary care sentinel surveillance Network) dataset and three machine learning methods to predict the diabetes Disses (DD) in early stage to safe human life at from early death .on this study Bagging ,Adaboost, and decision tree(J48) were used to predict the diabetes and the researcher was compare the result of those methods and concluded that Adaboost method was provide effective and better accuracy than the other methods in weka data mining tools

(h) Classification problems were identified in this study. one of the most problem in classification is data reduction .it has a vital role in prediction accuracy .to get better and efficient accuracy the data should be reduced as the researchers studied here. On this study PCA (principal component Analysis) for data pre-processing including data reduction for better accuracy. For prediction modified decision tree (DT) and Fuzzy were used for prediction purpose .finally it was concluded as to get better result the dataset should be reduced.

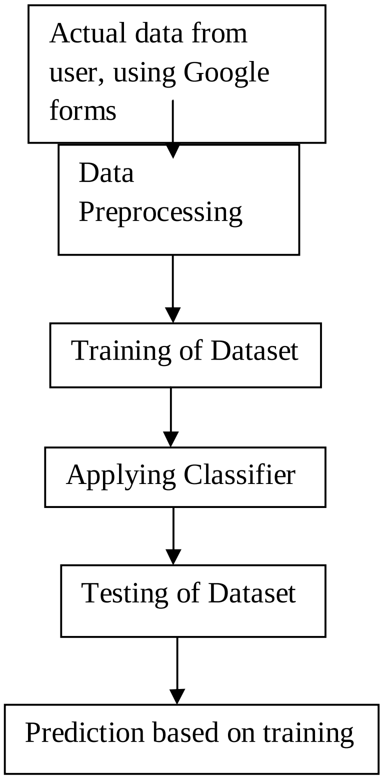
(i) In this study the performance of machine learning techniques were compared and measured based on their accuracy. The accuracy of the technique is varying from before pre-processing and after pre-processing as they identified on this study. This indicates the in the prediction of diseases the pre-processing of data set has its own impact on on the performance and accuracy of the prediction. Decision tree techniques provide better accuracy in this study before pre-processing to predict diabetes diseases. Random forest and support vector machine provide better prediction after pre-processing in this study using diabetes data set.

(j) K-means and Genetic algorithm used in this study for Dimension reduction in order to get better performance. The integration of support vector machine for prediction technique was used and provides better accuracy in small sample diabetes data set by selecting only five factors or parameters. 10 cross validation on this study used as evaluation method. finally reduced data set provide better performance than large dataset.

(k) In this study the researchers were use different data mining techniques to predict the diabetic diseases using real world data sets by collecting information by distributed questioner .in this study SPSS and weka tools were used for data analysis and prediction respectively .in this study the researchers compare three techniques ANN, Logistic regression, and j48 .finally it was concluded as j48 machine learning technique provide efficient and better accuracy.

(l) Oracle Data miner and Oracle Database 10g used for Analysis and storage respectively .the parameters or factors were identified in this study .the target variables were identified based on their percentage .this study concentrated on the treatment of the patient .the patient divided into two categories old and young based on their age and predict their treatment for both young and old diet control indicates high percentage on this study. The treatment predictive percentage done by support vector machine.

III PROPOSED METHODOLOGY



IV EXPECTED OUTCOME

User has to enter his information regarding height, weight, age, sex, type of exercise he/she performs and system will be able to predict whether that person can have diabetes in future or not? This will help user to take action at right time to prevent future mishap.

REFERENCES

[1] Song, Y., Liang, J., Lu, J., & Zhao, X. (2017). An efficient instance selection algorithm for k nearest neighbour regression. Neurocomputing, 251, 26-34.

[2] Abdar, M., Zomorodi-Moghadam, M., Das, R., & Ting, I. H. (2017). Performance analysis of classification algorithms on early detection of liver disease. Expert Systems with Applications, 67, 239-251.

[3] Zheng, T., Xie, W., Xu, L., He, X., Zhang, Y., You, M., ... & Chen, Y. (2017). A machine learning-based framework to identify type 2 diabetes through electronic health records. International journal of medical informatics, 97, 120-127.

[4] Mercaldo, F., Nardone, V., & Santone, A. (2017). Diabetes Mellitus Affected Patients Classification and Diagnosis through Machine Learning Techniques. Procedia Computer Science, 112(C), 2519-2528.

[5] Meza-Palacios, R., Aguilar-Lasserre, A. A., Ureña-Bogarín, E. L., Vázquez-Rodríguez, C. F., Posada-Gómez, R., & Trujillo-Mata, A. (2017). Development of a fuzzy expert system for the nephropathy control assessment in patients with type 2 diabetes mellitus. Expert Systems with Applications, 72, 335-343.

[6] Xu, W., Zhang, J., Zhang, Q., & Wei, X. (2017, February). Risk prediction of type II diabetes based on random forest model. In Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2017 Third International Conference on (pp. 382-386). IEEE.

[7] 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.

[8] Komi, M., Li, J., Zhai, Y., & Zhang, X. (2017, June). Application of data mining methods in diabetes prediction. In Image, Vision and Computing (ICIVC), 2017 2nd International Conference on (pp. 1006-1010). IEEE.

[9] Nilashi, M., bin Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). An Analytical Method for Diseases Prediction Using Machine Learning Techniques. Computers & Chemical Engineering.

[10] Balpande, V. R., & Wajgi, R. D. (2017, February). Prediction and severity estimation of diabetes using data mining technique. In Innovative Mechanisms for Industry Applications (ICIMIA), 2017 International Conference on (pp. 576-580). IEEE.

[11] Hashi, E. K., Zaman, M. S. U., & Hasan, M. R. (2017, February). An expert clinical decision support system to predict disease using classification techniques. In Electrical, Computer and Communication Engineering (ECCE), International Conference on(pp. 396-400). IEEE.

[12] Bashir, S., Qamar, U., Khan, F. H., & Naseem, L. (2016). HMV: a medical decision support framework using multi-layer classifiers for disease prediction. Journal of Computational Science, 13, 10-25.

[13] Mekruksavanich, S. (2016, August). Medical expert system based ontology for diabetes disease diagnosis. In Software Engineering and Service Science (ICSESS), 2016 7th IEEE International Conference on (pp. 383-389). IEEE.

[14] Rani, A. S., & Jyothi, S. (2016, March). Performance analysis of classification algorithms under different datasets. In Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on (pp. 1584-1589). IEEE.

[15] Pradeep, K. R., & Naveen, N. C. (2016, December). Predictive analysis of diabetes using J48 algorithm of classification techniques. In Contemporary Computing and Informatics (IC3I), 2016 2nd International Conference on (pp. 347-352). IEEE.

[16] Perveen, S., Shahbaz, M., Guergachi, A., & Keshavjee, K. (2016). Performance analysis of data mining classification techniques to predict diabetes. Procedia Computer Science, 82, 115-121.

[17] Kamadi, V. V., Allam, A. R., & Thummala, S. M. (2016). A computational intelligence technique for the effective diagnosis of diabetic patients using principal component analysis (PCA) and modified fuzzy SLIQ decision tree approach. Applied Soft Computing, 49, 137-145.

[18] Saravananathan, K., & Velmurugan, T. (2016). Analyzing Diabetic Data using Classification Algorithms in Data Mining. Indian Journal of Science and Technology, 9(43).

[19] Ramzan, M. (2016, August). Comparing and evaluating the performance of WEKA classifiers on critical diseases. In Information Processing (IICIP), 2016 1st India International Conference on (pp. 1-4). IEEE.

[20] Negi, A., & Jaiswal, V. (2016, December). A first attempt to develop a diabetes prediction method based on different global datasets. In Parallel, Distributed and Grid Computing (PDGC), 2016 Fourth International Conference on (pp. 237-241). IEEE.

[21] 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.

[22] Prajwala, T. R. (2015). A comparative study on decision tree and random forest using R tool. International journal of advanced research in computer and communication engineering, 4, 196-1.

[23] Vijayan, V. V., & Anjali, C. (2015, December). Prediction and diagnosis of diabetes mellitus—A machine learning approach. In Intelligent Computational Systems (RAICS), 2015 IEEE Recent Advances in (pp. 122-127). IEEE.

[24] Anand, A., & Shakti, D. (2015, September). Prediction of diabetes based on personal lifestyle indicators. In Next Generation Computing Technologies (NGCT), 2015 1st International Conference on (pp. 673-676). IEEE.

[25] Pavate, A., & Ansari, N. (2015, September). Risk Prediction of Disease Complications in Type 2 Diabetes Patients Using Soft Computing Techniques. In Advances in Computing and Communications (ICACC), 2015 Fifth International Conference on(pp. 371-375). IEEE.

[26] Nam, J. H., Kim, J., & Choi, H. G. (2015). Developing statistical diagnosis model by discovering principal parameters for Type 2 diabetes mellitus: a case for Korea. Public Health Prev. Med, 1(3), 86-93.

[27] Lukmanto, R. B., & Irwansyah, E. (2015). The Early Detection of Diabetes Mellitus (DM) Using Fuzzy Hierarchical Model. Procedia Computer Science, 59, 312-319.

[28] Kang, S., Kang, P., Ko, T., Cho, S., Rhee, S. J., & Yu, K. S. (2015). An efficient and effective ensemble of support vector machines for anti-diabetic drug failure prediction. Expert Systems with Applications, 42(9), 4265-4273.

[29] Kandhasamy, J. P., & Balamurali, S. (2015). Performance analysis of classifier models to predict diabetes mellitus. Procedia Computer Science, 47, 45-51.

[30] kumar Dewangan, A., & Agrawal, P. (2015). Classification of Diabetes Mellitus Using Machine Learning Techniques. International Journal of Engineering and Applied Sciences, 2(5), 145-148.

[31] Eswari, T., Sampath, P., & Lavanya, S. (2015). Predictive methodology for diabetic data analysis in big data. Procedia Computer Science, 50, 203-208.

[32] Mounika, M., Suganya, S. D., Vijayashanthi, B., & Anand, S. K. (2015). Predictive analysis of diabetic treatment using classification algorithm. IJCSIT, 6, 2502-2505.

[33] Nai-arun, N., & Moungmai, R. (2015). Comparison of classifiers for the risk of diabetes prediction. Procedia Computer Science, 69, 132-142.

[34] Wang, K. J., Adrian, A. M., Chen, K. H., & Wang, K. M. (2015). An improved electromagnetism-like mechanism algorithm and its application to the prediction of diabetes mellitus. Journal of biomedical informatics, 54, 220-229.

[35] Bashir, S., Qamar, U., Khan, F. H., & Javed, M. Y. (2014, December). An Efficient Rule-Based Classification of Diabetes Using ID3, C4. 5, & CART Ensembles. In Frontiers of Information Technology (FIT), 2014 12th International Conference on (pp. 226-231). IEEE.

[36] Lee, B. J., Ku, B., Nam, J., Pham, D. D., & Kim, J. Y. (2014). Prediction of fasting plasma glucose status using anthropometric measures for diagnosing type 2 diabetes. IEEE journal of biomedical and health informatics, 18(2), 555-561.

[37] Sankaranarayanan, S. (2014, March). Diabetic prognosis through Data Mining Methods and Techniques. In Intelligent Computing Applications (ICICA), 2014 International Conference on (pp. 162-166). IEEE.

[38] Varma, K. V., Rao, A. A., Lakshmi, T. S. M., & Rao, P. N. (2014). A computational intelligence approach for a better diagnosis of diabetic patients. Computers & Electrical Engineering, 40(5), 1758-1765.

[39] Li, L. (2014, November). Diagnosis of Diabetes Using a Weight-Adjusted Voting Approach. In Bioinformatics and Bioengineering (BIBE), 2014 IEEE International Conference on (pp. 320-324). IEEE.

[40] Aljumah, A. A., Ahamad, M. G., & Siddiqui, M. K. (2013). Application of data mining: Diabetes health care in young and old patients. Journal of King Saud University-Computer and Information Sciences, 25(2), 127-136.

[41] Kumari, V. A., & Chitra, R. (2013). Classification of diabetes disease using support vector machine. International Journal of Engineering Research and Applications, 3(2), 1797-1801.

[42] Meng, X. H., Huang, Y. X., Rao, D. P., Zhang, Q., & Liu, Q. (2013). Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. The Kaohsiung journal of medical sciences, 29(2), 93-99.

[43] Guo, Y., Bai, G., & Hu, Y. (2012, December). Using bayes network for prediction of type-2 diabetes. In Internet Technology And Secured Transactions, 2012 International Conference for (pp. 471-472). IEEE.

[44] Yıldırım, E. G., Karahoca, A., & Uçar, T. (2011). Dosage planning for diabetes patients using data mining methods. Procedia Computer Science, 3, 1374-1380.

[45] Al Jarullah, A. A. (2011, April). Decision tree discovery for the diagnosis of type II diabetes. In Innovations in Information Technology (IIT), 2011 International Conference on (pp. 303-307). IEEE.