International Journal of Scientific Research & Engineering Trends
Volume 4, Issue 6, Nov-Dec-2018, ISSN (Online): 2395-566X

A Survey Diabetes Prediction Using Machine Learning Techniques

M. Tech. Scholar Arvind Aada
Dept. of Computer Science &Engg
Astral Institute of Tech. & Research
Indore, Madhya Pradesh, India
Arvindaada.04@gmail.com

Prof. Sakshi Tiwari
Dept. of Computer Science &Engg.
Astral Institute of Tech. & Research
Indore, Madhya Pradesh, India
Sakshitiwari27@rediffmail.com

***Abstract*- Diabetes is a one of the main source of visual impairment, kidney disappointment, removals, heart disappointment and stroke. When we eat, our body transforms sustenance into sugars, or glucose. By then, our pancreas should discharge insulin. Insulin fills in as a "key" to open our cells, to enable the glucose to enter - and enable us to utilize the glucose for vitality. In any case, with diabetes, this framework does not work. A few noteworthy things can turn out badly – causing the beginning of diabetes. Type 1 and sort 2 diabetes are the most widely recognized types of the malady, yet there are additionally different sorts, for example, gestational diabetes, which happens amid pregnancy, just as different structures. This paper centers on ongoing advancements in AI which have had noteworthy effects in the recognition and analysis of diabetes.**

***Keywords*-Watchwords Diabetes, Type 1 Diabetes, Type 2 Diabetes, Insulin, Glucose, Machine Learning.**

I.INTRODUCTION

In understanding diabetes and how it creates, we have to comprehend what occurs in the body without diabetes. Sugar (glucose) originates from the sustenance's that we eat, explicitly starch nourishments. Sugar nourishments furnish our body with its primary vitality source everyone, even those individuals with diabetes, needs starch. Sugar nourishments incorporate bread, oat, pasta, rice, organic product, dairy items and vegetables (particularly bland vegetables). When we eat these nourishments, the body separates them into glucose. The glucose moves around the body in the circulation system. A portion of the glucose is taken to our mind to enable us to think obviously and work. The rest of the glucose is taken to the cells of our body for vitality and furthermore to our liver, where it is put away as vitality that is utilized later by the body.

All together for the body to utilize glucose for vitality, insulin is required. Insulin is a hormone that is created by the beta cells in the pancreas. Insulin works like a key to an entryway. Insulin appends itself to 'entryways' on the cell, opening the way to enable glucose to move from the circulation system, through the entryway, and into the cell. On the off chance that the pancreas can't create enough (insulin inadequacy) or if the body can - not utilize the insulin it produces (insulin opposition), glucose develops in the circulation system (hyperglycemias) and diabetes creates. Diabetes Mellitus implies elevated amounts of sugar (glucose) in the circulation system and in the pee. Signs or Symptoms of Diabetes Frequent Urination, Increased thirst, Increased yearning, Tired/Sleepiness, Weight misfortune, Blurred vision. Emotional episodes, Confusion and trouble concentrating, visit contaminations/poor mending.

1. Type 1 diabetes : In Type 1 diabetes the beta cells of the pancreas have been harmed or assaulted by the body's very own invulnerable framework (auto - insusceptibility). Because of this assault, the beta cells kick the bucket and are consequently powerless to make the required measure of insulin to move glucose into the cells, causing high glucose (hyperglycemias).

Type 1 diabetes happens in around 5 - 10% of those with diabetes and generally in individuals under 30 years old, however can happen at any age. The signs and side effects have a quick beginning and are typically exceptional in nature. As Type 1 diabetes is brought about by an absence of insulin, individuals need to supplant what the body can't deliver itself. As indicated by the most recent American Heart Association's Heart Disease and Stroke Statistics, around 8 million individuals 18 years and more established in the United States have type 2 diabetes and don't have any acquaintance with it.

Frequently type 1 diabetes stays undiscovered until indications end up extreme and hospitalization is required. Left untreated, diabetes can cause various wellbeing inconveniences. That is the reason it's so essential to both comprehend what cautioning signs to search for and to see a medicinal services supplier normally for routine wellbeing screenings [1]. PC Aided Diagnosis is a quickly developing unique region of research in medicinal industry. The ongoing scientists in

AI guarantee the improved precision of discernment and determination of sickness. Here the PCs are empowered to think by creating insight by learning. There are numerous sorts of Machine Learning Techniques and which are utilized to characterize the informational collections [1, 2]. They are Supervised, Unsupervised, Semi-Supervised, Reinforcement, Evolutionary learning, and profound learning calculations [2].

II. LITERATURE SURVEY

1. Machine Learning Techniques for Classification of Diabetes and Cardiovascular Diseases. Berina Et Al. [3]: Abstract: the outline of AI procedures in order of diabetes and cardiovascular infections (CVD) utilizing Artificial Neural Networks (ANNs) and Bayesian Networks (BNs). The near examination was performed on chosen papers that are distributed in the period from 2008 to 2017. The most normally utilized kind of ANN in chose papers is multilayer feed forward neural system with Lederberg-Marquardt learning calculation. Then again, the most generally utilized sort of BN is Naive Bayesian system which appeared most noteworthy precision esteems for arrangement of diabetes and CVD, 99.51% and 97.92% reflectively.

In addition, the estimation of mean precision of watched systems has demonstrated better outcomes utilizing ANN, which shows that higher probability to acquire increasingly exact outcomes in diabetes as well as CVD order is the point at which it is connected to ANN. In this paper creators intended to play out a survey of Artificial Neural Network and Bayesian Network and their application in order of diabetes and CVD illnesses. The object is to demonstrate the correlation of AI methods and to find the best alternative for accomplishing the most noteworthy yield exactness of the arrangement.

2. Methods- This paper speaks to the correlation of use of two AI methods, Artificial Neural Network and Bayesian Network in arrangement of diabetes and cardiovascular illnesses. Guided by understanding of specialists from the papers [3, 4] that additionally inspected AI methods yet in various field of studies, the writing audit was finished utilizing 20 distributed papers so as to acquire the applicable outcomes about diabetes and CVD characterization in the period from 2008 to 2017.

The review of Artificial Neural Networks utilized for order of diabetes and CVD (Table 1) demonstrates that the most regularly utilized sort of system in the two sicknesses is multilayer feed forward neural system. As preparing calculation, a large portion of creators of chose papers [5-11] have chosen to utilize Levenberg-Marquardt learning calculation.

Table 1 Ann Types for Classification of Diabetes and CVD

Paper	Type of ANN
DIABETES	
5	Multilayer feed forward neural network with sigmoid transfer function
6	Feed forward neural network using Lederberg-Marquardt method
7	Multilayer perception with back propagation learning algorithm and genetic algorithm
8	Two-laver feed forward neural network with sicrmoid function
9	Probabilistic neural network
CVD	
10	Multilayer neural network with statistical back propagation of error
11	Back propagation neural network with sigmoid transfer function
12	Feed forward neural networks with sigmoid transfer function using Liebenberg -Marquardt learning algorithm and SCG
13	Feed forward multilayer perception with sigmoid activation function trained with back propagation algorithm
14	MLP neural network with sigmoid transfer function

Each system utilizes blunder back propagation calculation to contrast the framework yield with the ideal yield esteem, and uses the determined mistake to coordinate the preparation. The distinction in the structures of these Networks is in exchange work where sigmoid exchange work is the most ordinarily utilized one.

Table2 Bn Types for Classification of Diabetes and Cvd

Paper	Type of BN
DIABETES	
15	Naive Bayesian Network
16	Naive Bayesian Network
17	Naive Bayesian Network
18	MLP + Naive Bayesian Network
19	Naive Bayesian Network
CVD	
20	Markov blanket estimation
21	Dynamic Bayesian network
22	Naive Bayesian network
23	Naive Bayesian network
24	Naive Bayesian network

The outline of Bayesian Networks utilized for characterization of diabetes and CVD (Table II)

demonstrates that the most regularly utilized sort of system in the two maladies is Naive Bayesian system. Credulous Bayesian systems are exceptionally straightforward BNs which are made out of coordinated non-cyclic charts with just a single in secret hub and a few watched hubs. This sort of BNs applies Bayes' hypothesis with solid freedom suppositions among highlights and does not require a long computational time for preparing which is its significant leverage.

3. Results- In the examination of utilization of Artificial Neural Network and Bayesian Network for grouping of diabetes and CVD, distinctive qualities for the system precision have been accomplished. The consequences of prepared ANN and BN for order of diabetes from chose papers [5-9, 15-19]. Creators saw that the exactness of diabetes order utilizing ANN shifts somewhere in the range of 72.2% and 99%. What's more, the exactness of diabetes arrangement utilizing BN fluctuates somewhere in the range of 71% and 99.51 %. As indicated by looked at results, the most noteworthy precision was accomplished in Bayesian Network yet in addition the littlest exactness was appeared Bayesian Network.

Conclusion: One of the greatest reasons for death worldwide is diabetes and cardiovascular infection. The early characterization of these maladies can be accomplished creating AI models, for example, Artificial Neural Network and Bayesian Network. In examination of mean precision of 10 logical papers about diabetes order and 10 papers about CVD characterization it was reasoned that the higher exactness was accomplished with ANN in the two cases (87.29 for diabetes and 89.38 for CVD). The utilized Naive Bayesian system, because of the supposition of autonomy among watched hubs, may be less exact than ANN approach. Thus, in understanding to got outcome it tends to be presumed that the higher plausibility to acquire better precision in characterization diabetes as well as CVD is the point at which it is connected to Artificial Neural Network.

4. Automatic Diagnosis of Diabetic Retinopathy, Dinu A.J Et Al. [25]: DME is one of the biggest reasons for visual misfortune in diabetes. There are different AI calculations that can be utilized to improve the exactness of determination of diabetic retinopathy. Iyer has played out a work to foresee diabetes malady by utilizing choice tree and Naive Bayes. J48 indicates 76.95% exactness by utilizing Cross Validation and Percentage Split Respectively [26]. Gullible Bayes presents 79.56% rightness by utilizing PS.

Calculations demonstrate the most astounding exactness by using rate split test. Sen and Dash created Meta-learning calculations for diabetes infection finding. Truck, Adaboost, Log boost, and evaluating learning calculations are utilized to foresee that understanding has diabetes [27]. From test results CART offers 78.64% precision. The Adaboost gets 77.86% precision. Logy boost offers the accuracy of 77.47%. Likewise Misclassification Rate of 21.35%, which is littler when contrasted with different systems. R. Catherine Silvia presented a component extraction strategy. This procedure is utilized to catch the worldwide attributes of the funds pictures and separate the ordinary from DME pictures. Diabetic macular edema location is done by means of managed learning. Malady seriousness is evaluated utilizing a rotational asymmetry metric by inspecting the symmetry of macular area [28]. A micro aneurysm is recognized utilizing Circular Hough Transform. The identification execution has particularity somewhere in the range of 74% and 90%.

5. a New Artificial Neural Networks Approach for Diagnosing Diabetes Disease Type II, Zahed Soltani Et Al. [29] Abstract: Diabetes is one of the real medical issues as it causes physical handicap and even passing in individuals. Along these lines, to analyze this risky malady better, strategies with least blunder rate must be utilized. Distinctive models of fake neural systems have the ability to determine this illness to have least mistake. Henceforth, in this paper creators have utilized probabilistic counterfeit neural systems for a way to deal with analyze diabetes illness type II. Creators exploited Pima Indians Diabetes dataset with 768 examples in their trials [30]. As indicated by this dataset, PNN is actualized in MATLAB. Moreover, expanding precision of diagnosing the Diabetes infection type II in preparing and testing the Pima Indians Diabetes dataset is the execution measure in this paper. At long last, creators inferred that preparation precision and testing exactness of the proposed technique is 89.56% and 81.49%, individually. Strategy:

In request to recognize diabetes and different illnesses, for example, heart infections, Parkinson's ailment, and lung malignant growth, having an informational collection is vital and important, since ANNs are prepared by these informational indexes and they can play out the analysis assignment. Thusly, in this paper, creators utilized Pima Indians Diabetes with 768 information test for diagnosing diabetes type 2. This informational collection comprises of 9 highlights for every datum test. Table (III) demonstrates these 9 highlights. As indicated by Table III, there are 9 highlights for every datum test. The initial 8 highlights are inputs, and the last component is the main yield. So as to characterize the 768 information tests, ninth component is utilized as it is grouped into two classes: class zero (solid) and class 1 (tolerant).

Table 3 Features of Pima Indians Diabetes for Diagnosing Diabetes Disease Type 2 [30].

No. of Attributes	Attributes	Descriptions and Attributes values
1	Number of Times Pregnant (NTP)	Numerical values
2	Plasma Glucose Concentration (PGC)	Numerical values
3	Diastolic Blood Pressure (DBP)	Numerical values in (mm Hg)
4	Triceps Skin Fold Thickness (TSFT)	Numerical values in mm
5	2-Hour Serum Insulin (2-HSI)	Numerical values in (mu U/ml)
6	Body Mass Index (BMI)	Numerical values in (weight in kg/(height in m)^2)
7	Diabetes Pedigree Function (DPF)	Numerical value
8	Age	Numerical values
9	Diagnosis of type 2 diabetes disease	Yes=1 No=0

Table 4 Statistical Analysis for Mean And Standard Deviation In Pima Indians Diabetes Data Set [30].

No. of Feature	Feature Name	Mean	Standard Deviation
1	Number of times pregnant	3.8	3.4
2	Plasma glucose concentration	120.9	32.0
3	Plasma glucose concentration	69.1	19.4
4	Triceps skin fold thickness	20.5	16.0
5	2-Hour serum insulin	79.8	115.2
6	Body mass index	32.0	7.9
7	Diabetes pedigree function	0.5	0.3
8	Age	33.2	11.3

The normal age of this informational index is somewhere in the range of 21 and 81 years. What's more, as indicated by Pima Indians Diabetes informational index which has 768 information tests, Table 2 demonstrates the Mean and standard deviation of the informational index. ANNs comprise of various models, for example, PNN, MLP, RBF, and GRNN. In this paper, Authors utilizes PNN demonstrate for diagnosing diabetes type 2. PNN display has a parallel structure and is exceptional for data grouping. As opposed to different ANNs, for example, MLP, PNN has a higher speed in preparing the information, and it discovers answers quicker than MLP. This model comprises of 3 layers: input layer, shrouded layer, and yield (aggressive) layer. The concealed layer is additionally called spiral base layer, as PNN show is a method of RBF demonstrate.

The normal age of this informational collection is somewhere in the range of 21 and 81 years. Likewise, as indicated by Pima Indians Diabetes informational index which has 768 information tests, Table 2 demonstrates the Mean and standard deviation of the informational index. ANNs comprise of various models, for example, PNN, MLP, RBF, and GRNN. In this paper, Authors utilizes PNN show for diagnosing diabetes type 2. PNN demonstrate has a parallel structure and is extraordinary for data grouping. As opposed to different ANNs, for example, MLP, PNN has a higher speed in preparing the information, and it discovers answers quicker than MLP. This model comprises of 3 layers: input layer, concealed layer, and yield (focused) layer. The shrouded layer is additionally called spiral base layer, as PNN show is a method of RBF display.

That the proposed model anticipated the class name of tuples accurately (above 80%). In excess of 80% of respondents (counting cardiologists and internists) who partook in the assessment session concur till unequivocally concurred that this examination pursued therapeutic techniques and that the outcome can bolster medicinal investigation identified with cardiovascular infection. Ends: The examination demonstrated that the proposed model accomplishes great execution for hazard level identification of cardiovascular disease.

REFERENCES

[1]. www.diabetesresearch.org/document.doc?id=284
[2]. D. Yu, and L. Deng, 2011, “Deep learning and its applications to signal and information processing,” IEEE Signal Process. Mag., vol. 28, no. 1, pp. 145-154.
[3]. Habibi, N., Hashim, S. Z. M., Norouzi, A., & Samian, M. R. (2014). A review of machine learning methods to predict the solubility of overexpressed recombinant proteins in Escherichia coli. BMC bioinformatics, 15(1), 134.
[4]. Langarizadeh, M., & Moghbeli, F. (2016). Applying Naive Bayesian Networks to Disease Prediction: a Systematic Review. Acta Informatica Medica, 24(5), 364.
[5]. Olaniyi, E. O., & Adnan, K. (2014). Onset diabetes diagnosis using artificial neural network. International Journal of Scientific and Engineering Research, 5(10).
[6]. Jaya lakshmi, T., & Santhakumaran, A (2010, February). A novel classification method for

diagnosis of diabetes mellitus using artificial neural networks. OSDE, 159-163. (2010)

[7]. Pradhan, M., & Sahu, R. K. (2011). Predict the onset of diabetes disease using Artificial Neural Network (ANN). International Journal of Computer Science & Emerging Technologies (E-ISSN: 2044-6004).

[8]. Sejdinovic, Dijana, et al. "Classification of Prediabetes and Type 2 Diabetes using Artificial Neural Network." Springer. CMBEBIH 2017.

[9]. Soltani, Z., & Jafarian, A (2016). A New Artificial Neural Networks Approach for diagnosing Diabetes Disease Type II. International Journal of Advanced Computer Science & Applications, 1(7), 89-94.

[10]. Atkov, O. Y., Gorokhova, S. G., Sboev, A G., Generozov, E. Y., Muraseyeva, E. v., Moroshkina, S. Y, & Cherniy, N. N. (2012). Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters. Journal of cardiology, 59(2), 190-194.

[11]. Olaniyi, E. O., Oyedotun, O. K., & Adnan, K. (2015). Heart diseases diagnosis using neural networks arbitration. International Journal of Intelligent Systems and Applications, 7(12), 72.

[12]. Colak, M. C. et. al., Predicting coronary artery disease using different artificial neural network modelslkoroner arter hastaliginin degisik yapay sinir agi modelleri lie tahmini. The Anatolian Journal of Cardiology (Anadolu Kardiyoloji Dergisi), 8(4), 249-255, (2008).

[13]. Can, M. (2013). Diagnosis of cardiovascular diseases by boosted neural networks.

[14]. Sayad, A T., & Halkarnikar, P. P. Diagnosis of heart disease using neural network approach. In Proceedings of IRF International Conference, 13th April-2014, Pune, India, ISBN (pp. 978-93).

[15]. 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.

[16]. Kumari, M., Vohra, R., & Arora, A (2014). Prediction of Diabetes Using Bayesian Network.

[17]. N. Sarma, S. Kumar, AK. Saini, A Comparative Study on Decision Tree and Bayes Net Classifier for Predicting Diabetes Type 2, 2014, ISSN: 2278-0882, ICRTIET-2014

[18]. Dewangan L. A., & Agrawal, P. Classification of Diabetes Mellitus Using Machine Learning Techniques

[19]. Nai-arun, N., & Moungmai, R. (2015). Comparison of Classifiers for the Risk of Diabetes Prediction. Procedia Computer Science, 69,132-142.

[20]. Elsayad, A, & Fakr, M. (2015). Diagnosis of cardiovascular diseases with Bayesian classifiers. 1. Comput. Sci., II (2), 274-282.

[21]. K. P. Exarchos, et al. Prediction of coronary atherosclerosis progression using dynamic Bayesian networks. IEEE EMBC, 2013.

[22]. D.S. Medhekar, M.P. Bote & Deshmukh, S. D., Heart disease prediction system using naive bayes. Int. J. Enhanced Res. Sci. Technol. (2013).

[23]. Patil, R. R., Heart disease prediction system using naive bayes and jelinek-mercer smoothing. International Journal of Advanced Research in Computer Science and Communication Engineering, (2014).

[24]. E. Miranda et. al., Detection of CYD Risk's Level for Adults Using Naive Bayes Classifier. Healthcare Informatics Research, (2016).

[25]. Dinu A.J., Ganesan R, Felix Joseph and Balaji V, "A study on Deep Machine Learning Algorithms for diagnosis of diseases." International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 17 (2017) pp. 6338-6346.

[26]. R. Catherine Silvia, R. Vijayalakshmi, 2013, "Detection of Non-Proliferative Diabetic Retinopathy in fundus images of the human retina", International Conference on Information Communication and Embedded Systems (ICICES).

[27]. Iyer A., Jeyalatha S. and Sumbaly R, 2015, "Diagnosis of Diabetes Using Classification Mining Techniques", International Journal of Data Mining & Knowledge Management Process (IJDKP), 5, 1-14.

[28]. Sen S.K. and Dash S, 2014, "Application of Meta Learning Algorithms for the Prediction of Diabetes Disease", International Journal of Advance Research in Computer Science and Management Studies, 2, 396-401.

[29]. Zahed Soltani and Ahmad Jafarian, "A New Artificial Neural Networks Approach for Diagnosing Diabetes Disease Type II." International Journal of Advanced Computer Science and Applications, Vol. 7, No. 6, 2016

[30]. Pima Indians Diabetes Data Set, https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diab etes [Last Available: February 2016].

[31]. E. Miranda et. al., Detection of CYD Risk's Level for Adults Using Naive Bayes Classifier. Healthcare Informatics Research, (2016).