



Research article

Prediction of diabetic patients in Iraq using binary dragonfly algorithm with long-short term memory neural network

Zaineb M. Alhakeem^{1,*}, Heba Hakim², Ola A. Hasan^{3,8}, Asif Ali Laghari⁴, Awais Khan Jumani^{5,6}, and Mohammed Nabil Jasm⁷

¹ Chemical Engineering and Oil Refining Department, Oil and Gas Engineering College, Basrah University for Oil and Gas, Basrah, Iraq

² Computer Engineering Department, Engineering College, University of Basrah, Basrah, Iraq

³ Real Estate Bank of Iraq, Basrah, Iraq

⁴ Software college, Shenyang Normal University, Shenyang, China

⁵ School of Electronic and Information Engineering, South China University of Technology, Guangzhou, Guangdong, China

⁶ Department of Computer Science, ILMA University Karachi, Sindh, Pakistan

⁷ Basra Oil Company, Basrah, Iraq

⁸ Computer Technology Engineering Department, Iraq University College, Basrah, Iraq

* **Correspondence:** Email: zainebalhakeem@buog.edu.iq.

Abstract: Over the past 20 years, there has been a surge of diabetes cases in Iraq. Blood tests administered in the absence of professional medical judgment have allowed for the early detection of diabetes, which will fasten disease detection and lower medical costs. This work focuses on the use of a Long-Short Term Memory (LSTM) neural network for diabetes classification in Iraq. Some medical tests and body features were used as classification features. The most relevant features were selected using the Binary Dragon Fly Algorithm (BDA) Binary version of the selection method because the features either selected or not. To reduce the number of features that are used in prediction, features without effects will be eliminated. This effects the classification accuracy, which is very important in both the computation time of the method and the cost of medical test that the individual will take during annual check ups. This work found out that among 11 features, only five features are most relevant to the disease. These features provide a classification accuracy up to 98% among three classes: diabetic, non diabetic and pre-diabetic.

Keywords: diabetes; binary dragon fly algorithm; long short term memory classifier; multi class classification

1. Introduction

Diabetes is a very serious disease that could lead to other complications if it is not diagnosed early. According to International Diabetics Federation (IDF) Diabetic Atlas website, there are increasing numbers of diagnosed patients in Iraq, with 10.7% of the population within ages 20-79 years being diagnosed in 2021, and is estimated to reach 12.2% within the next 20 years [1]. This percentage represents patients that have been diagnosed; however, there are patients who have not been diagnosed as of yet. Diagnosis is performed by either using medical tests such as urea, glucose, cholesterol, etc. or by bodily symptoms such as body weight, gender, etc. Doctors and researchers depend on their experience to choose which test is more relevant and appropriate to provide a good diagnosis alongside bodily symptoms.

Machine Learning is used in numerous application such as speech emotion recognition [2], hand gesture recognition [3] and medical applications found in the literature such as emotion recognition from electroencephalogram signals (EEG) [4]. In general, ML methods are used to classify the patients into diabetic, non-diabetic and pre-diabetic classes depending on the results of certain blood tests; however some features can decrease the accuracy of classification if they are either relevant or have poor relevancy to diagnosis. Therefore, a selection method should be used to reduce the number of features and to select strong features that will give an accurate diagnosis.

Mujumdar et al. proposed two types of classification procedures. ML methods and a pipeline of methods to obtain an improved classification accuracy [5]. Madan et al. introduced the features selection using a convolution neural network (CNN) and a bi-directional long-short term memory (bi-LSTM) as classifiers. These classifiers are optimized using the grid search technique to find the best hyper-parameters of the network [6]. Chang et al. compared three machine learning methods (J48 decision tree, Random Forest and Naive Bayes) as classifiers and the principal component analysis (PCA) as a feature selection method; they found that the most relevant features were glucose, Body Mass Index(BMI), age, insulin, and skin thickness. The only medical tests analyzed were the blood glucose and insulin ratios. the results of Random Forest is the best; however, the classification accuracy was no greater than 80% for any number of features [7]. Noori et al. compared five machine learning methods, and selected five features for classification. In this work, the Support Vector Machine (SVM) method was the best classifier with accuracy that no more than 83% [8]. Ahmed et al. used a Light Gradient Boosting Machine (LGBM) to classify in the absence of a features selection algorithm [9]. Butt et al. used a Long Short Term Memory (LSTM) neural network as a classifier for PIMA data; the network was fine tuned to achieve a classification accuracy no more than 88% [10]. Zou et al. used a neural network to classify the data alongside the Principle Component Analysis (PCA) as selection method. They used two datasets (Luzhou, China) and (PIMA, india) [11]. In [12], Deep Learning (DL) with two hidden layers was used. All previous studies used a PIMA diabetic dataset: this dataset focuses on diabetes in females for 768 woman aged between (21-81) years and the number of features are eight with only three medical tests and; the rest are body symptoms and features with only two classes of diabetes, with out any consideration of the pre-diabetic state. According to our knowledge, Olisah et al. was the first study that used a dataset from the Laboratory of Medical City Hospital (LMCH) in Iraq, which contains three classes alongside the PIMA dataset. For the first dataset, they used an Optimized twice growth deep neural network(O2GDNN) with feature selection and missing value imputation with the first dataset; for the second dataset, they only used the feature selection [13].

Table 1. Classes numerical values.

Class Label	Numerical Value
Non Diabetic	0
Diabetic	1
Pre Diabetic	2

Almost all of the studies used approximately as follows 80% training data and 20% testing data. The main contribution in this work are as follows:

1. High performance predictor of multi-class diabetic dataset in Iraq, with the prediction accuracy of three classes is of about 98% using 75% training and 25% testing sets.
2. Feature selection method of binary variables with an override condition of all features had a zero value.
3. The selection method decreases the number of features (medical tests) that were used in disease detection , which will lead to a decrease in medical bills.

2. Materials and method

Any prediction system should adopt the following steps: data collection, data pre-processing, features selection and classification. The following section describes the methods used in this work.

2.1. Dataset and pre-processing

In this work, the diabetes dataset was collected in the Laboratory of Medical City Hospital and (Specializes Center for Endocrinology and Diabetes-Al-Kindy Teaching Hospital); and the published dataset from 2020 contained 1000 patients (222 females and 778 males), with an age range of between 20-79 years old. Three classes are included diabetic (446), non-diabetic (52) and predict-diabetic (502). Eight medical tests were performed on each patient. The overall number of features were eight tests with three additional features (age, gender and BMI); these features are listed in Figure 1. The data was collected by asking the subjects questions about the age, weight and sex; this is a non-invasive method of collecting data. Additionally, a blood test was performed as additional data; this type of collecting data is invasive because the needles are used to take blood samples. A normalization process is performed using the zero score method to have all the features in the same range [14]. To simplify the classification process, a numerical value assigned to the outputs, as shown in Table 1.

2.2. Features selection method

Features selection is a step in classification process, and is usually used to either reduce the number of features or to select the most relevant features that could lead to an improved classification performance. Binary methods should be used in feature selection because the state of the feature is either selected (logic one) or not (logic zero); therefore, it is a binary decision. In this work, the Binary Dragonfly Algorithm was used as a pre-processing step to select the best features that will lead to high performance of classification [16].

In binary selection applications such as the features selection some applications (objective function), which does not accept the features being all zeros (no feature is selected), there is a probability to

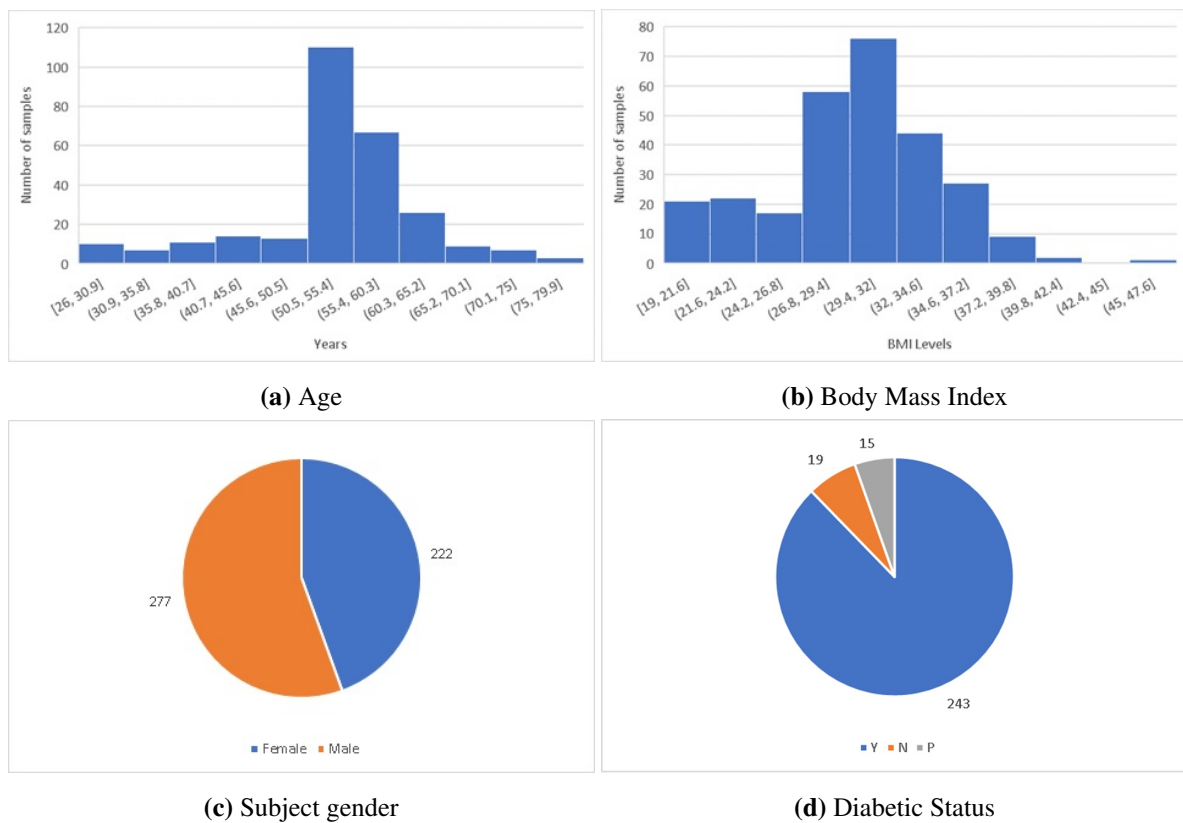


Figure 1. Noninvasive features and output of the LMCH dataset.

generate such an individual in the binary algorithm. This probability (P) increased with a decreasing number of features, as shown in Eq 2.1, where n is number of features which will break the selection algorithm.

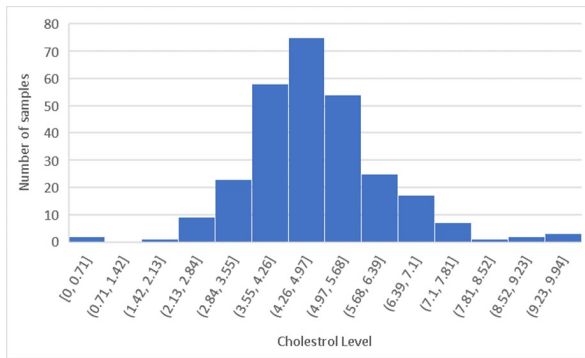
$$P\% = \frac{100}{2^n} \quad (2.1)$$

Binary dragonfly algorithm

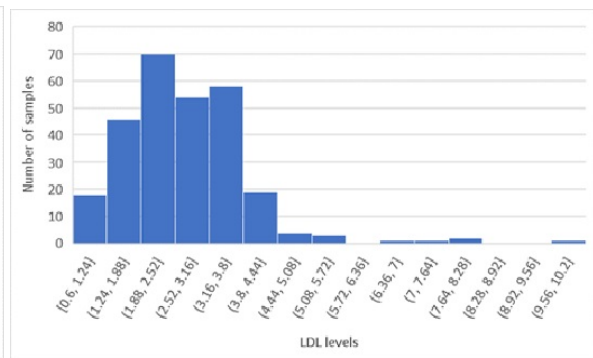
The dragon fly is a small predator that hunts small insects. Dragonflies have two behaviors: static and dynamic. These certain principles: separation (S), which is how the individuals avoid collisions with others; alignment (A), which is the matching velocity; and cohesion (C), which is the individual tendency towards the center of the neighborhood. to survive, usually focusing on finding food and avoiding enemies [15].

The Binary Dragonfly Algorithms [16] is shown in Figure 3. The first step is generating the first population; then initialize, the step vector is initialized randomly. This algorithm is a binary algorithm ,so all the individuals should only be a collection of zeros and ones. Dragonflies are evaluated using the objective function; in our case, the objective function is a classification error, so the individuals that represent the group of selected features will be applied to the classifier, as shown in Figure 6. After dragonfly evaluation, the food attraction and escaping from enemy are updated using Eqs. 2.2 and 2.3, respectively.

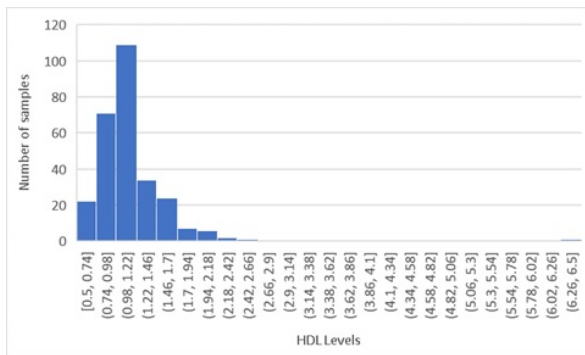
$$F_i = X^+ - X \quad (2.2)$$



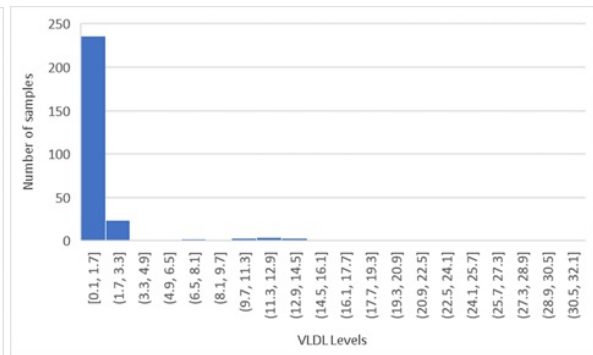
(a) Cholesterol Level



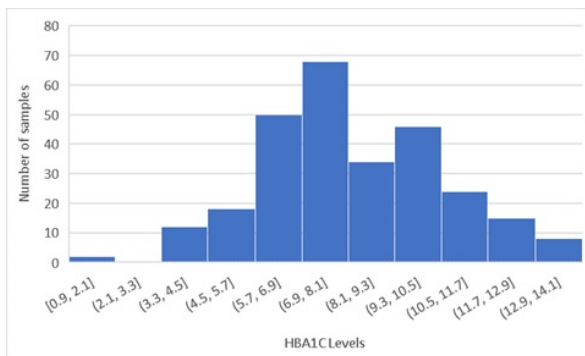
(b) LDL Cholesterol Level



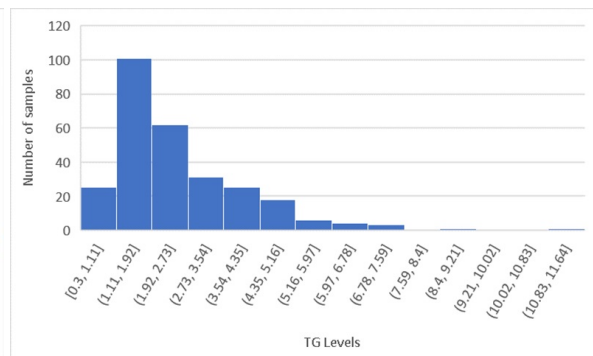
(c) HDL Cholesterol Level



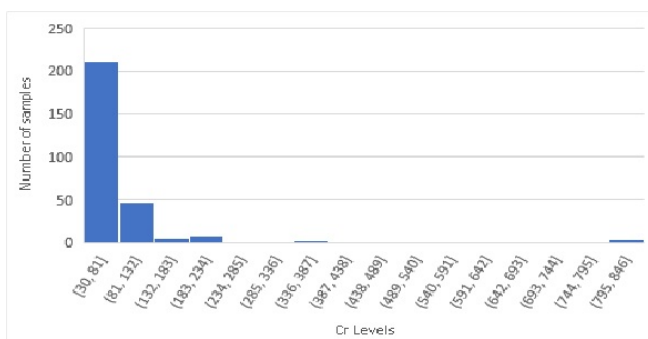
(d) VLDL Cholesterol Level



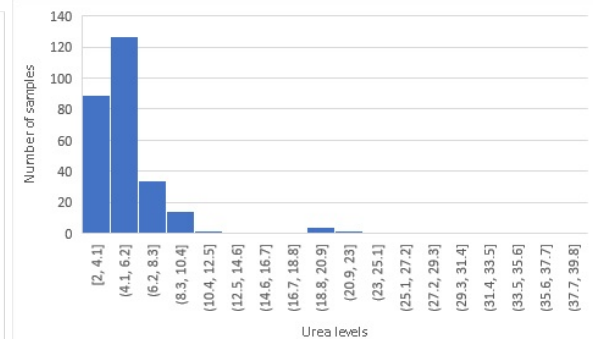
(e) HBA1C



(f) Triglycerides



(g) Creatinine Ratio



(h) Urea Level

Figure 2. Invasive features of the LMCH dataset.

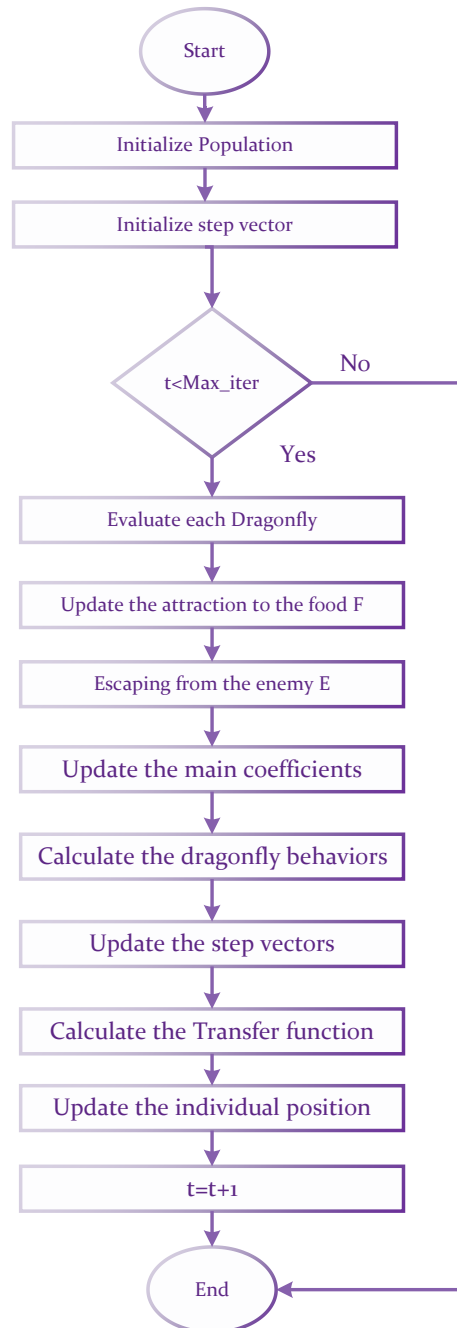


Figure 3. Binary Dragonfly Algorithm.

$$E_i = X^- + X \quad (2.3)$$

where X is the current individual, X^+ is the food position and X^- is the enemy position. The main coefficients (s, a, c, f, e and w) in Eq. 2.4 are updated using random values.

$$\Delta X_{t+1} = (sS_i + aAi + cCi + fFi + eE_i) + w\Delta X_t \quad (2.4)$$

After updating the coefficients, the behaviors of the dragonflies are calculated using the following equations:

$$A_i = \frac{\sum_{i=1}^N V_i}{N} \quad (2.5)$$

$$S_i = - \sum_{i=1}^N X - X_j \quad (2.6)$$

$$C_i = \frac{\sum_{i=1}^N X_i}{N} - X \quad (2.7)$$

where N is the neighborhood size, X_j is the j^{th} neighbor and V_j is the individual velocity in the j^{th} neighborhood. The step vectors ΔX_{t+1} are updated, and the transfer function is calculated by the probability of the changing position to keep the individuals in the binary range, as shown in Eqs. 2.8 and 2.9:

$$T(\Delta x) = \left| \frac{\Delta x}{\sqrt{\Delta x^2 + 1}} \right| \quad (2.8)$$

Finally, the individual position is updated using Eq. 2.8:

$$X_{t+1} = \begin{cases} -X_t, & x < T(\Delta x_{t+1}) \\ X_t, & x \geq T(\Delta x_{t+1}) \end{cases} \quad (2.9)$$

The algorithm is updated until the maximum number of iterations is reached.

2.3. Classification method

Classification is a process divide where the input features are divided into different groups according to a specific criteria. ML and neural networks achieve classification results in both binary and multi-classes.

Long-Short term memory

LSTM is a recurrent neural network that was introduced to overcome the problem of back-flow. This type of network has four vectors: forget activation vector f , input activation vector i , output activation vector o and the cell candidate vector $\widehat{c}(k)$. Figure 4 shows the structure of an individual cell of this network, where x_k, h_k and c_k are the input sequence, the cell output vector and the cell state vector to the network cell, respectively. These inputs will be applied to the sigma and hyperbolic tangent (tanh) functions; sigma is used as an activation function, while the tanh function is used to compute the cells state $\widehat{c}(k)$, as shown in the following equations (2.10-2.15):

$$c(k) = f(k) \odot c(k-1) + i(k) \odot \widehat{c}(k) \quad (2.10)$$

$$\widehat{c}(k) = \tanh[W_c x(k) + U_c h(k-1) b_c] \quad (2.11)$$

$$f(k) = \text{sigma}[W_f x(k) + U_f h(k-1) b_f] \quad (2.12)$$

$$i(k) = \text{sigma}[W_i x(k) + U_i h(k-1) b_i] \quad (2.13)$$

$$h(k) = f(k) \odot \tanh[c(k)] \quad (2.14)$$

$$o(k) = \text{sigma}[W_o x(k) + U_o h(k-1) b_o] \quad (2.15)$$

where the inputs weights, recurrent weights and bias are W_{input} , U_{input} and b_{input} , respectively. All of the multiplications are element by element, as denoted by the symbol \odot .

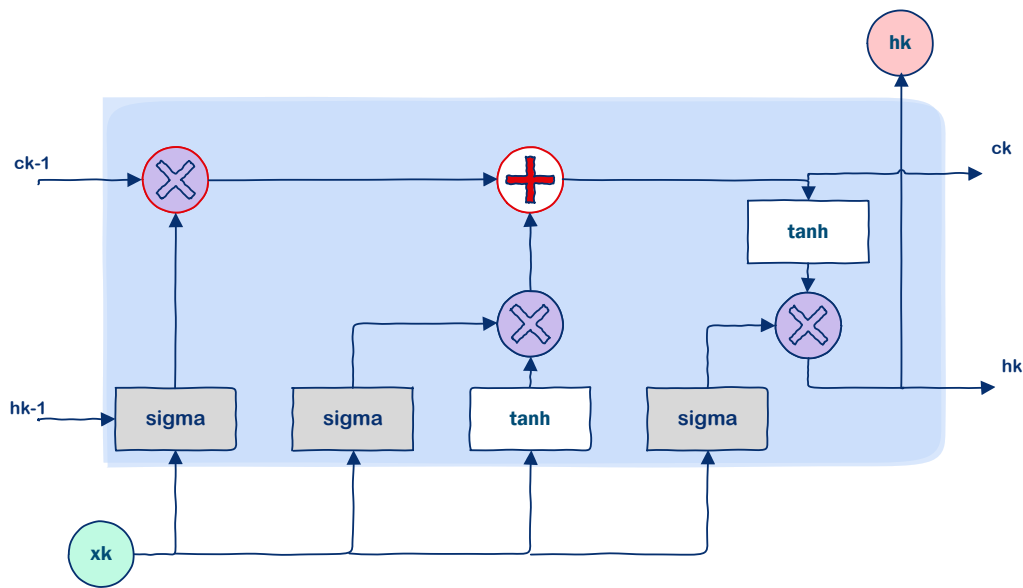


Figure 4. Structure of LSTM NN cell.

In this work, the network has six layers with a sigma activation function (input, two hidden layers, fully connected, softmax and classification) are, as shown in Figure 5. The training parameters are contain 100 cells in each hidden layer, 500 training epochs and 27 samples in the minibach segment [18].

2.4. Evaluation metrics

The classification method should be evaluated using some metrics such as accuracy, precision, recall and F-score [19]. The first metric is precision, which is the ratio between the correct classified outputs to and the summation of all samples that is classified as correct, whether they are classified correctly or not.

$$\text{Precision}(\%) = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} * 100(\%) \quad (2.16)$$

The second measurement is Recall or the sensitivity of the classifier to the class, which is the ratio between the correct classified samples to and the summation of all really correct classified samples for

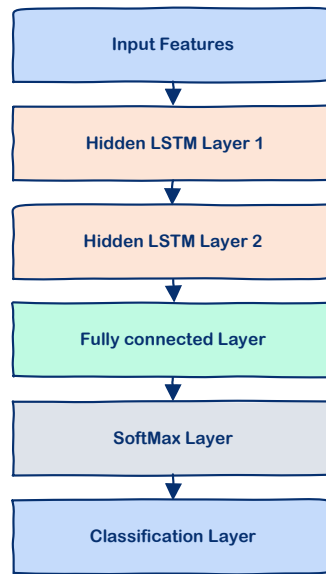


Figure 5. Architecture of LSTM NN.

all classes.

$$Recall(\%) = \frac{True\ positive}{True\ positive + False\ negative} * 100(\%) \quad (2.17)$$

These measurements are calculated for each class in a multi-class problem, to find out the total measurement of each one. The weighted average(WA) formulas are used, as shown in the Eqs. 2.18 and 2.19:

$$WAPrecision = \sum_{i=1}^N Actual\ Positives\ class_i * Precision_i \quad (2.18)$$

$$WARecall = \sum_{i=1}^N Actual\ Positives\ class_i * Recall_i \quad (2.19)$$

The F1- score is twice the ratio between the product of Precision and Recall and the summation of them.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (2.20)$$

The final and most common measurement is Accuracy, which is the ratio between the really correct classified samples and the overall number of samples.

$$Accuracy = \frac{True\ positive + True\ negative}{Total\ number\ of\ samples} \quad (2.21)$$

The classification error is another measurement of the performance, which is some form of accuracy, as shown in Eq. 2.21; this metric is used as the objective function during the features selection process.

$$Error = 1 - \frac{Accuracy(\%)}{100} \quad (2.22)$$

All of these metrics can provide a good judgment about the performance of the classification method that used.

3. The proposed method

The diabetes prediction system consists of two main phases: the first phase is the training phase, which consists of data collection, normalization, features selection and validation; the second method is the testing phase, which uses the best features found by the features selection method and applies them to the LSTM classifier to find the final class of the testing subject.

In this work, the probability of all features to be zeros is about 0.05%, although; this value is not very high but, it will break the selection algorithm loops if it occurs. The number of iterations is 500 with 20 individuals that means there is 10000 tests of the objective function through these tests. There is a probability that five individuals to could be zero and the objective function will not accept this individual.

This situation is solved in [17] by adding a step to the algorithm before applying the individual to the objective function to override this breaking by considering; all features are selected. The calculation is continued if the solution is not evaluated previously or if the algorithm is not generated before it could lead to a local minimum.

In this work, another solution is proposed using the previous individual instead of the current one; this solution will insure that the current solution will not break the calculation loop and is evaluated prior. The optimization algorithm used in this work is the Binary Dragonfly Algorithm. This work used the LSTM to classify the features into three classes: diabetic, pre-diabetic and non diabetic. The overall system is shown in Figure 6.

4. Results and discussion

In this paper, a classification method is used to further understand diabetes cases in Iraq. The data are divided into the training group, which is 75% of the overall dataset, and the testing group, which is 25% of the overall dataset. This dataset has three classes, each of which has a numerical value, as shown in Table 1. The input features that were used for the classification are the blood tests and some non medical features, the outputs of which are shown in Figures 1 and 2. To find out the best collection and the most relevant features, the Dragonfly Algorithm (DFA) was used; however, the selection problem here is binary and not an integer, so the binary version of the algorithm is used. During the execution of the selection algorithm, the DFA could reach a value of all features not selected (zero individual). This situation will break the evaluation loop because the objective function here is the classification error. The classifier could not work with without features, so the previous non-zero individual will replace the zero one to override this problem.

After the features selection step, the best collection of features in the training phase are used during classification process for the testing phase. First, the classifier is tested with all features in the absence of either features selection or reduction methods. After that, the selected features are used to evaluate the method using the testing dataset. Table 2 shows a comparison among different common classification methods. Table 3 shows the comparison between using all features that are shown in Figures 1 and 2 and the selected features by BDA. It is obvious that the BDA features gives a better performance than all features because there are irrelevant features that will reduce the accuracy. The confusion matrix shown in Table 4 highlights that there are only four missclassification cases and all the rest are correct.

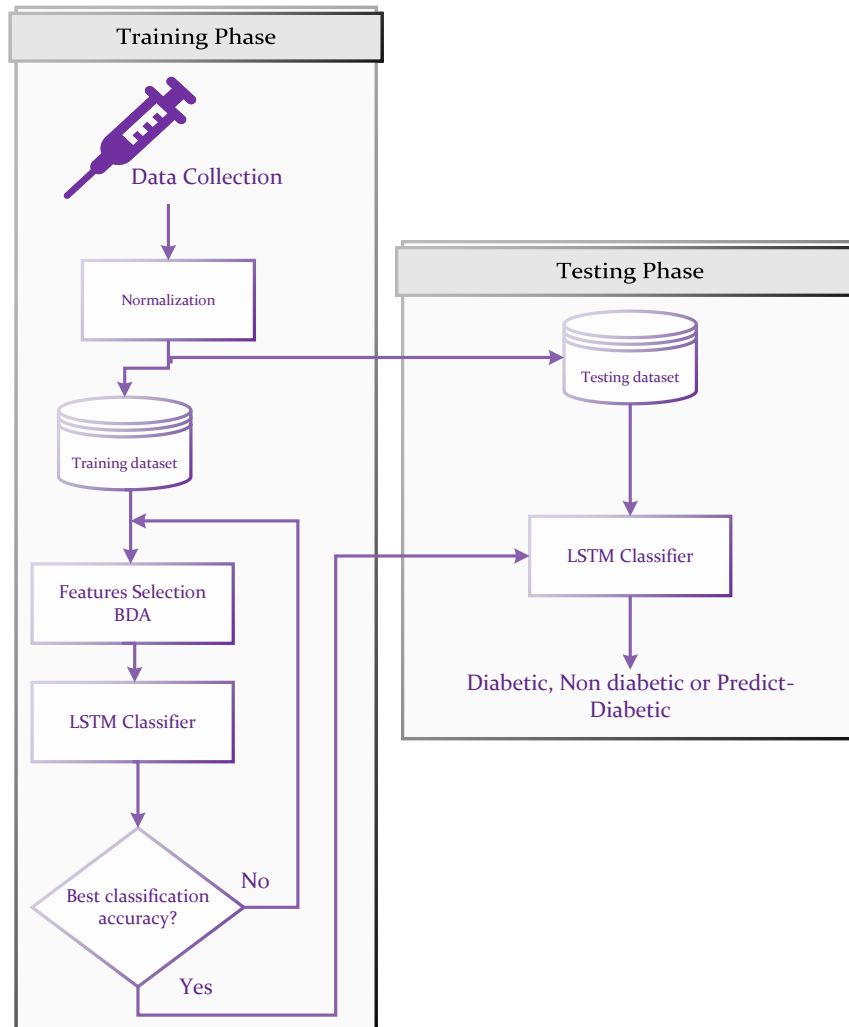


Figure 6. Diabetes prediction system using BDA and LSTM classifier.

Table 5 shows a comparison with other works, although the works applied the methods on binary classes data; our method shows the superiority on them by all performance metrics. The only work that has a better performance is CNN-BiLSTM [6]. In this work, they did not consider other classes in the classification process, which were diabetic and non-diabetic for only women. Alternatively, our work classified the inputs into three classes, diabetic, non-diabetic cases and pre-diabetic, which is a serious condition that will lead to diabetes if no action has been taken, such as diet or medications. Besides that, the performances are very close to each other, and the accuracy difference is only 85% between the two methods.

Table 2. Accuracy comparison using different methods.

Method	PIMA Dataset	LMCH Dataset
Decision Tree	79.2%	96%
Naive Bayse	78.1%	92.7%
SVM	82%	91.1%
LSTM	87.26	94.8%
Ensamble method	80.7%	96%

Table 3. Classification accuracy with different number of features.

All Features	BDA Features
94.8%	98.008%

Table 4. Confusion matrix of BDA LSTM classifier.

	Target Non-Diabetic	Target Diabetic	Target Pre-diabetic
Output Non-diabetic	35	4	0
Output Diabetic	0	195	0
Output Pre-diabetic	0	0	17

Table 5. Performance comparison with different works.

Method	Accuracy%	F1-Score%	Recall%	Precision%
PCA-Random Forest [7]	79.57	85.17	80.1	81.33
PCA-RF[11]	80	-	-	-
SVM[8]	83	70	63	79
LSTM[10]	87.26	-	85.1	86.6
LGML[9]	95.2	-	-	-
FS-MVI-O2GDNN[13]	97.248	97.255	97.245	97.342
DL[12]	98.08	96.81	98.46	95.22
Pipeline method[5]	98.8	-	-	-
CNN-BiLSTM [6]	98.85	96	94	98
FS-O2GDNN[13]	97.33	97.27	97.33	97.28
BDA-LSTM	98.008	97.35	99.16	95.83

5. Conclusions

The early detection of diabetes is a crucial situation because either a delay in treating the patient or taking medical actions could lead to very serious complications that affect the health of the patient. This work proposed a classification method that could detect three classes for the tested person: diabetic, non diabetic and pre diabetic. It can give an indicator to the patient to have a professional medical care after some blood tests. This work used LSTM as the multi-class classifier and BDA was used to reduce the number of tests that the subject needed to take. The selection process will reduce the number of features that are used for classification. In real life application, the selection process with a high classification performance will decrease the number of blood tests that should be taken to have an early prediction of diabetes. This will lead to a reduction in the cost of the periodic tests and check-ups. In many aspects, the method is superior compared to other works in the same field, such as the number of classes and the method performance, which they are 3% and 98%, respectively. A limitation of this work is that it used only one dataset with a limited number of features. To override this limitation , multiple datasets will be used in future work.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare that there is no conflict of interests.

References

1. Iraq diabetes report 2000-2045. Available from: <https://www.diabetesatlas.org/data/en/country/96/iq.html>
2. Tigga NP, Garg S (2023) Speech Emotion Recognition for multiclass classification using Hybrid CNN-LSTM. *International Journal of Microsystems and Iot* 1: 9–17. <https://doi.org/10.5281/zenodo.8158288>
3. Jaber HA, Rashid MT (2019) HD-sEMG gestures recognition by SVM classifier for controlling prosthesis. *Iraqi Journal of Computers, Communications, Control and System Engineering (IJCCCE)* 19: 10–19. <https://doi.org/10.33103/uot.ijccce.19.1.2>
4. Abgeena A, Garg S (2023) A novel convolution bi-directional gated recurrent unit neural network for emotion recognition in multichannel electroencephalogram signals. *Technol Health Care* 31: 1215–1234. <https://doi.org/10.3233/THC-220458>
5. Mujumdar A, Vaidehi V (2019) Diabetes prediction using machine learning algorithms. *Procedia Computer Science* 165: 292–299. <https://doi.org/10.1016/j.procs.2020.01.047>
6. Madan P, Singh V, Chaudhari V, Albagory Y, Dumka A, Singh R, et al. (2022) An optimization-based diabetes prediction model using cnn and bi-directional lstm in real-time environment. *Applied Sciences* 12: 3989. <https://doi.org/10.3390/app12083989>

7. Chang V, Bailey J, Xu QA, Sun Z (2023) Pima indians diabetes mellitus classification based on machine learning (ml) algorithms. *Neural Computing and Applications* 36: 16157–16173. <https://doi.org/10.1007/s00521-022-07049-z>
8. Noori NA, Yassin AA (2021) A comparative analysis for diabetic prediction based on machine learning techniques. *Journal of Basrah Researches (Sciences)* 47.
9. Ahamed BS, Arya MS, Nancy AO (2022) Prediction of type-2 diabetes mellitus disease using machine learning classifiers and techniques. *Frontiers in Computer Science* 4: 835242. <https://doi.org/10.1155/2022/9220560>
10. Butt UM, Letchmunan S, Ali M, Hassan FH, Baqir A, Sherazi H (2021) Machine learning based diabetes classification and prediction for healthcare applications. *Journal of healthcare engineering* 2021. <https://doi.org/10.1155/2021/9930985>
11. Zou Q, Qu K, Luo Y, Yin D, Ju Y, Tang H (2018) Predicting diabetes mellitus with machine learning techniques. *Frontiers in genetics* 9: 515. <https://doi.org/10.3389/fgene.2018.00515>
12. Naz H, Ahuja S (2020) Deep learning approach for diabetes prediction using pima indian dataset. *Journal of Diabetes and Metabolic Disorders* 19: 391–403. <https://doi.org/10.1007/s40200-020-00520-5>
13. Olisah CC, Smith L, Smith M (2022) Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective. *Comput Meth Prog Bio* 220: 106773. <https://doi.org/10.1016/j.cmpb.2022.106773>
14. Rashid A (2020) *Diabetes dataset*, Mendeley Data.
15. Mirjalili S (2015) Dragonfly algorithm: A new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. *Neural Comput Appl* 27: 1053–1073. <https://doi.org/10.1007/s00521-015-1920-1>
16. Mafarja MM, Eleyan D, Jaber I, Hammouri A, Mirjalili S (2017) Binary Dragonfly Algorithm for Feature Selection. *2017 International Conference on New Trends in Computing Sciences (ICTCS)*, 12–17.
17. Alhakeem ZM, Ali RS (2019) Fast channel selection method using crow search algorithm. *Proceedings of the International Conference on Information and Communication Technology*, 210–214. <https://doi.org/10.1145/3321289.3321309>
18. Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. *Neural Computation* 9: 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
19. Dalianis H (2018) *Evaluation Metrics and Evaluation*, Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-78503-5_6



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)