International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Classification of Diabetics and Cardiovascular Diseases using Machine Learning Frameworks

[1]Dr. Sudharson D, [2]Dr. K Vignesh, [3]Ms. Srinithi. J, [4]Mr. Aman Kumar Dubey, [5]Ms. Abhirami. K, [6]Ms. Thrisha Vaishnavi K S

[1]AP/AI&DS, Kumaraguru College of Technology, Coimbatore, India
[2]AP/ MBA, Kumaraguru College of Technology
[3]UG Scholar, Dept of AI&DS, Kumaraguru College of Technology, Coimbatore, India
[4]AP/MBA, Kumaraguru College of Technology
[5]UG Scholar Dept of AI&DS, Kumaraguru College of Technology, Coimbatore, India
[6]UG Scholar, Dept of AI&DS, Kumaraguru College of Technology, Coimbatore, India

ABSTRACT—

This article provides an overview of machine learning techniques in the classification of diabetes and cardiovascular disease (CVD) using artificial neural networks (ANN) and Bayesian networks (BN). A comparative study was conducted on selected articles published between 2008 and 2017 and the most used ANNs in articles using feedforward neural networks using the Levenberg-Marquardt learning algorithm. The type of Bayesian network used is a naive Bayesian network that gives 99.51. The accuracy of the classification of diabetics and cardiovascular disease is% and 97.92%, respectively. The implicit accuracy of localized networks has been shown to give higher-end results when using ANN.

Keywords-machine learning,' diabetes,' cardiovascular disease; Artificial Neural Network, Bayesian Network

I. INTRODUCTION:

Machine learning has evolved from research activities to an era in which computers can collect knowledge to imitate the brain. The first effort at ML began in 1952, when Arthur Samuel developed the first chess gameplay program to acquire enough skills to become a world champion [1]. Then, in 1957, Frank Rosenblatt created an electronic device that could be learned to solve complex problems by mimicking the processes of the human brain [2].

With ML, PC applications can search from data, understand common styles, and make decisions primarily based on accumulated knowledge, which poses a problem with the integrity of used medical databases. There is no [12]. Classification learning techniques are better known in healthcare because they are associated with normal life problems and the most used techniques are ANN and BN [15].

The application of ml in the disease class is very common, and researchers are receiving more attention in building systems that facilitate the tracking and diagnosis of diabetes and cardiovascular disease. According to the World Health Organization (WHO), diabetes and cardiovascular disease (CVD) are one of the top 10 causes of global decline [8]. A January 2017 study confirmed that CVD was the main motivation behind the end of global illness. The world's largest murderer is at the top of the list of top 10 causes of death over the last 15 years, with 15 million causes of death in 2015 [15]. Meanwhile, according to WHO's first global report on diabetes, the number of adults with diabetes increased from 108 million to 422 million between 1980 and 2014, and the number of diabetic victims. Increased from less than 1 million to 1 between 2000 and 20151, 6 million people [12]. Incidence and mortality from diabetes and cardiovascular disease indicate the need for early triage of potential patients who have successfully developed machine learning models. These models enable richer and more complex data analysis to obtain more accurate results and make better real-time decisions without human intervention [7].

This study was designed to study artificial and Bayesian neural networks and their applications in the classification of diabetes and CVD disorders. The purpose is to show a comparison of these machine learning techniques and find the best option for classification accuracy [9].

II. METHODS:

The application of two machine learning techniques, Artificial Neural Network, and Bayesian Network, in the classification of diabetes and cardiovascular disorders, are compared in this research. The literature study was conducted using 20 published papers to gain relevant data on diabetes and CVD classification from 2008 to 2017. It was guided by the experience of researchers from papers that also assessed machine learning approaches in various fields of study [4].

Each category compares the outcomes of five separate articles. Thus, among the 20 papers chosen, 5 papers represent diabetes classification using ANN, 5 papers represent CVD classification using ANN, 5 papers represent diabetes classification using Bayesian Network, and 5 papers represent CVD classification using Bayesian Network.

A. Artificial Neural Network

To classify the input data into the desired output, an artificial neural network employs supervised learning. It comprises weighted interconnections between artificial neurons that control the influence of the accompanying input signals [3]. In Figure 1, you can see how the ANN uses supervised learning to classify input characteristics for diabetes or cardiovascular disease.

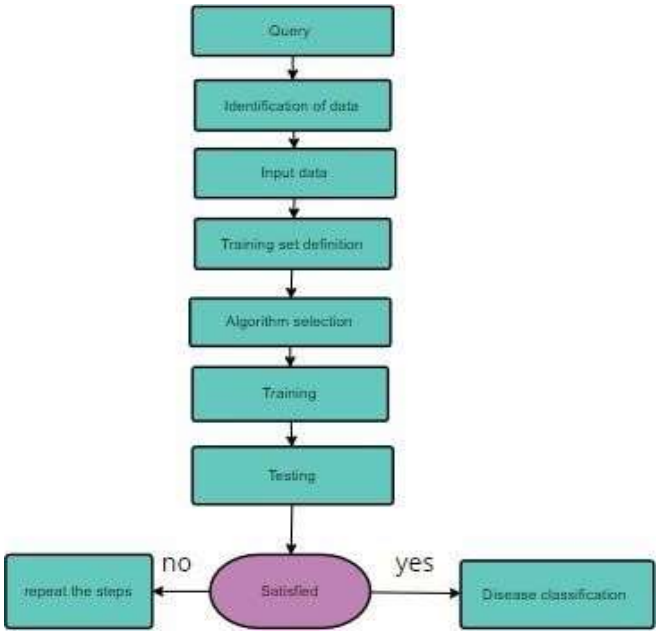

Fig 1: Classification of input into the desired output.

The first step in using ANNs to classify diabetes or cardiovascular disease is to collect and define the data to be fed into the network. The network is trained using a predefined training dataset and a predefined training process. Following the training procedure, the ANN is put to the test to see if it successfully classifies the condition [5]. Out of the 20 selected publications, 10 presented the results of ANN-based diabetes and cardiovascular classification. Table 1 shows the variety of neural networks that have been used to classify diseases. Out of the 20 publications chosen, ten present the findings of ANN-based diabetes and CVD classification. Table 1 shows the many types of neural networks that have been employed to do disease classification.

Table 1: Artificial neural network types used in classification

Type of ANN
DIABETES
Multilayer feedforward neural network with sigmoid transfer function
Feedforward neural network using Levenberg-Marquardt method
Multilayer perceptron with backpropagation learning algorithm and genetic algorithm
Two-layer feedforward neural network with sigmoid function
Probabilistic neural network
CVD
Multilayer neural network with statistical backpropagation of error
Backpropagation neural network with sigmoid transfer function
Feedforward neural networks with sigmoid transfer function using Levenberg -Marquardt learning algorithm and SCG
Feedforward multilayer perceptron with sigmoid activation function trained with backpropagation algorithm
MLP neural network with sigmoid transfer function

The most widely employed form of network in both diseases is a multilayer feedforward neural network, according to an overview of Artificial Neural Networks used for diabetes and CVD categorization (Table 1). Most writers in the selected articles [1] chose the Levenberg- Marquardt learning algorithm as their training algorithm. Each network compares the system output to the desired output value using the error backpropagation technique, and the

calculated error is used to steer the training [12]. The difference between these networks' topologies lies in the transfer function, with the sigmoid transfer function being the most widely utilized [10].

B. Bayesian Network

Bayesian networks (BNs) are probabilistic graphical models that can be used to reason about uncertainty. This model uses a directed rotation graph to represent a group of random variables (discrete or continuous), where arcs reflect direct links between them and their conditional dependencies. [2].

BN's use of algorithms based on probability theory is one of the main reasons why scientists are interested in using it to classify diabetes or CVD This theorem is clearly applicable to classification and regression issues [4].

Out of the 20 papers chosen, ten present the results of Bayesian Network-based diabetes and CVD classification. Table 2 shows the many forms of Bayesian networks that have been used to classify the diseases described.

Type of BN
DIABETES
Naïve Bayesian Network
Naïve Bayesian Network
Naïve Bayesian Network
MLP + Naïve Bayesian Network
Naïve Bayesian Network
CVD
Markov blanket estimation
Dynamic Bayesian network
Naïve Bayesian network
Naïve Bayesian network
Naïve Bayesian network

Table 2: Bayesian Networks types used in classification

The most widely used form of network in both conditions is the Bayesian network as an overview of the Bayesian network used to classify diabetes and cardiovascular disease [11]. Naive Bayesian networks (NBNs) are very basic BNs consisting of directed ac graphs with a single invisible node and many observable nodes [9]. This version of BN uses Bayes' theorem with strong assumptions about the independence of features and does not require too much computational time for training, which is one of its main advantages [2].

III. RESULTS

When evaluating the accuracy of an Artificial Neural Network with a Bayesian Network for diabetes and CVD class, diverse outcomes have been obtained. From decided on publications, Fig. 2 depicts the consequences of educated ANN and BN for diabetes categorization. The accuracy of diabetes class the usage of ANN tiers from 72.2% to 99 % at the left side. The accuracy of diabetes class the usage of BN tiers from 71% to 99.51 % at the proper side.

Fig 2: Accuracy score of classification of Diabetes using BNs.

According to the results, Bayesian network has the best accuracy but also has the lowest accuracy. Figure 3 illustrates the accuracy values of the two networks when aligned from lowest to highest to get more information from this comparison.

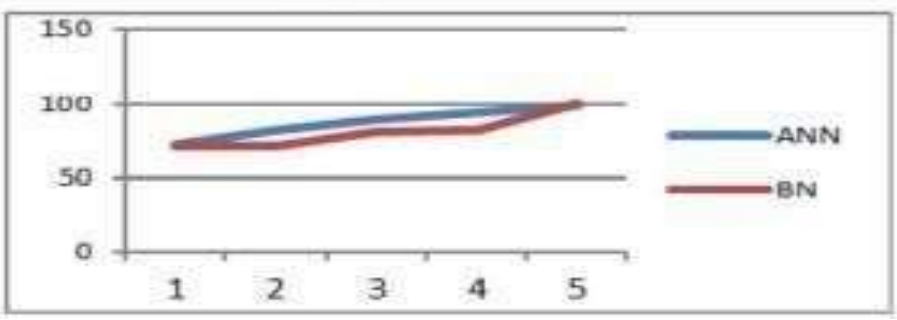

Fig 3: Accuracy Comparison of ANN and BNs for diabetes.

Fig 3 illustrates two curves, with blue representing ANN and red representing BN. Although the BN curve has the best accuracy, the ANN curve has higher accuracy in many cases than the BN curve. When we compare the mean accuracy of ANN and BN, we find that ANN has an accuracy of 87.29% and BN has an accuracy of 80.98%, indicating that ANN has a good chance of achieving higher accuracy in classifying diabetes.

Results of ANN and BN trained to classify CVD from selected articles are displayed in Fig. 4. The accuracy of CVD classifiers using ANNs varies between 80% and 95.91%, as seen on the left side. Accuracy of CVD classification using BN varied from 78% to 97.92% on the right side.

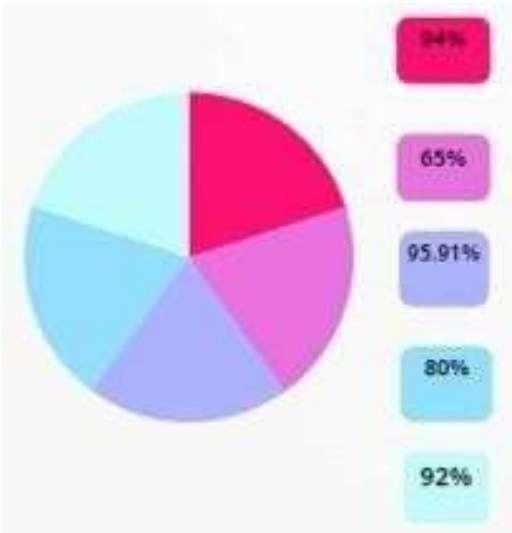

Fig 4.1: Accuracy Score of classification of CVD using ANN

Fig 4.2: Accuracy score of classification using BNs

According to Fig.4, the Bayesian Network acquired the best accuracy as well as the lowest accuracy for CVD classification. The accuracy values of both types of networks are lined up from lowest to highest and their values are shown in Fig. 5 for a clearer overview of this comparison.

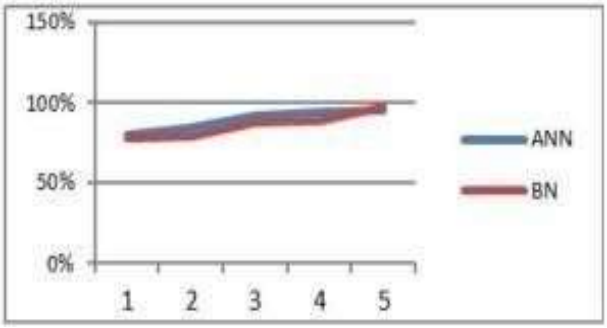

Fig 5: Accuracy comparison of ANN and BNs for CVD.

When these two curves are examined, it can be determined that, while BN has the maximum accuracy, ANN has greater accuracy in more cases than BN [9]. In addition, when the mean accuracy values are compared, ANN accuracy of 89.38 percent is greater than BN accuracy of 86.49 percent.

Furthermore, if we use the formula to estimate population standard deviation (a), we get:

$$\sigma = \sqrt{\frac{1}{N}\sum_{t=1}^{N}[x - \bar{x}]^2}$$

This will be able to obtain the data shown in Table 3

Table 3: Standard deviation values of ANN and BNs for diabetes and CVD

σ	ANN	BN
DIABETES	9,37	10,33
CVD	5,96	7,36

When comparing the standard deviations of ANN and BN for diabetes and cardiovascular disease, the highest value was obtained when BN was used for both. This suggests that in BN, the variation from the mean is larger. Therefore, it can be inferred that when the classification of diabetes and cardiovascular disease is performed by the ANN, there is a greater chance of obtaining more accurate and reliable results.

IV. CONCLUSION

Diabetes and cardiovascular disease are the two leading causes of death worldwide. Machine learning techniques such as Artificial Neural Networks and Bayesian Networks can be used to classify some disorders very early on. When comparing the mean accuracy of ten scientific articles on the classification f diabetes and ten scientific articles on the classification of cardiovascular disease, it was found that the ANN obtained higher accuracy in both cases. (87.29 for diabetes and 89.38 for cardiovascular disease). Due to the assumption of independence between observed nodes, the Bayesian network used alone may be less accurate than the ANN technique. As a result of the acquired results, it can be stated that when Artificial Neural Network is used to classify diabetes and/or CVD, there is a greater chance of obtaining higher accuracy.

V. REFERENCES

[1] Sudharson D, Divya P, Rajan DD, Ratheeshkumar AM. Performance Analysis of Enhanced Adaboost Framework in Multifacet medical dataset. NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO. 2021 Nov 7:1752-6.

[2] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[3] Sudharson D, Darshana S, Rohith SN, Priyanka K, Fathima A, Abdulla FM. Self-Reliant Dimensionality Reduction That Uses Improved Pareto Distribution PCA Framework. In2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA) 2021 Oct 8 (pp. 1-5). IEEE.

[4] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[5] Sudharson, D Prabha, D. A novel machine learning approach for software reliability growth modelling with pareto distribution function. Soft Comput 23, 8379–8387 (2019). https://doi.org/10.1007/s00500-019-04047-7

[6] S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, [6] G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271– 350.

[7] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[8] Sudharson, D Prabha, D "Improved EM algorithm in software reliability growth models" International Journal of Powertrains, Vol.9 Issue.3, pp.186-199, 2020,Inderscience Publishers (IEL).

[9] Dr.B.Arunkumar, D.Sudharson, "A Novel Approach for Boundary Line Detection using IOT During Tennis Matches", Advancement of Electrical, Information and Communication Technologies for Life Application, Volume.13, Issue.4, pp.243-246 2020.

[10] Rana Özakıncı, Ayça Tarhan 2018, “Early software defect prediction: A systematic map and review” Journal of Systems and Software, Vol. 144, Pp. 216-239 Nat. Volatiles & Essent. Oils, 2021; 8(5): 1752 - 1756 [12] Ramakanta Mohanty, V. Ravi, M.

[11] Sudharson, D., Prabha, D. "Hybrid software reliability model with Pareto distribution and ant colony optimization (PD–ACO)" International Journal of Intelligent Unmanned Systems, 2049-6427, 2020Emerald Publishing Limited

[12] R. Patra 2013, “Hybrid intelligent systems for predicting software reliability” Applied Soft Computing, Vol. 13, No. [12] 1, Pp. 189- 200

[13] Syed Wajahat Abbas Rizvi, Vivek Kumar Singh, Raees Ahmad Khan 2016, “Fuzzy Logic Based Software Reliability Quantification Framework: Early-Stage Perspective (FLSRQF)” Procedia .

[14] Sudharson D, Dr Prabha D. A contemporary survey on the aspects of software reliability growth models via size of the software, International Journal of Pure and Applied Mathematics, 119, Special Issue, 1253-1271, 2018.

[15] Qian Yuexia, Gu Weijie 2012, “The Research on Reliability Optimization of Software System Based on Niche Genetic Algorithm” AASRI Procedia, Vol. 1, Pp. 404- 409

[16] Sudharson D, Dr Prabha D. A REVIEW ON THE PROMINENT FACTORS OF SOFTWARE TESTING FOR OPTIMIZED OUTPUT THROUGH DATA ANALYTICS, International Journal of Pure and Applied Mathematics, vol. 115, issue. 6, pp no. 95-106,2017.

[17] Ratheesh Kumar A.M, Dr. Sudharson D, P.Divya, Dr. Sakthi Govindaraju, "A NOVEL AI AND RF TUTORED STUDENT LOCATING SYSTEM VIA UNSUPERVISED DATASET", Turkish Journal of Physiotherapy and Rehabilitation, Vol.32, Issue.2, pp.882-887 2021.