IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Binary Chaotic Optimized Fused Learning (Bcofl) Model For An Effective Diabetes Prediction Using Iot

S.Muthukumar, M.Jayakumar,

Research Scholar, Assistant Professor
Department of Computer Science,
Shri Nehru Maha Vidyalaya College of Arts and Science, Coimbatore, India

Abstract: According to the World Health Organization (WHO), there are 420 million people affected with diabetics worldwide, and this has caused an increase in the mortality rate to nearly one million per year. A serious situation has resulted from this exceptional rise in cases and fatalities, since the data statistics show a considerable rise in diabetic cases amid youngsters. For accurate diabetes categorization and prediction, this study makes use of a new framework called Binary Chaotic Optimized Fused Learning (BCoFL). In this case, feature selection and data dimensionality reduction are accomplished using the Binary Chaotic Hunger Game Search (BCHGS) optimization technique. For system validation and evaluation, the diabetes dataset from PIMA Indian patients was employed. Data cleaning and normalization processes are first used to preprocess the dataset. Additionally, the disease is predicted from the supplied data effectively and with little over fitting using the Fused Learning Classification Algorithm (FLCA). The performance and prediction results of the suggested CoFL technique are validated and compared using the accuracy, precision, prediction rate, and other parameters.

***Index Terms* - Diabetes Detection, Chronic Disease, Machine Learning, Binary Chaotic Hunger Game Search (BCHGS) Optimization, Fused Learning Classification Algorithm (FLCA), and Internet of Things (IoT).**

I. Introduction

Today's globe is widely familiar with the term "diabetes," which poses serious problems for both industrialized and developing nations[1, 2]. One of the chronic, fatal illnesses known as diabetes is brought on by an inadequate or insufficient insulin hormone. It is a vital hormone that the pancreas produces that enables cells to consume glucose from dietary sources to give them the energy they require. Hyperglycemia is the scientific name for the condition where there are high blood sugar levels. There are two basic causes for this situation: When the body is unable to produce the insulin needed by the cells in the bloodstream, the body's response to insulin is also impaired[3].Glucose can go from food into the bloodstream due to the pancreas' production of the hormone insulin. In the absence of that hormone due to pancreatic dysfunction, diabetes develops, which can cause unconsciousness, kidney and retinal failure, detrimental deterioration of pancreatic cells, cardiovascular dysfunction, cerebral blood vessels disorder, peripheral arterial disorders, cartilage failure, loss of weight, wounds, and dangerous immune impacts. According to statistics, 450 million individuals worldwide have diabetes as of 2017, and by 2045, that number will rise to 695 million. In general, Type I, type II, and diabetes during pregnancy are the three primary subtypes of diabetes[4]. Type I diabetes results from the pancreas' inability to produce insulin; type II diabetes results from the pancreas' inability to produce enough insulin or handle it appropriately. Type II diabetes is the most prevalent type of diabetes in adults and accounts for 90% of cases worldwide. This type of diabetes is typically referred to as insulin resistance since the body does not appropriately respond to the glucose released since insulin does not function properly[5, 6]. In certain type 2 diabetes, the pancreas is worn out, which causes less insulin to be produced and may lead to even higher blood sugar levels. A majority of older persons were diagnosed with type II diabetes in the past. Nevertheless, because of poor diet, a lack of exercise, and increasing obesity rates, it has become apparent that this is also happening in adolescents and youngsters. Fig 1 shows the statistical prediction rate of diabetes around the world [7], which illustrates that there will probably be 700 million patients can be affected by diabetes at 2045.

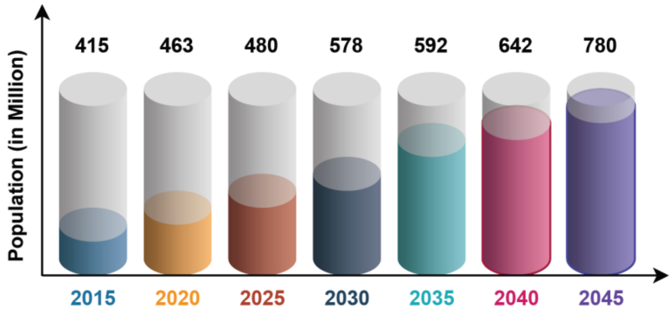

Fig 1. Global prevalence of diabetes

People with diabetes have a modest increased chance of getting various infections compared to healthy individuals, which has been recognized for years. As a result, diabetic patients must take all essential precautions to prevent further complications.Due to the fact that diabetes ignores national borders or socioeconomic condition of a country, it might be viewed as a severe worldwide threat[8]. Diabetes has been recognized as one of the top 10 causes of premature mortality on a global scale. But a lot of countries still don't have a national diabetes prevention strategy, therefore only a portion of the world's population has access to complete healthcare coverage as of yet.Information on health is vital to the survival of humanity. Even the smallest of patient facts can be preserved in healthcare databases for later diagnosis and analysis of difficulties the patient might encounter due to advancements in medical technology[9]. However, the abundance of data may result in inconsistent decisions. Healthcare organizations generate a ton of data, which results in cognitive saturation. Decisions based on this data can lead to errors in patient diagnosis and treatment.

Recent technological advancements have improved the use of machine learning and the Internet of Things (IoT) in numerous sectors[10]. It is a network environment where each connected device may speak with a variety of other components to transmit vital information for prompt and accurate decision-making. In the current wave of digital expansion, IoT represents a significant turning point. So, in critical cases like medical applications, IoT is crucial. IoT serves as a solid foundation for the development of sophisticated healthcare systems.Since it permits the creation of an adaptive and networked environment and offers both medical professionals and patients with a wide range of services, the merging of these two fields can be very helpful for medical IoT[11]. Medical personnel can take measures that might save a patient's life thanks to early disease forecasts. IoT sensors gather patient data, which is subsequently processed by machine learning algorithms to detect the presence of life-threatening conditions like diabetes.

The Knowledge Discovery in Databases (KDD) process, also known as data mining, is the process of computationally extracting information in the sense of structures and trends from sizable databases and modifying it in a suitable way for subsequent use[12]. Finding the hidden patterns of effective medical treatments for various disorders is helpful. The implementation of classification algorithms is thought to be a common DM methodology that is used to categorize and forecast the predefined data for a specific class.A subfield of information technology called "machine learning" analyses historical data to forecast responses to newly acquired data[13]. The training of a model for prediction can enhance its general efficacy and enable it to make conclusions using different parameters or circumstances by using artificial intelligence, identification of patterns, and statistical algorithms.Additionally, researchers demonstrated that ML algorithms typically result in better precise diagnoses for a variety of disorders.For diabetes detection and classification in earlier studies, a number of machine learning and deep learning algorithms are used[14, 15]. However, overfitting, class imbalance, a lack of resilience, and low efficiency limit the majority of approaches. Therefore, the goal of this research project is to develop, using cutting-edge data mining techniques, an intelligent system for the detection of diabetes using patient medical data. The core objectives of this work are given below:

- This study uses a cutting-edge framework called Binary-Chaotic Optimized Fused Learning (BCoFL) for precise diabetes classification and prediction.
- The PIMA Indian patients' diabetes dataset has been used for system validation and assessment. The dataset is first preprocessed with data cleaning and normalization operations.

- The Binary Chaotic Hunger Game Search (BCHGS) optimization technique is used here for feature selection and data dimensionality reduction.
- Additionally, the Fused Learning Classification Algorithm (FLCA) is utilized to accurately and with minimal overfitting predict the disease from the provided data.
- The accuracy, precision, prediction rate, and other metrics are used to validate and compare the performance and prediction outcomes of the proposed BCoFL approach.

The following sections are used to group the remaining portions of this paper: The comprehensive literature analysis in Section 2 looks at many known methods for predicting and categorizing chronic diseases. Additionally, it looks at each technique's benefits and drawbacks in relation to how well it performs and detects diseases. With a flow diagram and stage-by-stage explanations, Section 3 provides a thorough and understandable explanation of the proposed BCoFL approach. In Section 4, the effectiveness of the suggested method is verified and contrasted using a number of measures. In Section 5, the main conclusions, outcomes, and directions for future work are summarized.

II. Related Works

By giving the necessary data in real-time or on a historical basis and combining artificial intelligence and machine learning to make intelligent decisions, the fundamental goal of IoT is to make the world surrounding us intelligent. In the modern world, a lot of elements contribute to an unhealthy way of life and a decline in human health, including inconsistent eating patterns, poor nutrition, pollutants, insufficient physical activity, endless work, frustration, and high levels of stress. Around the world, up to 40% of young people, the elderly, and professional women lead sedentary lifestyles.Today, diabetes is a highly serious problem due to the fact that a large number of individuals die from the ailment each year. So, in order for the diabetic patient to live a normal life, regular monitoring is required. To avoid health issues, the diabetic patient's health must be regularly and continually evaluated. The information needed to support treatment isn't sufficiently provided by the available medical software platforms, such as patient monitoring systems and health management systems. The objective of this endeavor is to create a system for monitoring patient health, especially that of diabetics.*Hasan, et al* [16]conducted a new study to investigate the different types of machine learning based ensemble mechanisms used for diabetes prediction. The authors mainly intended to develop a new robust framework for identifying diabetes from the patients' medical record. This literature study includes the mechanisms of outlier rejection, missing value replacement, data standardization and classification. Here, the k-fold cross validation technique has been used for classification, which is one of the most popular techniques for model selection and classifier's error estimation.

Kee, et al [17]presented a systematic review to analyze the complications of diabetes with the use of machine learning techniques. In this study, several machine learning techniques have been used for discovering 10 different types of cardio vascular diseases. *Yahyaoui, et al* [18] developed a decision support system with the use of machine learning and deep learning techniques for an accurate prediction of diabetes from the patients' medical data. *Naz, et al* [7] implemented Sequential Mining Optimization (SMO) based expert system for predicting type II diabetes from PIMA dataset. Although there are numerous approaches for diagnosing diabetes, none of them are effective at uncovering hidden patterns with the precision required for sound decision-making. In order to forecast diabetes, this research provides an integrated strategy using the SMOTE and SMO algorithms. The SMOTE method is used in the beginning phase of this suggested classification model to preprocess the data. In order to improve the classifier's performance, SMO receives the output of the preprocessing.

Table 1. Survey on existing diabetes prediction models

Ref	Methods	Description	Observation
[19]	Machine learning model	This study applied a group of machine learning approaches for chronic disease detection.	RF performs well with the accuracy of 82%.
[20]	DT, NB, LR and ANN	It intends to use the supervised learning models for constructing an effective diabetes detection system.	The LR provides the 86.3% accuracy comparing to other models.
[21]	Filter based DT and ensemble model	It integrates RF and AB algorithms for diabetes prediction and applies wrapper model for feature selection.	It reduced the time up to 0.006 s.
[22]	LR, DT, DA, KNN and SVM	This study uses several machine learning models for diabetes prediction.	Accuracy is increased to 77.9%.
[23]	Hybrid ensemble model	It aims to identify the Diabetes Mellitus using an IoT based machine learning model.	Accuracy is maximized to 98.4%.

III. Proposed Methodology

The purpose of this endeavor is to develop a machine learning-based smart patient health monitoring system that can accurately and promptly identify the presence of a chronic disease such as diabetes in a patient. This monitoring system receives data from a number of medical wearable. The patient's health state is then assessed by this application using machine learning algorithms on raw data in order to make the proper judgments and diagnoses. The patient's condition is predicted by this system, and in the event of an emergency, a warning signal is issued to the doctor and guardians to request prompt supervision.

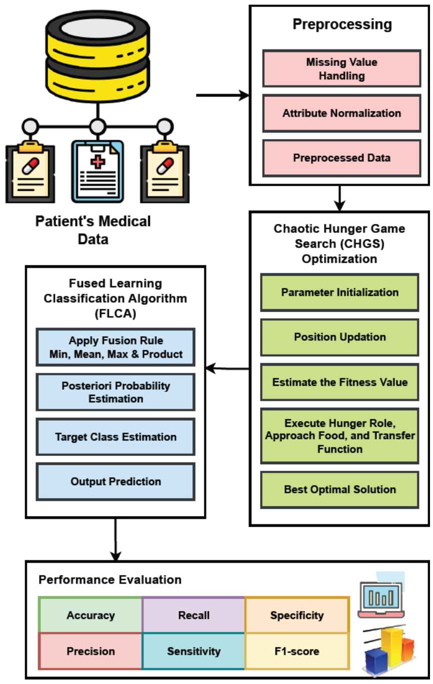

Fig 2. Flow of the proposed BCoFL based diabetes prediction system

The main objective of the present research is the development of an innovative framework that is referred to as Binary-Chaotic Optimized Fused Learning (BCoFL) for the identification and categorization of diabetes from patient medical records. Fig. 1 shows the proposed system's flow, which contains the following operational modules:

- Data preparation and handling of missing values
- Multi-Strategical Binary Chaotic Hunger Game Search (MSB-CHGS) based optimization for feature selection;
- Fused Learning Classification Algorithm (FLCA) for disease prediction;
- Performance evaluation and assessment

The system implementation and analysis in this work leverage the open PIMA dataset. The performance of the classifier can be affected by increased false prediction and error rate since the original dataset is frequently incomplete and may contain some irrelevant data. Thus, in order to produce normalized data, preprocessing and data normalization operations are carried out first. The necessary and useful characteristics are then selected from the preprocessed data using the MSB-CHGS optimization technique. In order to classify the provided data as healthy or diabetes illness afflicted, the FLCA is used. By employing significant measures, the effectiveness and outcomes of the suggested BCoFL technique are validated and evaluated.

Data Cleaning & Preprocessing

The first step in dataset preprocessing is the handling of missing values, which is followed by standardization. If missing or null values are not imputed, the machine learning classifier's ability to predict outcomes accurately will suffer. Here, the missing numbers are filled using the mean approach rather than dropping as represented in the following equation:

$$\rho(v) = \begin{cases} M(v), & if\ v = Null/Miss \\ v. & Otherwise \end{cases} \quad (1)$$

Where, ρ indicates the missing values of the data, v denotes the instances of feature vector, and $M(v)$ represents the mean value. The use of the mean approach to impute missing values is justified since it generates the necessary continuous data for algorithm training without adding anomalies.It is possible to attain a typical normal distribution with a zero mean and a unit variance by rescaling features using Z-score normalization. It helps to reduce the skewness of data, and is represented in the following equation:

$$\varpi(v) = \frac{v - M(v)}{\sigma(v)} \quad (2)$$

Where, $\varpi(v)$ is the normalized data, $\sigma(v)$ indicates the standard deviation, and $M(v)$ represents the mean value. At the end of preprocessing, the normalized and tuned dataset has been generated, which is further used for prediction.

Multi-Strategical BinarizedChaotic Hunger Game Search (MSB-CHGS) based Optimization

The most relevant and connected features from the preprocessed data are selected for feature selection using the MSB-CHGSbased optimization technique after preprocessing. Numerous optimization strategies are used in the existing studies to reduce the data dimensionality. The MSB-CHGS technique is typically used for its high convergence rate, searching effectiveness, and quick time to solution when compared to other optimization strategies. Additionally, it has several unique characteristics including improved accuracy, reduced iterations, and lower local optimum. The MSB-CHGSis one of the population based optimization technique that replicating sociable animals' cooperative foraging behaviors, which are correlated with the level of hunger in those species.In particular, MSB-CHGS uses an adaptive weight to imitate the consequences of hunger underneath logical constraints at every searching stage, offering a straightforward but dynamic framework with a rapid pace. As an animal becomes hungrier, their need for food increases since hunger is thought to be one of the most important aspects in preserving homeostasis and influencing their judgments and actions. Games also refer to the logical principles that almost all creatures follow in order to survive, such as the need to seek food and protect themselves from outside attackers. A reasonable game between hungry creatures that would battle to discover supplies of food and achieve the current competition would commence when they have only a small amount of food.

In this technique, the preprocessed dataset, number of population, data dimension, and maximum number of iterations are considered into account for parameter initialization. For the number of populations and dimensionality of data, the population $p(i,j)$ estimation is performed by using the following model:

$$p(i,j) = \left(\mathbb{ub}(j) - \mathbb{lb}(j)\right) \times r() + \mathbb{lb}(j) \tag{3}$$

Where, $\mathbb{ub}$ and $\mathbb{lb}$ are the upper and lower bound values, and r indicates the random number. Then, the sinusoidal map computation is performed and applied to the population P as shown in below:

$$\delta_{i+1} = g \times \delta_i^2 \times \sin(\pi\delta_i), \quad i = 1,2 \dots N_k - 1 \tag{4}$$

Where, δ_{i+1} indicates the sinusoidal map, g is the value set to 2.3, and N_k denotes the number of population. After that, the position of the agents are updated, during this process, the iteration count is initialized. Then, the fitness value is estimated for all the individuals in the population using the greedy selection model as represented in the following mathematical model:

$$\mathcal{f}\left(p_i(itr)\right) = \begin{cases} \mathcal{f}(p_i^*) & if\ \mathcal{f}(p_i^*) \le \mathcal{f}(p_i(itr)) \\ \mathcal{f}(p_i(itr)) & if\ \mathcal{f}(p_i^*) > f(p_i(itr)) \end{cases} \tag{5}$$

Where, p_i^* indicates the best value of positions, $\mathcal{f}(.)$ represents the fitness function, and $\mathcal{f}(p_i^*)$ denotes the best fitness value for each agent. Moreover, the hunger sensation parameter is estimated with respect to the lower bound of the hunger sensation as shown in below:

$$\mathfrak{H} = \begin{cases} L\mathfrak{H} \times (1+a) & T\mathfrak{H} < L\mathfrak{H} \\ T\mathfrak{H} & T\mathfrak{H} \ge L\mathfrak{H} \end{cases} \tag{6}$$

$$T\mathfrak{H} = \frac{\mathcal{f}(i) - \mathcal{f}_{best}}{\mathcal{f}_{worst} - \mathcal{f}_{best}} \times b \times 2 \times (\mathbb{ub} - \mathbb{lb}) \tag{7}$$

Where, $L\mathfrak{H}$ represents the lower bound of hunger sensation,a, b are the random numbers, $\mathcal{f}_{best}$ is the best fitness, $\mathcal{f}_{worst}$ is the worst fitness, $\mathbb{ub}$ and $\mathbb{lb}$ are the upper and lower bound values of the searching space. Moreover, the hungry level of each agent is estimated by using the following equation:

$$H(i) = \begin{cases} 0 & \mathcal{f}(i) = \mathcal{f}_{best} \\ H(i) + \mathfrak{H} & \mathcal{f}(i) \ne \mathcal{f}_{best} \end{cases} \tag{8}$$

Where, $H(i)$ is the hungry level of agent.Furthermore, the weight parameters such as ω_1 and ω_2 are estimated by using the following equations:

$$\overrightarrow{\omega_1(q)} = \begin{cases} H(i) \times \frac{N_k}{SH} \times m & n < q \\ 1 & n > q \end{cases} \tag{9}$$

$$\overrightarrow{\omega_2(q)} = (1 - \exp(-|H(i) - SH|)) \times v \times 2 \tag{10}$$

Where, $\overrightarrow{\omega_1(q)}$ and $\overrightarrow{\omega_1(q)}$ are the weight of hunger, SH represents the sum of the hungry of all agents, m, n, v are the random numbers from 0 to 1, and q indicates the algorithm design parameter. In addition, the variation control parameter γ is computed based on the following model:

$$\gamma = \text{sech}(|\mathscr{f}(i) - \mathscr{f}_{best}|) \tag{11}$$

Where, $\text{sech}(.)$ indicates the hyperbolic tangent function. Then, the range of activity $\vec{Y}$ is determined for all individuals as represented in the following model:

$$\vec{Y} = l \times (2 \times z - 1) \tag{12}$$

$$l = 2 \times (1 - \frac{itr}{\max\_itr}) \tag{13}$$

Where, z is the random number. After that, the position updation is performed by using the following model:

$$\overrightarrow{P(\imath tr + 1)} = \begin{cases} \overrightarrow{P(\imath tr)} \times (1 + r\_n(1) & r < q \\ \overrightarrow{\omega_1} \times \overrightarrow{P_b(\imath tr)} + \vec{Y} \times \overrightarrow{\omega_2} \times |\overrightarrow{P_b(\imath tr)} - \overrightarrow{P(\imath tr)}| & r > q, w > \gamma \\ \overrightarrow{\omega_1} \times \overrightarrow{P_b(\imath tr)} - \vec{Y} \times \overrightarrow{\omega_2} \times |\overrightarrow{P_b(\imath tr)} - \overrightarrow{P(\imath tr)}| & r > q, w < \gamma \end{cases}$$

(14)

Where, $\overrightarrow{P(\imath tr)}$ is the location of each individual, and r_n is the random number with normal distribution. Finally, the vertical crossover is estimated and the transfer position $p_i^d(itr)$ is computed as shown in below:

$$p_i^d(itr) = \begin{cases} 0 & if\ r \leq \delta(p_i^d(itr)) \\ 1 & if\ r > \delta(p_i^d(itr) \end{cases} \tag{15}$$

At the end, the best fitness value and best individual are obtained as the output of this procedure, which are used to choose the optimal features from the given dataset.

Algorithm 1 - Multi-Strategical Binary Chaotic Hunger Game Search (MSB-CHGS) based Optimization

Input: Preprocessed dataset P_{DS}, Number of population N_k, Dimension $\mathbb{D}$, and maximum number of iterations $\max\_itr$;
Output: Optimized feature set $\mathcal{O}_s$;
Step 1: Initialize the input parameters;
Step 2: for $i = 1: N_k$
for $i = 1: \mathbb{D}$
Population initialization;
$p(i,j) = (\mathbb{ub}(j) - \mathbb{lb}(j)) \times rand() + \mathbb{lb}(j);$
end for;
Compute sinusoidal map by using equ (4);
end for;
Step 3: Update the position of agents;
Step 4: Initialize the iteration count $itr = 1$;
Step 5: while $itr < \max\_itr + 1$
Estimate the fitness value for all individuals;
Perform greedy selection $\mathscr{f}(p_i(itr))$ operation using equ (5);
Update the parameters of best fitness, worst fitness and best individual;
Estimate the hungry role $\mathfrak{H}$ using equ (6) and (7);
Estimate the level of hungry $H(i)$ using equ (8);

Compute the weight values $\overrightarrow{\omega_1(q)}$ *and* $\overrightarrow{\omega_2(q)}$ *using equ (9) and (10);*

Step 6: for $i = 1: N_k$

Estimate the variation control parameter based on equ (11);

Compute the range of activity $\vec{Y}$ *using equ (12);*

Update the position $\overrightarrow{P(\iota tr+1)}$ *using equ (14);*

Perform vertical crossover;

Compute the transfer position $p_i^d(itr)$ *based on the binary model using equ (15);*

End for;

End while;

Step 7: Return the output as best fitness and best individual;

Fused Learning Classification Algorithm (FLCA)

The FLCA method is used to precisely forecast diabetes from the patient data after feature optimization. The FLCA is a machine learning method that was created by combining SVM and ANN functionality.The SVM algorithm has been widely applied to time series analysis, recognition, and classification processes. It divides the data into groups of data points with related characteristics. Additionally, the core tenet of SVMs is to estimate the best hyper-planes for the dataset's best adaptation.As the solution to a quadratic programming issue and the training of SVMs are comparable and produce a single solution, the objective function's convexity is a substantial advantage. In contrast to that, the ANN necessitates exponential optimization, which could cause the algorithm to become enslaved to local minimum. Compared to other existing methods used for prediction, the SVM algorithm has a higher level of precision. The SVM reduces systemic risk, whereas other machine learning techniques concentrate on reducing empirical risk. In order to minimize the training error, the SVM approach concentrates on lowering the optimum generalization error. Also, it efficiently handle a lot of data with minimized overfitting. The development of ideal hyper planes for data separation is another focus of the SVM approach.The structure and capabilities of the human brain system serve as foundation for ANNs. Although the ANN technique has its roots in computer science, it is currently widely applied across a wide range of academic fields.Input, hidden, and output nodes make up all three sorts of nodes found in typical ANN topologies. Each model has a different degree of attributes, while the previous one consists of the explanatory variables. The output nodes, whose total number is determined by decision probabilities, hold the dependent variables. Signals spread forward and are sent through links to connect nodes. Based on the information assigned to each link, various numerical weights are calculated. The conventional machine learning techniques are fused by several rules and formulation. But, in the proposed system, the fusion is performed with the rules of min, max, mean and product. In this model, the posterior probability computation is performed at first for determining the target class. Then, the mean fusion rule is estimated according to the probability of target class of two classifiers. Then, the decision is taken with the use of density function and class conditional probability values. At the end, the target class is produced as the output, which is used to predict diabetes from the provided data.

IV. RESULTS AND DISCUSSION

This section uses a number of metrics and open datasets to validate the effectiveness and outcomes of the suggested BCoFL process. Two different and well-known datasets, PIMA and BRFSS, have been employed in this study to validate the performance of the suggested model.Here, the PIMA Indians Diabetes dataset, which contains 768 female diabetic patients from the Indian population, and it has been employed to train and evaluate the machine learning models. This collection of data comprises of 268 people with diabetes and 500 patients without diabetes, each with eight unique features. The parameters used in this study for analysis are computed by using the following equations**:**

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{16}$$

$$\Pr ecision = \frac{TP}{TP+FP} \tag{17}$$

$$\operatorname{Re} call = \frac{TP}{TP+FN} \tag{18}$$

$$F1\_Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{19}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{20}$$

$$Specificity = \frac{TN}{TN + FP} \tag{21}$$

Where, TP – true positives, TN – true negatives, FP – false positives, and FN – false negatives. Fig. 3 verifies the efficacy of the proposed BCoFL, deep learning, and standard [24]machine learning techniques for diabetes prediction. In general, one of the most important parameters used to evaluate the classifier's prediction outcomes is accuracy. The results show that, in comparison to the earlier methods, the proposed BCoFL technique offers a higher accuracy value.

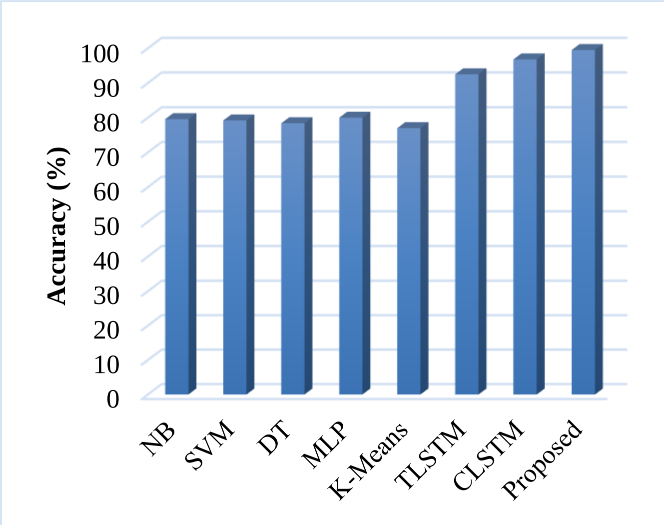

Fig 3. Accuracy using PIMA dataset

Additionally, the PIMA and BRFSS datasets are used to validate and evaluate the Receiver Operating Characteristics (ROC) of the conventional and proposed BCoFL approaches, as shown in Figs. 4 and 5, respectively. These methods are Extra Tree (ET), Decision Tree (DT), AdaBoost (AB), and Gradient Boost (GB) that were taken into consideration for this study. The outcomes show that the suggested BCoFL strategy outperforms the earlier methods with improved ROC results. The disease prediction performance of the FLCA in the proposed system is considerably enhanced by the use of the MSB-CHGS approach.

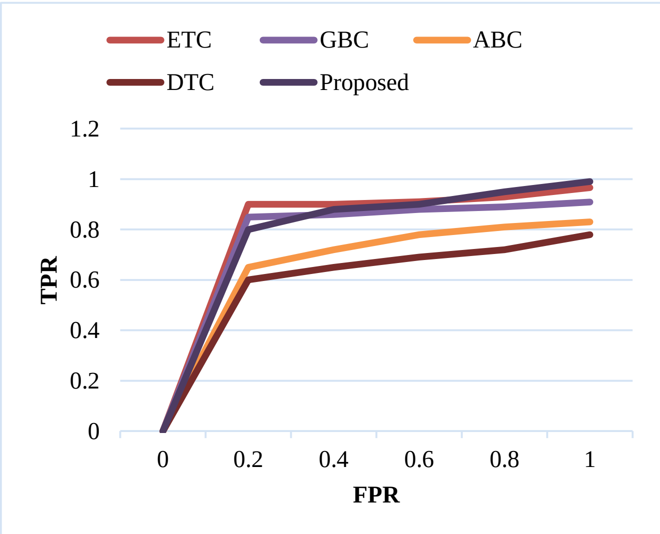

Fig 4. ROC plot for PIMA dataset

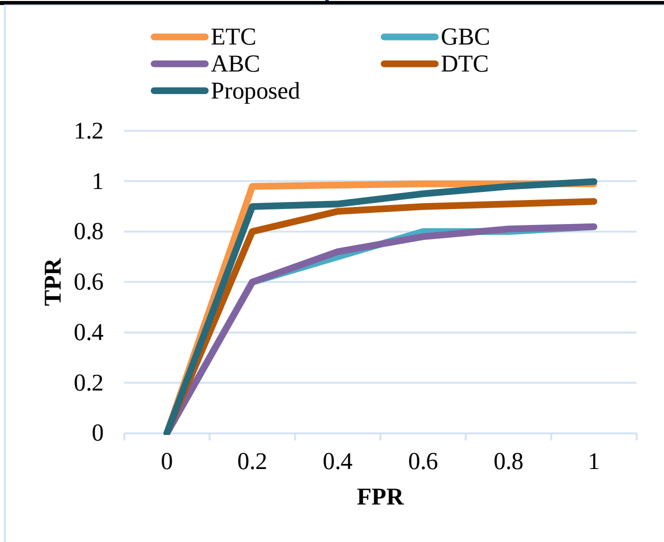

Fig 5. ROC plot for BRFSS dataset

Using the PIMA dataset, Table 2 and Fig. 6 compare and validate the precision, recall, f1-score, and accuracy values of the existing [25] and suggested diabetes detection techniques. The BRFSS dataset's validation and comparison of the same parameters is depicted in Table 3 and Fig 7. The study's findings show that the BCoFL technique is superior to other procedures with high performance values.

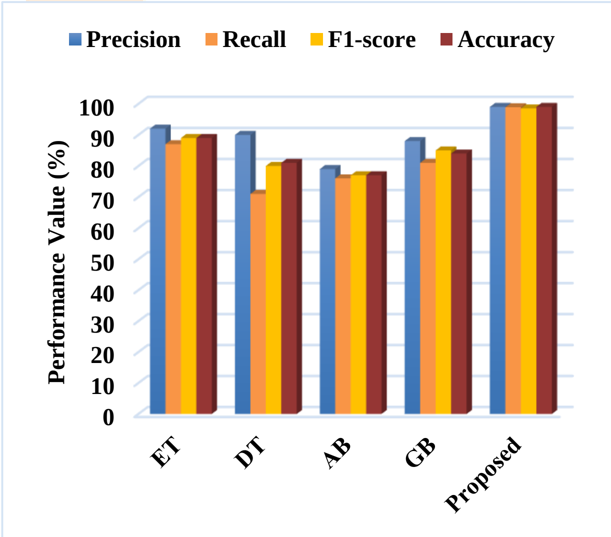

Fig 6. Performance comparison using PIMA dataset

Table 2. Comparative analysis using PIMA dataset

Methods	*Precision*	*Recall*	*F1-score*	*Accuracy*
Extra Tree (ET)	*92*	*87*	*89*	*89*
Decision Tree (DT)	*90*	*71*	*80*	*81*
AdaBoost (AB)	*79*	*76*	*77*	*77*
Gradient Boost (GB)	*88*	*81*	*85*	*84*

Proposed	*99*	*98.9*	*98.5*	*99*

Table 3. Comparative analysis using BRFSS dataset

Methods	*Precision*	*Recall*	*F1-score*	*Accuracy*
Extra Tree (ET)	*99*	*94*	*96*	*96*
Decision Tree (DT)	*99*	*86*	*92*	*92*
AdaBoost (AB)	*75*	*73*	*74*	*75*
Gradient Boost (GB)	*73*	*79*	*76*	*75*
Proposed	*98.8*	*98.5*	*99*	*99*

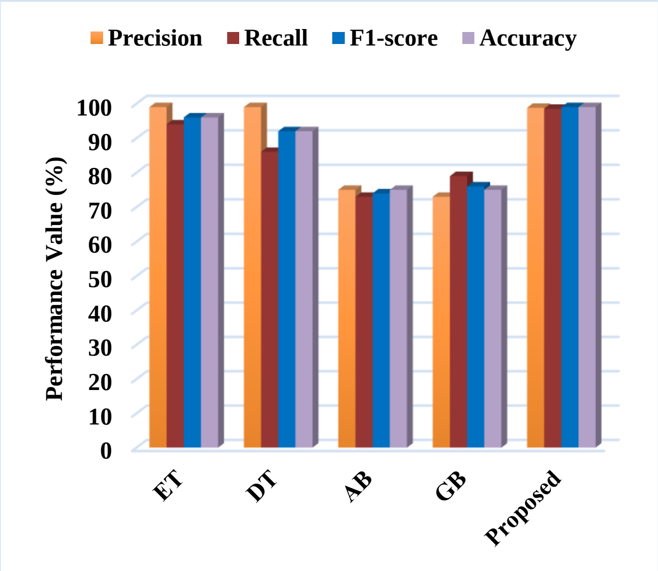

Fig 7. Performance comparison using BRFSS dataset

The overall comparative study of the suggested and conventional diabetes detection approaches[7]using the PIMA dataset is presented in Tables 4 and 5. Then, Fig. 8 to Fig. 10 give its pertinent graphical depictions. A wide range of evaluation indicators are taken into account for this assessment. When compared to existing learning models, the estimated output findings show that the proposed BCoFL mechanism offers better prediction outcomes. Since the introduction of MSB-CHGS is the main factor enhancing the FLCA's prediction outcomes.

Table 4. Performance comparison with machine learning models

Methods	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F-Measure (%)*

SMO-SMOTE	*99.07*	*96.23*	*98.24*	*97.71*
J48	*96.62*	*95.06*	*96.35*	*93.21*
RF	*90.34*	*89.09*	*84.43*	*86.78*
Part	*76.33*	*60.72*	*65.81*	*62.76*
NB	*77.43*	*60.23*	*65.09*	*62.88*
Proposed	*99.5*	*99*	*98.8*	*99*

Table 5. Detection performance

Methods	***Specificity (%)***	***Sensitivity (%)***	***Execution time (ms)***
SMO-SMOTE	*99.14*	*95.52*	*0.1*
J48	*97.86*	*94.03*	*0.2*
RF	*91.43*	*88.06*	*3.66*
Part	*84.29*	*59.70*	*0.25*
NB	*85.28*	*60.65*	*0.5*
Proposed	*99.2*	*99*	*0.09*

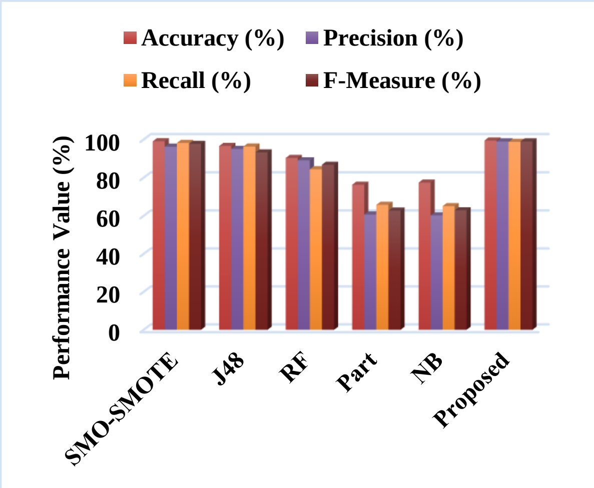

Fig 8. Prediction performance analysis using PIMA dataset

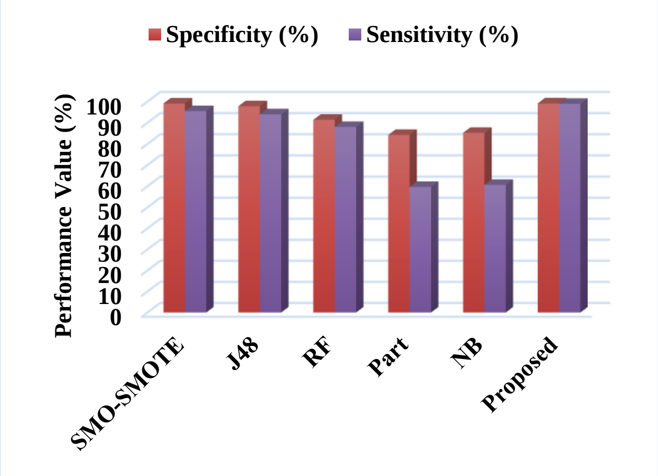

Fig 9. Sensitivity and Specificity comparison

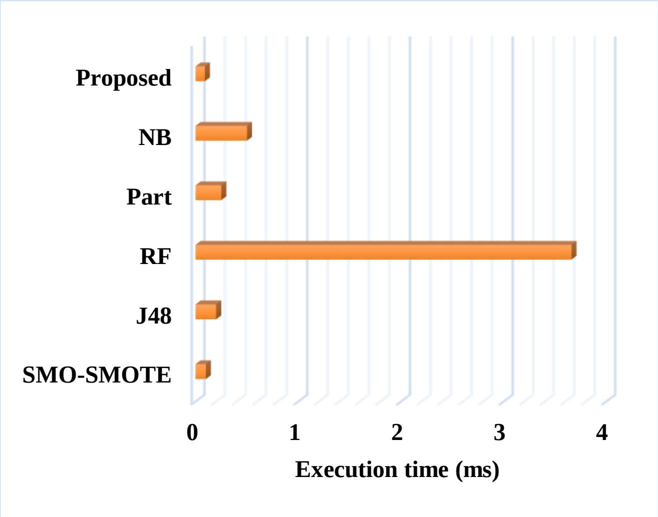

Fig 10. Time analysis

IV. ACKNOWLEDGMENT

The goal of this work is to create a smart patient health monitoring system based on machine learning that can quickly and reliably detect a patient's existence of a chronic condition like diabetes. The data that this monitoring system collects comes from several medical wearable. This program then evaluates the patient's health status using machine learning algorithms on the raw data in order to get the correct conclusions and diagnoses. This technology forecasts the patient's status, and in the event of an emergency, a signal is sent to the doctor and guardians to request immediate care. The development of a novel framework known as BCoFL for the identification and categorization of phenomena is the primary goal of the current research.The open PIMA and BRFSS datasets are used in this work's system development and analysis. Increased false prediction and error rates can have an impact on the classifier's performance because the initial dataset is frequently insufficient and may contain some irrelevant data. Thus, preprocessing and data normalization processes are carried out initially in order to obtain normalized data. Using the MSB-CHGS optimization technique, the essential and beneficial features are then chosen from the preprocessed data. The FLCA is used to categorize the supplied data as healthy or diabetes disease affected. The effectiveness and results of the suggested BCoFL technique are validated and assessed using meaningful measures.The results show that for the datasets employed in this work, the suggested BCoFL technique offers the maximum accuracy up to 99.2% with the least amount of processing time of 0.09 ms. The current study can be improved in the future by incorporating a hyper-parameter tuning model with an advanced deep learning algorithm for diabetes prediction.

REFERENCES

[1] A. Pyne and B. Chakraborty, "Artificial Neural Network based approach to Diabetes Prediction using Pima Indians Diabetes Dataset," in *2023 International Conference on Control, Automation and Diagnosis (ICCAD)*, 2023, pp. 01-06.

[2] M. A. Uddin, M. M. Islam, M. A. Talukder, M. A. A. Hossain, A. Akhter, S. Aryal, *et al.*, "Machine Learning Based Diabetes Detection Model for False Negative Reduction," *Biomedical Materials & Devices*, pp. 1-17, 2023.

[3] G. Rajput and A. Alashetty, "Diabetes Classification Using ML Algorithms," in *Inventive Systems and Control: Proceedings of ICISC 2023*, ed: Springer, 2023, pp. 867-877.

[4] G. V. S. Aripirala and P. Ramalingam, "Diabetes mellitus (DM) detection using decision tree and naïve bayes algorithm for accuracy, specificity and sensitivity improvement," in *AIP Conference Proceedings*, 2023.

[5] M. F. Aslan and K. Sabanci, "A novel proposal for deep learning-based diabetes prediction: Converting clinical data to image data," Diagnostics, vol. 13, p. 796, 2023.

[6] M. F. Aslan and K. Sabanci, "A novel proposal for deep learning-based diabetes prediction: Converting clinical data to image data," *Diagnostics*, vol. 13, p. 796, 2023.

[7] H. Naz and S. Ahuja, "SMOTE-SMO-based expert system for type II diabetes detection using PIMA dataset," *International Journal of Diabetes in Developing Countries*, vol. 42, pp. 245-253, 2022.

[8] A. Doğru, S. Buyrukoğlu, and M. Arı, "A hybrid super ensemble learning model for the early-stage prediction of diabetes risk," *Medical & Biological Engineering & Computing*, vol. 61, pp. 785-797, 2023.

[9] A. Perdana, A. Hermawan, and D. Avianto, "Analyze Important Features of PIMA Indian Database For Diabetes Prediction Using KNN," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, pp. 70-75, 2023.

[10] R. Rajalakshmi, P. Sivakumar, L. K. Kumari, and M. C. Selvi, "A novel deep learning model for diabetes melliuts prediction in IoT-based healthcare environment with effective feature selection mechanism," *The Journal of Supercomputing*, pp. 1-21, 2023.

[11] M. Zahedi, S. Bhajna, and A. Das, "Novel Machine Learning Techniques for Diabetes Prediction," in *Doctoral Symposium on Human Centered Computing*, 2023, pp. 274-288.

[12] A. Hennebelle, H. Materwala, and L. Ismail, "HealthEdge: a machine learning-based smart healthcare framework for prediction of type 2 diabetes in an integrated IoT, edge, and cloud computing system," *Procedia Computer Science*, vol. 220, pp. 331-338, 2023**.**

[13] K. Azbeg, M. Boudhane, O. Ouchetto, and S. Jai Andaloussi, "Diabetes emergency cases identification based on a statistical predictive model," *Journal of Big Data*, vol. 9, pp. 1-25, 2022.

[14] K. MANOHARI, "An Efficient Exploration on Big Data Analysis in Adolescent Diabetic Prediction with Deep Learning Techniques: Big Data Analysis," *International Journal of Information Technology, Research and Applications*, vol. 2, pp. 33-40, 2023.

[15] S. A. V. Jesuraj, R. Ganesh, V. Manjramkar, M. Sridharan, V. R. Dubey, and M. Arun, "A Novel Machine Learning Technique for Diabetic Prediction in IoT-based Healthcare Monitoring System," in *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, 2023, pp. 1000-1005.

[16] M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers," *IEEE Access*, vol. 8, pp. 76516-76531, 2020.

[17] O. T. Kee, H. Harun, N. Mustafa, N. A. Abdul Murad, S. F. Chin, R. Jaafar, *et al.*, "Cardiovascular complications in a diabetes prediction model using machine learning: a systematic review," *Cardiovascular Diabetology*, vol. 22, p. 13, 2023.

[18] A. Yahyaoui, A. Jamil, J. Rasheed, and M. Yesiltepe, "A decision support system for diabetes prediction using machine learning and deep learning techniques," in *2019 1st International informatics and software engineering conference (UBMYK)*, 2019, pp. 1-4.

[19] N. Abdulhadi and A. Al-Mousa, "Diabetes detection using machine learning classification methods," in *2021 International Conference on Information Technology (ICIT)*, 2021, pp. 350-354.

[20] A. Sharma, K. Guleria, and N. Goyal, "Prediction of diabetes disease using machine learning model," in *International Conference on Communication, Computing and Electronics Systems: Proceedings of ICCCES 2020*, 2021, pp. 683-692.

[21] A. U. Haq, J. P. Li, J. Khan, M. H. Memon, S. Nazir, S. Ahmad, *et al.*, "Intelligent machine learning approach for effective recognition of diabetes in E-healthcare using clinical data," *Sensors*, vol. 20, p. 2649, 2020.

[22] A. Al-Zebari and A. Sengur, "Performance comparison of machine learning techniques on diabetes disease detection," in *2019 1st international informatics and software engineering conference (UBMYK)*, 2019, pp. 1-4.

[23] S. Padhy, S. Dash, S. Routray, S. Ahmad, J. Nazeer, and A. Alam, "IoT-based hybrid ensemble machine learning model for efficient diabetes mellitus prediction," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[24] P. B. K. Chowdary and R. U. Kumar, "An effective approach for detecting diabetes using deep learning techniques based on convolutional LSTM networks," *International Journal of Advanced Computer Science and Applications*, vol. 12, 2021.

[25] M. A. Hama Saeed, "Diabetes type 2 classification using machine learning algorithms with up-sampling technique," *Journal of Electrical Systems and Information Technology*, vol. 10, p. 8, 2023/02/08 2023.