An Optimal Framework for Intelligent Prediction of Prediabetes and Type-2 Diabetes Using Genomic Data

Sreenivas Pratapagiri[1,*] , Shanker Chandre[2] and Dr. Balaji Maram[3]

[1]Research Scholar, School of Computer Science and Artificial Intelligence, SR University, Warangal, Telangana, India, 506371.

Email: psv.kits@gmail.com

[2]Assistant Professor, Supervisor, School of Computer Science and Artificial Intelligence, SR University, Warangal, Telangana, India, 506371.

Email:Shanker.chandre@gmail.com

[3]Professor, Supervisor,School of Computer Science and Artificial Intelligence, SR University, Warangal, Telangana, India, 506371.

Email:balajee.maram@sru.edu.in

ABSTRACT

The most important study in medical research is the prediction of type 2 diabetes in patients. As far as type 2 diabetes prediction models go, there are a number of options. Nevertheless, due to subpar quality, the desired outcome was not achieved. Nan characteristics in the gene data make type 2 diabetes prediction as complicated as possible. Scores for both performance and prediction were low due to the flaws. Consequently, the plan is to create a new method for predicting type 2 diabetes using chimpanzees and functional link neural architecture (CbFLNA). Methods such as feature selection, categorization, and gene expression have been carried out as part of the pre-processing methodology. The first steps were to import the genomic database, preprocess the data, and extract the meaningful features. Next, sort the individuals' illnesses according to the likelihood that they have type 2 diabetes. When all was said and done, the model's performance was evaluated, and it achieved a very high accuracy score in the prediction.

Keywords: Genomic data, Functional link neural approach, Normalization, Chimp optimization, Feature selection.

1. INTRODUCTION

When insulin secretion or production is inadequate, a metabolic illness known as diabetes develops [1]. This leads to high blood sugar levels. Worldwide, 285 million people were predicted to suffer from hyperglycemia in 2010 [2]. According to the present rate of evolution of the sickness, this figure has been increased to 552 million through 2030 [3]. About 10% of the population will have diabetes by the year 2040 [4]. Diabetes is becoming more common due to people's increasingly individualistic behaviors, varied diets, and higher standard of living [5]. Research developing fast and easy ways to diagnose and cure diabetes is, thus, worthwhile [6-7]. Genomic sequences are the basis of a diabetes diagnosis. It leads to a more accurate and suitable result, which encourages the development of better habits that are less likely to require insulin soon [8]. Furthermore, Fig. 1 reveals the diabetes prognosis that is based on genomic data.

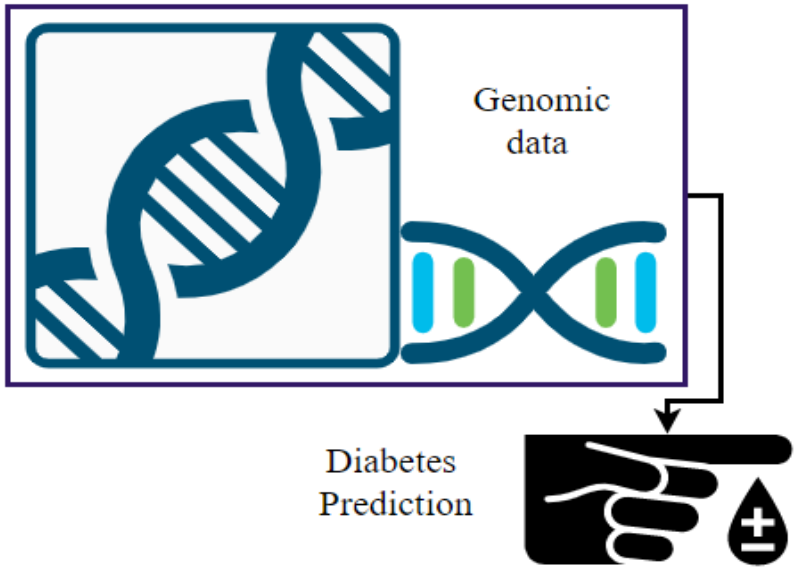


Fig.1 Genomics-based diabetes prediction

Patients can slow the advancement of their diseases and enhance their overall health with a very accurate diagnostic [9]. Two applications of AI exist: condition prediction [10] and anomaly diagnosis [11]. Diabetes might be predicted using forward-looking approaches based on present and past health problems [12]. Multiple complications and an increased risk of early death can arise from this genetic disorder [13]. Some disease-related features have been discovered, albeit tentatively, in previous association and candidate gene studies [14-15]. The results of those genetic studies were considered useless all the way through, regardless of where the genes were located [16-17]. Recently, genes associated with type 2 diabetes sickness have been discovered by genome-based broadest association techniques [16]. Also included in the genome data are single-molecule polymorphisms [18-20], which are common in human chromosomes and have been shown to characterize a variety of disorders [21].

Models trained using deep networks [22-24], spatial-temporal systems, etc., are adequate for predicting type-2 diabetes when other feature selection methods are fine-tuned in the classification layer. However, due to its restricted features, it is neither acceptable or adequate for prediabetes prediction. Determining the best neural system to use as a diabetes recognition framework was thus the primary goal of the current investigation.

An analysis of the study's primary finding is presented below.

- Genomic data was used mostly for training the system.
- Therefore, in order to identify the forms of diabetes mentioned, a new CbFLNA is developed with the necessary functional properties.
- The second layer of CbFLNA is responsible for filtering the noise characteristics, which is done during the pre-processing stage.

- Now that the data has been updated, the classification layer can take it as input and choose the necessary elements.
- Additionally, prediabetes and type 2 diabetes were identified and categorized from the defined segments.
- After all that, we checked the suggested model's accuracy, recall, precision, f-score, and error rate, and we saw that it performed better than the previous models.

Section two presents the present work paper as linked work, and Section three reveals the traditional method's problem statements. Section four then elaborates on the remedy for the problem that has been defined. In Section 5, we talk about the new solution's validated result. Section 6 served as the paper's conclusion.

2. RELATED WORKS

Below, we will describe a few of the recent studies that are relevant to our current work,

Different types of diabetes are defined according to the severity of the condition. Similarly, features for detecting and diagnosing diabetes and its severity have been presented by Joseph Bamidele Awotunde et al. [25] using deep neural mechanisms. To assess the dependability of the deep network, genetic data is used here. In addition, the genetic algorithm problem governed the operating levels of the deep network. The results were then compared both before and after the optimization function was used. On the other hand, incorrect diagnosis of type 2 diabetes does occur occasionally.

The most pressing need is a prediabetes detection system in order to lessen the effects of diabetes. Accordingly, an intelligent network-based spatial-temporal system has been presented by Maher Abdallah et al. [26]. Prior to deployment, the prediabetes prediction had a high sensitivity score, which was enough to estimate the likelihood of diabetes. To get the relevant characteristics, nevertheless, an extra feature selection model using optimization features was necessary.

Diabetes detection makes good use of smart technologies in addition to novel, intelligent methods. By turning on several smart devices within the human body, Saiteja Prasad Chatrati et al. [27] have successfully established the framework for diabetes identification. It routinely monitors the biological parameters of the human body and provides the predictive results. But this prediction device is pricey in comparison to other novel methods. Additionally, no feature selection or classification is carried out.

In order to routinely assess the body's condition using the primary biological indicators, such as insulin level and blood pressure, Markku Laakso et al. [28] presented the biomarker approach. It is believed that this procedure can detect diabetes automatically whenever it occurs. The medical data cloud was also informed of the tracked results. Here, the signal of the physical parameter anticipated the biological changes. But prediabetes prediction and classification are not its strong

suits.

A methodology for analyzing the severity of renal disease using diabetes prediction was presented by Heejin Jin et al. [29]. Here, both the genetic and chronic disorder data sets were taken into account. This analysis allowed for more accurate prediction of the effects of diabetes on renal illness. Consequently, the medical application's most pressing need is a diabetes prediction framework to assess and analyze the many aspects of the condition. Having said that, it is not possible to accurately forecast diabetes outcomes on chronic data alone.

3. SYSTEM MODEL AND PROBLEM STATEMENT

The main objective of this research study is to detect type 2 and prediabetes. But it's not easy to tell how bad the diabetes is just by looking at the genes. In the past, many feature selection strategies were used to uncover diabetes traits in genetic data. But it won't be enough to analyze and detect every aspect of diabetes. In light of this, the current study has used an optimal deep network architecture to forecast the various forms of diabetes using gene feature analysis.

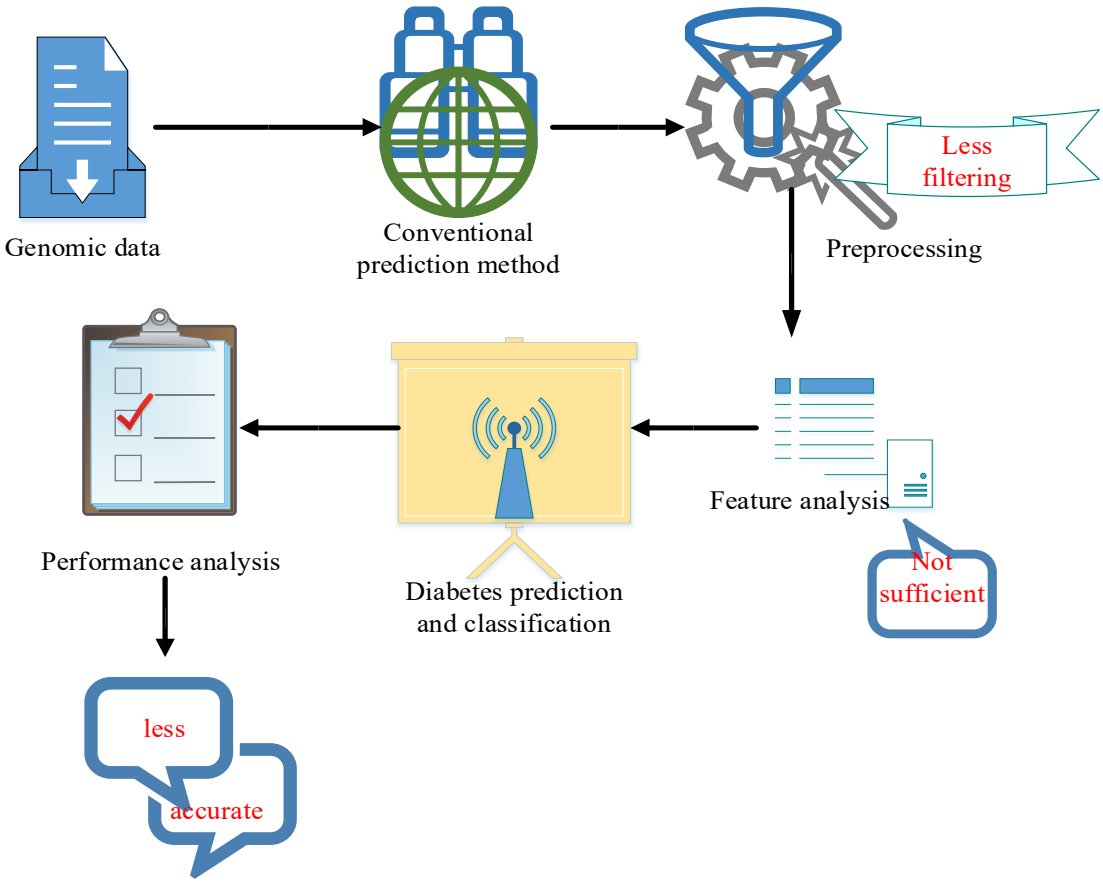


Fig: 2 Conventional prediction issues

As shown in Figure 2, the conventional standard method's prediction system is detailed. The process's pre-processing mechanism has filtered out features with low noise levels. The illness aspects that aren't relevant have been chosen by the feature selection. As a result, significant feature extraction requires more than just feature selection. It optimizes algorithm complexity at the expense of accuracy. In order to circumvent these issues, a functional link neural technique based on Chimp was used for type 2 diabetes prediction.

4. PROPOSED METHODOLOGY

Analyzing and predicting Type-2 diabetes and prediabetes is the goal of a new Chimp-based Functional link neural approach (CbFLNA). The genomic information used in this study is also taken into account when predicting the prevalence of prediabetes and type 2 diabetes. Using insulin and blood pressure rates, prediabetes and type 2 diabetes can be determined in this setting. The feature analysis, diabetes prediction, and classification processes began with pre-processing the genetic data, which was subsequently used as input to the classification layer. As a last step, we compared and contrasted the performance measures across different models.

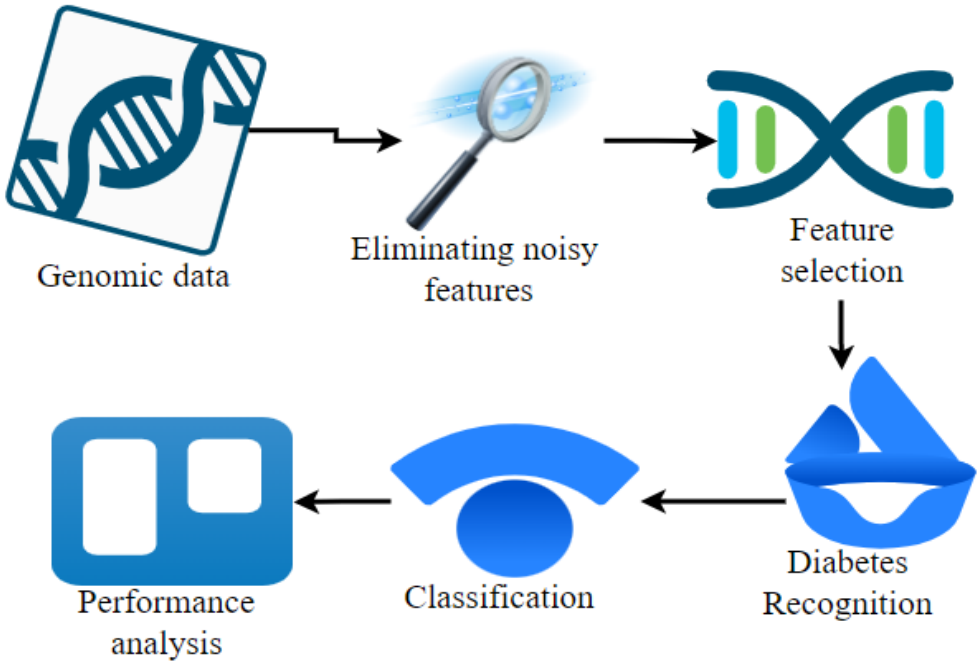


Fig.3 Proposed architecture

Figure 3 depicts the framework of the suggested methodology. Analyzing and predicting type 2 diabetes and prediabetes need this methodology. By conducting comparative studies, we can provide the performance evaluation of gene prediction in each dataset.

4.1 Proposed methodology

There are five levels to the process of the suggested methodology: input, filtering, categorization, optimization, and output. The CbFLNA method, which relies on Chimpanzee, can carry out the operation.

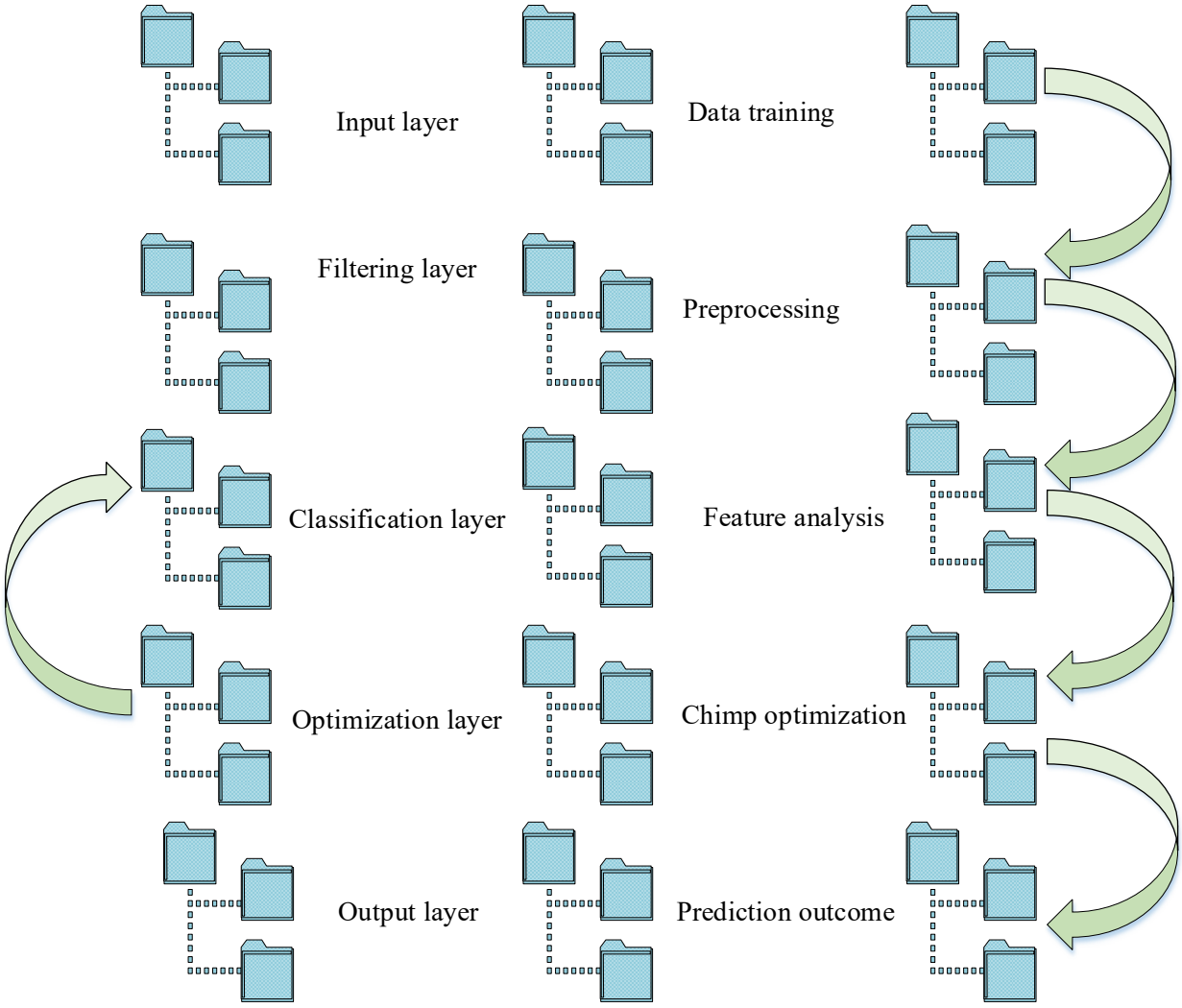


Fig: 4 CbFLNA processing layer

As shown in Figure 4, the unique CbFLNA's functional layers are detailed. In this step, the input layer received the acquired dataset. The filtering layer has already performed the pre-processing. The data that had been filtered for noise was then input into the classification layer. The next step is to sort the parameters using chimp's optimization tool, and then pass the optimized result on to the output layer.

4.1.1 Pre-processing

The genomic data was initially loaded into the Python environment after collection. Data training was defined by Eqn 1.

$$G[d_n] = (d1, d2, ...d_k) \quad (1)$$

In this case, G stands for gene expression, and and are the data labels for the gene expression, with and being the total amount of data points. Despite reducing the algorithm's complexity, the data quality is huge, which impacts the forecast outcome. It simplifies the dataset while maintaining excellent accuracy. Therefore, a crucial part of the process is data pre-processing.

$$P[d_n] = x[(d_n - n_a] \quad (2)$$

Both the normal and Nan characteristics are present in the input dataset. To make the dataset more straightforward, the pre-processing function removes the normalization and Nan characteristics. Applying Eqn. (2) accomplishes the pre-processing purpose. This is where the pre-processing variable is represented. n_a stands for both Nan characteristics, while x is the tracking variable. The pre-processing function produces an uncomplicated dataset.

4.1.2 Feature selection

Feature selection of the most relevant features in the dataset is used to select the subset. It optimizes prediction accuracy while decreasing execution time. Feature selection is picking out the most relevant disease characteristics from the dataset and extracting them. The disease's maximum and minimum matching attributes are both included in the dataset.

$$F = G(d_n) + a[d_n - f_s] \quad (3)$$

The feature selection procedure has been completed by Eqn 3. F stands for the feature selection function variable. There are desirable and undesirable features in the dataset. Obtaining the desired features completes the feature selection process. A represents the tracking variable, while fs stands for the desired features.

4.1.3 Classification and Prediction

Both normal and disease-specific characteristics are defined by the categorization function. The initial dataset typically includes both typical and out-of-the-ordinary elements. Consequently, gene expression and illness feature classification were accomplished using the disease feature dataset. Lastly, the classification layer was used to classify the illness features.

$$C(dis\_feature) = \begin{Bmatrix} if\,(G=0) & Normal \\ else & Abnormal \end{Bmatrix} \quad (4)$$

The condition in the classification function is in the form of $if\,(G[0,1])$. The Eqn 4 represented the classification function of the disease features. Two instances [0, 1] were run through the categorization function. The normal features are denoted by 0, and the aberrant features are denoted by 1.

4.1.4 Disease prediction

Make predictions about the examined disease samples using the current gene features after feature selection and the classification function. Using the Eqn 5, one can anticipate the occurrence of an illness.

$$\Pr obability = 1 \times \frac{ds}{\max - Gn} + 0.1 \tag{5}$$

A probability value of one indicates that the maximum number of genes have been considered. In the samples of this gene that have been analyzed, more than one likelihood equals 100%. The illness samples that were analyzed can be characterized as follows: ds. Some essential features include glucose, protein, blood cholesterol levels and blood pressure $\max\_Gn$. The trained genomic disease database has shown that certain gene characteristics are different.

Algorithm1: CbFLNA

Start

{

Data initialization()

{

$\text{int}\, G(d_n) = 1,2,3,....n;$

//initialize the disease database

}

Pre-processing()

{

$\text{int}\, P, x, n_a;$

//initialize the pre-processing variables

$P(d_n) \longrightarrow d_n - n_a$

// Removing the noise features

}

Feature selection()

{

$\text{int}\, F, f_S;$

//initialize the feature selection variables

$Select \longrightarrow d_n - f_S$

//selecting the maximum matching features

}

Classification()

{

$if\,(G(d)) = 0$

{

Normal

}**else (Abnormal)**

}

Gene expression()

{

$\Pr ob \longrightarrow Classification(\max\_Gn)$

// classified the maximum number of genes

}

}

Stop

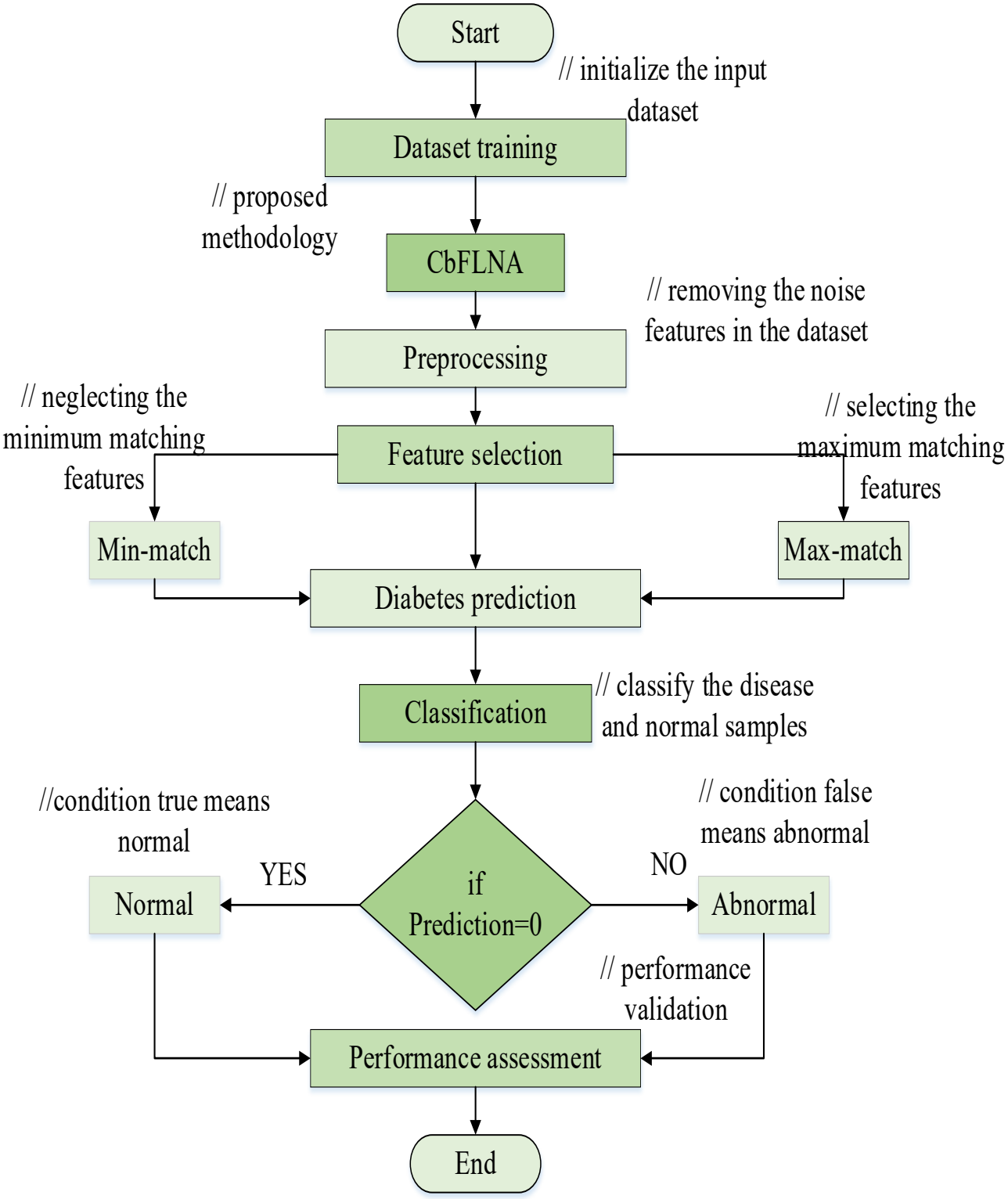


Fig.5: A schematic of the CbFLNA process

In Fig. 5, we can see the detailed procedures for applying the suggested model. The algorithm then describes the pseudo-code for the detailed mathematical formulation. There are both healthy and unhealthy qualities in the dataset. 1. Feature selection and illness sample classification were done in the first step. In addition, only samples from patients with diseases had their gene expression profiles determined. Several categorization metrics were used to determine the CbFLNA's performance after the indicated stages were processed.

5. RESULT AND DISCUSSION

We used the Windows 10 framework to implement the innovative CbFLNA approach, which was validated in the Python environment. Genomic data is the database that is considered for testing and validation. "Normal" and "abnormal" data points coexist in the dataset.

Table.1: Detailing the parameters for execution

Description of parameters	
Programming environment	Python
Total sample count	1512
Operating system	Windows 10
Optimization	Chimp optimization
Database	Genomic data
Deep network	Functional link neural network

You may see the details of the execution parameters in Table 1. It was in the pre-processing step that the training errors were first eliminated. The classification step then made use of the data that had been corrected for errors. Not only that, gene expression was predicted, and performance evaluations can be based on a number of other criteria.

5.1 case study

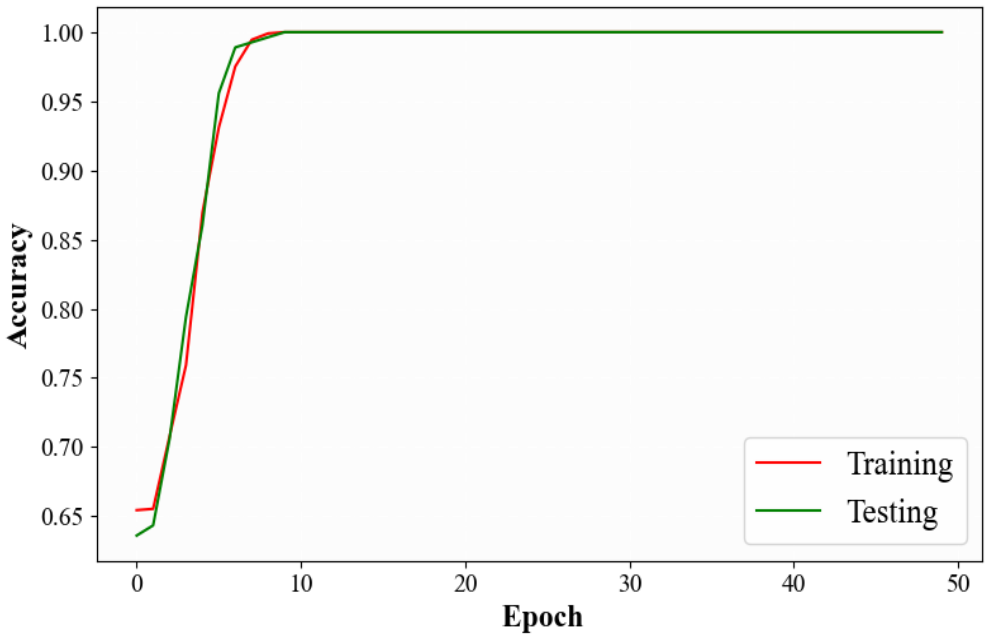
We conducted some test validation to ensure the proposed methodology was working properly, and we reported the results in a methodical way. The dataset on diabetes and genes was used to validate the tests. The total number of samples is 1,701. A total of 1,360 samples were used for training purposes, whereas 341 samples were used for testing.

Table.2: Database details

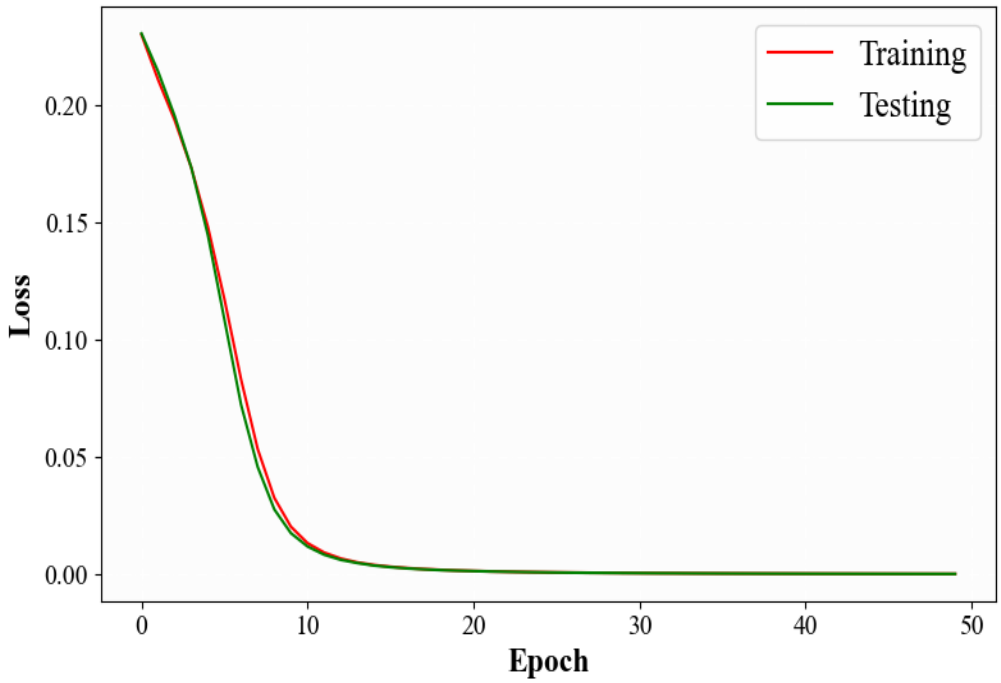
Total no of samples: 1701

Abnormal	583
Normal	1118
Training(80%):1360	
Abnormal	476
Normal	884
Testing(20%): 341	
Abnormal	107
Normal	234

You may find the database's details in Table 2. Also, samples are taken into account for training, with 476 samples being aberrant and 884 being standard. Additionally, out of 341 samples, 234 were found to be normal and 107 to be abnormal.



a)



b)

Fig: 6 a) Training accuracy b) Training loss

Over the training epoch, the CbFLNA's accuracy and loss were evaluated, as shown in Fig. 6. One way to measure how well a diabetes prediction system works is by looking at its train-test validation exactness score. By utilizing loss metrics and evaluating the framework via its dual training and test validation phases, we can determine its failure ratio.

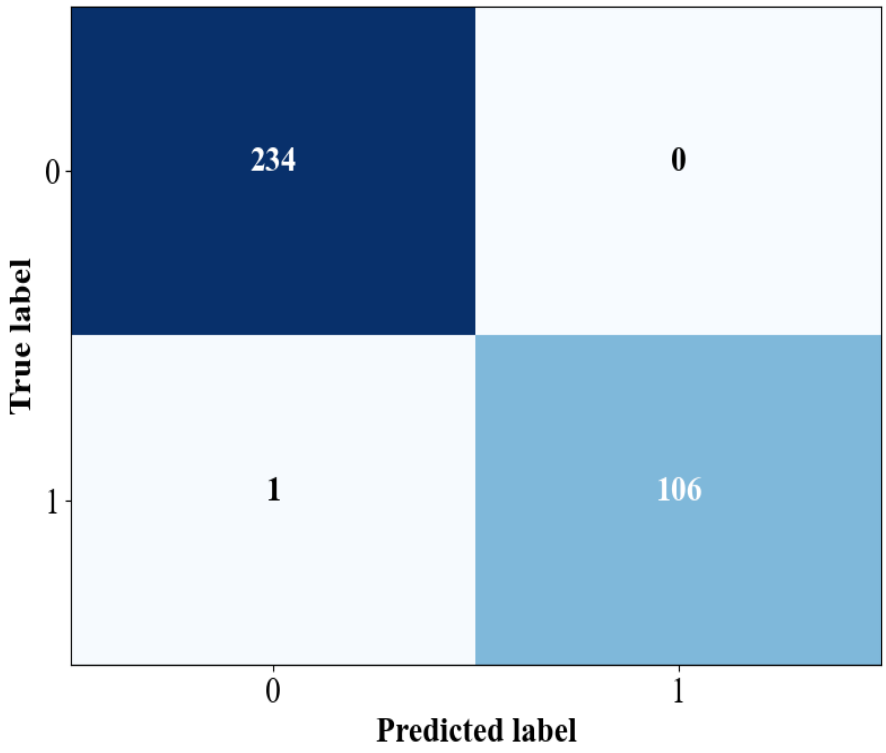


Fig.7: Confusion matrix

As shown in Figure 7, the projected effect manifested as a network confusion. Achieving a positive or negative score for each true or false class represented the categorization outcome. In this scenario, the forecast was split into two categories: zero and one. The normal value is 0, and the abnormal value is 1.

5.2 Performance assessment

By computing the main metrics used for validation, such as accuracy, recall, precision, and f-score, the improvement score of the unique CbFLNA that was constructed could be supported. Utilize the newly linked model to examine the enhancement in performance. Some examples of current models are Random Forest (RF) [30], Bagged decision tree (BDT) [31], Multilayer perceptron (MLP) [32], and Edited nearest neighbour (ENN) [33].

5.2.1 Precision: Precision is the measure of the optimistic forecast that characterizes the proportion of outliers in the total samples. We used Eqn (6) to determine the metrics' precision measure.

$$\Pr ecision = \frac{TP}{FP + TP} \tag{6}$$

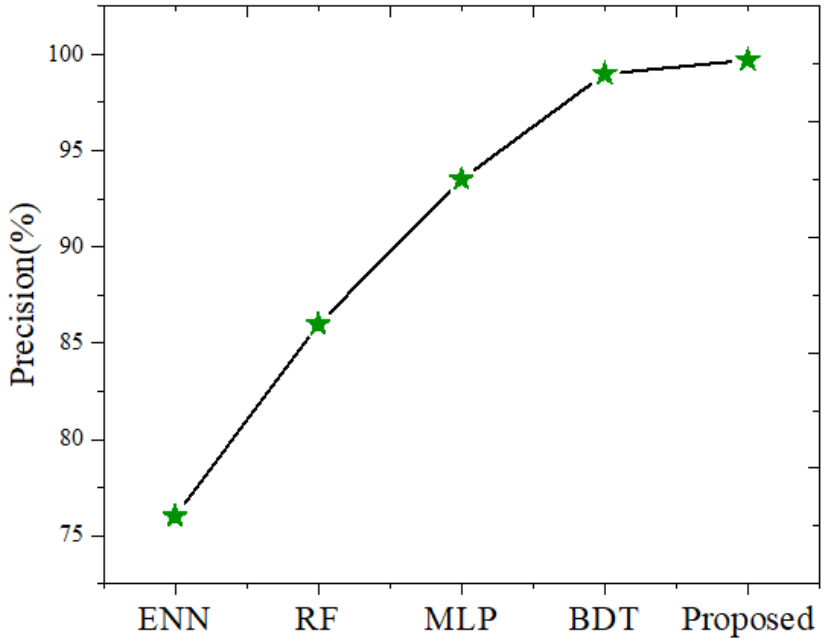


Fig. 8 Precision assessment

The ENN approach achieved a precision of 76%, the RF method 84.97%, the BDT method 99.998%, and the MLP method 93.5 %. The suggested innovative CbFLNA technique achieved 99.70% accuracy when the comparable mechanism was taken into account. The data is shown in Figure 8.

5.2.2 Accuracy: When determining whether data is aberrant or normal, accuracy is the defining factor. Taking the average of the positive and negative scores for the forecast is how accuracy is calculated. Equation (7) was used to calculate the accuracy measure.

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} \tag{7}$$

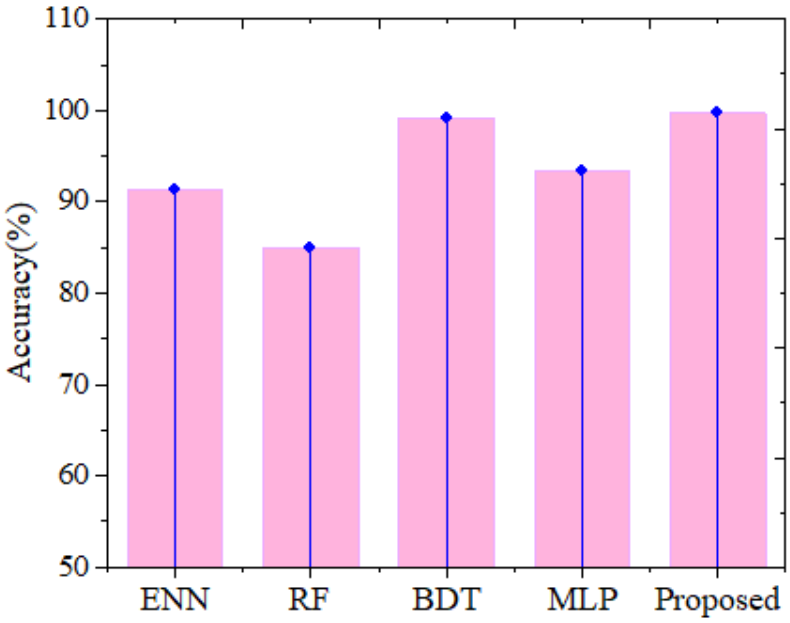


Fig.9 Accuracy assessment

The ENN approach achieved 91.4% Accuracy, the BDT method 99.14%, the RF method 84.95%, and the MLP method 93.46%. The suggested innovative CbFLNA technique achieved a 99.70% Accuracy when the comparison mechanism was taken into account. Figure 9 displays the data.

5.2.3 Recall: The section of the anomalous case that has successfully retrieved the whole anomalous sample is called recall. The metrics recall are defined by Eqn 8.

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{8}$$

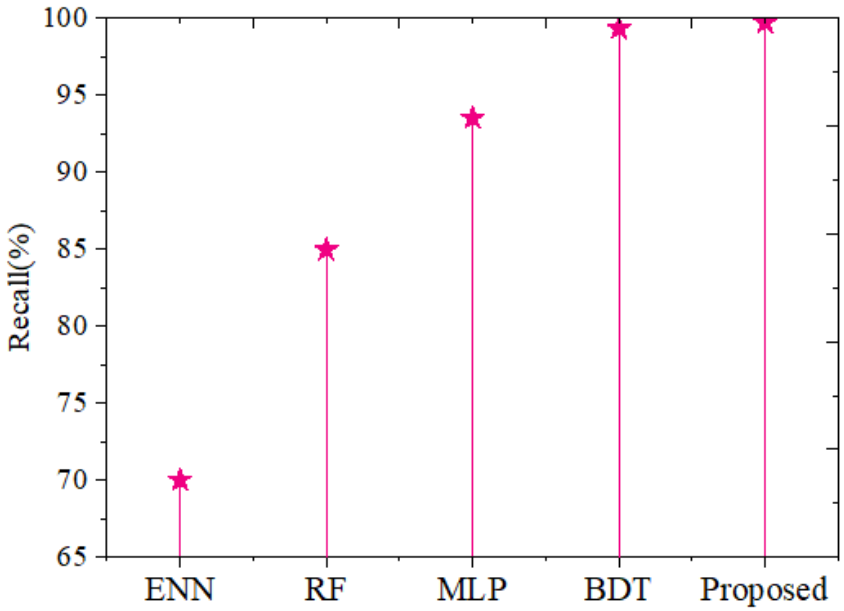


Fig.10 Recall assessment

The ENN approach achieved a recall of 70%, the BDT method 99.32%, the RF method 84.95%, and the MLP method 93.5 %. The suggested innovative CbFLNA technique achieved a recall of 99.70% when taking the comparison mechanism into account. In Fig.10, the data is shown.

5.2.4 F-score: The F-score is a measure of the system's performance that depicts the relationship between recall and precision. Prediction statistics of true and false scores are averaged to get the f-score. The f-score measurements were checked using Eqn (9)

$$F-score = 2 \times \frac{\mathrm{Re}\,call \times \mathrm{Pr}\,ecision}{\mathrm{Pr}\,ecision + \mathrm{Re}\,call} \tag{9}$$

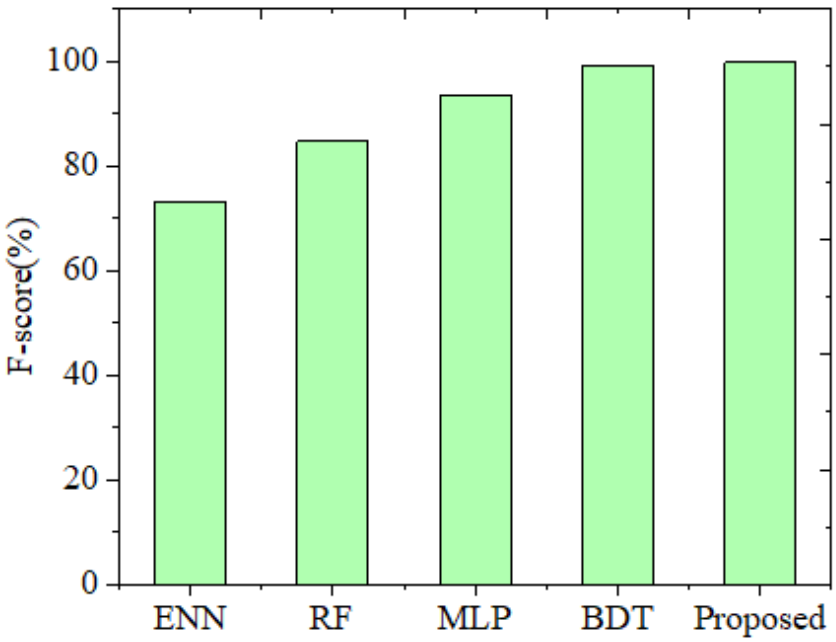


Fig.11 F-score assessment

The ENN technique achieved an f-score of 73%, the RF method an f-score of 84.77%, the BDT method a 99.15%, and the MLP method a 93.4 %. The suggested innovative CbFLNA technique achieved an f-score of 99.70% when taking the comparison mechanism into account. Figure 11 displays the data.

5.2.5 Error rate: When comparing a technique to an accurate one, the metrics error rate indicates how far off the mark the methodology's predictions are. Equation (10), which measures the error rate,

$$Error\ rate = \frac{FP + FN}{TP + FP + FN + TN} \tag{10}$$

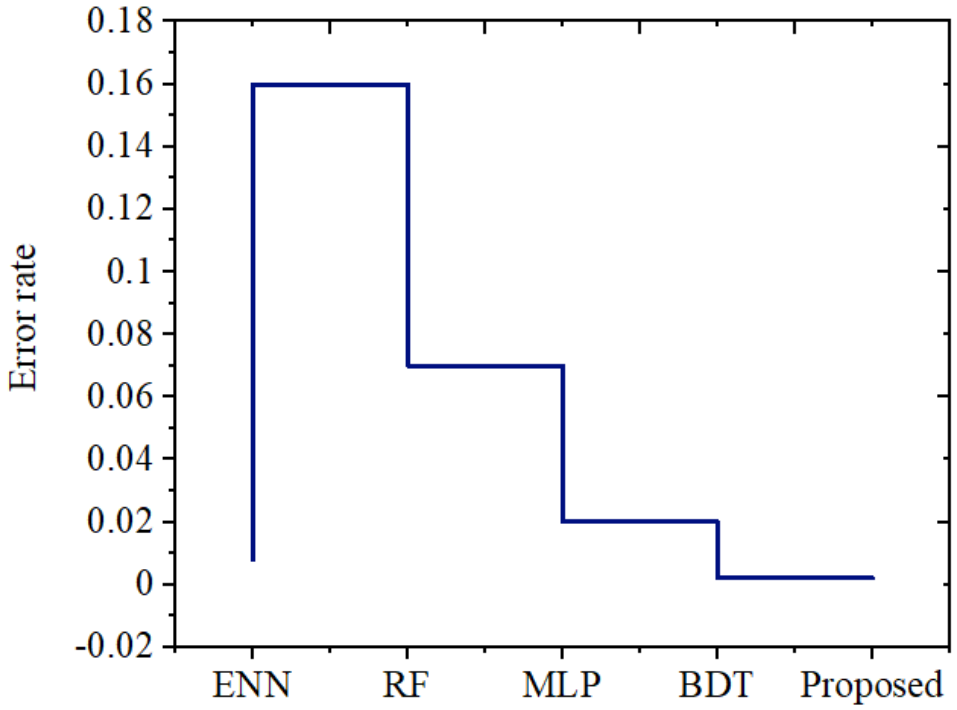


Fig.11 Error rate assessment

There is an error rate of 0.008 with the ENN approach, 0.02% with the BDT method, 0.16 with the RF method, and 0.07 with the MLP method. An error rate of 0.002 was recorded while comparing the mechanisms of the proposed novel CbFLNA approach. Figure 12 displays the data.

5.2.6 Time: Duration needed to finish the whole procedure of 2D diabetes prediction. Time spent training is 4.757 seconds, time spent testing is 0.175 seconds, and time spent executing is 74.1547 seconds according to the suggested technique.

Table.3: Comparison assessments

Methods	**Precision (%)**	**Error rate**	**Recall (%)**	**F-score (%)**	**Accuracy (%)**
ENN	76	0.008	70	73	91.4
BDT	98.98	0.02	99.32	99.15	99.14

RF	85.97	0.16	84.95	84.77	84.95
MLP	93.5	0.07	93.5	93.4	93.46
Proposed	99.70	0.002	99.70	99.70	99.70

5.3 Discussion

All performance analyses yielded a very accurate score for the unique CbFLNA, confirming that the proposed model worked as expected. Results show that the suggested approach works well for predicting the onset of type 2 diabetes. Table 4 shows the overall success of the unique CbFLNA approach.

Table.4: Overall performance of CbFLNA

Performance of CBRDTF	
Efficiency parameters	**Performance (%)**
Precision	99.70
f-score	99.70
Recall	99.70
Time	74.1547s
Error rate	0.002
Accuracy	99.70

In Table 4 we can see how well the new CbFLNA technique worked in general. The recognition effectiveness has been reported to be 99.70% across all prediction measures. The accuracy achieved by the suggested method was 99.70%. When compared to the current process, the accuracy rate is high. By achieving the best possible outcome in the prediction parameters, the suggested model demonstrated the exceptional performance of the new CbFLNA.

6. CONCLUSION

In order to forecast patients' risk of developing type 2 diabetes and prediabetes using genomic data, this study employed the innovative CbFLNA methodology. The input data is first cleaned up by removing the Nan characteristics. Following the termination of the Nan features, the feature selection phase is employed to choose the most relevant features for categorization. Afterwards, the classification layer was thought to take the error-free data as input. To forecast T2DM with an accuracy score, the Chimp optimization technique is employed. Two scenarios, normal and abnormal, are included in the forecast. Then, the methodology's performance was determined using the classification. The accuracy of the technique for predicting type 2 diabetes is 99.70%, and the prediction score is 1% higher than that of the traditional method. Also, the error rate for the proposed task is 0.002%. The result was a 1% improvement over the conventional method. A high performance was achieved by the suggested methods. This effort does not, however, implement security. This approach, when used in conjunction with future security implementation designs, will provide superior outcomes.

REFERENCES

1) Shojima, Nobuhiro, and Toshimasa Yamauchi. "Progress in the genetics of type 2 diabetes and diabetic complications." Journal of Diabetes Investigation (2023).
2) Jin, Heejin, et al. "Identification of genetic variants associated with diabetic kidney disease in multiple Korean cohorts via a genome-wide association study meta-analysis." BMC Medicine 21.1 (2023): 16.
3) Rastogi, Rashi, and Mamta Bansal. "Diabetes prediction model using data mining techniques." Measurement: Sensors 25 (2023): 100605.
4) Elashi, Asma A., et al. "The Genetic Spectrum of Maturity-Onset Diabetes of the Young (MODY) in Qatar, a Population-Based Study." International Journal of Molecular Sciences 24.1 (2023): 130.
5) Ghalwash, Mohamed, et al. "Islet autoantibody screening in at-risk adolescents to predict type 1 diabetes until young adulthood: a prospective cohort study." The Lancet Child & Adolescent Health (2023).
6) Ahmadizar, Fariba, et al. "Association between Arterial Stiffness/Remodeling and New-onset Type 2 Diabetes Mellitus in General Population." Diabetes Research and Clinical Practice (2023): 110237.
7) Khan, Natasha, et al. "Role of 19 SNPs in 10 genes with type 2 diabetes in the Pakistani population." Gene 848 (2023): 146899.
8) Abdollahi, Jafar, and Babak Nouri-Moghaddam. "Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction." Iran Journal of Computer Science 5.3 (2022): 205-220.
9) Florez, Jose C., and Ewan R. Pearson. "A roadmap to achieve pharmacological precision medicine in diabetes." Diabetologia 65.11 (2022): 1830-1838.
10) Deutsch, Aaron J., Emma Ahlqvist, and Miriam S. Udler. "Phenotypic and genetic classification of diabetes." Diabetologia 65.11 (2022): 1758-1769.

11) Srinivasa Sai Abhijit Challapalli. (2024). Optimizing Dallas-Fort Worth Bus Transportation System Using Any Logic. Journal of Sensors, IoT & Health Sciences, *2*(2024): 40-55.
12) Whitfield, John B., et al. "A genetic risk score and diabetes predict the development of alcohol-related cirrhosis in drinkers." Journal of Hepatology 76.2 (2022): 275-282.
13) Yang, Hui, et al. "Risk prediction of diabetes: big data mining with the fusion of multifarious physical examination indicators." Information Fusion 75 (2021): 140-149.
14) Naz, Huma, and Sachin Ahuja. "Deep learning approach for diabetes prediction using PIMA Indian dataset." Journal of Diabetes & Metabolic Disorders 19 (2020): 391-403.
15) Tigga, Neha Prerna, and Shruti Garg. "Prediction of type 2 diabetes using machine learning classification methods." Procedia Computer Science 167 (2020): 706-716.
16) Srinivasa Sai Abhijit Challapalli. Sentiment Analysis of the Twitter Dataset for the Prediction of Sentiments. Journal of Sensors, IoT & Health Sciences, 2(2024), 1-15.
17) Zhou, Huaping, Raushan Myrzashova, and Rui Zheng. "Diabetes prediction model based on an enhanced deep neural network." EURASIP Journal on Wireless Communications and Networking 2020 (2020): 1-13.
18) K. Vijay Kumar, S. Sravanthi, Syed Shujauddin Sameer, & K. Anil Kumar. Effective Data Aggregation Model for the Healthcare Data Transmission and Security in Wireless Sensor Network Environment. *Journal of Sensors, IoT & Health Sciences*, *1*(2023): 40-50.
19) Swapna Saturi, Rafia Adiba, I. Yeshwanth Reddy, Md.Arsalan Shareef, & D.Prathyush. Advanced MRI-Based Alzheimer ’s disease Classification with Hybrid Convolutional Neural Networks. Journal of Computer Allied Intelligence, 3(2025): 31-39.
20) Rai, Vivek, et al. "Single-cell ATAC-Seq in human pancreatic islets and deep learning upscaling of rare cells reveals cell-specific type 2 diabetes regulatory signatures." Molecular Metabolism 32 (2020): 109-121.
21) Redondo, Maria J., et al. "The clinical consequences of heterogeneity within and between different diabetes types." Diabetologia 63 (2020): 2040-2048.
22) Alex, Suja A., et al. "Deep LSTM Model for Diabetes Prediction with Class Balancing by SMOTE." Electronics 11.17 (2022): 2737.
23) Ellahham, Samer. "Artificial intelligence: the future for diabetes care." The American journal of medicine 133.8 (2020): 895-900.
24) Park, Byoungjin, Hye Sun Lee, and Yong-Jae Lee. "Triglyceride glucose (TyG) index as a predictor of incident type 2 diabetes among nonobese adults: a 12-year longitudinal study of the Korean Genome and Epidemiology Study cohort." Translational Research 228 (2021): 42-51.
25) Awotunde, Joseph Bamidele, et al. "Prediction and classification of diabetes mellitus using genomic data." Intelligent IoT systems in personalized health care. Academic Press, 2021. 235-292.

26) Abdallah, Maher, et al. "Diagnostic accuracy of the Finnish Diabetes Risk Score for the prediction of undiagnosed type 2 diabetes, prediabetes, and metabolic syndrome in the Lebanese University." Diabetology & Metabolic Syndrome 12 (2020): 1-11.
27) Chatrati, Saiteja Prasad, et al. "Smart home health monitoring system for predicting type 2 diabetes and hypertension." Journal of King Saud University-Computer and Information Sciences 34.3 (2022): 862-870.
28) Laakso, Markku. "Biomarkers for type 2 diabetes." Molecular Metabolism 27 (2019): S139-S146.
29) Jin, Heejin, et al. "Identification of genetic variants associated with diabetic kidney disease in multiple Korean cohorts via a genome-wide association study meta-analysis." BMC Medicine 21.1 (2023): 16.
30) Lu H, Uddin S, Hajati F, Moni MA, Khushi M. A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus. Applied Intelligence. 2022 Feb;52(3):2411-22.
31) Ganie SM, Malik MB. An ensemble machine learning approach for predicting type-II diabetes mellitus based on lifestyle indicators. Healthcare Analytics. 2022 Nov 1;2:100092.
32) Shuja M, Mittal S, Zaman M. Effective prediction of type II diabetes mellitus using data mining classifiers and SMOTE. Advances in Computing and Intelligent Systems: Proceedings of ICACM 2019 2020 (pp. 195-211). Springer Singapore.
33) Balasubramanian S, Kashyap R, CVN ST, Anuradha M. Hybrid prediction model for type-2 diabetes with class imbalance. 2020 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT) 2020 Dec 20 (pp. 1-6). IEEE.