DIABETES PREDICTION USING DEEP LEARNING

Prince Gupta
School of Computing Science and Engineering (B-Tech)
Galg0tias University
Greater Noida, Uttar Pradesh, India
prince2001gupta02@gmail.com

Shashi Kiran
School of Computing Science and Engineering (B-Tech)
Galgotias University
Greater Noida, Uttar Pradesh, India
S0431kiran@gmail.com

Abstract- Diabetes Mellitus is among the most dangerous illness, and many people suffer from it. Obesity, low physical activity,generic diabetes, lifestyle, bad eating, high cholesterol, and other variables can all contribute to diabetes mellitus. Diabetes patients are more likely to suffers heart disease, Kidney disease, stroke, vision issues, nerve damage, and other complications. Local hospitals procedure is to obtain relevant diabetes diagnostic information through a range of tests, and now to give suitable therapy based on the diagnosis. In the industry of health care, artificial intelligence plays a vital role. Large volume of data exist industry of health care. Using machine learning, one may examine large data-sets and discover hidden information, hidden patterns, and anticipate outcome based on the data. The accuracy of classification and prediction in present methods is not very great. In this study, we propose a diabetes prediction improved diabetes classification model that contain a few extrinsic causes responsible for diabetes, additionally regular factor such as glucose, BMI, Age, Insulin, and so on. When a fresh data-set is compared to an existing data-set, the accuracy of classification is tested. In addition, a diabetes prediction pipeline model was created with the purpose of enhancing classification accuracy.

Keyword : Diabetes, Deep Learning, Prediction, Data-set.

I. EVIDENCE BEFORE STUDY

According to National Diabetes Report, 2017[1], persons in distress with diabetes affects 30.3 million people in united states, or 9.4 percent of the total population. Moreover,china is worlds top country as indicates in Fig 1, India is the world's second most populous country, with 65.1 million diabetics as of 2013[2].In the United States, diabetic complications are routinely recognized as the top of death[3].Officials claim that figures from 2017,diabetes cause 8.8 percent of the worldwide population, by 2045, this number is predicted to increase by 9.9percent[4].Diabetes has been increasingly common in china in recent years of development, ans this has had a significant impact in everyone's life. Females have a higher rate of people affected by diabetes than males.according to official figures, almost 110 million people are impacted by this disease in 2017[5].

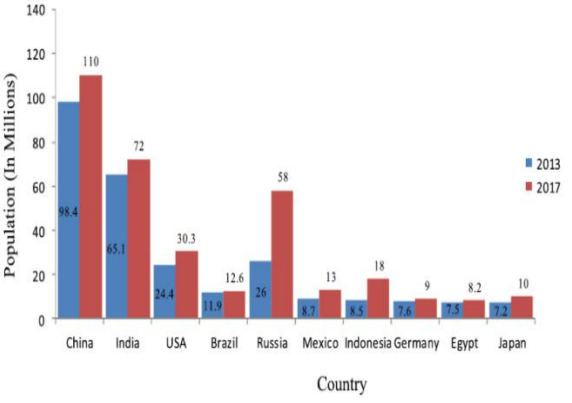

Fig.1: Globally, numbers of persons with diabetes.

II. INTRODUCTION

Diabetes is regarded as one of the most significant concerns facing the global health care community, and its impact is growing at an alarming rate. As a result, it is the world health organization (WHO eight)'s the leading cause of premature death worldwide in 2016[1]. According to global prevalence of diabetes, 1.6 million people die each year as the result of the diabetes[2]. In its first international report, WHO showed that the population with diabetes claimed that at the end of 2014, form 108 million (4.2 percent) to 0422 million(8.2 percent)[6]. On commemorate World Diabetes Day 2018, WHO has decided to join forces with partners from all cross the world to highlight the diabetes consequences. As state by the WHO, one out of every three adults is overweight or obese. The situation worse day by day. Diabetes is a long term disorder in which glucose(blood sugar) is not effectively digested(glucose is produced from the food we eat), causing blood sugar to rise over tolerable levels. The health care industry collects a large amount if data,but that data is unable to uncover undetected trends in order to make successful decisions[7]. A range of data extraction and machine algorithms can yield fast and reliable result due to various undiscovered mistakes and hidden patterns[8]. This research represent a collection of investigations conducted on the PIMA data-set employing DT,NB,ANN and DL data mining algorithms[9]. The representations of algorithms is rational and very well, while DL is a self-learn from a data technique that is currently commonly used for diabetes prediction[4]. a DL network is a method that makes advantage of ANN property to connect cells with a variety of representational layers [4,6,10]. DL increases precision of information representation by widening amount from the layer to the next, take into account[9]. The model has a 98.07 percent overall accuracy by incorporating DL into rapid minor tools, This provide a medical officials and practitioners with a well structured diabetic information base. More importantly, our goal is to lower costs and produce good result as compared to conventional methods[11].These machines learning techniques are used to increase the precision of the existing approaches. In terms of illness prediction,however,DL and ANN offer the best results since they are more dependable,robust,and accurate.The remainder of the research is structured as follows: The final section reviews recent significant work on diabetes prediction using data mining methods.In the fourth portion of the study, the data set description and recommended methods are described.The result and discussion section is covered in the fifth section. In the sixth section , the study concludes with the discussion on the future research.

III. LITERATURE SURVEY

Data mining techniques have surpassed prior methods in terms of predictability, accuracy, and precision. Machine learning, on the other hand, is an artificial intelligence approach for learning and teaching connections between nodes[12].he capacity of machine learning approaches to drive prediction models without substantial training on the underlying mechanisms is their most essential attribute.Data mining and machine learning approaches aid in the detection of buried data when using the cutting-edge approach[13]. In this section, we will examine numerous prior research to demonstrate the use of data mining appoaches in the driving prediction model, primarily for diabetes. Nasreen Samer EI Jerjawi and Samy S.Abu-Naser[14] used ANN(Artificial Neural Network) to create a diabetes prediction model that can be highly valuable for health care officials and

predictions.the Exceedingly dangerous complication of the stickness motivated the author. He devised an ANN model decreasing the training's error function. As a result, the computed average error function was 0.01 percent, and ANN had an accuracy of 87.3 percent. Sajida Perveen et al.[15]endorsed Ada Boost technology. An Ada Boost ensemble model outperforms bagging j48 in the categorization of diabetic patients.The author is inspired by the growing global effect of diabetes;as a result, diabetes mellitus prediction and prevention are becoming more significant in the healthcare community[16].Nahla H.Barakat et.al[17]introduced an intelligent SVM model for diabetes diagnosis. Diabetes is a serious health condition that effects individuals all over the world,according to doctors, and 80 percent of type 2 diabetes problems may be prevented if detected early.In the hypothetical case, many data mining and machine learning techniques were investigated for diabetes prediction.

The authors developed a module to the SVM model that converts the "black box" concept of an SVM into an understandable representation . On SVM classification, the machinemakes a highly accurate judgement. Stefan Ravizza et.al[18]show how data mining techniqyes may be used to healthcare and provide a model for estimating the probablity of unrelived illness. Instead of adopting the deep patient method, they employeed a healthcare-support, statistics and distinctive evaluation method on genuine data compared it to clinical evidence using straight algorithms. Miotto et al.[19] proposed Deep patient, an unsupervised technique for predicting the risk of numerous illness based in a variety of variables. In 2017, Carrera et al.[21] suggested a computer-assisted technique for diagnosing diabetic retinopathy using digital signals in retinal images. The main goal of this suggested technique is to categorize the location of non-proliferative diabetic retinopathyatany of the retinal picture. The main advantage of this method is that it is quite durable, yet it require consistency and accuracy to enhance for the documented method application. Diabetic retinopathy is a long term disorder that has involved into the most frequent lifestyle disease. Cardiac arrest, renal failure, malfunction stomach, and persistently high blood sugar levels, and other complications might occur after a long period of time with this disease. SVM and entropy approaches were employed by Huang et al.[22] to access the accuracy of three independent database (diabetic retinopathy debrecen, Vertebral column,and mammographic mass).

III. SYSTEM ARCHITECTURE

-

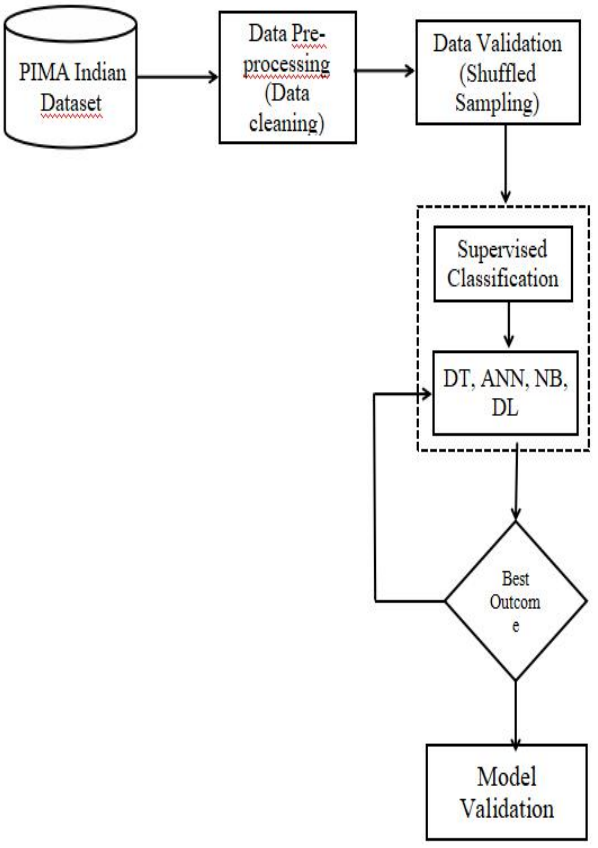

Fig.2: System Architecture

IV. IMPLEMENTATION

Module 1(Deep Learning Architecture):

Dl is a multi-layer graze perceptrone based system which also helps with ANN properties and therefore is trained using stochastic gradient descent with back-propagation. The system is divided into four layers, each of which replicates nodes and neurons and is oriented in the same manner (one-way connection). Each node has a one way link to another node and two hidden layers which each node uses its local data to train a copy of the worldwide design variables.

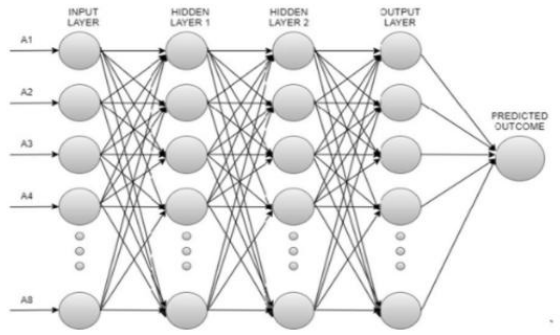

Fig.3: Deep Learning Architecture

Module 2(Decision tree) :

A decision tree (DT).is a network used in judgment that acts as a splitting rule for each distinct attribute. It's branched network that can be utilized both visually and explicitly to make decisions.Every characteristic is a branched network, and when a branch comes to a close, it construct a rule that splits values belonging to distinct classes. As the name implies, it is a tree-like structure that includes a decision at the conclusion, which is referred to as the tree,s leaf. The root is the most powerful attribute that may be used to forecast the outcome of rule formation.

Layer	Units	Type	Dropout	L1	L2	Mean	Momentum	Mean Weight	Weight RMS	Mean Bias	Bias RMS
1	8	Input	0.00%	-	-	-	-	-	-	-	-
2	50	Rectifier	0.00%	0.000010	0.000000	0.002799	0.000000	0.000422	0.193671	0.463731	0.052644
3	50	Rectifier	0.00%	0.000010	0.000000	0.015552	0.000000	-0.005877	0.145696	0.985745	0.024486
4	2	Soft-max	-	0.000010	0.000000	0.001496	0.000000	-0.050042	0.430350	0.000000	0.004501

Table 1: The parameters that were utilized to optimize of DL model.

Module3-(Artificial neural network):

In comparison to existing algorithms, another categorization approach is ANN . It is a machine learning method that generate extremely accurate outcomes. It is a mathematical concept based on biological neurons structural and function, as a result neural network is a collection of interconnected neurons. Mesh connectivity is the interconnected neurons is operational connectivity with each neuron having the same importance.

Methods	Accuracy obtained (in %)
Firefly and Cuckoo Search Algorithms	81%
Feed forward NN	82%
NB	79.56%
SVM	78%
LDA - MWSVM	89.74%
Neural Network with Genetic Algorithm	87.46%
K - means and DT	90.03%
PCA , K-Means Algorithm	72%
DL,ANN,SVM and DT (Highest accuracy achieved using DT)	98.07%

Table 2: A comparative assessment of relevant research work for detection of diabetes the Indian PIMA data-set.

V. RESULT AND DISCUSSION

In this study, outcome were accomplished by applying four classification algorithms (DL, ANN, NB, and DL) to demonstrate maximum precision in diabetes detection. Out of these four classifiers, DL and DT have the greatest accuracy (98.07%) and have been shown to be a valuable tool for early diabetes detection. On the prepared system, we employee the PIMA data-set and apply it to a DL approach. Furthermore, it can help in health care practitioners and function as a backup estimate when making decision based on retrieved attributes. Several scholars have previously used the PIMA data-set to predict diabetes utilizing various algorithms. Consequently, using their applicable methodologies, part of researcher's work has been correctly depicted. Table 3 summarize all of potential PIMA data-set research up to this point, and

our proposed method obtained the maximum correctness, 98.7%, on the Indian PIMA data-set.

VI. CONCLUSION AND FUTURE WORK

The purpose of this study was to create a diabetes threat assessment prediction model. Like previously said, diabetes disease affects a substantial portion of the human population. If left untreated, it will pose a serious threat to the entire world. Therefore in our previous research, we used the PIMA data-set to test the variety of classifiers and found that algorithms for machine learning and data mining can lowering risk factors while also enhancing efficiency and precision. As demonstrate in Table no. 1, the result of PIMA Indian data is better than other offered techniques on the same data-set utilizing data mining algorithms, like shown in Table 1. The correctness of four classifiers(DT, ANN, NB, and DL) is within the scope of 90-98%, which is significantly higher than existing methods. With an accuracy rate of 98.07 percent,the most productive and interesting of the classification techniques for diabetic research in DL. our future planes, we want to build a comprehensive system, such as an app or a website, that can use the well known DL algorithm to help health care professionals diagnose diabetes early.

VII. REFERENCE

1. “Global Report on Diabetes, 2016”. Available at: https://apps.who.int/iris/bitstream/handle/10665/204871/9789241565257_eng.pdf;jsessionid=2BC28035503CFAFF295E70CFB4A0E1DF?Sequence=1.

2. “Diabetes: Asia's 'silent killer'”, November 14, 2013”. Available at: www.bbc.com/news/world-asia24740288.

3. Mathers CD, Loncar D. Projections of global mortality and burdenof disease from 2002 to 2030. 2015;3(11). https://doi.org/10.1371/journal.pmed.0030442.

4. Swapna G, Vinayakumar R, Soman KP. Diabetes detection usingdeep learning algorithms. ICT Express. 2018;4(4):243–6. https://doi.org/10.1016/j.icte.2018.10.00.5.Elsevier B.V.

5. Wu H, et al. Type 2 diabetes mellitus prediction model based ondata mining. Informatics in Medicine Unlocked. 2018;10:1007. https://doi.org/10.1016/j.imu.2017.12.006. Elsevier Ltd.

6.Emerging T, Factors R. Diabetes mellitus , fasting blood glucose concentration , and risk of vascular disease : a collaborative meta-analysis of 102 prospective studies. The Lancet. 2010;375(9733):2215–22. https://doi.org/10.1016/S0140-6736(10)60484-9Elsevier Ltd.

7. Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. 2008 IEEE/ACS International Conference on Computer Systems and Applications 2008;108–15.https://doi.org/10.1109/AICCSA.2008.4493524.

8. Huang CL, Chen MC, Wang CJ. Credit scoring with a data mining approach based on support vector machines. Expert System Appl.2007;33(4):847–56. https://doi.org/10.1016/j.eswa.2006.07.007.

9. Zhang LM. Genetic deep neural networks using different activation functions for financial data mining. In: Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015; 2015.p. 2849–51. https://doi.org/10.1109/BigData.2015.7364099.

10. Grundy SM. obesity, Metabolic Syndrome , and Cardiovascular Disease.

2004;89(6):2595–600. https://doi.org/10.1210/jc.2004-0372.

11. Palaniappan S. Intelligent heart disease prediction system using data mining techniques, (march 2008). 2017. https://doi.org/10.1109/AICCSA.2008.4493524

12. Craven MW, Shavlik JW. Using neural networks for data mining.Futur Gener Comput Syst. 1997;13(2–3):211–29. https://doi.org/10.1016/s0167-739x(97)00022-8.

13. Radhimeenakshi S. Classification and prediction of heart disease risk using data mining techniques of support vector machine and artificial neural networks. In: 2016 International Conference on Computing for Sustainable Global Development (INDIACom);2016;3107–11.

14. El-Jerjawi NS, Abu-Naser SS. Diabetes prediction using artificial neural network. International Journal of Advanced Science and Technology.2018;121:55–64. https://doi.org/10.14257/ijast.2018.121.05.

15. Perveen S, Shahbaz M, Keshavjee K, Guergachi A. Metabolic syndrome and development of diabetes mellitus: predictive modelingbased on machine learning techniques, IEEE Access. IEEE. 2019;7:1365–75. https://doi.org/10.1109/ACCESS.2018.2884249.

16. Perveen S. et al. Performance analysis of data mining classificationtechniques to predict diabetes. Procedia Computer Science.2016;82:115–21. https://doi.org/10.1016/j.procs.2016.04.016 Elsevier Masson SAS.

17. Barakat N, Bradley AP, Barakat MNH. Intelligible support vector machines for diagnosis of diabetes mellitus. IEEE Trans Inf TechnolBiomed. 2010;14(4):1114–20. https://doi.org/10.1109/TITB.2009.2039485.

18. Ravizza S, Huschto T, Adamov A, Böhm L, Büsser A, Flöther FF,et al. Predicting the early risk of chronic kidney disease in patients with diabetes using real-world data. Nature Medicine. 2019;25(1):57–9. https://doi.org/10.1038/s41591-018-0239-8. Springer US.

19. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning forhealthcare: review, opportunities and challenges. Brief Bioinform.2017;19(6):1236–46. https://doi.org/10.1093/bib/bbx044.

20. Alade oM, Sowunmi oY. Information technology science.2018;724:14–22. https://doi.org/10.1007/978-3-319-74980-8.

21. Carrera EV, Carrera R. Automated detection of diabetic retinopathyusing SVM, 2017. pp. 6–9.

22. Huang YP, Nashrullah M. SVM based decision tree for medical knowledge representation. In: 2016 International Conference on Fuzzy Theory and Its Applications, I Fuzzy 2016; 2017https://doi.org/10.1109/iFUZZY.2016.8004949