*[1]Sreenivas Pratapagiri,

[2]C. V. Guru Rao

Intelligent Prediabetes and Type-2 Diabetes Prediction from the Genomic Data using an Optimal Framework

Journal of Electrical Systems

***Abstract:* - Objective**: Diabetes is a chronic illness with a high prevalence that has a negative impact on people's quality of life and a high death rate in both industrialized and developing nations. The patient's diabetes type 2 prediction is the most critical study in medical research. Several prediction models exist to predict type 2 diabetes. However, the relevant result was not satisfied because of poor quality. The gene data contains Nan features, which maximize the complexity of predicting type 2 diabetes. The demerits resulted in low prediction and performance scores.

Material and Methods: The proposed work aims to develop a novel chimp-based functional link neural approach (CbFLNA) to predict type 2 diabetes. The proposed methodology is pre-processing; feature selection, classification, and gene expression have been performed. Initially, the genomic database was imported, the data were pre-processed, and the meaning features were extracted. Then, predict the type 2 diabetes of the patients and classify the conditions.

Results: The performance was measured and the model attained 99.7% accuracy, 99.7% precision, 99.7% recall, 99.7% F-score, 0.002 error rate and execution time of 74.1547s.

Conclusion: The presented model attained a high exactness score in the prediction.

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

1 INTRODUCTION

Diabetic is a metabolism disease characterized by elevated blood sugar content caused by inadequate insulin secretion or production [1]. In 2010, hyperglycaemia was expected to strike 285 million individuals worldwide [2]. This number has been raised to a total of 552 million through 2030, based on the illness's current rate of progression [3]. Diabetes will affect approximately one out of ten persons by 2040 [4]. Individualized behaviours, diverse habits, and rising living levels are increasing the prevalence of diabetes [5]. Therefore, research into the practical and quick diagnosis and treatment of diabetes is beneficial [6]. The diagnosis of diabetes is founded on genomic sequences [7]. It produces a more reliable and appropriate outcome that facilitates the adoption of healthier behaviours, which are more unlikely to end up in insulin shortly [8]. Moreover, the diabetes prediction using genomic data is exposed in Fig. 1.

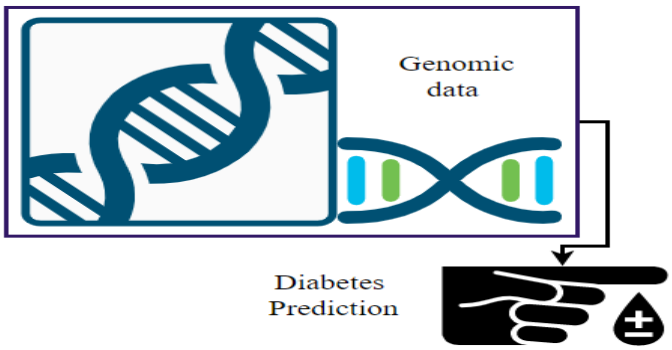


Fig. 1 Diabetes prediction using genomic data

[1] Research Scholar, Department of Computer Science & Artificial Intelligence,
SR University, Anantha Sagar, Hasanparthi, Hanumakonda - 506371, Telangana, India.
E-mail: sreenivaspratapagirisp.kits@gmail.com
[2]Professor & Dean, Department of Computer Science and Artificial Intelligence,
SR University, Anantha Sagar, Hasanparthi, Hanumakonda - 506371, Telangana, India.
E-mail: cvgururaogr@outlook.com

A highly accurate illness diagnosis permits people with subsequent diseases to delay the disease's progression and experience improved general well-being [9]. There are two artificial intelligence techniques [10]: predicting future conditions and diagnosing abnormalities [11]. Based on current and previous health issues, forward forecasting methods may indicate the onset of diabetes [12]. The condition is affected by hereditary factors and can result in multiple problems and a higher likelihood of premature mortality [13]. Previous potential gene and association research has identified several disease-related traits on a limited basis [14]. Regardless of the localization of an array of genetics, those genetic investigations were deemed ineffective throughout their entirety [15]. Genome-based broadest association approaches have recently facilitated the discovery of genes linked to type-2 diabetes illness [16]. Moreover, the genome data includes single-molecule polymorphisms [17] that have been demonstrated to describe various diseases and are prevalent in people's chromosomes within a group of organisms [18].

Several neural approaches were implemented in the past, such as deep networks [21], spatial-temporal systems [22], etc., and those models are sufficient to predict type-2 diabetes by tuning the classification layer by other feature selection methods [19]. However, it is not suitable and adequate for prediabetes prediction because of its limited features. Hence, the present study has intended to design the optimal neural system as the diabetes recognition framework.

The following is an examination of the study's essential contribution,

- Primarily, the system was trained using the genomic data.
- As a result, a novel CbFLNA is designed with needed functional parameters to recognize the described diabetes types.
- The pre-processing stage involved filtering the noise features, which is performed in the second layer of CbFLNA.
- Henceforth, the classification layer receives the modified data as input, and then the required elements are selected.
- Moreover, from the specified segments, type-2 diabetes and prediabetes were recognized and classified.
- Finally, the performance of the proposed model was verified in terms of recall, precision, Accuracy, f-score and error rate, and improvement score than the existing models were noted.

The current work paper is presented as related work in the second section, and the problem statements of the conventional method are exposed in section three. Then, the solution for the defined problem is elaborated in section four. The validated result of the novel solution is discussed in section 5. In the end, the research paper was concluded in section 6.

2 RELATED WORKS

Some of the recent papers related to this current work are described below,

The Diabetes disease severity is classified into different classes based on the severity. Likewise, Awotunde et al. [21] have introduced deep neural mechanism features for detecting and identifying diabetes and its severity. Here, the genomic data is utilized to value the deep network reliability. In addition, the deep network functioning layers were regulated by the genetic algorithm problem. Finally, the outcomes were compared separately, that is, before and after applying the optimization function. However, in some cases, type-2 diabetes is wrongly detected.

To reduce the diabetes impact, the prediabetes recognition system is the most needed task. So, Abdallah et al. [22] have introduced a spatial-temporal system based on the intelligent network. By preceding the implementation, a high sensitivity score was reported for the prediabetes prediction, sufficient to estimate the possibility of diabetes occurrence. However, it required an additional feature selection model based on the optimization features to extract the needed meaningful features.

Besides the innovative, intelligent approaches, smart devices are effectively utilized for diabetes detection. So, Chatrati et al. [23] have implemented the diabetes recognition framework by activating smart devices in the human body. It gives the predictive outcomes regularly by monitoring the human body's biological parameters.

But considering the other innovative approaches, this prediction gadget is expensive. Also, the classification and different feature selection are not performed.

Laakso et al. [24] have introduced the biomarker strategy for analyzing the body status regularly through the chief biological parameters like blood pressure and insulin level. This process is considered the automatic process for finding diabetes at any time. Also, the monitored outcomes were shared with the medical data cloud. Here, the biological changes were predicted by the signal of the physical parameter. However, it is not suitable for the prediabetes prediction and classification.

Jin et al. [25] have introduced the kidney disease severity analysis framework based on diabetes prediction. Here, the genomic data and the chronic disorder data were considered. From this evaluation, the impacts of diabetes on kidney disease were predicted more precisely. Thus, the diabetes prediction framework is the most needed task for the medical application for analyzing and evaluating the different disease features. However, the chronic data alone does not give an accurate diabetes prediction outcome.

3 SYSTEM MODEL AND PROBLEM STATEMENT

The current research study's primary goal is to identify type-2 and prediabetes. However, it is not easy to detect the diabetes severity from the gene data [20]. Several feature selection methods were executed in the past for the genomic data for finding the diabetes features. However, it is insufficient for analyzing and detecting all diabetes parts. Considering this, the present work has executed the optimized deep network framework as the gene feature analysis model to predict diabetes types.

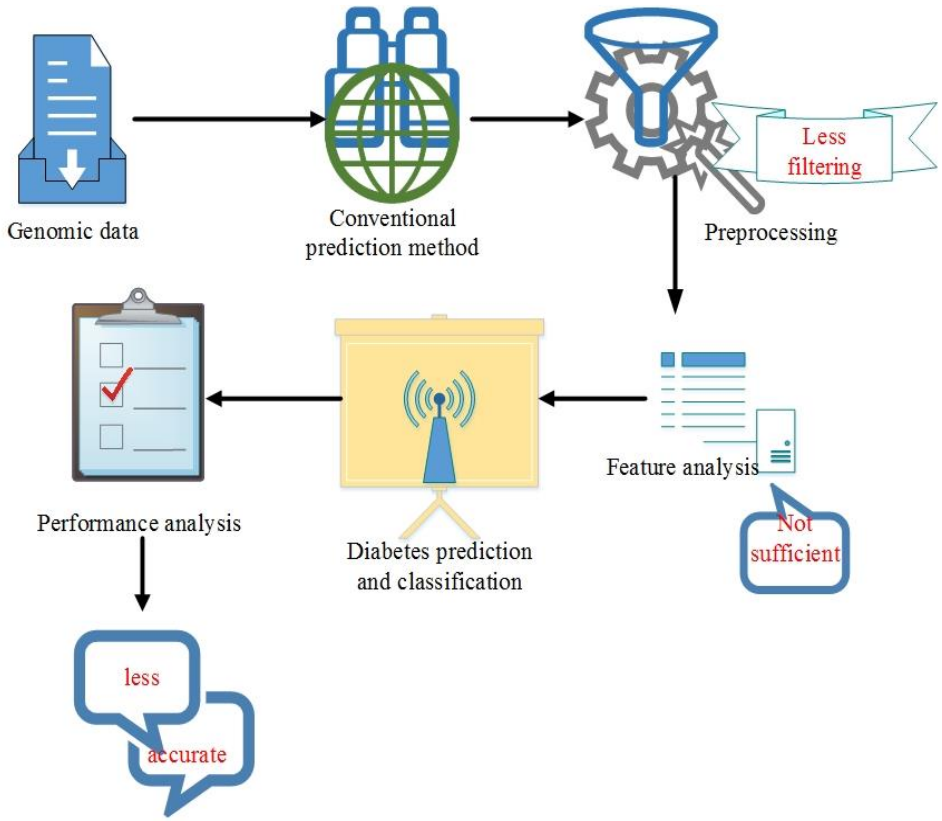


Fig. 2 Difficulties in the conventional prediction method

The prediction system of the conventional standard method is described in Fig. 2. The pre-processing system in the process has filtered the lesser amount of noise features. The feature selection has been selecting the irrelevant disease features. Hence, feature selection is not sufficient for extracting meaningful features. It maximizes the complexity of the algorithm and minimizes the accuracy rate. To overcome these problems, the Chimp-based functional link neural approach was implemented to predict type-2 diabetes.

4 PROPOSED METHODOLOGY

A novel Chimp-based Functional link neural approach (CbFLNA) is designed to analyze and predict Type-2 diabetes and prediabetes. Moreover, this study's genomic database is considered for estimating type-2 diabetes and prediabetes. Here, prediabetes and type-2 diabetes are calculated based on the insulin and the blood pressure rate. Primarily, the genomic data was pre-processed and taken as the input of the classification layer, and then feature analysis and diabetes prediction and classification were performed. Finally, the performance metrics were evaluated and contrasted with each other models.

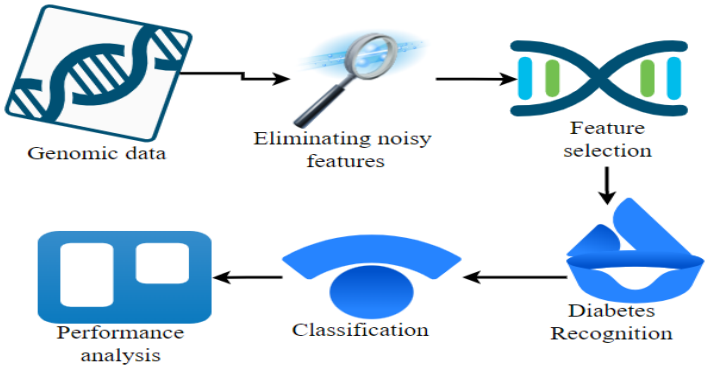


Fig. 3 Proposed architecture

The building structure of the proposed methodology is described in Fig. 3. The need for this methodology is to analyze and predict diabetes type-2 and prediabetes. The performance assessment in the gene prediction in each dataset is described by performing the comparison studies.

4.1 Process of the proposed methodology

The process of the proposed methodology contains five layers: input layer, filtering layer, classification layer, optimization layer and output layer. The process can be done using the Chimp-based functional link neural approach (CbFLNA).

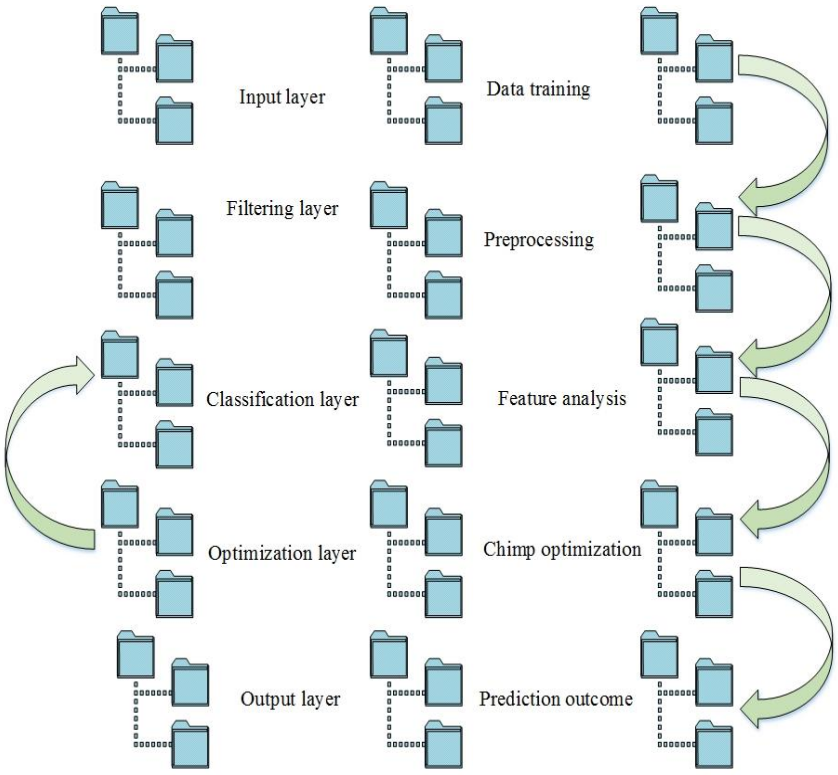


Fig. 4 Processing layer of CbFLNA

The functioning layers of the novel CbFLNA are described in Fig. 4. Here, the collected dataset was imported into the input layer. The pre-processing has been done in the filtering layer. The noise-filtered data was obtained and imported to the classification layer. Then, the parameters are classified and adopted using the chimp optimization function, and the optimized result is received in the output layer.

4.1.1 Pre-processing

Initially, the genomic data was collected and imported into the Python environment. The Eqn. (1) described the function of data training.

$$G[d_n] = (d1, d2, ...d_k) \tag{1}$$

Where G is represented as the gene expression, the data was denoted as d and $G(d)$ is designated as the data of the gene expression and k is the number of data. The data quality is enormous, affecting the prediction result, and it has minimized the algorithm's complexity. It provides high Accuracy, and the complexity of the dataset was reduced. Hence, the pre-processing of data plays a significant role in the method.

$$P[d_n] = x[(d_n - n_a] \tag{2}$$

The input dataset contains both normal and the Nan features. The pre-processing function eliminates the Nan features and normalization in the dataset and prepares a straightforward dataset. The pre-processing function is done by using the Eqn. (2). Where P is denoted as the pre-processing variable. Both Nan features are denoted as n_a , and x is the tracking variable. The outcome of the pre-processing function is a straightforward dataset.

4.1.2 Feature selection

Selecting the subset is the feature selection function of the maximum relevant features in the dataset. It reduces the running time and maximizes the prediction accuracy. The feature selection process involves selecting and extracting the maximum matched disease features from the dataset. The dataset contains both maximum and minimum matching features of the disease.

$$F = G(d_n) + a[d_n - f_s] \tag{3}$$

The Eqn. (3) has done the feature selection process. Where the variable of the feature selection function is denoted as F . The dataset contains both wanted and unwanted features. The feature selection process is done by getting the wanted features. The tracking variable is denoted as a , and the wanted features are denoted as f_s .

4.1.3 Classification and Prediction

The classification function determines the disease features and the normal data. Usually, the initial dataset consists of both normal parts and abnormal features. Hence, the disease feature dataset was used to find the gene expression and classify the disease features. Finally, the disease features were categorized in the classification layer.

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

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. The classification function was executed in two cases [0, 1]. Where 0 indicates the normal features and 1 indicates the abnormal features.

4.1.4 Disease prediction

After the feature selection and the classification function, predict the tested disease samples in the present gene features. The disease prediction can be done by using the Eqn. (5).

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

The probability score is denoted as one, which has analyzed the maximum current genes. More than one probability represents 100% in the tested samples of the present gene. The tested disease samples can be described as follows ds. Some essential features include glucose, protein, blood cholesterol levels and blood pressure $\max\_Gn$. These gene features have been different based on the trained genomic disease database.

Algorithm 1: CbFLNA

Start
{
Data initialization()
{
$\text{int } G(d_n) = 1,2,3,\ldots n;$
//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()
{
$\text{Pr}ob \longrightarrow Classification(\max\_Gn)$
// classified the maximum number of genes
}
}
Stop

The step-by-step processes in implementing the proposed model are described in Fig. 5. Then, the pseudo-code for the elaborated mathematical formulation is described in the Algorithm 1. The dataset has both normal and disease features. In the initial phase, feature selection was done, and the disease samples were classified.

Moreover, the gene expression was identified only for the disease samples. After processing the described steps, the performance of the CbFLNA has been calculated by various classification metrics.

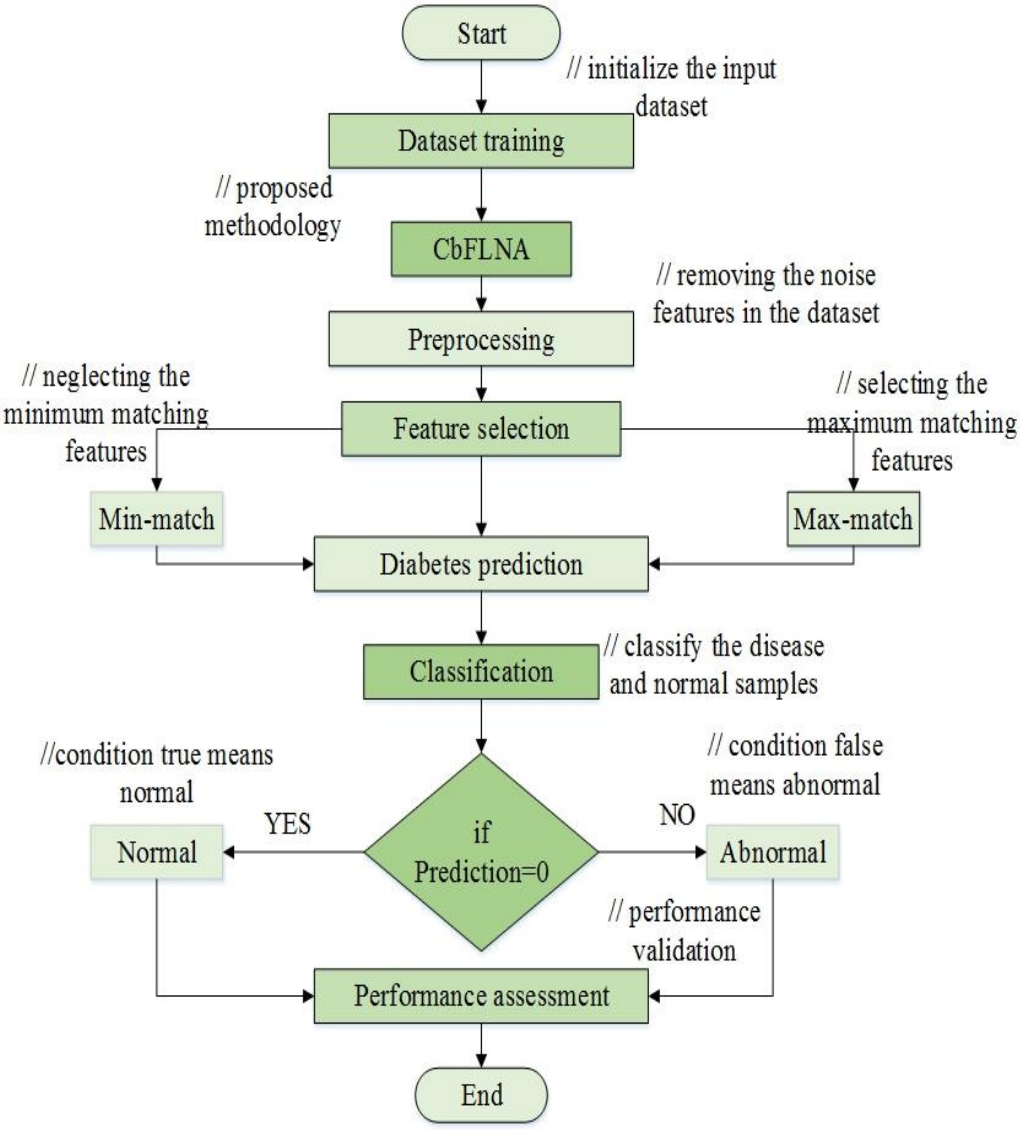


Fig. 5 The flow diagram of CbFLNA

5 RESULT AND DISCUSSION

The novel CbFLNA methodology was validated in the Python programming environment and carried out in the Windows 10 framework. The database taken into account for testing and validation is genomic data. The data contains both normal and abnormal data.

Table 1 Specification of Execution parameters

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

The execution parameters are specified in Table 1. Initially, the training flaws were removed in the pre-processing phase. Then, the error-cleaned data was imported into the classification phase. In addition, the gene expression was forecasted, and the performance assessment can be calculated using various metrics.

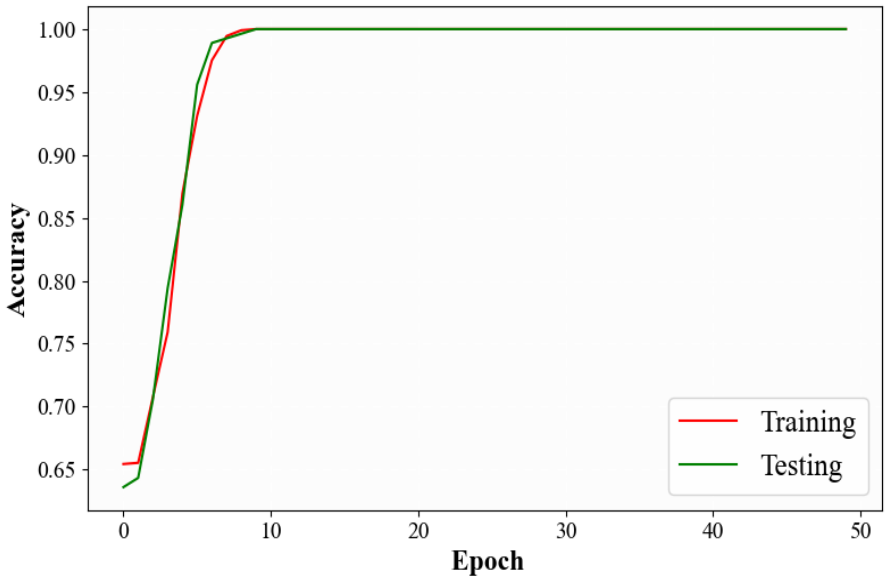
5.1 Case Study

To check the functioning performance of the proposed methodology, some test validation was executed, and the results were described systematically. For the test validation, the diabetes and gene dataset was adopted. It contains a total of 1701 samples. Of the total samples, 1360 are training samples, and 341 are testing samples.

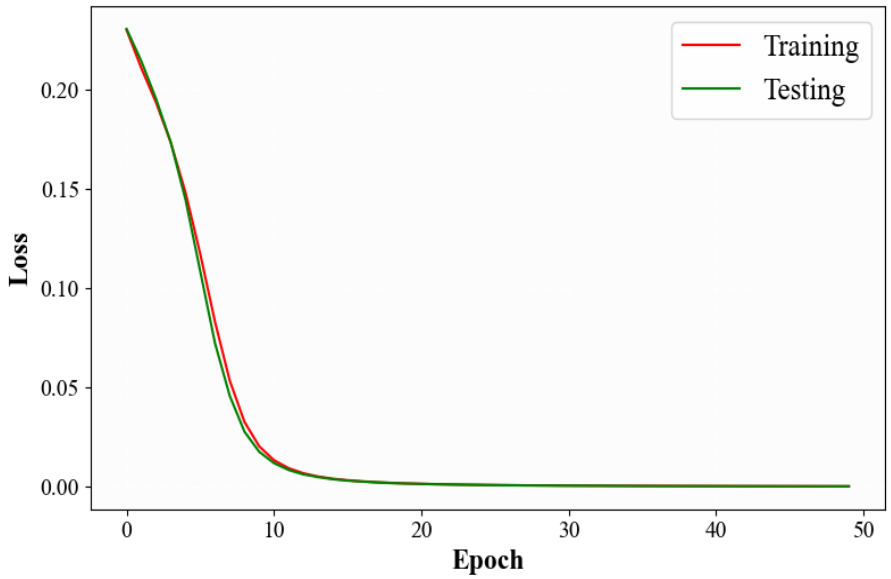
Table 2 Database details

Total no of samples: 1701	
Normal	1118
Abnormal	583
Training (80%):1360	
Normal	884
Abnormal	476
Testing (20%): 341	
Normal	234
Abnormal	107

Table 2 describes the details of the database. Moreover, samples are considered for training, in that 884 are the standard samples, and 476 are the abnormal samples. Also, the samples considered for testing is 341, in that 234 are normal and 107 are abnormal.



a)



b)

Fig. 6 a) Training accuracy b) Training loss

Fig. 6 represents the accuracy and loss assessment of the CbFLNA over the training epoch. The train and test validation exactness score is used to assess the accuracy of the diabetes prediction system. The failure ratio of the implemented framework is calculated using loss measures and assessed using dual training and test validation phases.

The outcome of the predicted result was exposed as network confusion in Fig. 7. The classification outcome was attained in the form of positive and negative scores for both true and false classes. Here, the prediction was classified into two cases: 0 and 1. 0 indicates the normal and 1 shows the abnormal.

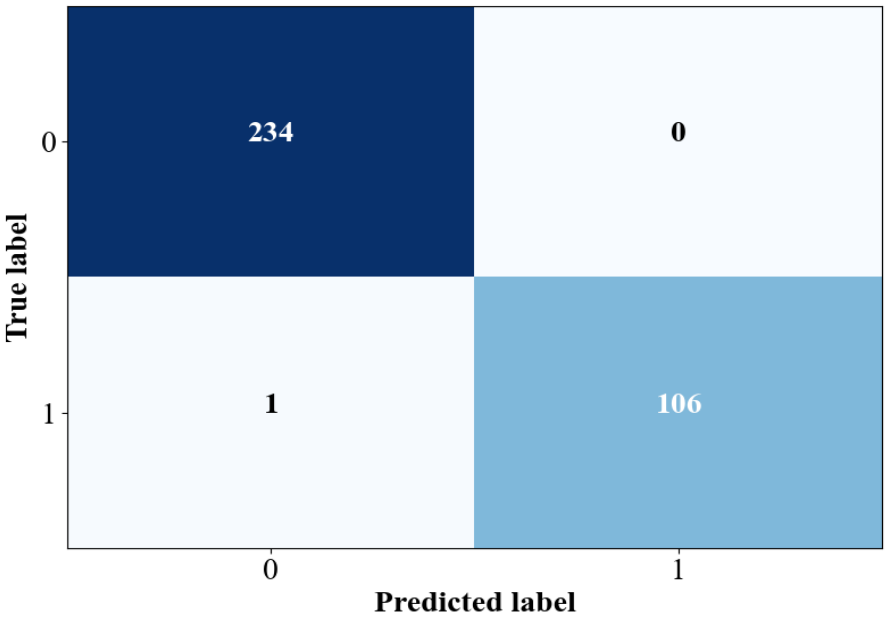


Fig. 7 Confusion matrix

5.2 Performance assessment

The improvement score of the designed novel CbFLNA was justified by calculating the chief metrics like Accuracy, recall, precision and f-score taken for the validation. For analyzing the performance improvement,

take the recently associated model. The existing models such as Random forest (RF) [26], Bagged decision tree (BDT) [27], Multilayer perceptron (MLP) [28], and Edited nearest neighbour (ENN) [29].

5.2.1 *Precision*

The precision denotes the value of the optimistic prediction that describes the part of the abnormal case among the entire samples. The metrics precision measure was calculated by using the Eqn. (6).

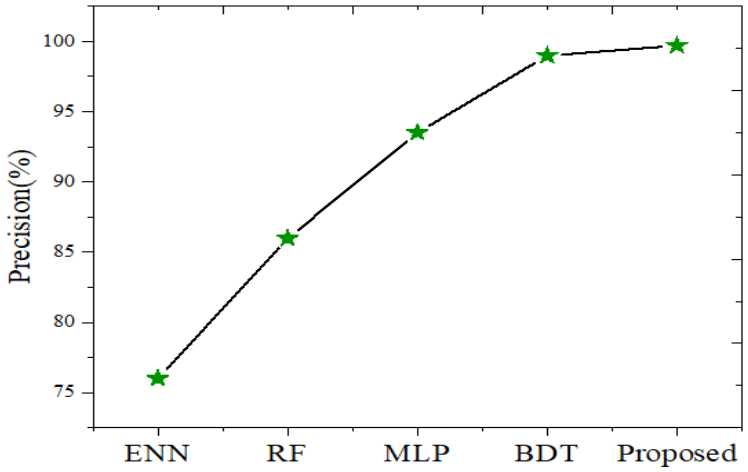


Fig. 8 Precision assessment

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

The precision earned by the ENN method is 76%, the RF method gained 85.97%, the BDT method gained 98.98%, and the MLP method gained 93.5%. Considering the compared mechanism, the proposed novel CbFLNA methodology earned 99.70% precision. The statistics are revealed in Fig. 8.

5.2.2 *Accuracy*

Accuracy defines the correct value of prediction instances as abnormal or normal data. Accuracy is measured by taking the average score of the prediction's positive and negative scores. The accuracy measure was calculated by using the Eqn. (7).

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

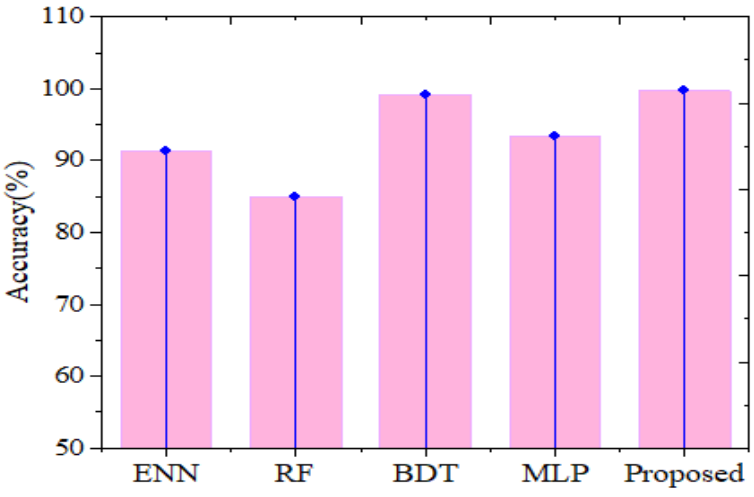


Fig. 9 Accuracy assessment

The Accuracy earned by the ENN method was 91.4%, The BDT method gained 99.14%, the RF method earned 84.95%, and the MLP method earned 93.46%. Considering the compared mechanism, the proposed novel CbFLNA methodology made 99.70% Accuracy. The statistics are revealed in Fig. 9.

5.2.3 *Recall*

Recall denotes the part of the abnormal case that has retrieved the entire abnormal sample. Eqn. (8) formulates the metrics recall.

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

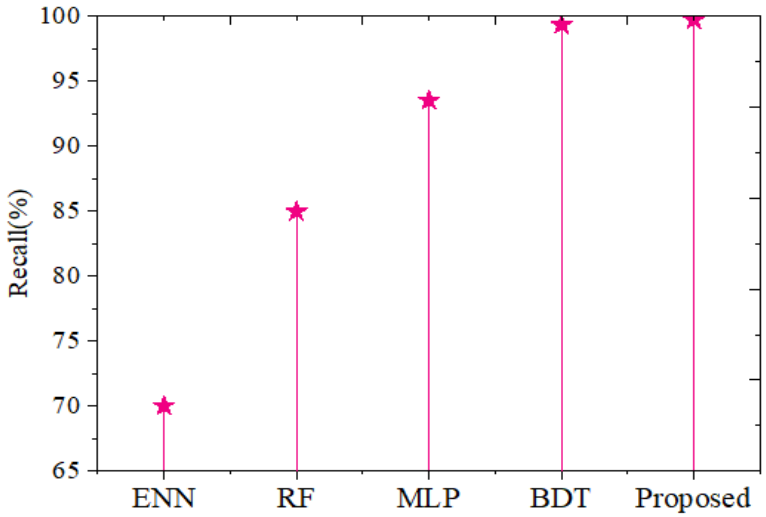


Fig. 10 Recall assessment

The recall earned by the ENN method was 70%, the BDT method made 99.32%, the RF method earned 84.95, and the MLP method earned 93.5%. Considering the comparison mechanism, the proposed novel CbFLNA methodology gained 99.70% of recall. The statistics are revealed in Fig. 10.

5.2.4 *F-score*

The F-score shows the system performance by relating the elements of precision and recall. Hence, the f-score is the average prediction statistics of true and false scores. The metrics f-score was validated by using the Eqn. (9).

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

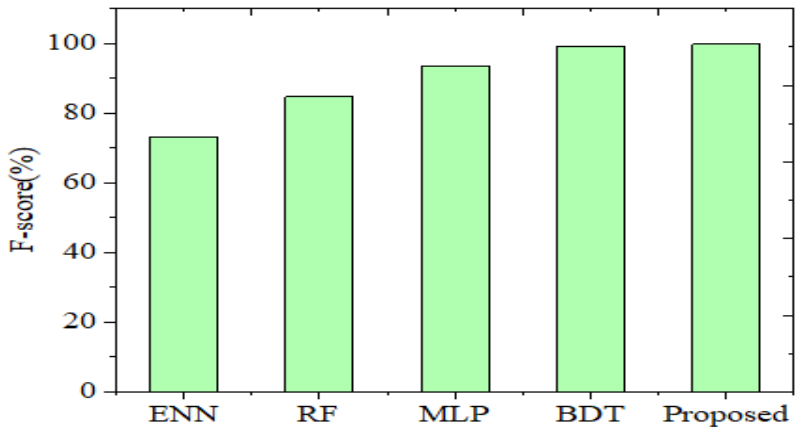


Fig. 11 F-score assessment

The f-score earned by the ENN method is 73%, the RF method gained 84.77%, the BDT method earned 99.15, and the MLP method gained 93.4% of the f-score. Considering the compared mechanism, the proposed novel CbFLNA methodology gained 99.70% of the f-score. The statistics are revealed in Fig. 11.

5.2.5 Error rate

The metrics error rate measures the methodology's prediction error degree concerning the accurate methodology. The error rate can be measured by using the Eqn. (10).

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

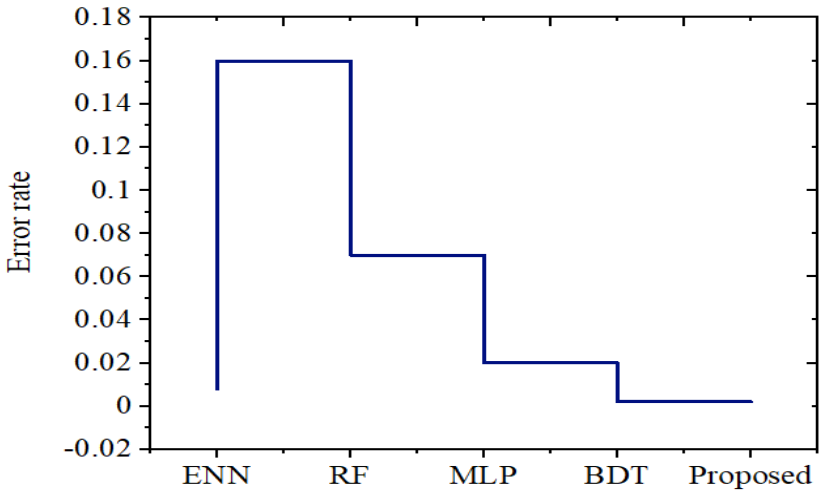


Fig. 12 Error rate assessment

The error rate of the ENN method is 0.008, the BDT method is 0.02, the RF method is 0.16, and the MLP method is 0.07. Considering the compared mechanism of the proposed novel CbFLNA methodology, the error rate is 0.002. The statistics are revealed in Fig. 12.

5.2.6 Time

Time required to complete the overall process of predicting type 2 diabetes. In the proposed methodology, the time taken for the training is 4.757 seconds, the testing time is 0.175 seconds, and the execution time is 74.1547 seconds. The Comparison assessments is demonstrated in Table 3.

Table 3 Comparison assessments

Comparison statistics					
Methods	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)	Error rate
ENN	91.4	76	70	73	0.008
BDT	99.14	98.98	99.32	99.15	0.02
RF	84.95	85.97	84.95	84.77	0.16
MLP	93.46	93.5	93.5	93.4	0.07
Proposed	99.70	99.70	99.70	99.70	0.002

5.3 Discussion

The unique CbFLNA received a highly accurate score from all performance analyses, which supported the successful operation of the suggested model. It has been confirmed that the proposed methodology is ideally suited for type-2 diabetes prediction. The innovative CbFLNA methodology's overall effectiveness is demonstrated in Table 4.

Table 4 Overall performance of CbFLNA

Performance of CBRDTF	
Efficiency parameters	Performance (%)
Accuracy	99.70
precision	99.70
recall	99.70
f-score	99.70
Error rate	0.002
Time	74.1547s

The overall performance of the novel CbFLNA methodology is described in Table 4. All predicting metrics have reported 99.70% recognition effectiveness. The proposed method earned 99.70% of Accuracy. The accuracy rate is high when compared to the existing process. The proposed model made the optimum consequence in the prediction parameters, revealing the novel CbFLNA's high performance.

6 CONCLUSION

The novel CbFLNA methodology was implemented in this study to predict both type-2 diabetes and prediabetes of the patients from the genomic data. First, the Nan features are eliminated from the input data. After stopping the Nan features, they select the meaningful features for classification in the feature selection phase. Then, the error-free data was considered as the input of the classification layer. The Chimp optimization algorithm is utilized to predict type 2 diabetes with the exactness score. The prediction contains two cases, normal and abnormal. After that, based on the classification, the methodology's performance was calculated. The type 2 diabetes prediction system's Accuracy is 99.70%, and the prediction score improved by 1% compared to the conventional method. Also, the error rate of the proposed work is 0.002%. When compared to the traditional way, the score improved by 1%. The proposed methodology earned a high performance. However, the security implementation is not done in this work. In future, designing the security implementation along with this method will give better results.

ACKNOWLEDGEMENT

None

Compliance with Ethical Standards

1. Disclosure of Potential Conflict of Interest:

   The authors declare that they have no potential conflict of interest.

2. Statement of Animal and Human Rights

i. Ethical Approval

All applicable institutional and/or national guidelines for the care and use of animals were followed.

ii. Informed Consent

For this type of analysis formal consent is not needed.

DATA AVAILABILITY STATEMENTS

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHORS' CONTRIBUTION

Authors S.P. and C.V.G.R. have contributed equally to the work.

REFERENCES

[1] Nnamudi AC, Orhue NJ, Ijeh II, Nwabueze AN. Finnish diabetes risk score outperformed triglyceride-glucose index in diabetes risk prediction. J Diabetes Metab Disord. 2023;22:1337–1345. https://doi.org/10.1007/s40200-023-01252-y

[2] Valipour M, Khalili D, Solaymani-Dodaran M, Motevalian SA, Khamseh ME, Baradaran HR. External validation of the UK prospective diabetes study (UKPDS) risk engine in patients with type 2 diabetes identified in the national diabetes program in Iran. J Diabetes Metab Disord. 2023;22:1145–1150. https://doi.org/10.1007/s40200-023-01224-2

[3] Ganie SM, Malik MB, Arif T. Performance analysis and prediction of type 2 diabetes mellitus based on lifestyle data using machine learning approaches. J Diabetes Metab Disord. 2022;21:339–352. https://doi.org/10.1007/s40200-022-00981-w

[4] Elashi AA, Toor SM, Diboun I, Al-Sarraj Y, Taheri S, Suhre K, Abou-Samra AB, Albagha OME. The Genetic Spectrum of Maturity-Onset Diabetes of the Young (MODY) in Qatar, a Population-Based Study. Int J Mol Sci. 2023;24(1):130. https://doi.org/10.3390/ijms24010130

[5] Ghalwash M, Anand V, Lou O, Martin F, Rewers M, Ziegler AG, Toppari J, Hagopian WA, Veijola R. Islet autoantibody screening in at-risk adolescents to predict type 1 diabetes until young adulthood: a prospective cohort study. Lancet Child Adolesc Health 2023. https://doi.org/10.1016/S2352-4642(22)00350-9

[6] Ahmadizar F, Wang K, Roos M, Bos M, Mattace-Raso F, Kavousi M. Association between Arterial Stiffness/Remodeling and New-onset Type 2 Diabetes Mellitus in General Population. Diabetes Res Clin Pract. 2023;196:110237. https://doi.org/10.1016/j.diabres.2023.110237

[7] Khan N, Paterson AD, Roshandel D, Maqbool S, Fazal N, Ali L, Khurram R, Maqsood SI, Ali SHB, Khan H, Tariq K, Bano S, Azam A, Muslim I, Maria M, Agha Z, Babar M, Waheed NK, Azam M, Qamar R. Role of 19 SNPs in 10 genes with type 2 diabetes in the Pakistani population. Gene 2023;848:146899. https://doi.org/10.1016/j.gene.2022.146899

[8] Abdollahi J, Nouri-Moghaddam B. Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction. Iran J Comput Sci. 2022;5(3):205-220. https://doi.org/10.1007/s42044-022-00100-1

[9] Florez JC, Pearson ER. A roadmap to achieve pharmacological precision medicine in diabetes. Diabetologia 2022;65(11):1830-1838. https://doi.org/10.1007/s00125-022-05732-3

[10] Deutsch AJ, Ahlqvist E, Udler MS. Phenotypic and genetic classification of diabetes. Diabetologia 2022;65(11):1758-1769. https://doi.org/10.1007/s00125-022-05769-4

[11] Whitfield JB, Schwantes-An TH, Darlay R, Aithal GP, Atkinson SR, Bataller R, Botwin G, Chalasani NP, Cordell HJ, Daly AK, Day CP, Eyer F, Foroud T, Gleeson D, Goldman D, Haber PS, Jacquet JM, Liang T, Liangpunsakul S, Masson S, Seth D. A genetic risk score and diabetes predict the development of alcohol-related cirrhosis in drinkers. J Hepatol. 2022;76(2):275-282. https://doi.org/10.1016/j.jhep.2021.10.005

[12] Yang H, Luo Y, Ren X, Wu M, He X, Peng B, Deng K, Yan D, Tang H, Lin H. Risk prediction of diabetes: big data mining with the fusion of multifarious physical examination indicators. Inf Fusion 2021;75:140-149. https://doi.org/10.1016/j.inffus.2021.02.015

[13] Naz H, Ahuja S. Deep learning approach for diabetes prediction using PIMA Indian dataset. Journal of Diabetes & Metabolic Disorders 2020;19:391-403. https://doi.org/10.1007/s40200-020-00520-5

[14] Tigga NP, Garg S. Prediction of type 2 diabetes using machine learning classification methods. Procedia Comput Sci. 2020;167:706-716. https://doi.org/10.1016/j.procs.2020.03.336

[15] Zhou H, Myrzashova R, Zheng R. Diabetes prediction model based on an enhanced deep neural network. Eurasip J Wirel Commun Netw. 2020;2020:1-13. https://doi.org/10.1186/s13638-020-01765-7

[16] Rai V, Quang DX, Erdos MR, Cusanovich DA, Daza RM, Narisu N, Zou LS, Didion JP, Guan Y, Shendure J, Parker SCJ, Collins FS. Single-cell ATAC-Seq in human pancreatic islets and deep learning upscaling of rare cells reveals cell-specific type 2 diabetes regulatory signatures. Mol Metab. 2020;32:109-121. https://doi.org/10.1016/j.molmet.2019.12.006

[17] Redondo MJ, Hagopian WA, Oram R, Steck AK, Vehik K, Weedon M, Balasubramanyam A, Dabelea D. The clinical consequences of heterogeneity within and between different diabetes types. Diabetologia 2020;63:2040-2048. https://doi.org/10.1007/s00125-020-05211-7

[18] Alex SA, Jhanjhi NZ, Humayun M, Ibrahim AO, Abulfaraj AW. Deep LSTM Model for Diabetes Prediction with Class Balancing by SMOTE. Electronics 2022;11(17):2737. https://doi.org/10.3390/electronics11172737

[19] Ellahham S. Artificial intelligence: the future for diabetes care. Am J Med. 2020;133(8):895-900. https://doi.org/10.1016/j.amjmed.2020.03.033

[20] Park B, Lee HS, Lee YJ. 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. Transl Res. 2021;228:42-51. https://doi.org/10.1016/j.trsl.2020.08.003

[21] Awotunde JB, Ayo FE, Jimoh RG, Ogundokun RO, Matiluko OE, Oladipo ID, Abdulraheem M. Prediction and classification of diabetes mellitus using genomic data. Intelligent IoT systems in personalized health care, Academic Press, 2021, pp. 235-292. https://doi.org/10.1016/B978-0-12-821187-8.00009-5

[22] Abdallah M, Sharbaji S, Sharbaji M, Daher Z, Faour T, Mansour Z, Hneino M. Diagnostic accuracy of the Finnish Diabetes Risk Score for the prediction of undiagnosed type 2 diabetes, prediabetes, and metabolic syndrome in the Lebanese University. Diabetol Metab Syndr. 2020;12:1-11.

[23] Chatrati SP, Hossain G, Goyal A, Bhan A, Bhattacharya S, Gaurav D, Tiwari SM. Smart home health monitoring system for predicting type 2 diabetes and hypertension. J King Saud Univ – Comput Inf Sci. 2022;34(3):862-870. https://doi.org/10.1016/j.jksuci.2020.01.010

[24] Laakso M. Biomarkers for type 2 diabetes. Mol Metab. 2019;27:S139-S146. https://doi.org/10.1016/j.molmet.2019.06.016

[25] Jin H, Kim YA, Lee Y, Kwon S, Do AR, Seo S, Won S, Seo JH. Identification of genetic variants associated with diabetic kidney disease in multiple Korean cohorts via a genome-wide association study meta-analysis. BMC Medicine 2023;21(1):16. https://doi.org/10.1186/s12916-022-02723-4

[26] 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. Appl Intell. 2022;52(3):2411-22. https://doi.org/10.1007/s10489-021-02533-w

[27] Ganie SM, Malik MB. An ensemble machine learning approach for predicting type-II diabetes mellitus based on lifestyle indicators. Healthcare Analytics. 2022;2:100092. https://doi.org/10.1016/j.health.2022.100092

[28] 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, Springer Singapore; 2020. pp. 195-211. https://doi.org/10.1007/978-981-15-0222-4_17

[29] 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), IEEE; 2020. DOI: 10.1109/ICMLANT50963.2020.9355975