A Chaotic Levy Flights Bat Algorithm for Diagnosing Diabetes Mellitus

Omar S. Soliman
Faculty of Computers and Information,
Cairo University
5 Ahmed Zewail Street, Orman, Giza, Egypt

Eman Abo ElHamd
Faculty of Computers and Information,
Cairo University
5 Ahmed Zewail Street, Orman, Giza, Egypt

ABSTRACT
Bat algorithm is a meta-heuristic algorithm that is based on the echolocation behavior of bats. The searching behavior of the algorithm depends on generating uniformly distributed random walks in the search space. Hence, it may suffer from being tapped in local optima. In this paper, a classification using Bat inspired algorithm with chaotic levy flight variable is proposed. The chaotic variable has set of characteristics that enable it to enrich the searching behavior and prevent the Bat algorithm from being trapped into local optimum. The chaotic sequence and a chaotic Levy flight are incorporated with Bat algorithm for many purposes including, efficiently generating new solutions via randomization, increase the diversity of the solutions, avoid trapping in a local optimum and increase the chances of finding global optimum solution. The proposed algorithm aims to help physicians in early diagnosis and treatment of Diabetes Mellitus (DM). DM is a major health problem in both industrial and developing countries and its incidence is rising. The proposed algorithm is applied on Pima Indians Diabetes data set from UCI repository of machine learning data bases. The experimental results prove the superiority of the proposed algorithm over the traditional Bat algorithm as well as different classifiers which were implemented on the same data set and within the same environment.

Keywords
Bat Inspired Algorithm (BIA), Levy Flight, Chaotic Variable, Diabetes Mellitus (DM).

1. INTRODUCTION
Diabetes Mellitus is a group of metabolic diseases in which a person has high blood sugar that produces the classical symptoms of polyuria (frequent urination), polydipsia (increased thirst) and polyphagia (increased hunger). This high blood sugar might be either because the pancreas does not produce enough insulin, or because cells don't respond to the produced insulin. There are 3 major types of DM, "Type-I DM", also called "Insulin Dependent Diabetes Mellitus", which results from the body's failure to produce insulin. The second type of DM is called "Type-II DM" or "Non-Insulin Dependent Diabetes Mellitus" which results from insulin resistance as the cells fail to use insulin properly. Sometimes it is combined with an absolute insulin deficiency. The third type of DM is "Gestational Diabetes" which occurs when pregnant women without a previous diagnosis of diabetes develop a high blood

glucose level; it may precede development of Type-I DM [1], [2].

Although both types [I & II] are chronic conditions that cannot be cured, Early Diagnosis has a great effect in controlling them. Gestational DM usually resolves after delivery [2]. Untreated DM causes many complications either acute or series. Acute complications include diabetic ketoacidosis and non-ketotic hyperosmolarcoma. While, series long term complications include cardiocascular disease, chronic renal failure, and diabetic retinopathy. Since the cells can't take in the glucose, it builds up in the blood. High levels of blood glucose can damage the tiny blood vessels in the kidneys, heart, eyes or nervous system. That's why diabetes can eventually cause heart disease, stroke, kidney disease, blindness and nerve damage to nerves in the feet (especially if left untreated) [3], [4]. Early Diagnosis and adequate treatment of DM is very important, as well as blood pressure control and lifestyle factors such as stopping smoking and maintaining a healthy body weight. In light of this introduction, the aim of this paper is to propose a classification algorithm using (chaotic levy flight Bat algorithm) that helps physicians in early diagnosis and treatment of DM patients.

Bat-inspired algorithm is a Bio-Inspired optimization algorithm developed by Xin-She Yang in 2010 [5]. It's based on the echolocation behavior of micro-bats with varying pulse rates of emission and loudness. The efficiency of Bat algorithm stems from making balance between local intensive exploitation and global diverse exploration. Although it's proved to be better than many global search algorithms such as Particle Swarm Optimization algorithm, the algorithm may stuck in a local optimum solution while depending of random walk [27]. Hence, increasing the diversity of the solutions by levy flight walks with chaotic variable could increase the diversification of the algorithm and preventing it from trapping in a local optimum. The rest of this paper is organized as follows; section 2 describes the problem background and related work. The proposed algorithm is introduced in section 3; while experimental results are presented in section 4. The last section is devoted to the conclusion and further research.

2. BACKGROUND AND RELATED WORK
Diabetes Mellitus (DM) is a worldwide health problem that preoccupies many researchers [1]. Different classification algorithms have been applied on this area trying to classify the patients or predict their future state. This section introduces some of these works.

Artificial Neural Network (ANN) is one of the most effective classifiers that are widely used in the diagnosis of DM. A brief review and discussion of the philosophy, capabilities, and limitations of ANN in medical diagnosis through selected examples including DM was introduced in [10]. A hybrid binary classification model using the basic concepts of soft computing and ANN was proposed in [11]. While a multilayer perception NN and a Conditional Logistic Regression were used to predict albuminuria in type II DM through work presented in [12]. In [13], a model using ANN with RBF kernel and one hidden layer was introduced. The Artificial Meta-Plasticity on Multilayer Perceptron was used as prediction model for diabetes with classification accuracy of 89.93% [14]. ANN and Multivariate Logistic Regression model was proposed in [15].

On the other hand, a proposed machine learning algorithm termed "Mixture of Expert" was used for the determination of a patient's diabetic state in [16]. A survey of more than one supervised and unsupervised algorithms was introduced in [17]. Support Vector Machine (SVM) is one of the most important classifiers in this area that is proved to have effective results. SVM technique was proposed for classification of DM patients. The results showed a sensitivity of 99.45% for the classifier and specificity of 100% [18]. A robust version of SVM based on Value-at-Risk measure referred to as VaR-SVM was proposed in [19].

Hybrid algorithms are also used for the purpose of DM classification and proved to have robust results. In [20] a classification algorithm based on Fuzzy Systems, Evolutionary Algorithms and ANN was proposed. A study in [21] that has a main outcome measure of age specific mortality rates due to cardiovascular disease and all causes was presented. A hybrid model that integrates Genetic Algorithm and Back Propagation Network was proposed in [22]. Also, a hybrid binary classification model was proposed for type II DM classification, based on the basic concepts of soft computing and artificial intelligence techniques [23].

2.1 Meta-Heuristic Bat Inspired Technique

Bat-inspired technique is a Bio-Inspired, meta-heuristic technique developed by Xin-She Yang [5]. Bat algorithm is based on the echolocation behavior of micro-bats with varying pulse rates of emission and loudness. The main idea of bats is as follows; Bat sends signal with loudness of frequency 20 kHz to 200 kHz. This signal returns back to the bat as echo signal after striking the object as demonstrated in Figure 1. The received echo-signal allows the bat to calculate the distance between its current location and the surrounding objects; then the bat flies towards the minimum distance object [6].

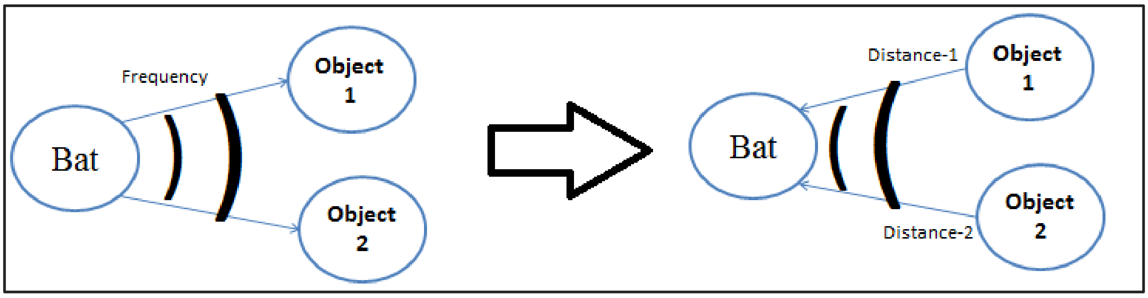

Figure 1: Bat sends signal of frequency f and receives echo-signal

In Bio-Inspired bat algorithm, each virtual bat flies randomly with a velocity $\overleftarrow{v_k}$ at position $\overleftarrow{x_k}$. Each position represents a solution for the classification problem. Every Bat has a varying frequency f_k (or wavelength λ_k) and loudness A_k. When Bat searches and finds its prey, it updates its frequency, loudness and pulse emission rate. Bat algorithm is based on "Exploitation" by using local random walk as the selection for the best solution continue until reaching stopping criteria [7]. The basic idea behind the Bat Algorithm (Algorithm-1) is that a population of n bats (possible solutions) in d dimensions uses echolocation to sense distance and fly randomly through a search space, updating their positions $\overleftarrow{x_i}$ and velocities $\overleftarrow{v_k}$. Each solution $\overleftarrow{x_k} = (x_1, x_2 \dots, x_d)^T$ is evaluated by a fitness function $f(\overleftarrow{x_k}), k = 1,2 \dots n$. Bat's loudness and pulse rate is guided by two main parameters, Loudness Decay Factor (α) and Pulse

Increase Factor (γ). The goal of each bat is to find prey (best solution). To achieve this goal, each bat iteratively updates pulse rate and loudness to make balance between exploitation and exploration respectively. Bat decreases the loudness once finding its prey/solution (to avoid losing the prey), and it increases the rate of pulse emission in order to raise the attack accuracy. The frequencies f_k, solutions $\overleftarrow{x_k}$ and velocities $\overleftarrow{v_k}$ are calculated at time step t using Eqs.(1),(2),(3) respectively.

$$f_k = f_{min} + (f_{max} - f_{min}).\beta \qquad (1)$$

$$\overrightarrow{v_k}^t = \overrightarrow{v_k}^{t-1} + (\overrightarrow{x_k}^t - \overrightarrow{x_*}).f_k \qquad (2)$$

$$\overrightarrow{x_k}^t = \overrightarrow{x_k}^{t-1} + \overrightarrow{v_k}^t \qquad (3)$$

Where $\beta \in [0,1]$ is a random vector follows uniform distribution. $\overrightarrow{x_*}$ is the current global best location (solution) that is located after comparing all the solutions among all the n bats. Once a solution is selected among the current best solutions, random walk is required in a local search process. The new solution for each bat is generated locally using Eq.(4)

$$\overrightarrow{x_{new}} = \overrightarrow{x_{old}} + \varepsilon A^t \qquad (4)$$

Where $\varepsilon \epsilon [-1,1]$ is a random number, A^t is the average loudness of all the bats at this time step. For each bat, the loudness A_k and the pulse rate r_k are updated using Eqs.(5),(6) respectively

$$A_k{}^{t+1} = \propto A_k{}^t \qquad (5)$$

$$r_k{}^{t+1} = r_k{}^0.(1 - e^{-\gamma t}) \qquad (6)$$

Where $0 < \alpha < 1$, while $\gamma > 0$. In fact each bat should have different values of loudness and pulse emission rate and this is achieved by randomization.

Algorithm-1: Bat Inspired Algorithm

1. Initialize the bat population $\overrightarrow{x_i}(i = 1, .., n)$ and the velocities $\overrightarrow{v_i}$ with the objective function $f(\vec{x})$, $\vec{x} = (x_1, \dots x_d)^T$
2. Determine the pulse frequency f_i for each $\overrightarrow{x_i}$
3. Initialize pulse rates r_i and loudness A_i
4. **While** (*Stopping criteria not met*)
5. Adjust frequencies, update velocities and locations (solutions) to generate new solutions using Eqs.(1),(2),(3)

6. **if** $(rand > r_i)$
7. Select a solution among the best solutions
8. Generate a local solution around the selected best solutions
9. **end if**
10. Generate a new solution by flying randomly using Eq.(4)
11. **if** $(rand < A_i \,\&\, f(\overrightarrow{x_i}) < f(\overrightarrow{x_*})$
12. Accept the new solutions
13. Use Eqs.(5),(6) to increase r_i and reduce A_i
14. **end if**
15. Rank the bats and find the current best $\overrightarrow{x_*}$
16. **end while**
17. Output the results

The initial loudness and pulse rate are $A_k{}^0 \epsilon\, [1,2]$ and $r_k{}^0 \epsilon\, [0,1]$ respectively. It's important to highlight that the values of A_k and r_k will be updated only if the solutions are improved; as this is an indicator that the bats are moving toward the optimal solution [7]. In fact $A_k{}^t \to 0,\ r_k{}^t \to r_k{}^0, as\ t \to \infty$. Algotithm-1 states the pseudo code of Bat algorithm with random walk. It's important to mention that Bat algorithm has some similarities to PSO algorithm. Actually, Bat Algorithm can be considered as a balanced combination of the standard PSO as a global search optimization technique and the intensive local search controlled by the loudness and pulse rate. In Bat algorithm, f_k essentially controls the pace and range of the movement of the swarming particles [10]. If $A_k = 0$ & $r_k = 1$, Bat algorithm becomes the standard PSO algorithm [11].

3. PROPOSED ALGORITHM

The main goal of meta-heuristic algorithms is to maintain balance between diversification and intensification. Too little exploration and too much exploitation may cause the system to be trapped in local optima, which makes it very difficult or even impossible to find the global optimum. On the other hand, with too much exploration and too little exploitation, it may be difficult for the system to converge and thus slows down the overall search performance [7]. Hence, maintaining a certain degree of diversity is proved to help in obtaining this balance and avoid the tradeoff between exploration and exploitation. The uniform random movements of bats limit achieving this balance. For this purpose, a chaotic variable is incorporated with bat algorithm (Instead of uniform random variable). Actually this combination has been introduced in [7] and is implemented in this work on Pima Indians Diabetes dataset. The chaotic variable has many characteristics including ergodicity, pseudo-randomness and irregularity [7]. These characteristics enrich the searching behavior and avoid trapping in local optimum solution. In general Levy flight is a random walk whose step length is drawn from the Levy distribution. The levy flight with chaotic variable depends on generating chaotic sequence C_s that is calculated using Eq.(7)

$$C_s(t+1) = 4 \times C_s(t) \times \big(1 - C_s(t)\big),\ \ 0 \le C_s(t) \le 1 \qquad (7)$$

Levy distribution has infinite variance that occasionally allows long steps far from the neighborhood of the previous sample. Hence, setting a smaller neighborhood range and making small jumps, is helpful for finding optimum solution in the region (Exploitation). However, large jumps are needed to avoid local solutions (Exploration). In fact, there is no perspective to specify these regions. There is no specific definition of the mean and the variance of chaotic sequences. Thus, a combination of a chaotic sequence and Levy random process may result in better answers. In the proposed algorithm, the new neighbor is generated using Eq.(8)

$$x_{new} = x_{old} + C_s \otimes \text{Levy}(\lambda) \qquad (8)$$

Where Levy distribution and a chaotic sequence are used to generate Levy (λ) and C_s respectively. The product $\otimes$ means entry-wise multiplications. Levy flights essentially provide a random walk, while their random steps are drawn from a Levy distribution for large steps, that has an infinite variance with an infinite mean as mentioned in Eq.(9)

$$Levy \sim u = t^{-\lambda}, (1 \le \lambda \le 3) \qquad (9)$$

For maintaining variability in the solutions, the chaotic sequence generates several neighborhoods of suboptimal solutions; Hence, preventing the search process from becoming premature. Ergodicity feature helps chaotic sequence in generating several neighborhoods of near-optimal solutions. The algorithm probably converges to a space in the search space where good solutions are denser. The new frequencies f_k , solutions $\overleftarrow{x_k}$ and velocities $\overleftarrow{v_k}$ are calculated at time step t using Eqs.(10),(11),(12) respectively.

$$f_k = f_{min} + (f_{max} - f_{min}).\, C_s \qquad (10)$$

$$\overrightarrow{v_k}^t = \overrightarrow{v_k}^{t-1} + \big(\overrightarrow{x_k}^t - \overrightarrow{x_g}^t\big).\, f_k + \big(\overrightarrow{x_k}^t - \overrightarrow{x_i}^t\big).\, f_k \qquad (11)$$

$$\overrightarrow{x_k}^t = \overrightarrow{x_k}^{t-1} + \overrightarrow{v_k}^t \qquad (12)$$

Where $\overrightarrow{x_g}^t$ is the global best solution of all bats, while, $\overrightarrow{x_i}^t$ is the local best solution of each bat. Each bat follows the best hunting position founded by not only taking all bats into consideration, but also its own preference when searching for food (Prey) [7]. The main steps of the proposed algorithm are stated in the pseudo code in Algorithm-2.

Algorithm-2: Chaotic Levy Flight Bat Algorithm

1. Initialize the bat population $\overrightarrow{x_k}(i = 1,..,n)$ and the velocities $\overrightarrow{v_k}$ with the objective function $f(\vec{x}),\ \vec{x} = (x_1, \ldots x_d)^T$
2. Determine pulse frequency f_k for each $\overrightarrow{x_k}$
3. Initialize pulse rates r_k and loudness A_k
4. **while** $(Stopping\ criteria\ not\ met)$
5. Generate the chaotic sequence C_s using Eq.(7)
6. Compute $f(\overrightarrow{x_k})$
7. Find the current best position $\overrightarrow{x_*}$
8. Adjust frequencies, update velocities and locations (solutions) to generate new solutions using Eqs.(10),(11),(12)
9. **if** $(C_s > r_k)$
10. Select a solution among the best solutions
11. Generate a local solution around the selected best solutions by flying randomly via chaotic Levy flights using Eq.(8)
12. **end if**

13. Generate a new solution by flying randomly via chaotic Levy flights using Eq.(8)
14. **if** $(C_s < A_k \ \& \ f(\overrightarrow{x_k}) < f(\overrightarrow{x_*}))$
15. Accept the new solutions
16. Use Eqs.(5),(6) to increase r_k and reduce A_k
17. **end if**
18. Rank the bats and find the current best $\overrightarrow{x_*}$
19. **end while**
20. Output the results

3.1 Pima Indians Diabetes Dataset

Pima Indians Diabetes dataset from UCI repository of machine learning data bases contains a total of 768 cases and 8 features as shown in Table 1. In addition to the class label ('1' Healthy and '0' means Infected). Among those 768 cases, there are 500 healthy cases and 268 suffered from DM. All patients in this dataset are Pima-Indian women whose age is at least 21 years old [26].

The proposed algorithm, chaotic levy flight bat algorithm is applied on Pima Indians diabetes dataset with its 8 features. It's important to mention that no feature reduction is made as all of them are proved to be statistically significant and directly affect the class determination. This dataset contains 'zero' values in cells where it's biologically impossible, such as the Blood Pressure attribute. These cells are considered to be missing and are filled using Local Mean algorithm, Algoritm-3; where the empty cell is filled by the average of the n previous cells.

Table 1. Pima Indians Diabetes Dataset

Feature Name	Mean	Standard Deviation
1. Number of pregnancy times.	3.8	3.4
2. Plasma glucose concentration a 2 h in an oral glucose tolerance test.	120.9	32.0
3. Diastolic Blood Pressure.	69.1	19.4
4. Triceps skin fold thickness.	20.5	16.0
5. 2-h Serum Insulin.	79.8	115.2
6. Body Mass Index.	32.0	7.9
7. Diabetes Pedigree Function.	0.5	0.3
8. Age.	33.2	11.8

Algorithm-3: Local Mean Method

1. Input the parameter (n).
2. Given an empty cell, that has an initial value (U = 0).
3. Calculate the average of its n previous cells using Eq.(13)

$$U_t = \frac{U_{n-1}+U_{n-2}\ldots U_0}{n} \quad (13)$$

4. Return number in the empty cell.

4. EXPERIMENTAL RESULTS

As mentioned earlier, the aim of this paper is to help physicians in early diagnosis and treatment of DM patients. For this purpose, chaotic levy flights bat inspired algorithm, Algorithm-2 is utilized to classify the cases into their appropriate class. The proposed algorithm is applied on Pima Indians diabetes dataset from UCI repository of machine learning data bases. The missing cells are filled using Local Mean algorithm, Algorithm-3. The implementation was done using MATLAB software. The parameters of the proposed algorithm are adjusted as stated in Table 2.

Table 2. Parameters' values of chaotic levy flight bat algorithm

Parameter	Symbol	Value
Minimum Frequency	f_{min}	0
Maximum Frequency	f_{max}	2
No. of Records	n	764
No. of Iterations	T_{max}	1000
Pulse Increasing Factor	γ	0.9
Cooling Factor	$\propto$	0.9
Minimum of levy distribution parameter	λ_{min}	1
Maximum of levy distribution parameter	λ_{max}	3

In order to evaluate the performance of the proposed algorithm, the classification accuracy was calculated using Eq.(14). The classification accuracy is based on 4 main terms (TP, TN, FP and FN). Table 3 sates the abbreviation and definition of each term [13].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

Table 3. Classification Accuracy Parameters

Abbreviation	Definition
True Positive (TP)	The proportion of positive cases that are correctly identified.
True Negative (TN)	The proportion of negative cases that are correctly identified.
False Positive (FP)	The proportion of negative cases that are incorrectly classified as positive.
False Negative (FN)	The proportion of positive cases that are incorrectly classified as negative.

The classification process is implemented using Cross Validation (CV) method. CV method means partitioning the data set into complementary subsets, performing the analysis on one subset (called the training set), and test the classification algorithm on the other subset (called the testing set) [24]. In this work, different types of CV are applied on different numbers of iterations. Table 4 states the classification accuracy resulted from each CV type, whereas Figure 2 demonstrates the average classification accuracy of each CV type over 4 different numbers of iterations. Figure 2 shows that the average classification accuracy in all the CV types reaches its maximum limits when running the proposed algorithm for 1000 iteration. After 1000 iteration the average classification accuracy decreases slowly and after 10000 iterations it's almost stable.

This is might be because the algorithm already reached a global optimal solution.

It's obvious from Table 3 that 50-50 CV type has the highest average classification accuracy value among all other CV types. Hence, it's considered to be the average classification accuracy of the proposed algorithm while comparing it with other classifiers as shown in Table 5 and Figure 3.

The proposed algorithm is compared with both traditional Bat Inspired algorithm (with uniform random variable) as well as some other classification algorithms that were applied on the same data set as illustrated in Table 5 as well as Figure 3.

Table 4. Classification accuracy of each CV type over 1000 iteration

CV Type	Average Classification Accuracy
70-30	97.93%
50-50	**98.65%**
80-20	97.778%

Table 5. Average classification accuracy of the proposed algorithm

CI Technique	Average Accuracy
ANN and AIS [16]	76.00%
MLP/BN/J48graft/JRip and FLR [28]	81.33%
MLP, SVM, KNN, QDA and LDA [11]	82.40%
GA and ANN [29]	84.71%
AMMLP [14]	89.93%
Bat Algorithm with random walk	88.36%
The Proposed Algorithm	**98.65%**

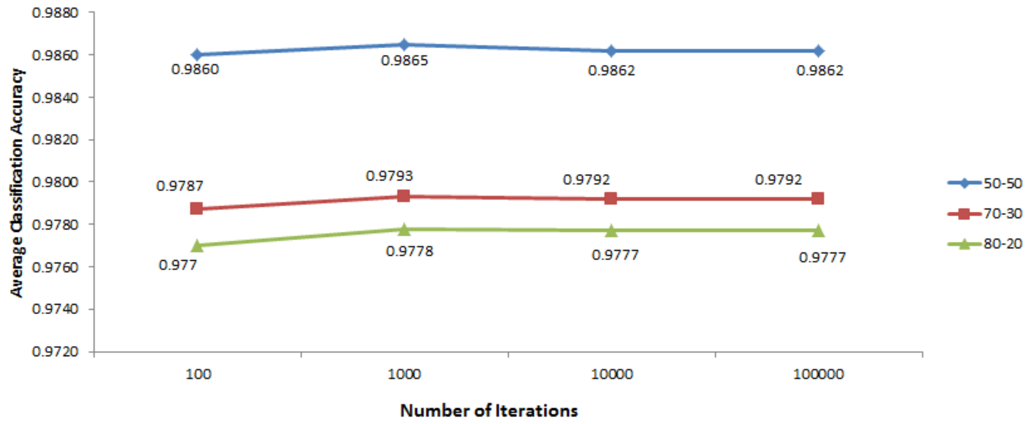

Figure 2: Average classification accuracy of the proposed algorithm for different CV types over different iterations

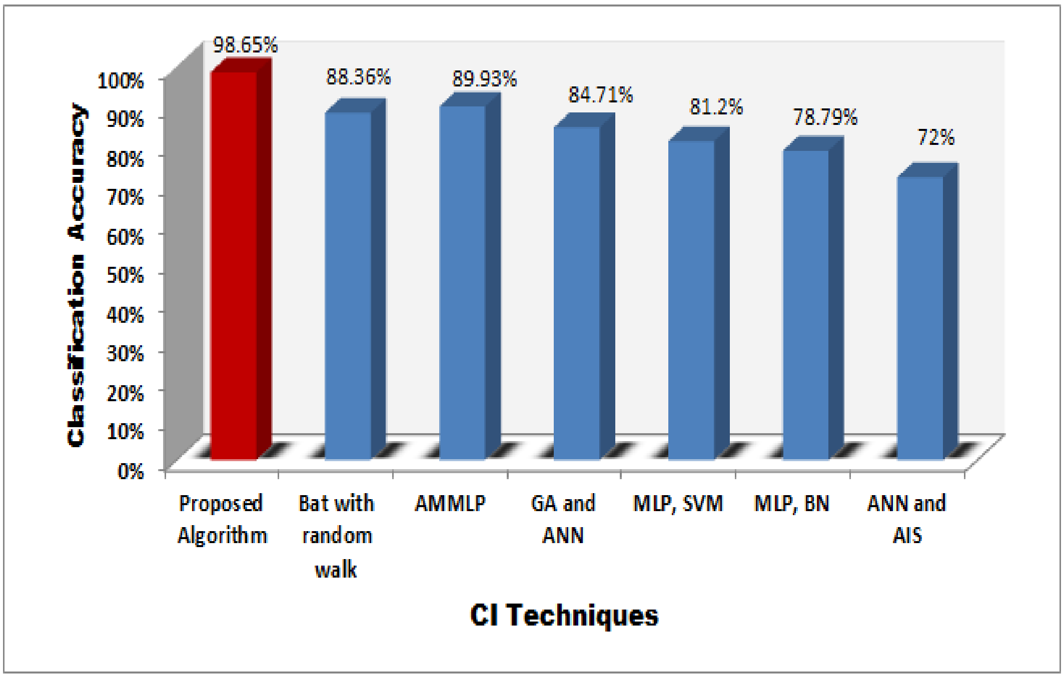

Figure 3: Average classification accuracy of the proposed algorithm VS different CI Techniques

5. CONCLUSION AND FUTURE WORK

This paper introduced classification algorithm that integrates Bat inspired algorithm with chaotic levy flights variable. The proposed algorithm may help physicians in early diagnosis and treatment of DM patients. It maintained a certain degree of diversity that helped in obtaining balance between exploration and exploitation. Chaotic levy flights sequence could enrich the searching behavior and avoid trapping in local optima. Chaotic Levy flights bat algorithm was applied on Pima Indians Diabetes dataset from UCI repository of machine learning data bases. Missing cells were filled using Local Mean method. Classification accuracy was calculated as a performance measurement. Different types of CV were applied on different numbers of iterations. The highest classification accuracy resulted from 50-50 CV type which was equals to 98.65%. The comparison between the proposed algorithm and traditional Bat Inspired algorithm with random walk as well as other classifiers proved the superiority of the proposed algorithm over all these classifiers. As a future work, the proposed algorithm might be tested on other chronic diseases. Also, other types of distribution variables may be used to calculate the step length of each bat instead of levy flight.

6. REFRENCES

[1] World health organization, diabetes center, fact sheet n312, www.who.int/mediacentre/factsheets/fs312/en (2011).

[2] American Diabetes Association. "Diagnosis and classification of diabetes mellitus." Diabetes care 31.Supplement 1 (2008): S55-S60.

[3] SETTING IN DIABETES MELLITUS CARE. Guidelines for improving the care of the older person with diabetes mellitus. Hypertension, 16:2, 2003.

[4] Jaakko Tuomilehto, Jaana Lindstrom, Johan G Eriksson, Timo T Valle, Helena Hamalainen, PirjoIlanne-Parikka, Sirkka Keinanen-Kiukaanniemi, Mauri Laakso, Anne Louheranta, Merja Rastas, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. New England Journal of Medicine, 344(18):1343-1350, 2001.

[5] Yang, Xin-She, Suash Deb, and Simon Fong. "Bat Algorithm is Better Than Intermittent Search Strategy." JOURNAL OF MULTIPLE-VALUED LOGIC AND SOFT COMPUTING 22.3 (2014): 223-237.

[6] Mishra, Sashikala, Kailash Shaw, and Debahuti Mishra. "A new meta-heuristic bat inspired classification approach for microarray data." Procedia Technology 4 (2012): 802-806.

[7] Lin, Jiann-Horng, et al. "A chaotic Levy flight bat algorithm for parameter estimation in nonlinear dynamic biological systems, J." Computer and Information Technology 2.2 (2012): 56-63.

[8] Yang, Xin-She. "Bat algorithm and cuckoo search: a tutorial." Artificial Intelligence, Evolutionary Computing and Metaheuristics. Springer Berlin Heidelberg, 2013. 421-434.

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

[10] Qeethara Kadhim Al-Shayea. Artificial neural networks in medical diagnosis. International Journal of Computer Science, 8(2):150–154, 2011.

[11] Mehdi Khashei, Saeede Eftekhari, and Jamshid Parvizian. Diagnosing diabetes type ii using a soft intelligent binary classification model. Review of Bioinformatics and Biometrics, 1(1), 2012.

[12] Afsaneh Morteza, Manouchehr Nakhjavani, Firouzeh Asgarani, Filipe LF Carvalho, Reza Karimi, and Alireza Esteghamati. Inconsistency in albuminuria predictors in type 2 diabetes: a comparison between neural network and conditional logistic regression. Translational Research, 2013.

[13] P Venkatesan and S Anitha. Application of a radial basis function neural network for diagnosis of diabetes mellitus. CURRENT SCIENCE-BANGALORE-, 91(9): 1195, 2006.

[14] Alexis Marcano-Cede˜no, Joaqu´ın Torres, and Diego Andina. A prediction model to diabetes using artificial metaplasticity. In New Challenges on Bioinspired Applications, pages 418–425. Springer, 2011.

[15] Chongjian Wang, Linlin Li, Ling Wang, Zhiguang Ping, Muanda Tsobo Flory, Gaoshuai Wang, Yuanlin Xi, and Wenjie Li. Evaluating the risk of type 2 diabetes mellitus using artificial neural network: An effective classification approach. Diabetes research and clinical practice, 2013

[16] Mehmet Recep Bozkurt, Nil¨ufer Yurtay, Ziynet Yılmaz, and Cengiz Sertkaya. Comparison of different methods for determining diabetes disease, 2013.

[17] Devang Odedra Shankaracharya, Subir Samanta, and Ambarish S Vidyarthi. Computational intelligence in early diabetes diagnosis: a review. The review of diabetic studies: RDS, 7(4):252, 2010.

[18] Elodie Scaglia, Ganesh D Sockalingum, Juergen Schmitt, Cyril Gobinet, Nathalie Schneider, Michel Manfait, and G´erard Thi´efin. Noninvasive assessment of hepatic fibrosis in patients with chronic hepatitis c using serum fourier transform infrared spectroscopy. Analytical and bioanalytical chemistry, 401(9):2919–2925, 2011.

[19] Peter Tsyurmasto, Michael Zabarankin, and Stan Uryasev. Value-at-risk support vector machine: Stability to outliers. 2013.

[20] Javad Haddadnia Hadi Varharam Mohammad Fiuzy, Azam Qarehkhani. Introduction of a method to diabetes diagnosis according to optimum rules in fuzzy systems based on combination of data mining algorithm (d-t), evolutionary algorithms (aco) and artificial neural networks (nn). The Journal of Mathematics and Computer Science (JMCS), 6(4):272–285, 2013.

[21] Elin Olafsdottir, Thor Aspelund, Gunnar Sigurdsson, Rafn Benediktsson, Bolli Thorsson, Tamara B Harris, Lenore J Launer, Gudny Eiriksdottir, and Vilmundur Gudnason. Similar decline in mortality rate of older persons with and without type 2 diabetes between 1993 and 2004 the icelandic population-based reykjavik and ages-reykjavik cohort studies. BMC public health, 13(1):36, 2013.

[22] R Karakı¸s, M Tez, YA Kılı¸c, Y Kuru, and I G¨uler. A genetic algorithm model˙ based on artificial neural network for prediction of the axillary lymph node status in breastcancer. Engineering Applications of Artificial Intelligence, 26(3):945–950, 2013.

[23] Mehdi Khashei, Saeede Eftekhari, and Jamshid Parvizian. Diagnosing diabetes type ii using a soft intelligent binary classification model. Review of Bioinformatics and Biometrics, 1(1), 2012.

[24] Rachel O'Reilly, Cross-validation for model selection in model-based clustering, 2012.

[25] Catalin Stoean, Ruxandra Stoean, Monica Lupsor, Horia Stefanescu, and Radu Badea. Feature selection for a cooperative coevolutionary classifier in liver fibrosis diagnosis. Computers in Biology and Medicine, 41(4):238–246,2011.

[26] Kayaer, Kamer, and Tulay Yıldırım. "Medical diagnosis on Pima Indian diabetes using general regression neural networks." Proceedings of the international conference on artificial neural networks and neural information processing (ICANN/ICONIP). 2003.

[27] Latif, Aadil, and Peter Palensky. "Economic Dispatch Using Modified Bat Algorithm." Algorithms 7.3 (2014): 328-338.

[28] Rashedur M Rahman and Farhana Afroz. Comparison of various classification techniques using different data mining tools for diabetes diagnosis. 2013

[29] Asha Gowda Karegowda, AS Manjunath, and MA Jayaram. Application of genetic algorithm optimized neural network connection weights for medical diagnosis of pima indians diabetes. International Journal on Soft Computing, 2(2):15–23, 2011