Classification of Cardiovascular Complications in A Diabetes Prediction Model Using Hybrid Neural Network Algorithms

Ms. Vidya Priya Darcini S[1], Shrimathi C M[2], Saveetha V[3]
[1]Assistant Professor, Dept of CSE
[2, 3]Dept of CSE
[1, 2, 3] Sri Ramakrishna Institute of Technology, Coimbatore ,Tamil Nadu, India

Abstract- *Diabetes and Heart Diseases are widespread health issues around the world. The prevalence of heart disease varies by region, but it affects people of all ages, genders, and backgrounds. The main risk factors for heart disease include high blood pressure, high cholesterol levels, and diabetes. Neural networks and machine learning techniques have been applied to heart disease in various ways to assist in diagnosis, management, and treatment. Neural networks can be used to develop predictive models that analyse patient data, to diagnose heart disease or predict the risk of developing heart disease in the future. Neural networks can be trained to predict the risk of diabetes and cardiovascular complications. The focus of this research is to diagnose coronary artery disease among patients based on their clinical data using a neural network. The paper focuses on the dual approach where in the first phase diagnosis of diabetes is carried out using a neural network with accuracy in Random Forest algorithm- 98.67%.*

Keywords- Neural network Algorithms; Machine Learning; Diabetes; Cardiovascular complications;

I. INTRODUCTION

Machine learning, an area of computer science, aims to boost computer innovation. Machine learning has several applications in our day-to-day life applications, particularly in the domain of healthcare. Machine learning is important in the health sector because of its tremendous data analysis skills. Patients with diabetes continue to be at significant risk for severe Cardio Vascular (CV) events despite significant advancements in medical treatment for these conditions. Cardiovascular disease (CVD) is a leading cause of mortality in both type 1 diabetes (T1DM) and type 2 diabetes (T2DM). Over 50% of deaths in T2DM are related to CVD, demonstrating its significant impact. Furthermore, in T1DM, CVD is the main cause of death.The co-occurrence of cardiovascular problems and diabetes highlights the urgent need for efficient interventions and preventive measures to reduce the high mortality rates associated with these conditions.

II. METHODOLOGIES

To improve the accuracy of anticipating the early onset of diabetes and heart disease, combinational and more sophisticated models are required. The more data that is added to the database, the more intelligent the system will become. The goal of the suggested system is to create an effective prediction model that can foretell the start of heart disease and diabetes in people using sophisticated models that work with algorithms that yield more precise findings. The diabetes and heart disease datasets are fed into the proposed system as CSV files. The next important phase in the process is data pre-processing, which is applied to the dataset. This covers feature engineering, data cleansing, and data selection. The most significant features are then extracted using the pre-processed data that was obtained. Then, based on the specifications in place at that point in the procedure, the dataset is divided into the training and testing sets. During the learning phase, when multiple iterations are necessary, the training data is employed. Using the test data, the model's performance is determined. Initially, training dataset is utilized to fit the parameters of the model. Using the fitted model, the answers for the observation in second dataset called validation dataset are predicted.

An objective assessment of a model fit on the training dataset is given by the validation dataset. Lastly, a final model fit on the training dataset is objectively assessed using the test dataset. The hybrid model uses features extracted from datasets of diabetes and heart disease. It combines these features to predict the underlying causes of heart disease in individuals with diabetes. The extracted features are fed into various algorithm to predict the accuracy of the system. The algorithms used are random forest algorithm and back propagation neural network with particle swarm optimization. The algorithm classifies whether the presence of diabetes in individuals will have an impact on the development or progression of heart disease.

III. BLOCK DIAGRAM OF THE SYSTEM MODEL

The diabetes data are collected. The data is preprocessed. The important features are extracted using appropriate algorithm. The extracted features are fed into a hybrid model which contains Random Forest algorithm and Back Propagation Neural Network with Particle Swarn Optimization. The dataset is split into training and testing data. Training data is used to train the model. After training, the performance of the system is tested using the test data. Thus, the goal of our work is to develop a method that, given a set of chosen symptoms, can determine if a patient has heart disease or diabetes by evaluating the symptoms and making predictions. It is put into practice by building a fully connected neural network model on a machine learning repository including datasets related to diabetes and heart disease. In order to increase the model's accuracy, included a number of optimization techniques in addition to creating training and testing examples then constructed the fully connected model.

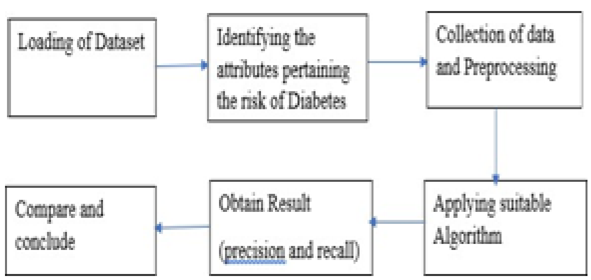

Fig. 1. System Architecture

IV. PROPOSED SYSTEM

DIABETES PREDICTION SYSTEM

The diabetes data are collected. The data are preprocessed to check and eliminate the missing values and fill them with median values. For diabetes the features are extracted using Principal Component Analysis. Using the extracted features the model moves forward for next step. The extracted features of diabetes are fed into Random Forest algorithm and also the features are fed into Back Propagation Neural Network with Particle Sparm Optimmization. Meanwhile the data is split into training and testing. The dataset is split into 70% for training and 30% for testing.

Machine Learning offers a wide range of prediction methods. One of the most common and well-liked methods for optimizing feed forward neural network is Back-propagation Neural Network (BPNN) algorithm.

In order to anticipate the outcomes of disease detection, the system uses Random Forest , a well-liked machine learning method for classification and regression tasks brcause of its high accuracy, robustness, feature importance, versatility, and scalability. Despite its apparent simplicity, this approach performs better than lot other prediction analysis techniques. The medical field clinical professionals and health industry see great value in this method. The application works hard to provide consumers of medical data with best possible experience. It is possible to teach the classifier to recognize the disease's presence. The output may be obtained quickly because the entire process is automated.

V. SYSTEM DESIGN

Datasets - Diabetes-

The data set has 1000 observations and 11 variables The data were collected from the Iraqi society, as they data were acquired from the laboratory of Medical City Hospital and (the Specializes Center for Endocrinology and Diabetes-Al-Kindy Teaching Hospital).

TABLE 1. DIABETES ATTRIBUTES

Name	Labels	Units	Levels
Age	Age	years	
Urea	Urea	mmol/l	
Cr	Creatinine ratio	mg/dl	
HbA1c	Hemoglobin A1c	mg/dl	
Chol	Cholesterol	mg/dl	
TG	Triglycerides	mg/dl	
HDL	High- Density Lipoprotein	mg/dl	
LDL	Low-Density Lipopprotein	mg/dl	
VLDL	Very-Low-Density Lipoprotein	mg/dl	
BMI	Body Mass Index	kg/m^2	
CLASS	CLASS		2

VI. ALGORITHMSUSED

A. Principal Component Analysis:

One unsupervised learning technique that lowers dimensionality in machine learning is Principal Component Analysis. It is a statistical process that converts a set of correlated feature observations into set of linearly uncorrelated features via orthogonal transformation. The most current update to the features are called Principal Components. It is a popular tool for exploratory data analysis as well as predictive modeling. It is a technique for reducing variations so that significant patterns in the dataset may be found.

B. Random Forest Algorithm:

One popular machine learning approach is Random Forest approach. It is an algorithm for Supervised Machine learning.It can be used to solve machine learning challenges that combine classification and regression. The random forest algorithm is based on the concept of ensemble learning, involves in combining multiple classifiers to improve the model's performance and solve an issue that is challenging.
The Random Forest Algorithm makes use of a number of decision tree-based classifiers. Each decision tree is first trained on an individual basis. Based on these trees' predictions, the random forest then predicts the average of these results.

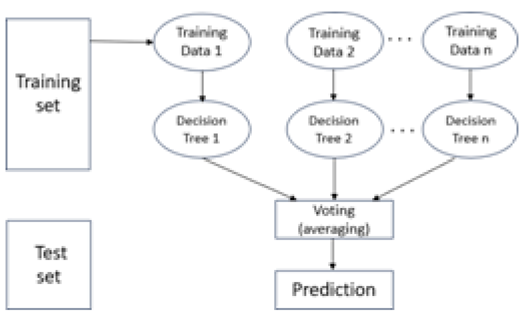

Fig. 2. Working of Random Forest Algorithm

C. Particle Swarm Optimization

An effective optimisation method for adjusting neural network parameters and enhancing performance is called Particle Swarm Optimisation or PSO. It is modelled after the social behaviour of schools of fish or flocks of birds. PSO involves searching a search space with a collection of possible solutions, referred to as particles, in order to identify the best answer. The objective is to identify the parameter values that minimise an error or expense function, with every particle representing a set of neural network parameters. The first step in the process is to initialise a swarm of particles.

The performance of each particle's corresponding neural network is evaluated using a cost function. This cost function measures how well the network performs on a specific task, such as classification or regression. The lower the cost, the better the network's performance.

Global Search: PSO explores a large search space efficiently, making it suitable for finding global optima in complex neural network parameter spaces.

Easy Implementation: PSO is relatively simple to implement and requires few hyperparameters, making it a practical choice for optimization.

Versatility: PSO can be used for various types of neural networks, including feedforward, recurrent, and convolutional neural networks.

D. Back Propagation Neural Network

One of the fundamental ideas of a neural network is backpropagation. To categorise our data as best we can is our aim. The linear regression model's parameter is optimised by the use of gradient descent. Here, we follow a similar approach and use the gradient descent technique with backpropagation. It is a method for adjusting the weights of a neural network based on the error rate observed in the previous epoch, or iteration. In neural networks, backpropagation is known as "backward propagation of errors" or BPE.

Artificial neural networks are usually taught in this manner. The gradient of the error function is calculated for a single training sample using backpropagation.

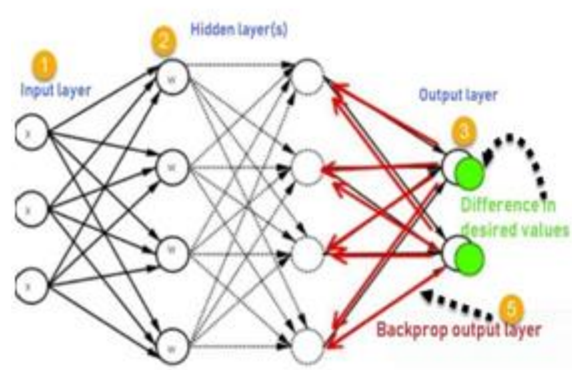

Fig. 3. Working of Back Propagation Neural Network

VII. WORKING OF THE SYSTEM

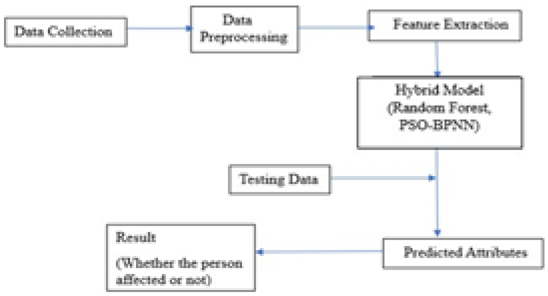

Fig. 4. Working of the system

A. Diabetes Dataset:

1) Data Collection: The diabetes raw data of patients from the the laboratory of Medical City Hospital and (the Specializes Center for Endocrinology and Diabetes-Al-Kindy Teaching Hospital). The number of attributes contained was 11. The data sethad around1000records.

2) Preprocessing: The data is preprocessed to eliminate the missing or null values using median values. The median values are calculated and replaces the null value present in the dataset. Then the correlation matrix was obtained for the attributes in the dataset.

3) Feature Extraction: The Principal Component Analysis technique is used for the feature extraction technique. By using this we can identify the most important and required features with higher efficiency.

4) Preparing the data for training: The model is trained using the training data and performance of the system is evaluated and accuracy calculated using the test data.

5) Training of the Model:

Building the model: The Random Forest Algorithm is a machine learning algorithm that is often utilized. Simply Random Forest is applied to the whole dataset without any extracted features. The accuracy obtained was noted. The features extracted from PCA are fed into the Random Forest algorithm. The model trained with training dataset. The model learns from the training data.

After training, the system is tested using the test data. The accuracy obtained using Random Forest algorithm was 98.66%.

Another neural network algorithm is used to train and test the model. **The Back Propagation Neural Network** is applied to the system. The model learns from the training data. The accuracy is obtained from test data. The obtained accuracy was 78.50%.

To increase the accuracy, the same data is again optimized using **Particle Swarm Optimization** to adjust the parameters. The optimized data is fed into the Back Propagation Neural Network again. The accuracy obtained this time was 96.50%.

VIII. RESULT ANALYSIS

TABLE 2. ACCURACY OBTAINED FOR DIABETES DATASET

Algorithms used	Accuracy
Random forest with PCA	98.66%
Random Forest without PCA	98.66%
Back Propagation Neural Network with Paricle Swarm Optimization	96.50%
Back Propagation Neural Network without Paricle Swarm Optimization	78.50%

IX. CONCLUSION

The system aims in predicting the accuracy of the diabetes dataset. The system eliminates the null value using median value and precedes with algorithm. The Random Forest algorithm gives high accuracy for the Diabetes Disease dataset when compared with Back Propagation Neural Network with Particle Swarm Optimization.

Diabetes is prevalent global health concerns. Machine learning, specifically neural networks, has been instrumental in diabetes diagnosis and management. This research adopts a dual-phase approach: first, diabetes diagnosis using a neural network, followed by coronary artery disease diagnosis. Clinical data, collected from Medical City Hospital and the Specialized Centre for Endocrinology and Diabetes-Al-Kindy Teaching Hospital, undergoes preprocessing and feature extraction. Notably, the Random Forest Algorithm achieves an accuracy of 98.67%, while Particle Swarm Optimization with Back Propagation Neural Network yields 96.50%.

These results underscore the potential of machine learning in healthcare, improving diagnostics for critical conditions like diabetes and heart disease.

REFERENCES

[1] Bala Manoj Kumar P, Srinivasa Perumal R , Nadesh R K, Arivuselvan K ,"Type Diabetes mellitus prediction using Deep Neural Networks Science and Computational Intelligence 2022, ID: 216 (2023) 21–30.

[2] Muhammad Exell Febriana, Fransiskus Xaverius Ferdinana, Gustian Paul Sendania, Kristien Margi Suryanigrum, RezkiYunanda," Diabetes prediction using supervised machine learning",7th International Conference on Computer classifier",International Journal of Cognitive Computing in Engineering 1 (2020) 55–61.

[3] MD. Shafiqul Islam, Marwa K. Garaqe (Member IEEE), Samir Brahim Belhaouari (Senior Member,IEEE), and Muhammad A. Abdul-Ghani "Advanced Techniques for Predicting the Future Progression of Type 2 Diabetes",IEEEACCESS.2020.3005540,VOLUME 8, 2020

[4] Varun Sapra, Luxmi Sapra, Akashdeep Bhardwaj, Salil Bharany, Akash Saxena, Faten Khalid Karim, Sara Ghorashi, Ali Wagdy Mohamed," Integrated approach using deep neural network and CBR for detecting severity of coronary artery disease ", doi:10.1016/j.aej.2023.01.029 (2023) 1110-0168.

[5] Yue Huang, YingBo Ren, Hai Yang, YiJie Ding, Yan Liu, YunChun Yang, AnQiong Mao, Tan Yang, YingZi Wang, Feng Xiao, QiZhou He, Ying Zhang, "Using a machine learning-based risk prediction model to analyze the coronary artery calcification score and predict coronary heart disease and riskassessment",doi:10.1016/j.compbiomed.2022.1 06297, ID- 106297, Volume 151,2022

[6] Hanna Kwiendacz, Agata M. Wijata, Jakub Nalepa, Julia Piaśnik, Justyna Kulpa, Mikołaj Herba, Sylwia Boczek, Kamil Kegler, Mirela Hendel, Krzysztof Irlik, Janusz Gumprecht, Gregory Y. H. Lip and Katarzyna Nabrdalik," Machine learning profles of cardiovascular risk in patients with diabetes mellitus: the Silesia DiabetesHeartProject",(2023)

[7] Jobeda Jamal Khanam, Simon Y. Foo, "A comparison of machine learning algorithms for diabetes prediction",ICT Express Volume 7, Issue 4,432-439 (2021).

[8] Liangjun Jiang A, Zhenhua Xia B, Ronghui Zhu C, Haimei Gong A, Jing Wang D, Juan Li E, Lei Wang, "Diabetes risk prediction model based on community follow-up data using machine learning" Volume 35,Article ID:102358,(2023).

[9] Chollette C. Olisah, Lyndon Smith, Melvyn Smith, "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective",Article ID:220 (2022) 106773.

[10] MrsP.Renukadevi, M.Savitha, R.M.Soundharya, " Diabetics Prediction using Back Propagation Neural Network algorithm", Volume 12, Issue 2, Article ID: 2250- 1770 (2022).

[11] Nesreen Samer El_Jerjawi, and Samy S. Abu-Naser, "Diabetes Prediction Using Artificial Neural Network" (2018).

[12] Mythili Thirugnanam, Dr Praveen Kumar, S Vignesh Srivatsan, Nerlesh C R, "Improving the Prediction Rate of Diabetes Diagnosis using Fuzzy,Neural Network, Case based (FNC) approach, "Procedia Engineering ,Volume 38 ,1709 – 1718.

[13] Putri Sari Asiha ,Yufis Azhara , Galih WasisWicaksonoa, Denar RegataAkbia , "Interpretable Machine Learning Model For Heart Disease Prediction", 8th International Conference on Computer Science and Computational Intelligence, doi:10.1016/j.procs.2023.10.544 (2023).

[14] Mandakini Priyadarshani Behera, Archana Sarangi, Debahuti Mishra , Shubhendu Kumar Sarangi, "A HybridMachine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine" (2023).

[15] Senthilkumar Mohan, Chandrasegar Thirumalai, and Gautam Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" in IEEE special section on smart caching, communications, computing and cybersecurity for information-centric internet of things July 3, 2019, pp.81542-81554.