Enhancing Healthcare Data Classification: Leveraging Machine Learning on ChatGPT-Generated Datasets

Basheer Abd Al Rida Sadiq[1,*], Murhaf Obaidi[2]

[1]Al-Imam Al-Kadhum College for Islamic Science University, Iraq

[2]Mustansiriah University, Department of Mathematics, Iraq

Emails: basheer.abdrida@alkadhum-col.edu.iq, Red7obaidi756@gmail.com

Abstract

With the large-scale language methods namely ChatGPT, there is a chance to explore the use of machine learning (ML) methods on ChatGPT-generated data for classifying healthcare data. Healthcare data classification gains more significance in extracting and organizing useful insights from the huge volume of medical data available. The ChatGPT-generated data has realistic and different healthcare-based text datasets that can be applied to training classification methods. ML approaches include supervised learning methods as support vector machines (SVMs), and random forests (RF), which can be implemented for classifying the healthcare data. The methods were trained on the ChatGPT-generated data that can be carefully validated and labelled with suitable classes related to the healthcare field. With this motivation, this article presents an automated healthcare data classification-using barnacles mating optimizer with a pyramid neural network (AHDC-BMOPNN) technique. The presented AHDC-BMOPNN technique examines the healthcare data effectually using an ML model with a feature selection process. Primarily, the AHDC-BMOPNN technique exploits min-max data normalization for scaling the input dataset. In addition, the butterfly optimization algorithm-based feature selection (BOA-FS) method is deployed for the selection of optimum feature subset. In this work, the PNN algorithm was utilized for the classification of medical data. Ultimately, the BMO-based hyperparameter tuning process takes place to boost the overall classifier results of the PNN technique. The empirical findings of the AHDC-BMOPNN approach was validated on ChatGPT generated dataset. The simulation values highlight that the AHDC-BMOPNN method and the diverse healthcare text data generated by ChatGPT enhance the ability to extract valuable insights and organize medical information effectively.

Keywords: Healthcare data analysis; ChatGPT; Feature selection; Artificial intelligence; Machine learning; Metaheuristics

1. Introduction

The OpenAI has developed ChatGPT, a state of art language method that can generate human-like text. It depends on the Generative Pre-training Transformer (GPT) structure, which can be a type of neural network (NNs), devised for generating natural language text [1]. It can be trained on large datasets of human-generated texts and can produce human-like responses to an extensive range of natural language inputs. It answers the query, language translation, language translation, technical writing, text summarization, and customer care services [2]. Nowadays, it is trending because of high-quality text generation, versatility and customizable, easy use, development in Natural Language Processing (NLP), and higher demand for conversational AI. ChatGPT utilizes ML to produce natural language [3]. It is pretrained initially on large dataset of human-generated texts, like websites, books, and articles. The method learns to generate and understand natural language texts by examining these datasets. After, the method is fine-tuned on particular domains or tasks by training it on small datasets that are related to that

DOI: https://doi.org/10.54216/IJAACI.050203
Received: August 09, 2023 Revised: December 11, 2023 Accepted: March 10, 2024

domain or task [4]. Once fine-tuned, it could accept natural language inputs, like a prompt or a question, and produce a fluent and coherent response. The input was sent through the feed-forward neural network of the model, which has many layers of artificial neurons. The method generates output texts by sampling from a probability distribution of the next word in the series, given by the earlier words [5]. The last layer of the network termed the decoder made the output. ChatGPT enhances adherence to treatment regimens and presents accessible and practical care, improves the care presented by human health care providers, and enhances the results of patients [6]. A transformer method would identify features in the training dataset and implement that inference knowledge. Transformer methods can hallucinate forecasts in clinical summaries as they were rewarded by generating predictions and finding patterns depending on them [7]. Patients who live in rural or underserved areas have struggles in meeting a qualified diabetes educator or other healthcare experts. Those patients can utilize ChatGPT to get knowledge and help from dependable sources, while they could not visit medical centres physically. Patients can suffer from confusion and anxiety as they change to their new diabetic analysis [8]. Few patients can identify ChatGPT as an approachable and handy technique to get data and support while finding out how to control their diseases. For some persons to adequately control their diabetes, intensive or recurrent assistance is required. Such patients can access extra assistance and data using

ChatGPT that aids them in better controlling their conditions. It lowers the necessity of human care, answers patient questions, and enhances happiness. ChatGPT can make interesting and appropriate content through NLP, based on the user and input preferences [9]. It promotes interaction between patients, healthcare professionals, and insurance providers. ChatGPT help provide quick access to relevant healthcare data to suitable parties. Artificial intelligence (AI) is the multidisciplinary method of linguistics and computer science that aims to construct machines likely to execute tasks that need human intelligence [10]. Such tasks contain the ability to understand, learn, rationalize, and adapt extracted methods along with the reactivity to complicated human features namely creativity, emotion, etc.

This article presents an automated healthcare data classification-using barnacles mating optimizer with a pyramid neural network (AHDC-BMOPNN) technique. The presented AHDC-BMOPNN technique examines the healthcare data effectually using an ML model with a feature selection process. Primarily, the AHDC-BMOPNN technique exploits min-max data normalization for scaling the input dataset. In addition, the butterfly optimization algorithm-based feature selection (BOA-FS) method is deployed for the selection of optimal feature subset. In this work, the PNN model was utilized for the classification of medical data. At last, the BMO-based hyperparameter tuning process takes place to improve the overall classification accuracy of the PNN method. The empirical findings of the AHDC-BMOPNN system was validated on ChatGPT generated dataset.

2. Literature Review

Mageshkumar and Lakshmanan [11] establishes an Intelligent DD with Deep TL Aided Classification Model for Cloud-based Healthcare System (IDDTLC-CHS) approach. The neighbourhood correlation sequence (NCS) technique was utilized For DD that creates optimal code words and compresses by Deflate method. The SGO technique was executed to optimum change of parameters contained in the BiGRU approach. Hoang et al. [12] examine an innovative approach exploiting a novel segmentation approach and wide-ShuffleNet for the classifier of skin tumor. Primary, the research workers compute the entropy-based weighting and first-order cumulative moment (EW-FCM) of skin tumor images. Afterwards, the authors input the segmentation results as innovative DL infrastructure wide-ShuffleNet and define the skin cancer types.

Li et al. [13] presented a robust and accurate system to analyze HD and this method depends on ML techniques. This method was introduced by using the ANN, SVM, KNN, LR, DT, and NB classification techniques but standard feature selection (FS) approaches are leveraged like Relief, Minimal redundancy maximal relevance, least absolute shrinkage selective operator, and local learning to eradicate unwanted and essential features. The authors presented a conditional mutual FS approach to resolve the FS problem. In [14], the authors examine the possibility of utilizing the CNN-based classification method as a sample of DL approaches in this design.

Sahoo et al. [15] presented a precise categorization of daily human actions from gyroscope sensor data and accelerometers after transforming them into spectrogram imageries. The feature extraction was followed by using

DOI: https://doi.org/10.54216/IJAACI.050203
Received: August 09, 2023 Revised: December 11, 2023 Accepted: March 10, 2024

pre-trained weights of 2 efficient and popular TL-CNN methods. Eventually, a wrapper-related FS approach was used to choose the best feature subsets that reduced the training period and enhances the final classification performances. Mir and Dhage [16] intend at constructing a classifier method utilizing WEKA tools for forecasting diabetes disease using RF, NB, Simple CART, and SVM algorithms. The research hopes to suggest the optimal method depends on effective performance outcomes for forecasting diabetes diseases.

Stephen et al. [17] develop a CNN method for the detection and classification the presence of pneumonia in the sequence of chest X-ray image (CXR) instances. Unlike other techniques that depend on TL or handcrafted approaches to attain notable classification performances, the authors built a CNN method for extracting features from specified CXR and categorizing it to determine if an individual was diseased with pneumonia. In [18], develops a smart healthcare mechanism for predicting HD employing cooperative DL and feature fusion algorithms. The feature fusion approach combines the attributes abstracted from sensor datasets and electronic clinical reports for generating useful healthcare data. Then, the data gain method eradicates unwanted and essential features and chooses the noteworthy ones that decline the computational burden and enhance the model accuracy.

3. The Proposed Model

In this work, we design the AHDC-BMOPNN technique for healthcare data analysis on ChatGPT-generated data. The projected AHDC-BMOPNN technique examines the ChatGPT-generated healthcare data effectually using an ML model with a feature selection process. It encompasses a set of sub-processes like min-max scaling, BOA-FS-based feature subset selection, PNN classification, and BMO-based hyperparameter tuning. Fig. 1 demonstrates the overall flow of the AHDC-BMOPNN method.

3.1. Data Creation

In this work, the ChatGPT is used to generate a set of medical records for the diabetes disease classification process. The dataset is generated using the following query: Create a csv format dataset with 500 rows for diabetes disease with two classes. Factors involving the dataset are age, gender, BMI, blood_pressure, glucose_level, insulin_level, and class.

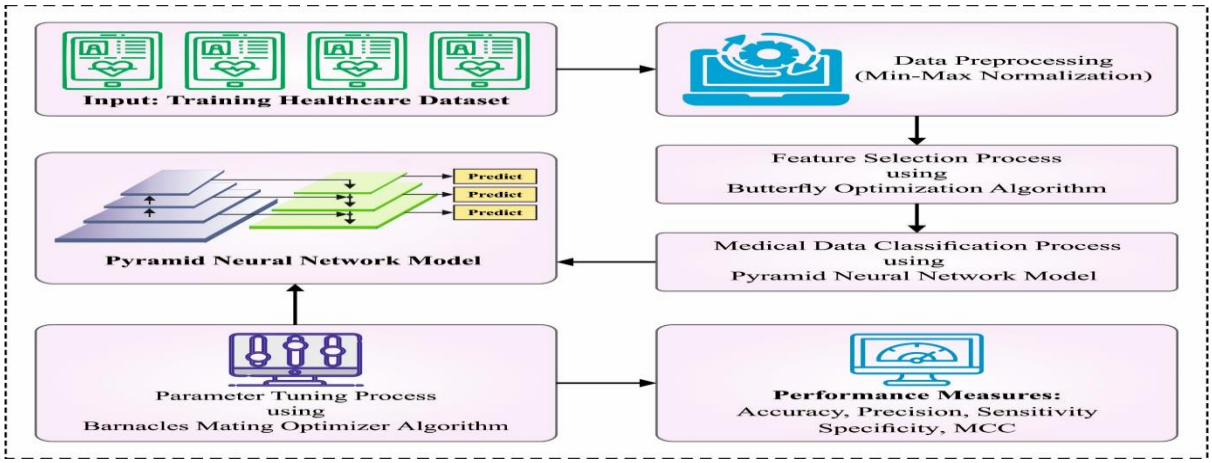

Fig. 1. Overall flow of AHDC-BMOPNN approach

3.2. Min-max Normalization

The min-max normalization technique is used for the data feature normalization, causing biases in their values [19]. Also, it is applied for the scaling values of normalized attributes within [01]; for achieving uniformity in data features. Eq. (1) normalizes the values of data feature in a specific interval.

$$\text{Min Max} = \frac{X_i - \min(A)}{\max(A) - \min(A)} \qquad (1)$$

DOI: https://doi.org/10.54216/IJAACI.050203
Received: August 09, 2023 Revised: December 11, 2023 Accepted: March 10, 2024

3.3. Feature Selection Process

The BOA-FS technique is applied for the optimal selection of feature subset. BOA is a metaheuristic optimization algorithm motivated by the food-searching behaviour of butterflies (BF) [20]. In general, the BF finds the food utilizing the sense receptor that can be employed for smelling the fragrance of flowers and food. In BOA, it can be considered that the BFs themselves that is similar to FF generate the fragrance. Therefore, the BFs can sense 2 fragrances, one from the flower or food and the other from the neighbouring BF. This helps to frame the social learning system viz., the BFs walking towards the fittest BF that produces a strong odour smell. Afterwards, once a BF is incapable of detecting the odour from others, it randomly strides and these movements are named a local search. The fragrance level (pf_i) in BOA generated by the i^{th} BFs are shown as follows:

$$pf_i = cI^a \qquad (2)$$

In Eq. (2), c refers to the sensory system, I indicate the stimulus concentration, and a denotes the power exponent that relies on modality. The c is set to 0.01 and the power exponent is diverse between 0.1 and 0.3 as $= 0.1 + 0.2\left(\frac{r}{T}\right)$, where T denotes the maximal iteration amount and r shows the existing iteration. In the optimizer process, the location vector was upgraded as follows:

$$x_i^{t+1} = x_i^{\mathrm{t}} + F_i^{t+1} \qquad (3)$$

In Eq. (3), x_{if} and x_{if+1} vector characterizes the solution vector of i^{fh} BFs at r and $r+1$ iterations correspondingly. F_i shows the fragrance which is utilized by i^{fh} BF for upgrading the position at a specific iteration. In the global search phase, the movement of i^{th} BF nearby the fittest BF is shown below:

$$F_i^{t+1} = (r^2 \times g^* - x_i^t) \times pf_i \qquad (4)$$

In Eq. (4), F_i^{t+1} shows the amount of movement by i^{fh} BF for updating the location, the x_i^t shows the solution vector for i^{th} BFs at t iteration number, g^* shows the fittest BF (optimum solution) amongst the solution at iteration t (existing iteration) and r implies the uniformly distributed random integer. The local arbitrary search phase of i^{th} BF is modelled by Eq. (5):

$$F_i^{t+1} = \left(r^2 \times x_j^{\mathrm{t}} - x_{\mathrm{k}}^t\right) \times pf_i \qquad (5)$$

In Eq. (5), the x_j^t and x_{k}^t represent the j^{th} and k^{th} BFs correspondingly at iteration numbers r, and r denotes the uniformly distributed random integer. Fig. 2 displays the steps included in BOA.

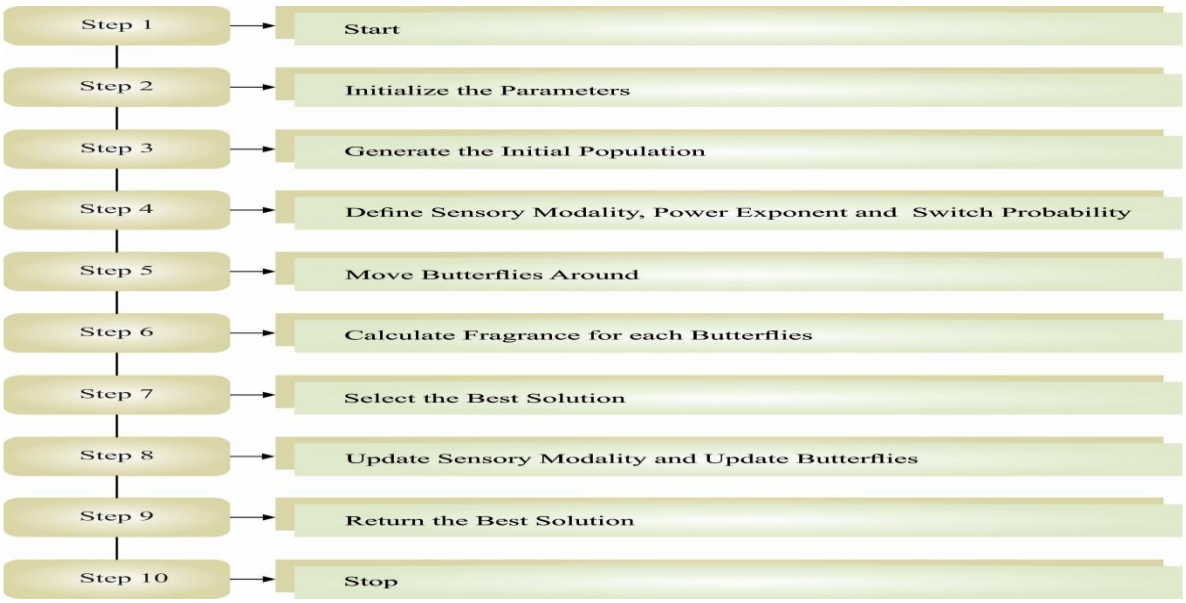

Fig. 2. Steps included in BOA

DOI: https://doi.org/10.54216/IJAACI.050203
Received: August 09, 2023 Revised: December 11, 2023 Accepted: March 10, 2024

The BOA-FS technique can be accomplished by the transfer operator that transforms continuous fragrance value within [0,1]. The sigmoid transfer function has been widely applied for mapping purposes, as follows:

$$S\left(F_i^k(t)\right) = \frac{1}{1 + e^{-F_i^k(t)}} \quad (6)$$

Where $F_i^{\mathrm{k}}(t)$ denotes the continuous value of fragrances of i^{th} BF at t iteration from the k^{th} dimensional.

To obtain the final binarization objective, the thresholding method was utilized over S-shaped curved attained in Eq. (6). Therefore, the thresholding was expressed as follows:

$$x_i^{\mathrm{k}}(t+1) = \{0 if rand \ < S\left(F_i^k(t)\right) lif \ rand \geq S\left(F_i^k(t)\right). \quad (7)$$

Where the $x_i^{\mathrm{k}}(t+1)$ denotes the location of i^{th} BF at $(t+1)$ iteration in k^{th} dimensional and $F_i^{\mathrm{k}}(t)$ indicates the fragrance of i^{th} BF at r iteration in k^{th} dimensional. The feature set provided by the BOA-FS is termed P_1.

3.4. Healthcare Data Analysis using PNN

In this work, the PNN model is used for the disease detection process. The architecture of PNN model comprises exposed and concealed layers [21]. The exposed layer comprises the input and output layers. The concept behindhand this encoded-decoded model is to transform low into higher dimension space via encoded for increasing the amount of degrees of freedom and facilitating the control of non-linear input-output relationship related to the PMSM structure. Then, using the decoder, the data was submitted to low-dimensional space. This procedure is considered as dimensional evolutions among input as well as output parameters.

The study implemented PNN has multi-level feature and context pyramid networks, where all the layers are characterized as weight W multiplied with the x vector having neurons, along with the b deviation. Then, assume a standard grid of N components, where the parameter value i said for $y_i \in [0,1]$, and the aim is to enhance the parameters $y_j \in [0,1]$ for obtaining the optimum structure.

The activation function was added to all the layers of the PNN for realizing the mapping of PNN mechanism in the parameters to the main function, and to efficiently model the highly nonlinear and simple linear mapping problems:

$$y_i = \sigma\left(\chi_i + \bar{b}(X)\right), i \in \{1,2,\dots,N\} \quad (8)$$

$$\sum_{i=1}^{N} y_i = V_0 \quad (9)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (10)$$

Where $X = \{x_1, x_2, \dots, x_N\} \in R^N$, the function σ ensures the magnetic potential and the function $b(x)$ ensures that the PNN satisfies the volume constraints of the optimizer problems.

The study intends to adopt a DNN with layer $L+1$, and the amount of neurons in all the layers $(n_0, n_1, \dots, n_L)$ and $\mu: R \to R$. The pre-activation function $\tilde{a}^{\mathrm{l}} \in R^{n_l}$ and the activation function $a^l \in R^{n_l}$ are recursively defined on each l^{th} layer, as follows:

$$a^0(x) = x$$

$$\tilde{a}^{\mathrm{l}+1}(x) = \frac{\alpha}{\sqrt{n_l}} W^l(x) a^l(x) + \beta b^l \quad (11)$$

$$a^{l+1}(x) = \mu(\tilde{a}^{\mathrm{l}+1}(x))$$

DOI: https://doi.org/10.54216/IJAACI.050203
Received: August 09, 2023 Revised: December 11, 2023 Accepted: March 10, 2024

Where super parameter $\alpha, \beta \in [0,1]$ indicates the contribution of weighted and biased items, correspondingly. Uniform distributed normal random variable $(0,1)$, parameter $\theta = (\theta_p)_p$ of the weighted matrix W^l and the deviation vector b^l. The outcome of the network was $f_\theta(x) = \tilde{\alpha}^{\mathrm{l}}(x)$. To make sure that the difference of neurons at all the layers is equivalent to 1 while initializing, we set α and β to fulfil $\alpha^2 + \beta^2 = 1$ and $E_{X\sim N(0,1)}[\mu(X)^2] = 1$.

The loss function is described as follows:

$$\theta \mapsto C\left(\gamma^{NN}(\theta)\right) = C\left(\sum(X(\theta)\right)) \tag{12}$$

Meanwhile, a primary value in the optimizer technique was generally specified, however, the NN was randomly initialized, and the primary density field in the DL-based optimizer technique was non-constant and random. To evaluate (θ) for preventing these problems, rather than utilizing the initial density field:

$$\forall i \in \{1, \dots, N\} \forall i \in \{1, \dots, N\},$$

$$X_i(\theta) = \overline{f}_{\theta(t)}(z_i) = f_{\theta(t)}(z_i) - f_{\theta(t=0)}(z_i) + \log\left(\frac{V_0}{N - V_0}\right) \tag{13}$$

3.5. Parameter Tuning using BMO Technique

Finally, the BMO technique chooses the parameters of the PNN model [22]. The process of BMO was discussed below:

(a) Initializing:

Assume the barnacles (initial individual) for the solution as a matrix:

$$X = \begin{bmatrix} x_1^1 & \dots & x_1^{\mathrm{N}} \\ \vdots & \ddots & \vdots \\ x_{\mathrm{n}}^1 & \dots & x_{\mathrm{n}}^{\mathrm{N}} \end{bmatrix} \tag{14}$$

In Eq. (14), N defines the count of decision variables which are subjected to the lower and upper boundaries, and n refers to the number of barnacles (population) as follows:

$$u_b = \left[u_b^1, \dots, u_b^i\right] \tag{15}$$

$$l_b = \left[l_b^1, \dots, l_b^i\right] \tag{16}$$

Where l_b and u_b indicate the lower and upper bounds of the i^{th} variables. It is sorted as an initial iteration solution after evaluating the objective function for all the barnacles.

(b) Selection procedure

Based on the length of their penises, $p_1$the selection of barnacles for mating was made. This selection was stimulated by the subsequent consideration:

- The choice is an arbitrary procedure restricted to the penis length of the barnacles.
- All the barnacles might contribute or receive their sperm from other individuals. But every barnacle was fertilized by a single barnacle.
- If at a definite point, then a similar barnacle was chosen (self-mating), it is not to be assumed in the method and the procedure was continual without the newest offspring generation.
- If the selection procedure, it is exceeding the pl value, then the sperm cast will have occurred.

The abovementioned consideration indicates that the BMO involves exploration and exploitation.

DOI: https://doi.org/10.54216/IJAACI.050203
Received: August 09, 2023 Revised: December 11, 2023 Accepted: March 10, 2024

The offspring can be maintained by the sperm cast procedure as follows:

$$b_D = rand(n) \quad (17)$$

$$b_M = rand(n) \quad (18)$$

In Eq. (18), b_D and b_M describes mated parents. Eqs. (17) & (18) show that the selection was randomly made.

(c) Reproduction

The BMO technique emphases on the inheritance features or genotype frequencies of the parents in offspring generation dependent upon HardyWeinberg principles since there is no mathematical process proposed for the reproduction of barnacles. By using the following equation, the new offspring can be accomplished:

$$X_i^{N_{new}} = pX_{b_D}^N + qX_{b_M}^N \quad (19)$$

In Eq. (19), $X_{b_M}^N$ and $X_{b_D}^N$ define the variable of Mum and Dad of barnacles, correspondingly. p is a uniform distribution random number within [0,1], $q = (1 - p)$.

The sperm cast occurs once the selection of barnacle that mated surpasses the pl values that is initially set:

$$X_i^{N_{new}} = rand \times X_{b_M}^n \quad (20)$$

In Eq. (20), $rand$ shows the arbitrarily created number in zero and one. The newest offspring for the exploration technique was created by the Mum barnacle.

The offspring are evaluated and united with the parents to manage the solution matrix expansion in the size of populations. Next, the sorting procedure is implemented for selecting the best solution that fits the population size and the worse outcomes were removed.

Fitness choice is a primary features of the BMO system. The encoded solution can be exploited to assess the goodness of solution candidate. At present, the accuracy values are the major criteria employed to design FF.

$$Fitness = \max(P) \quad (21)$$

$$P = \frac{TP}{TP + FP} \quad (22)$$

Here TP and FP are the true and false positive values.

4. Results and Discussion

The experimental validation of the AHDC-BMOPNN approach is tested on the ChatGPT-generated diabetes dataset. The dataset is generated using the following features: Features: age, gender, BMI, blood_pressure, glucose_level, insulin_level, and class. Among the available features, the presented model has chosen 4 features. The dataset contains 500 samples with two classes as determined in Table 1. A few sample instances are given in the following:

35, Female,26.8,70,100,125,0

52,Male,31.2,90,160,215,1

48,Male,23.5,80,120,100,0

53,Female,25.1,75,145,180,1

30,Male,26.6,70,110,120,0

41,Female,27.3,72,130,150,1

DOI: https://doi.org/10.54216/IJAACI.050203
Received: August 09, 2023 Revised: December 11, 2023 Accepted: March 10, 2024

Table 1: Details of the database

Class	No. of Samples
Non-Diabetic	250
Diabetic	250
Total Samples	**500**

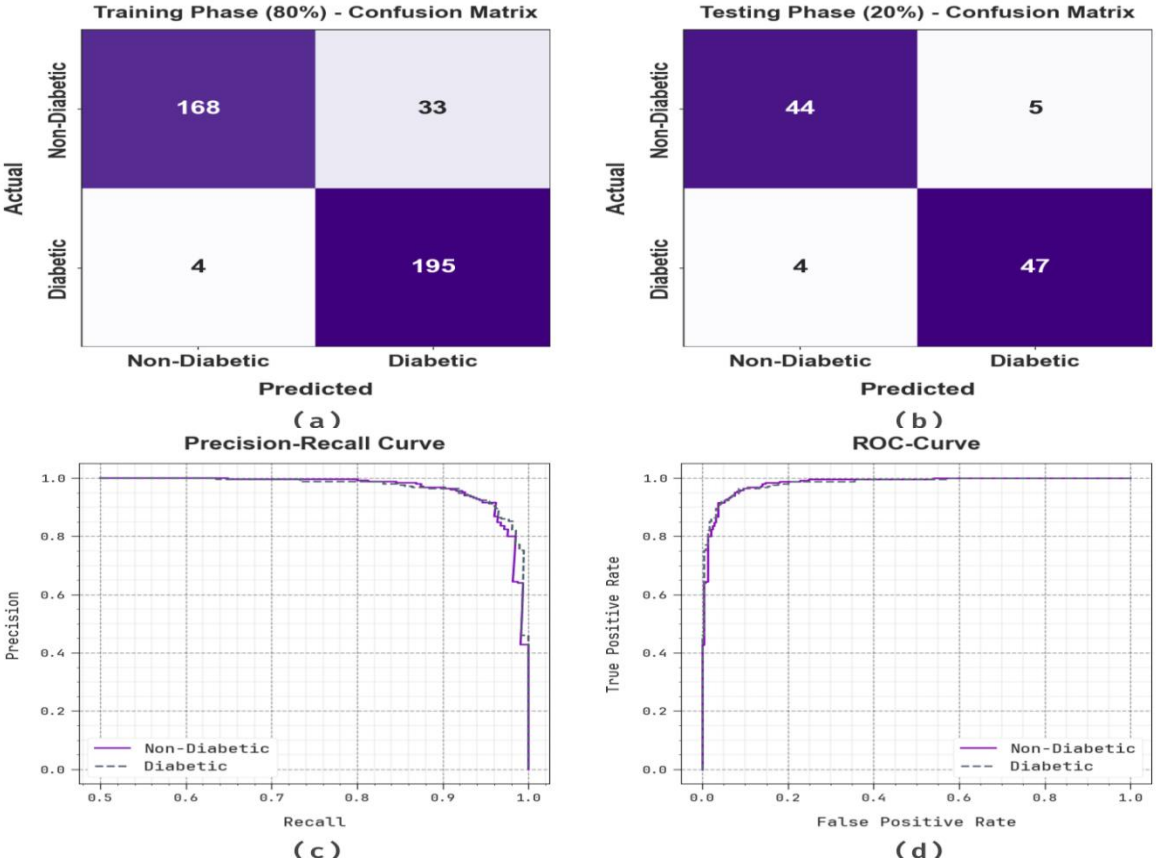

Fig. 3. Classifier result of (a-b) 70%of TRP: 30%of TSP, (c) PR-curve, and (d) ROC-curve

Fig. 3 shows the performance of the AHDC-BMOPNN method under test dataset. Fig. 3a depicts the confusion matrix presented by the AHDC-BMOPNN approach on 80% of TRP. The figure denoted that the AHDC-BMOPNN algorithm has detected 168 instances of Non-Diabetic and 195 instances of Diabetic. Likewise, Fig. 3b depicts the confusion matrix presented by the AHDC-BMOPNN method on 20% of TSP. The outcome represented that the AHDC-BMOPNN approach has identified 44 instances of Non-Diabetic and 47 instances of Diabetic. In addition, Fig. 3c demonstrates the PR curve of the AHDC-BMOPNN method. The figures reported that the AHDC-BMOPNN approach has gained high performance of PR on different class labels. Eventually, Fig. 3d displays the ROC examination of the AHDC-BMOPNN model. The figure depicted that the AHDC-BMOPNN technique has productive outcomes with greater values of ROC on different classes.

The overall diabetic disease detection outcomes of the AHDC-BMOPNN method are depicted in Table 2 and Fig. 4. The figure implies that the AHDC-BMOPNN methodology effectually recognized the diabetic and non-diabetic classes. For example, on 80% of TRP, the AHDC-BMOPNN algorithm gains average $accu_y$, $prec_n$, $sens_y$, $spec_y$, and MCC of 90.79%, 91.60%, 90.79%, 90.79%, and 82.38% correspondingly. Conversely, on 20% of TSP, the AHDC-BMOPNN method offers average $accu_y$, $prec_n$, $sens_y$, $spec_y$, and MCC of 90.98%, 91.03%, 90.98%, 90.98%, and 82% correspondingly.

DOI: https://doi.org/10.54216/IJAACI.050203
Received: August 09, 2023 Revised: December 11, 2023 Accepted: March 10, 2024

Table 2: Diabetic disease detection outcome of AHDC-BMOPNN system on 80% of TRP: 20% of TSP

Class	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	MCC
Training Phase (80%)					
Non-Diabetic	83.58	97.67	83.58	97.99	82.38
Diabetic	97.99	85.53	97.99	83.58	82.38
Average	**90.79**	**91.60**	**90.79**	**90.79**	**82.38**
Testing Phase (20%)					
Non-Diabetic	89.80	91.67	89.80	92.16	82.00
Diabetic	92.16	90.38	92.16	89.80	82.00
Average	**90.98**	**91.03**	**90.98**	**90.98**	**82.00**

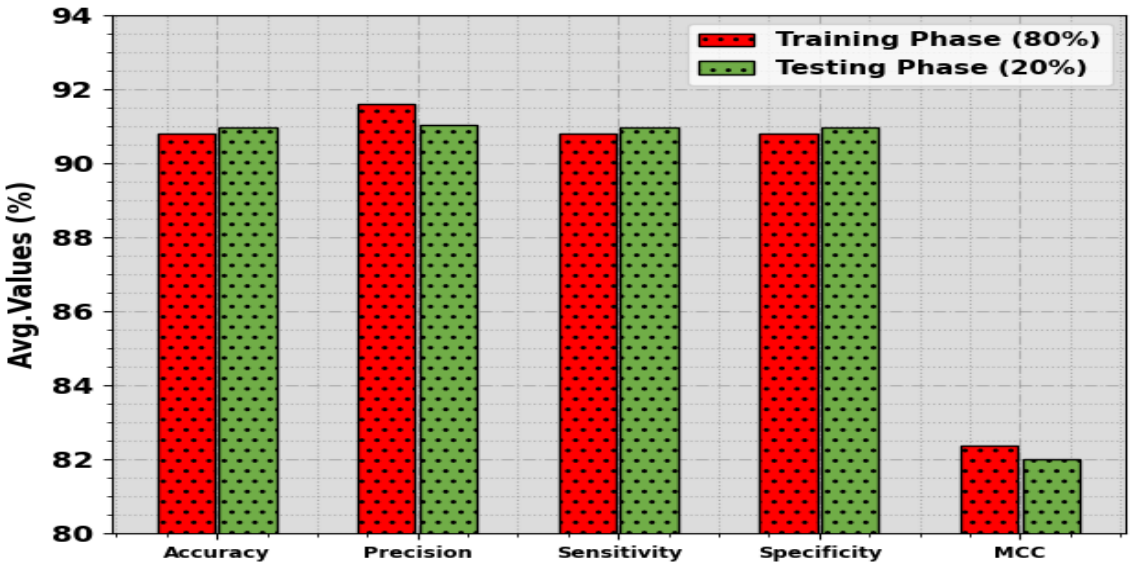

Fig. 4. Average outcome of AHDC-BMOPNN system on 80% of TRP: 20% of TSP

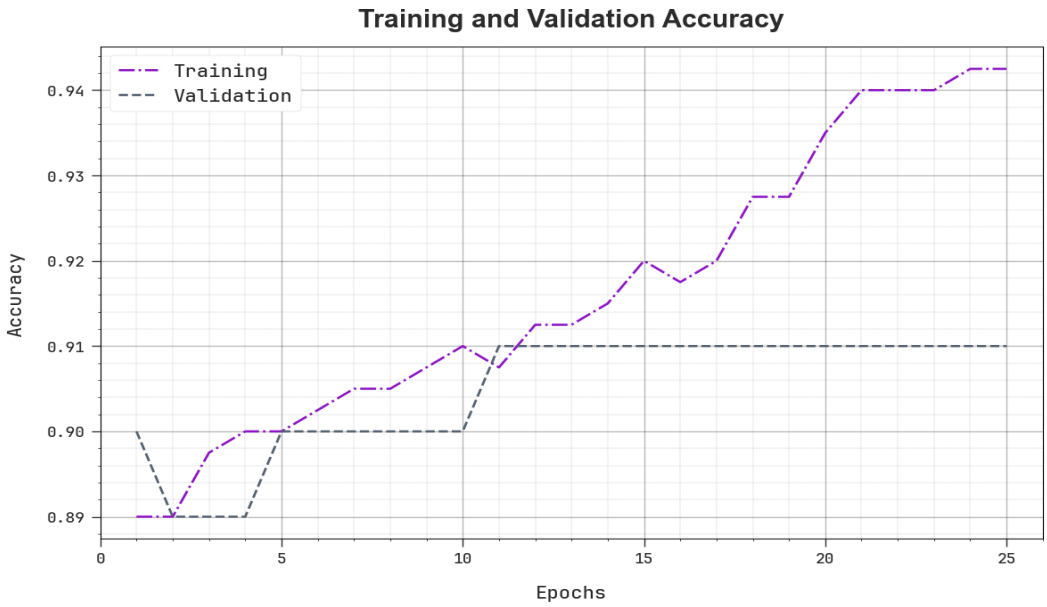

Fig. 5. Accuracy curve of the AHDC-BMOPNN approach

DOI: https://doi.org/10.54216/IJAACI.050203
Received: August 09, 2023 Revised: December 11, 2023 Accepted: March 10, 2024

Fig. 5 examines the $accu_y$ of the AHDC-BMOPNN technique in the training and validation on the test database. The result exhibits that the AHDC-BMOPNN approach obtains greater $accu_y$ values over higher epochs. In addition, the greater validation $accu_y$ over training $accu_y$ shows that the AHDC-BMOPNN method learns productively on the test database.

The loss analysis of the AHDC-BMOPNN method in the training and validation is shown on the test database in Fig. 6. The figure specifies that the AHDC-BMOPNN algorithm reach adjacent values of training and validation loss. The AHDC-BMOPNN method learns productively on a test database.

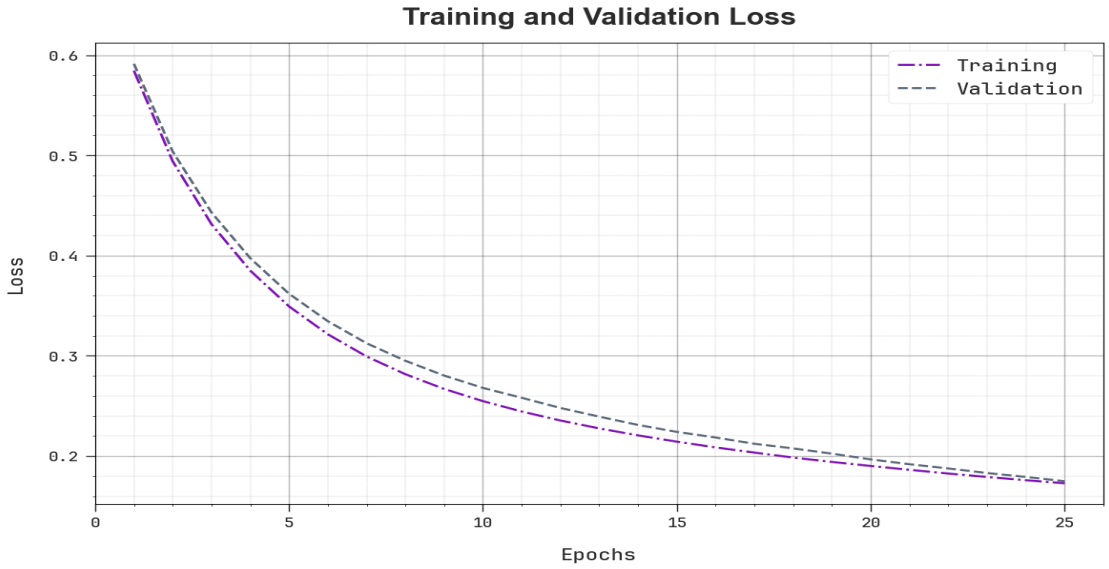

Fig. 6. Loss curve of the AHDC-BMOPNN method

The outperforming outcomes of the AHDC-BMOPNN technique can be assured by a comparative analysis with recent models in Table 3 and Fig. 7. The experimental values stated that the AHDC-BMOPNN method obtains superior outcomes under various measures. Based on $accu_y$, the AHDC-BMOPNN technique provides increasing $accu_y$ of 90.98% while the CNN, XGBoost, ELM, LSTM, NB, and K-means models reached decreasing $accu_y$ of 84.14%, 79.14%, 80.12%, 81.33%, 79.89%, and 80.61% correspondingly. Meanwhile, based on $prec_n$, the AHDC-BMOPNN method offers enhanced $prec_n$ of 91.03% while the CNN, XGBoost, ELM, LSTM, NB, and K-means models gained decreased $prec_n$ of 82.36%, 79.14%, 83.14%, 80.26%, 81.27%, and 81.32% respectively. Simultaneously, based on $sens_y$, the AHDC-BMOPNN algorithm reaches $sens_y$ of 90.98% while the CNN, XGBoost, ELM, LSTM, NB, and K-means models provides reduced $sens_y$ of 81.35%, 84.24%, 83.65%, 80.65%, 79.88%, and 84.29% correspondingly. Finally, based on $spec_y$, the AHDC-BMOPNN method resulted in $spec_y$ of 90.98% while the CNN, XGBoost, ELM, LSTM, NB, and K-means approaches reached decreasing $spec_y$ of 80.16%, 84.90%, 81.96%, 82.07%, 81.18%, and 83.15% respectively. These outcomes stated the supremacy of the AHDC-BMOPNN method over other approaches.

Table 3: Comparative outcome of AHDC-BMOPNN system with existing algorithms

Methodology	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	**MCC**
AHDC-BMOPNN	90.98	91.03	90.98	90.98	82.00
CNN Model	84.14	82.36	81.35	80.16	77.68
XGBoost	79.14	79.14	84.24	84.90	71.95
ELM Model	80.12	83.14	83.65	81.96	77.39
LSTM Model	81.33	80.26	80.65	82.07	76.59
Naïve Bayes	79.89	81.27	79.88	81.18	73.60
K-means	80.61	81.32	84.29	83.15	77.55

DOI: https://doi.org/10.54216/IJAACI.050203
Received: August 09, 2023 Revised: December 11, 2023 Accepted: March 10, 2024

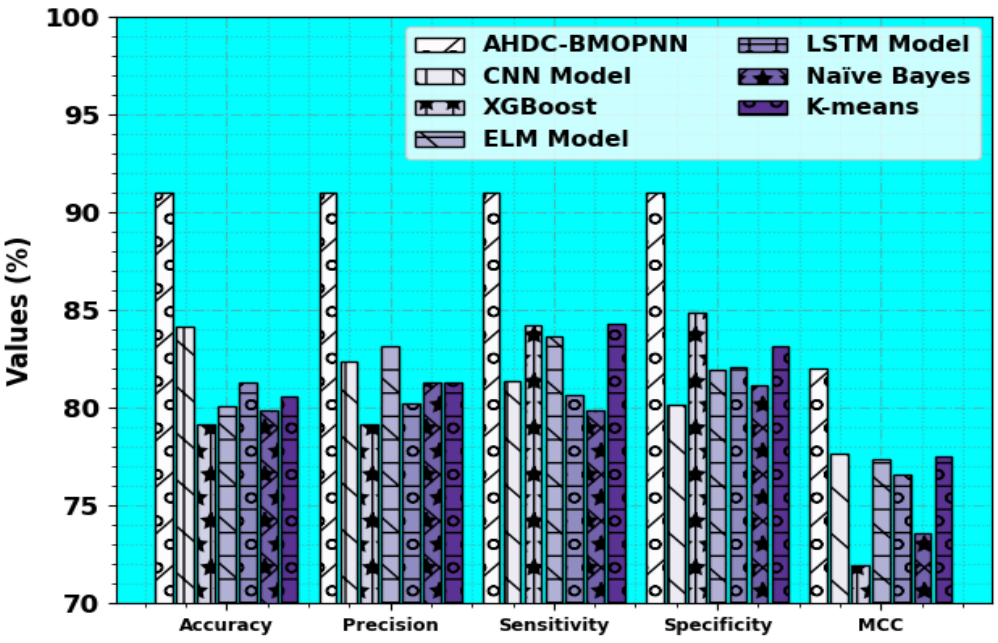

Fig. 7. Comparative outcome of AHDC-BMOPNN method with existing algorithms

5. Conclusion

In this article, we design the AHDC-BMOPNN technique for healthcare data analysis on ChatGPT-generated data. The projected AHDC-BMOPNN technique examines the ChatGPT-generated healthcare data effectually using an ML model with a feature selection process. It encompasses a set of sub-processes such as min-max scaling, BOA-FS-based feature subset selection, PNN classification, and BMO-based hyperparameter tuning. In this study, the AHDC-BMOPNN technique uses the BOA-FS technique for selecting an optimum subset of features. For medical data classification, the PNN model was used in this work. At last, the BMO-based hyperparameter tuning process takes place to improve the overall classification accuracy of the PNN approach. The experimental values of the AHDC-BMOPNN method was validated on ChatGPT generated database. The simulation result analysis indicate that the AHDC-BMOPNN algorithm and the diverse healthcare text data generated by ChatGPT enhance the ability to extract valuable insights and organize medical information effectively.

References

[1] Xue, V.W., Lei, P. and Cho, W.C., 2023. The potential impact of ChatGPT in clinical and translational medicine. *Clinical and Translational Medicine*, *13*(3).

[2] Aydın, Ö. and Karaarslan, E., 2022. OpenAI ChatGPT generated literature review: Digital twin in healthcare. *Available at SSRN 4308687*.

[3] Sallam, M., 2023, March. ChatGPT utility in healthcare education, research, and practice: a systematic review on the promising perspectives and valid concerns. In *Healthcare* (Vol. 11, No. 6, p. 887). MDPI.

[4] Li, J., Dada, A., Kleesiek, J. and Egger, J., 2023. ChatGPT in Healthcare: A Taxonomy and Systematic Review. *medRxiv*, pp.2023-03.

[5] Islam, N., Sutradhar, D., Noor, H., Raya, J.T., Maisha, M.T. and Farid, D.M., 2023. Distinguishing Human Generated Text From ChatGPT Generated Text Using Machine Learning. *arXiv preprint arXiv:2306.01761*.

[6] Javaid, M., Haleem, A. and Singh, R.P., 2023. ChatGPT for healthcare services: An emerging stage for an innovative perspective. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *3*(1), p.100105.

[7] Cascella, M., Montomoli, J., Bellini, V. and Bignami, E., 2023. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *Journal of Medical Systems*, *47*(1), p.33.

DOI: https://doi.org/10.54216/IJAACI.050203
Received: August 09, 2023 Revised: December 11, 2023 Accepted: March 10, 2024

[8] Liao, W., Liu, Z., Dai, H., Xu, S., Wu, Z., Zhang, Y., Huang, X., Zhu, D., Cai, H., Liu, T. and Li, X., 2023. Differentiate ChatGPT-generated and Human-written Medical Texts. *arXiv preprint arXiv:2304.11567*.

[9] Wang, D.Q., Feng, L.Y., Ye, J.G., Zou, J.G. and Zheng, Y.F., 2023. Accelerating the integration of ChatGPT and other large-scale AI models into biomedical research and healthcare. *MedComm–Future Medicine*, *2*(2), p.e43.

[10] De Angelis, L., Baglivo, F., Arzilli, G., Privitera, G.P., Ferragina, P., Tozzi, A.E. and Rizzo, C., 2023. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Frontiers in Public Health*, *11*, p.1567.

[11] Mageshkumar, N. and Lakshmanan, L., 2023. Intelligent data deduplication with Deep Transfer Learning Enabled Classification Model for Cloud-based Healthcare Systems. *Expert Systems with Applications*, *215*, p.119257.

[12] Hoang, L., Lee, S.H., Lee, E.J. and Kwon, K.R., 2022. Multiclass skin lesion classification using a novel lightweight deep learning framework for smart healthcare. *Applied Sciences*, *12*(5), p.2677.

[13] Li, J.P., Haq, A.U., Din, S.U., Khan, J., Khan, A. and Saboor, A., 2020. Heart disease identification method using machine learning classification in e-healthcare. *IEEE Access*, *8*, pp.107562-107582.

[14] Azimi, I., Takalo-Mattila, J., Anzanpour, A., Rahmani, A.M., Soininen, J.P. and Liljeberg, P., 2018, September. Empowering healthcare IoT systems with hierarchical edge-based deep learning. In *Proceedings of the 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies* (pp. 63-68).

[15] Sahoo, K.K., Ghosh, R., Mallik, S., Roy, A., Singh, P.K. and Zhao, Z., 2023. Wrapper-based deep feature optimization for activity recognition in the wearable sensor networks of healthcare systems. *Scientific Reports*, *13*(1), p.965.

[16] Mir, A. and Dhage, S.N., 2018, August. Diabetes disease prediction using machine learning on big data of healthcare. In *2018 fourth international conference on computing communication control and automation (ICCUBEA)* (pp. 1-6). IEEE.

[17] Stephen, O., Sain, M., Maduh, U.J. and Jeong, D.U., 2019. An efficient deep learning approach to pneumonia classification in healthcare. *Journal of healthcare engineering*, *2019*.

[18] Ali, F., El-Sappagh, S., Islam, S.R., Kwak, D., Ali, A., Imran, M. and Kwak, K.S., 2020. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. *Information Fusion*, *63*, pp.208-222.

[19] Iqbal, N., Jamil, F., Ahmad, S. and Kim, D., 2021. A novel blockchain-based integrity and reliable veterinary clinic information management system using predictive analytics for the provisioning of quality health services. *IEEE Access*, *9*, pp.8069-8098.

[20] Qaisar, S.M., Khan, S.I., Srinivasan, K. and Krichen, M., 2023. Arrhythmia classification using multirate processing metaheuristic optimization and variational mode decomposition. *Journal of King Saud University-Computer and Information Sciences*, *35*(1), pp.26-37.

[21] Liu, X., Peng, W., Xie, L. and Zhang, X., 2023. Optimization of a Multi-Type PMSM Based on Pyramid Neural Network. *Applied Sciences*, *13*(11), p.6810.

[22] Li, H., Guo, H. and Yousefi, N., 2020. A hybrid fuel cell/battery vehicle by considering economy considerations optimized by Converged Barnacles Mating Optimizer (CBMO) algorithm. *Energy Reports*, *6*, pp.2441-2449.

DOI: https://doi.org/10.54216/IJAACI.050203
Received: August 09, 2023 Revised: December 11, 2023 Accepted: March 10, 2024