ENHANCING DIABETES MELLITUS PREDICTION AND SURVEILLANCE: A HYBRID MACHINE LEARNING APPROACH

Hamzeh Ghorbani*[1], Mehrdad Babak Rad[2], Harutyun S. Hovhannisyan[3], Parvin Ghorbani[4], Lilit Sukiasyan[5], Delaram Ansari[6], Simin Ghorbani[7], Arsen Minasyan[7], Anahit Mkrtchyan[8], Natali Minasian[9], Aregnazan Ikilikyan[10], Rozi Yeremyan[11], Alla Krasnikova[12], Eduard Avagyan[13]

[1,7,8,10,11,13]Faculty of General Medicine, University of Traditional Medicine of Armenia (UTMA), 38a Marshal Babajanyan St., Yerevan 0040, Armenia

[2,6]Department of Dentistry, University of Traditional Medicine of Armenia (UTMA), 38a Marshal Babajanyan St., Yerevan 0040, Armenia

[3]Department of Internal Disease Propaedeutics, Yerevan State Medical University, Yerevan, Armenia
[4]Department of Cardiology, Faculty of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

[7]Department of Nursing and Midwifery, Faculty of Nursing, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

[9]Faculty of General Medicine, Yaroslavl State Medical University, Yaroslavl, Russia

[12]Department of Cardiology, Faculty of Medicine, Yerevan State Medical University, Yerevan, Armenia

hamzehghorbani68@yahoo.com*

ABSTRACT

Diabetes mellitus (DM) is a group of metabolic disorders characterized by elevated blood glucose levels, presenting a significant global health burden affecting both developed and developing countries. Many complications which may develop due to this disease such as stroke, renal failure (RF), neuropathies and vascular disorders (N&VD), myocardial infarction (MI), cardiovascular-disease (CVD), and peripheral artery disease (PAD) underscore its paramount importance. In this study 768 data points sourced from an open-access database of various input variables including skin thickness (ST), body mass index (BMI), glucose level (GL), blood pressure (BP), age (A), insulin level (IL), and diabetes pedigree function (DPF). For prediction of DM, three novel hybrid algorithms (HA) with combination of Multi-Layer Perceptron (MLP) with two robust optimized algorithms like *Cauchy Crazy Particle Swarm Optimization (CCPSO) and Particle Swarm Optimization (PSO) used. Findings show that MLP-CCPSO higher accuracy than other models in terms of accuracy (for testing subset; Recall=0.96, Accuracy=0.95, Precision=0.97, and F1-Score=0.96). The robustness of this hybrid approach is evidenced by its enhanced accuracy, improved convergence, greater robustness, increased efficiency, and optimized parameter control. Notably, our results establish the superiority of MLP-CCPSO over MLP-PSO and MLP algorithms in predicting DM.*

Keywords: Diabetes Mellitus, Metabolic Disorders, Hybrid Algorithms, Prediction and Surveillance, Prediction.

INTRODUCTION

Diabetes mellitus (DM), often known simply as diabetes, comprises a range of metabolic disorders distinguished by elevated levels of blood glucose [1-3]. This disease is a significant health challenge worldwide, impacting both developed and developing countries. The DM can precipitate numerous intricate long-term complications, including [4]:

- Stroke
- Kidney failure (KF)
- Neuropathies and vascular disorders (N&VD)
- Myocardial infarction (MI)

- Peripheral artery disease (PAD)

Insulin, synthesized by pancreas is critical for transporting intracellular glucose to insulin-dependent cells/tissues, such as liver, muscle, and adipose tissue, thereby regulating blood glucose levels within the body [5]. Table 1, show the cause of the insulin deficiency and results of them:

Table 1: Cause, results and treatments of the insulin deficiency.

No.	Cause	Mechanism	Results	Treatment
01	DM Type 1 (T1DM)	The immune system mistakenly attacks and degredate the insulin-producing β cells located in the pancreas.	Over the course of time, this immune reaction ultimately results in a complete depletion of insulin production	Manage the blood glucose levels by receiving insulin injections or using an insulin pump daily
02	DM Type 2 (T2DM)	The body develops resistance to insulin. Despite the pancreas continuing to produce insulin, the body's cells are unable to utilize it effectively	Inadequate insulin activity results in excess glucose remaining in the bloodstream, depriving cells of sufficient energy	Lifestyle changes, oral hypoglycemic medications, and drugs for treating insulin resistance
03	Chronic pancreatitis (CP)	Pancreatogenic diabetes arises. Inflammation injures the pancreas, impairing its functions in nutrient digestion, absorption, and utilization	This damage leads to insufficient production of insulin, glucagon, and other vital hormones essential for digestion	Lifestyle changes, pain management, surgical option

The insulin deficiency, due to pancreatic malfunction can lead to the following diseases [6]:

- Retinal detachment
- Sexual dysfunction
- Joint failure
- Diabetes mellitus
- Coma
- Pathological degradation of pancreatic β cells
- Peripheral vascular diseases
- Weight loss
- Immunological problem
- Cardiovascular dysfunction

Based on available statistics, it is evident that in 1980, approximately 122 million individuals were afflicted with diabetes [7]. However, by 2014, this figure surged to approximately 422 million [7]. In 2017, approximately 451 million individuals were affected by diabetes, and according to forecasts, this number is projected to escalate to 693 million by 2045 [8]. In general, diabetes mellitus (DM) is classified into three types: T1DM, T2DM, and gestational diabetes (GD).

- T1DM, which is associated with genetic autoimmune factors, has some of its causes and some treatment mentioned in Table 1. It generally occurs in young people under 30 years of age and is associated with the symptoms outlined in Table 2 [1].

- T2DM develops over time and typically affects individuals over 45 years old. Causes and some treatment of this disease it's mentions in the Table 1. This type of diabetes is also associated with the symptoms listed in Table 2 [9].
- GD primarily affects pregnant women (rare case). Causes of this disease are maternal overweight and obesity, later age for pregnancy, previous history in pregnancy, family history of T2DM. This disease can be treatment with Metformin. The symptoms detailed in Table 2 [10].

Table 2: Symptoms for different types of DM.

No.	Type	Age	Symptoms	Reference
01	T1DM	< =30	• Fatigue • Weight loss • Thirst • Polyuria • Constant hunger • Vision changes	[1]
02	T2DM	>= 45	• Obesity • Hypertension • Dyslipidemia • Arteriosclerosis	[9]
03	GD	Pregnant women	• No any symptoms • Excessive thirst • Sweating and frequent urination	[10]

To diagnose DM based on its type, you can refer to the symptoms outlined in Table 2. However, for confirmation of the disease, common investigations such as blood tests, urine tests, and genetic tests are typically conducted. One of the emerging techniques that has recently gained traction across various disciplines, including medical sciences, is the utilization of artificial intelligence methods for the potential detection of various parameters. Many researchers have sought to predict this disease quickly by using machine learning classification algorithms (MLCA). The methods utilized by previous researchers to predict DM are presented in Table 3.

Table 3: Researchers who predict the DM by classifier machine learning algorithm.

No.	Year	Techniques	Reference
01	2019	• K-Nearest Neighbor (KNN) • Support Vector Machine (SVM) • Naive Bayes (NB) • C4.5 Decision Tree (DT)	[11]
02	2020	• KNN • Linear Discriminant Analysis (LDA) • Random Forest (RF) • SVM	[12]
03	2021	• Adaptive Neuro-Fuzzy Inference System (ANFIS) • SVM • DT • NB • Artificial Neural Network (ANN) • KNN • Deep Neural Network (DNN)	[13]

		• RF	
04	2022	• ANN • SVM	[14]
05	2023	• SVM • NB • KNN • RF • LR • DT	[15]

Many researchers have conducted extensive work on predicting DM using artificial intelligence algorithms, as depicted in Table 3. This article utilizes 768 data points collected from an opensource database to predict DM. The information used in this article includes data such as body mass index (BMI), glucose level (GL), age (A), skin thickness (ST), insulin level (IL), blood pressure (BP), and Diabetes Pedigree Function (DPF). The primary distinction between this article and others lies in the integration of new hybrid algorithms (HA) combination of the Multi-Layer Perceptron (MLP) and robust optimizer Cauchy Crazy Particle Swarm Optimization (CCPSO) and Particle Swarm Optimization (PSO) algorithms to predict DM.

MATERIALS AND METHODS

Workflow

In order to present this article quickly, can utilize the information provided in Figure 1. In this figure, it is evident that data from an open-source database was collected initially. Subsequently, the data should be normalized and divided into three categories: training, testing, and validation. In the next step, the structure of the MLP algorithm was built using the majority of the data. Finally, the control parameters of the MLP algorithm were determined through the use of optimization algorithms. These combined algorithms were then used to predict the DM parameter. Ultimately, statistical criteria are utilized to compare these algorithms, and the best HA is presented.

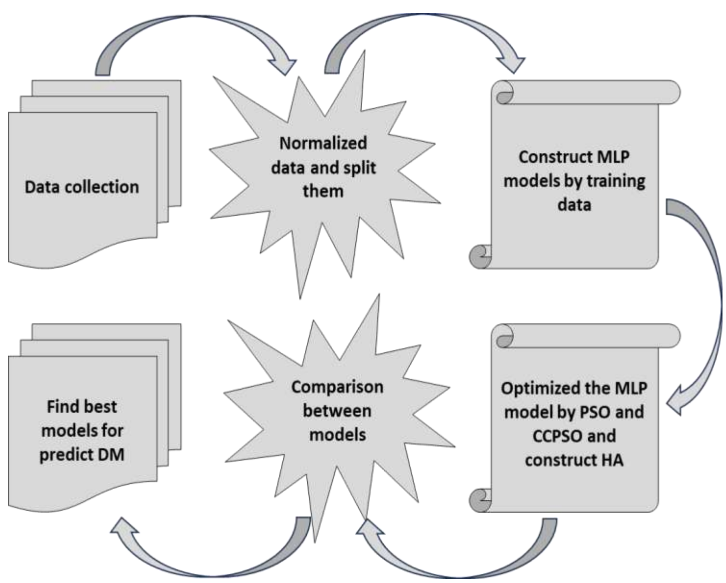

Figure 1: Workflow for prediction of DM by different HA.

Data Collection

According to the criteria presented by WHO, many factors can affect DM. In this article, only a portion of these influential parameters has been mentioned, which can help for predicting the occurrence of DM easily. Table 2 displays the information utilized in this article. The data consists of 768 data points sourced from an open-access database, including parameters such as blood pressure (BP), body mass index (BMI), skin thickness (ST), glucose level (GL), age (A), insulin level (IL), and diabetes pedigree function (DPF).

Multi-Layer Perceptron (MLP)

One of the most useful algorithms for predicting and classifying datasets with structural complexities is the MLP algorithm [16]. This advanced ANN algorithm comprises three layers: input, hidden, and output. With this algorithm, parameters and highly complex mathematical problems can be solved easily. The hidden layer, one of the most important layers of this algorithm, can simplify the solution of complex problems by smoothing the gradient [17]. Figure 2 illustrates the structure of the MLP algorithm utilized in this article.

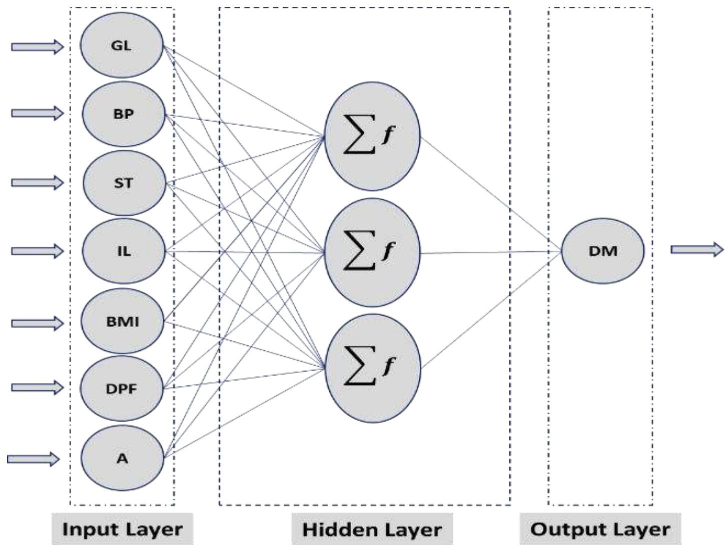

Figure 2: Architecture of the MLP for prediction of DM.

Particle Swarm Optimization (PSO)

One of the globally significant strategic algorithms that has effectively aided in solving complex problems is the PSO algorithm [18]. Inspired by the collective intelligence and organized behavior of birds and fish, this algorithm enables the easy and highly accurate resolution of complex problems [19]. It operates by maneuvering within a safe range to locate the minimum and maximum of the objective function. Figure 2 illustrates the graphical architecture of the PSO algorithm.

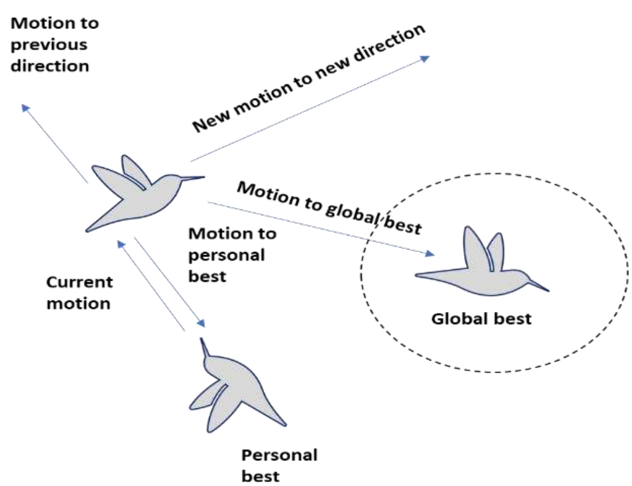

Figure 3: Architecture of the PSO for prediction of DM.

Cauchy Crazy Particle Swarm Optimization (CCPSO)

One of the novel concepts in the realm of artificial intelligence optimizations is the utilization of the combination of the Cauchy jump operator algorithm with the PSO algorithm to create a new optimizer algorithm [20]. The Cauchy mutation operator algorithm operates on a continuous probability distribution. By using this operator, diversity can be easily preserved, and convergence can be enhanced [20].

Hybrid MLP-PSO

To attain higher accuracy compared to the MLP algorithm, it can be integrated or hybridized with the PSO algorithm. Through the hybridization of these two algorithms, PSO can enhance the accuracy of the algorithm by optimizing the control of MLP parameters, leading to better and improved results [21]. The flowchart depicted in Figure 4 illustrates the performance of the MLP-PSO algorithm.

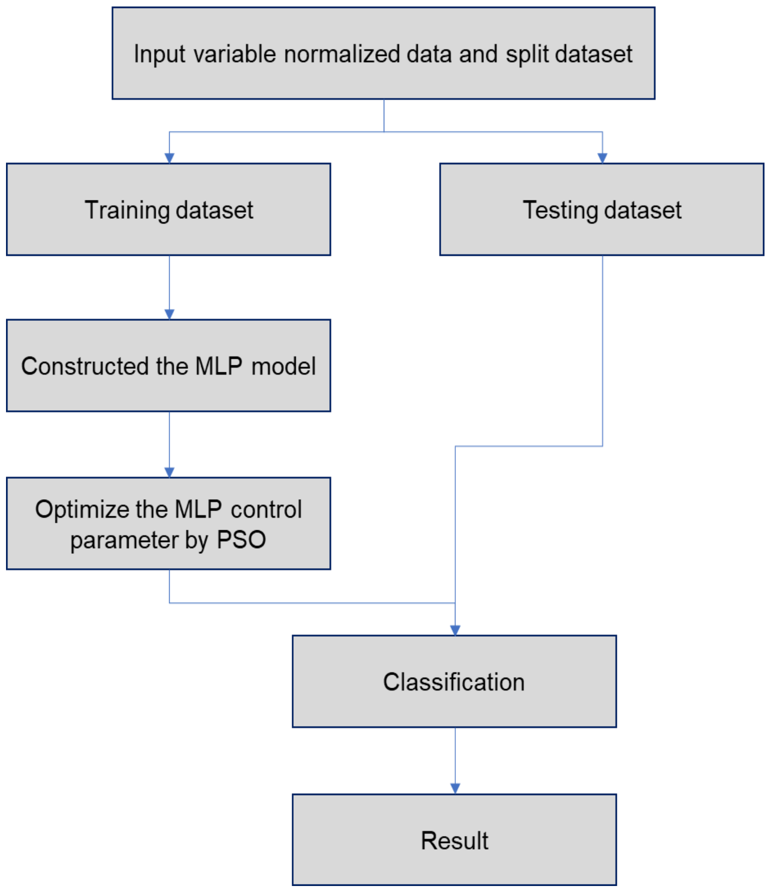

Figure 4: Flowchart of the MLP-PSO for prediction of DM.

Hybrid MLP-CCPSO

To construct this algorithm, the MLP algorithm was initially used as a network. Subsequently, the PSO algorithm was utilized to control the parameters associated with the MLP algorithm and optimize it. Finally, by implementing the Cauchy distribution mutation operator, the accuracy of this algorithm is enhanced [22]. Through the use of this Cauchy distribution, it becomes feasible to maintain the probability distribution among the data and its diversity [22]. Figure 5 illustrates the flowchart of the new MLP-CCPSO recombination algorithm.

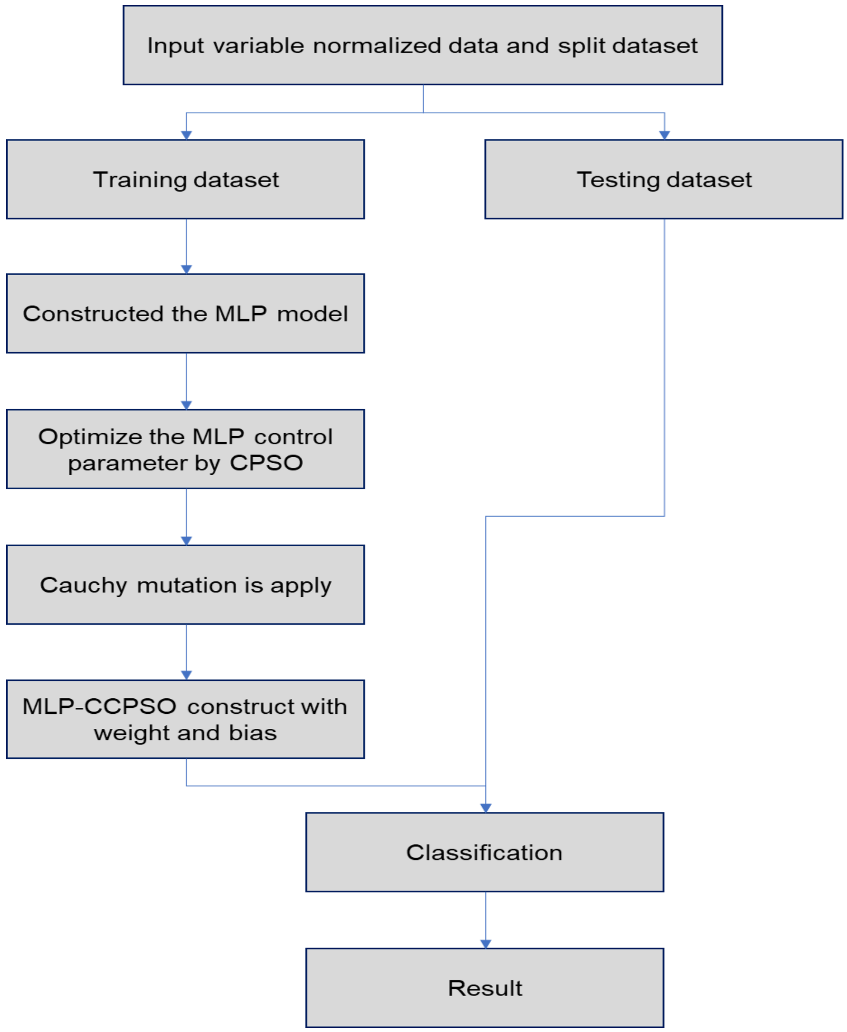

Figure 5: Flowchart of the MLP-PSO for prediction of DM.

RESULT AND DISCUSSION

To determine and compare new HA algorithms, statistical parameters and important criteria of statistical classification are utilized. These statistical parameters enable the comparison of HA algorithms (Equations 1-4).

$$Accuracy = \frac{TN_{DM} + TP_{DM}}{FN_{DM} + FP_{DM} + TP_{DM} + TN_{DM}} \times 100 \quad (1)$$

$$Recall = \frac{TP_{DM}}{TP_{DM} + FN_{DM}} \times 100 \quad (2)$$

$$Precision = \frac{TP_{DM}}{TP_{DM} + FP_{DM}} \times 100 \quad (3)$$

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (4)$$

Table 4 presents the results of the algorithms utilized to predict DM. The data information in this table is reported in two subsets, train and test, separately in Table 2. In this article, three algorithms—MLP, MLP-PSO, and MLP-CCPSO—are used to predict DM. The results depicted in this table indicate that the MLP-CCPSO algorithm shows higher performance accuracy compared to the other algorithms used in this article.

Table 4: Comparative analysis of ML and HA models for DM Prediction.

Split dataset	Models	Precision	Recall	Accuracy	F1-Score
Training dataset	MLP	0.85	0.84	0.83	0.80
	MLP-PSO	0.90	0.89	0.91	0.90
	MLP-CCPSO	0.98	0.97	0.96	0.97
Testing dataset	MLP	0.82	0.81	0.81	0.79
	MLP-PSO	0.89	0.88	0.90	0.88
	MLP-CCPSO	0.97	0.96	0.95	0.96

The two most important parameters used to compare the performance of classification algorithms are Precision and Accuracy, as illustrated in Figure 6. Based on the results presented in Figure 6 and Table 4, it is evident that the MLP-CCPSO algorithm exhibits greater accuracy than the other algorithms used in this article. Following that, the MLP-PSO algorithm shows higher performance accuracy.

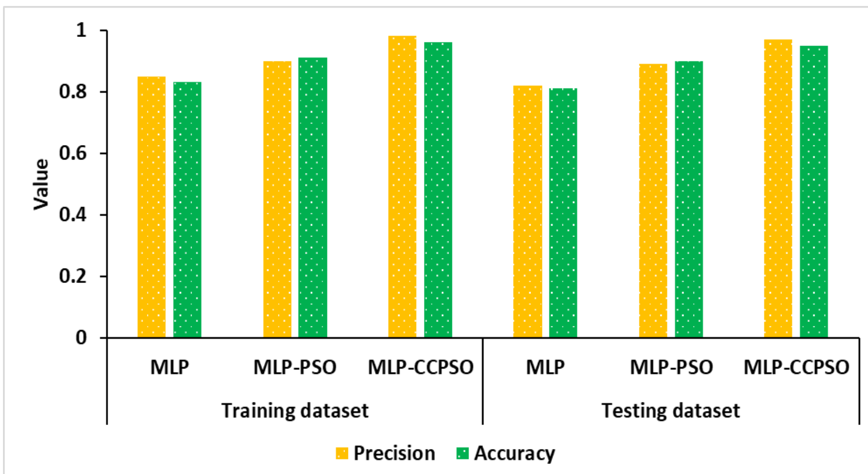

Figure 6: Rectangular diagram for important two metrics Precision and Accuracy for prediction of DM.

With other metrics such as Recall and F1-Score, it makes suitable comparisons among classification algorithms for DM prediction. Upon comparing these values, the results indicate that the MLP-CCPSO algorithm is more accurate than other algorithms, and the performance accuracy of the algorithms is ranked as MLP-CCPSO > MLP-PSO > MLP.

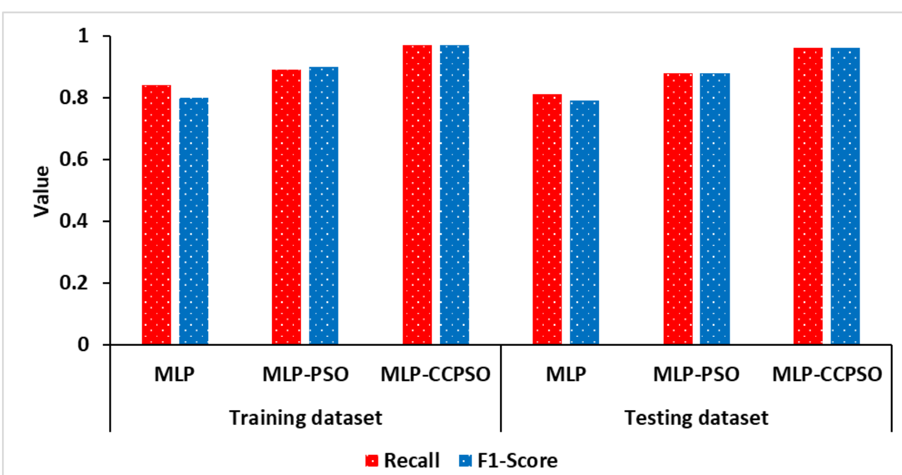

Figure 7: Rectangular diagram for two important metrics Recall and F1-Score for prediction of DM.

CONCLUSION

Diabetes mellitus (DM), often known simply as diabetes, comprises a range of metabolic disorders distinguished by heightened levels of blood glucose. This condition poses a significant health challenge worldwide, impacting both developed and developing countries. From the perspective of its impact on other diseases such as stroke, renal failure (RF), neuropathies and vascular disorders (N&VD), myocardial infarction (MI), cardiovascular-disease (CVD), and peripheral artery disease (PAD), this disease is of utmost importance. This study utilizes 768 data points collected from an open-source database. The input variables used in this article include body mass index (BMI), glucose level (GL), blood pressure (BP), skin thickness (ST), age (A), insulin level (IL), and diabetes pedigree function (DPF). To predict DM, new hybrid algorithms (HA) with combination of the Multi-Layer Perceptron (MLP) with two robust optimized algorithms like Particle Swarm Optimization (PSO), and Cauchy Crazy Particle Swarm Optimization (CCPSO) algorithms are used. The results show that MLP-CCPSO exhibits higher accuracy than other algorithms used in this article. The advantages of this robust hybrid algorithm are enhanced accuracy, improved convergence, greater robustness, increased efficiency, and optimized parameter control. Results indicate that the accuracy of the three algorithms for the prediction of DM follows the order: MLP-CCPSO > MLP-PSO > MLP.

DATA AVAILABILITY

The data is available within this link: https://www.kaggle.com/datasets/mrsimple07/diabetes-prediction.

CONFLICT OF INTEREST

There is no conflict for this article.

REFERENCES

[1] American Diabetes A. Diagnosis and classification of diabetes mellitus. Diabetes care 2010;33(Supplement_1):S62-S9.

[2] Emerging Risk Factors C. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. The lancet 2010;375(9733):2215-22.

[3] Lonappan A, Bindu G, Thomas V, Jacob J, Rajasekaran C, Mathew KT. Diagnosis of diabetes mellitus using microwaves. Journal of Electromagnetic Waves and Applications 2007;21(10):1393-401.

[4] Kumar DA, Govindasamy R. Performance and evaluation of classification data mining techniques in diabetes. International Journal of Computer Science and Information Technologies 2015;6(2):1312-9.

[5] Krasteva A, Panov V, Krasteva A, Kisselova A, Krastev Z. Oral cavity and systemic diseases—diabetes mellitus. Biotechnology & Biotechnological Equipment 2011;25(1):2183-6.

[6] Vaishali R, Sasikala R, Ramasubbareddy S, Remya S, Nalluri S. Genetic algorithm based feature selection and MOE Fuzzy classification algorithm on Pima Indians Diabetes dataset. IEEE; 2017:1-5.

[7] Nathan DM. Long-term complications of diabetes mellitus. New England journal of medicine 1993;328(23):1676-85.

[8] Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, Malanda B. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes research and clinical practice 2018;138:271-81.

[9] Robertson G, Lehmann ED, Sandham W, Hamilton D. Blood glucose prediction using artificial neural networks trained with the AIDA diabetes simulator: a proof-of-concept pilot study. Journal of Electrical and Computer Engineering 2011;2011:2-.

[10] McIntyre HD, Catalano P, Zhang C, Desoye G, Mathiesen ER, Damm P. Gestational diabetes mellitus. Nature reviews Disease primers 2019;5(1):47.

[11] Faruque MF, Sarker IH. Performance analysis of machine learning techniques to predict diabetes mellitus. IEEE; 2019:1-4.

[12] Tripathi G, Kumar R. Early prediction of diabetes mellitus using machine learning. IEEE; 2020:1009-14.

[13] Hussain A, Naaz S. Prediction of diabetes mellitus: comparative study of various machine learning models. Springer; 2021:103-15.

[14] Ahmed U, Issa GF, Khan MA, Aftab S, Khan MF, Said RAT, et al. Prediction of diabetes empowered with fused machine learning. IEEE Access 2022;10:8529-38.

[15] Kangra K, Singh J. Comparative analysis of predictive machine learning algorithms for diabetes mellitus. Bulletin of Electrical Engineering and Informatics 2023;12(3):1728-37.

[16] Brownlee J. Deep learning with Python: develop deep learning models on Theano and TensorFlow using Keras. Machine Learning Mastery; 2016.

[17] Al Bataineh A, Kaur D. A comparative study of different curve fitting algorithms in artificial neural network using housing dataset. IEEE; 2018:174-8.

[18] Eberhart RC, Shi Y, Kennedy J. Swarm intelligence. Elsevier; 2001.

[19] Blondin J. Particle swarm optimization: A tutorial. from site: http://cs armstrong edu/saad/csci8100/pso tutorial pdf 2009.

[20] Sarangi SK, Panda R, Sarangi A. Crazy firefly algorithm for function optimization. IEEE; 2017:1-5.

[21] Al Bataineh A, Manacek S. MLP-PSO hybrid algorithm for heart disease prediction. Journal of Personalized Medicine 2022;12(8):1208.

[22] Wang H, Li C, Liu Y, Zeng S. A hybrid particle swarm algorithm with Cauchy mutation. IEEE; 2007:356-60.