2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) | 979-8-3503-9926-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICACITE57410.2023.10182961

Artificial Neural Network Based Categorization of Diabetes Mellitus and its Comparison with K-Nearest Neighbors Algorithm

Raju Ranjan
Professor, Department of CSE
Galgotias University
Greater Noida, Uttar Pradesh, India
ranjanr.cs@gmail.com

Dr. Sachin Gupta
Chancellor
Department of Management
Sanskriti University
Mathura, Uttar Pradesh, India
chancellor@sanskriti.edu.in

Sachin Sharma
Assistant professor
Department of CSE
Dev Bhoomi Uttarakhand University
Dehradun, Uttarakhand, India.
socse.sachin@dbuu.ac.in

Sura Rahim Alatba
Computer Science Department
Al-TurathUniversiy College
Baghdad, Iraq
sura.raheem@turath.edu.iq

Dr. Anusha Sreeram
Assistant Professor
Operations & IT
ICFAI Business School (A Constituent of IFHE University)
Hyderabad, India
seeramanusha@gmail.com

Sowmya Fernandez
Assistant Professor
Computer Science
A.P.C. Mahalaxmi College for Women
Thoothukudi, India
sfsoris@gmail.com

Abstract: - **The categorization of diabetes mellitus in humans with a Novel Artificial Neural Network (ANN) algorithm and it'sassessment with K-Nearest Neighbors to improve the performance in terms of precision and accuracyis proposed in this paper. Kaggle, a repository for machine learning algorithms has been used as a source of Pima Indian diabetes dataset. This dataset having 8 attributes in 6 different categories are considered as input attributes for the classifiers. 20 samples have been used for the testing of both the Novel artificial neural network algorithm and K-nearest neighborsAccuracy and precision of Artificial neural network algorithm is calculated and compared with K-nearest neighbors algorithm. The significance value is less than 0.05. The G power is taken as 0.8. In diabetes mellitus classification of the considereddataset, the proposed work shows that performance of the Novel artificial neural network algorithm is better as compared to the K-nearest neighbors algorithm.**

Keywords: - Diabetes mellitus ,Novel artificial neural network, Classification of diabetes,K-nearest neighbors algorithm, Machine learning,.

I. Introduction

Diabetes is a category of metabolic illnesses in which a patient has excessive blood sugar levels as a result of either insufficient insulin production by the pancreas or improper insulin response by cells. Problems of Diabetes comes because ofmacro and micro vascular injury both. Thedanger of cardiovascular illness increases by three to four times due to Diabetes. Different type of Injuries, failure, and malfunction of numerous organs likeeyes ,kidney, nerves, core and blood vessels are due to Diabetic hyperglycemia[1]. Diabetes management and artificial pancreas systems are the key areas of medical domain where applications of the study isused [2].

In the past few yearsseveral research papersarepublished on a variety of diabetes diseases using machine learning techniques for diabetes detection. During 2015 to 2021 approximately 1140 articles published in google scholar and 10 articles in IEEE Xplore. In the past few years, surveys of for diabetes illnessusing machine learning algorithms were explored frequently. Diabetes mellitus occurs when we have low amounts of insulin and do not produce insulin properly [3]. Various fragments of the body which jeopardyhypertension, kidney, heartand other eye problems due to Diabetes mellitus. The researcher [4] and it also can be recognized by losing weight, height and lack of insulin [5]. Most people are affected by Type-2 disease [6]. This will lead to blurred vision, tiredness, weight loss, hunger [7]. Several specific data-driven models are proposed to predict the blood glucose levels such as regression models, support vector machines [8]. From the past few years to cure diabetes, Artificial Intelligence (AI) was usedmostly[9]. Various algorithms based on AI like machine learning, fuzzy logic, deep learning and combination of any two or morealgorithms are used to solve the problems of diabetes[10].

Majority of the present research are proposed with reference to the data collected from a specific national group. One of the biggest limitation of the published research articles is the use of limited data and the subjects considered for the detection of diabetes is limited. A large dataset with several subjects are considered in the proposed research work to overcome the limitations published in the earlier research papers for the detection of diabetes. In the proposed work, Novel Artificial neural network performance is analyzed with reference to a large dataset. These datasets are collected from Pima Indian's dataset.

II. Preparation Of The Data

In order to do the classification with the more accuracy the preprocessing of the data and preparation of the data is more important. This step helps to prepare the data which is raw data and received from the actual patient's data (hospitals) to ready to feed to the device. As the data received data is the raw data hence it is incomplete and also irregular. Hence to make the data suitable for the system and to avoid the less accuracy results, it is required to prepare the data. The data generally has the two types of attributes they are nominal and

real. Nominal and real valued attributes. In general, the age of the patients is divided in to three segments they are namely Kid (15-25 years), Adult (26- 40 years) and Aged (above 41 years). In the same manner the weight of is also classified in to three segments and also as per the age category every age category has the three divisions such as Underweight, Normal and Overweight, this weight classification is different for different age groups. As the main causes of diabetes is also the Blood pressure, hence the last parameter which is classified is blood pressure. Again, this classification is also different for the different age groups. So, this may also called as the sub classification under the age classification.

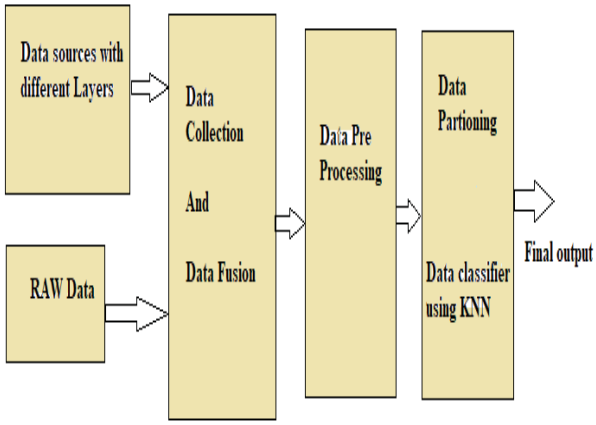

Fig. 1. Block Diagram of KNN Classifier

Figure 1 shows the block diagram of the classification method with KNN. From the figure it can be understood that, Pima Indian diabetes dataset is collected from the Kaggle resource library and divided into 2 groups namely group 1 and group 2. Dataset of group 1 is used for KNN while group 2 dataset is used for novel ANN algorithms. Now, load the dataset into the google Colab for thecalculationofaccuracy and precision for severaliterations. After that values of 20 samples from each accuracy as well as precision are exported into the excel sheet. By using these exported values graphs have been drawn for accuracy and precision separately for the comparison of novel ANN and KNN classifier. Results are shown in Figure 3 and Figure 4.

```
Name       : Optimal K Nearest Neighbor
Data:  Traning : TrD    Test Data: TtD
Result:  Classification : COD
//  Intiate the value of M and K
   K = 1 and M = 1
foreach points set TrD and TtD
   // Analyse the test data wrto set value
     Calculate distance TrD and TtD
      Sorting out
     Find hiigherst value of K
     Find nigh K
   // Find repeated class
      Find repeated neighbours
   //  Calculate error
       Comapre
     If TtD initial is less than the actual
      Store the data
      end
```

Fig. 2. K-Nearest Neighbors algorithm

.The algorithm for the data preparation and for the classification is given in the figure 2. The table 1, shows the data set preparations in sequence.

III. Results

Artificial neural network has significantly($p<0.05$) better accuracy and precision than the K-nearest neighbors algorithm. Statistical features extracted from the raw data is shown in Table 1 is used for the training and learning the algorithm. The accuracy comparison is shown in Figure 3 for ANN and KNNalgorithms on the diabetes dataset. Accuracy rates for ANN and KNN were 95% and 94%, respectively. Comparative study of ANN and KNN algorithmson the diabetes datasetfor precision is shown in Figure 4. Precision of 98% is achieved by using ANN algorithm while KNN algorithmgives 97% of precision value in the diabetes dataset [13]. From the results shown in figure 3 and figure 4 reflects that the performance of ANN is pointedlysuperiorto the KNN in terms of accuracy and precision.

TABLE I. Data Set Preparation Table Representation

	Age	Wt	BP	BI	Function	Classification
count	45	65	120	65	2000.00	2000.00
mean	3.7035	59.4	69.1	20.93	0.47093	0.34200
std	3.3060	30.5	19.1	16.10	0.32355	0.47449
min	0.0000	0.00	0.00	0.000	0.07800	0.00000
25%	1.0000	37.	63.0	0.000	0.24400	0.00000
50%	3.0000	55.0	72.0	23.00	0.37600	0.00000
75%	6.0000	79.0	80.0	32.00	0.62400	1.00000
max	17.000	135.	122.	110.0	2.42000	1.00000

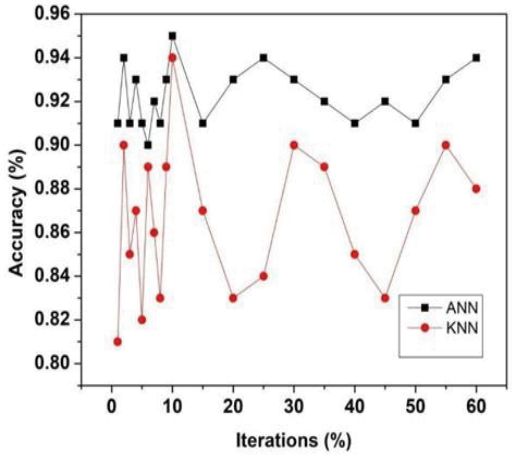

Fig. 3. Comparison of ANN and KNN at different accuracy values.

IV. Discussion

With a 95 percent accuracy and a 98 percent precision in this study, it was found that ANN appears to be superior to KNN (p 0.05). The effectiveness of ANN and KNN in classifying the diabetic mellitus disease using the dataset received from Kaggle is examined in this investigation.

According to the suggested work, ANN performs better in classification than KNN classifier.

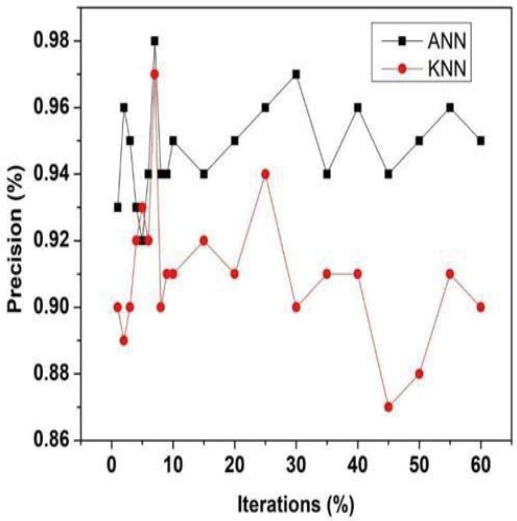

Fig. 4. Comparison of ANN and KNN at different precision values.

When comparing this study with other related articles, the authors have achieved by calculating the certainty factor selected on the calculation of the Novel Artificial Neural Network. Data of 100 medical records from RSUD BendanPekalongan, taken and divided into two parts. 70 data has been used totrainand rest 30 for the testing the proposed algorithms. It is found that the accuracylevel of the analysisscheme is almost 100% accurate in accordance with the expert's diagnosis [14-18].

When the target classes are overlapping, KNN fails to performand alsotakes maximum time to train for a large size of dataset. When the dataset is really huge, ANN is not computationally efficient. Future research should focus on using multiclass ANN, providing more data to the training sets, and merging ANN with other KNN techniques to increase ANN accuracy. Additionally, binary data can be created from multidimensional data to increase accuracy.

V. Conclusion

Novel Artificial Neural Networks (ANN) provide suggestivelybetterresult in detectionof diabetes ailmentas compared to K-Nearest Neighbors (KNN). Accuracy and precision of 95% and 98% respectively are produced by ANN while 94% and 97% is produced by the KNN which shows significant improvement of results by ANN over KNN.

References

[1] A. Goswami, A. Bora, G. Kundu, S. Ghosh, and A. Goswami, "Bleeding disorders in dental practice: A diagnostic overview," Journal of the International Clinical Dental Research Organization, vol. 6, no. 2. p. 143, 2014. doi: 10.4103/2231-0754.143529.

[2] Bourouis, S., Band, S. S., Mosavi, A., Agrawal, S., & Hamdi, M. (2022). Meta-Heuristic Algorithm-Tuned Neural Network for Breast Cancer Diagnosis Using Ultrasound Images. Frontiers in Oncology, 12, 834028.

[3] G. Tripathi and R. Kumar, "Early Prediction of Diabetes Mellitus Using Machine Learning," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). 2020. doi: 10.1109/icrito48877.2020.9197832.

[4] N. Sneha and T. Gangil, "Analysis of diabetes mellitus for early prediction using optimal features selection," Journal of Big Data, vol. 6, no. 1. 2019. doi: 10.1186/s40537-019-0175-6.

[5] D. Sisodia and D. S. Sisodia, "Prediction of Diabetes using Classification Algorithms," Procedia Computer Science, vol. 132. pp. 1578–1585, 2018. doi: 10.1016/j.procs.2018.05.122.

[6] O. M. Omisoreet al., "An affective learning-based system for diagnosis and personalized management of diabetes mellitus," Future Generation Computer Systems, vol. 117. pp. 273–290, 2021. doi: 10.1016/j.future.2020.10.035.

[7] W. Zhong et al., "Gestational Diabetes Mellitus Prediction Based on Two Classification Algorithms," 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). 2019. doi: 10.1109/cisp-bmei48845.2019.8965819.

[8] Fakih, A. H., Venkatesh, A. N., Vani, Naved, M., Kshirsagar, P. R., & Vijayakumar, P. (2022, May). An efficient prediction of diabetes using artificial neural networks. In AIP Conference Proceedings (Vol. 2393, No. 1, p. 020071). AIP Publishing LLC.

[9] S. Abhari, S. R. NiakanKalhori, M. Ebrahimi, H. Hasannejadasl, and A. Garavand, "Artificial Intelligence Applications in Type 2 Diabetes Mellitus Care: Focus on Machine Learning Methods," Healthc. Inform. Res., vol. 25, no. 4, pp. 248–261, Oct. 2019.

[10] N. Hong, H. Park, and Y. Rhee, "Machine Learning Application in Diabetes and Endocrine Disorders," The Journal of Korean Diabetes, vol. 21, no. 3. pp. 130–139, 2020. doi: 10.4093/jkd.2020.21.3.130.

[11] S.-C. Chow, J. Shao, H. Wang, and Y. Lokhnygina, Sample Size Calculations in Clinical Research. CRC Press, 2017.

[12] Z. Larissa, N. Tamara, P. Alexander, G. Varvanina, and S. Mikhail, "Retinol-binding protein in the diagnosis nafld and type 2 diabets," Endocrine Abstracts. 2021. doi: 10.1530/endoabs.73.aep262.

[13] M. M. Shabtari, V. K. Shukla, H. Singh, and I. Nanda, "Analyzing PIMA Indian Diabetes Dataset through Data Mining Tool 'RapidMiner,'" 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). 2021. doi: 10.1109/icacite51222.2021.9404741.

[14] Y. Gao and F. Chai, "Risk of non-vertebral fractures in men with type 2 diabetes: A systematic review and meta-analysis of cohort studies," Exp. Gerontol., p. 111378, Apr. 2021.

[15] Ali A. Abaker, Fakhreldeen A. Saeed "A COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS TO BUILD A PREDICTIVE MODEL FOR DETECTING DIABETES COMPLICATIONS" informatica vol45(2021)

[16] A. P, A. Sharma, S. R. Kawale, S. P. Diwan and D. G. V, "Intelligent Breast Abnormality Framework for Detection and Evaluation of Breast Abnormal Parameters," 2022 International Conference on Edge Computing and Applications (ICECAA), Tamilnadu, India, 2022, pp. 1503-1508, doi: 10.1109/ICECAA55415.2022.9936206.

[17] A. P, A. Sharma, S. B. M, P. Pavankumar, N. K. Darwante and D. G. V, "Performance Monitoring and Dynamic Scaling Algorithm for Queue Based Internet of Things," 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2022, pp. 1-7, doi: 10.1109/ICSES55317.2022.9914108.

[18] A. P, A. Sharma, S. Reddy P, P. S. Patwal and D. Gowda V, "Data Analytics and Cloud-Based Platform for Internet of Things Applications in Smart Cities," 2022 International Conference on Industry 4.0 Technology (I4Tech), Pune, India, 2022, pp. 1-6, doi: 10.1109/I4Tech55392.2022.9952780.