Performance Study of Different Metaheuristics for Diabetes Diagnosis

Fatima Bekaddour[1], Mohamed Ben Rahmoune[2], Chikhi Salim[1], and Ahmed Hafaifa[2(✉)]

[1] SCAL Team, MISC Laboratory, Constantine, Algeria
fatima.bekaddour@gmail.com, slchikhi@yahoo.com
[2] Applied Automation and Industrial Diagnostics Laboratory,
Djelfa University, Djelfa, Algeria
{BR.Mohamed,hafaifa.ahmed.dz}@ieee.org

Abstract. The problem of medical data classification involves an optimization phase that may be solved through metaheuristic approaches. In this work, we evaluate the performance in diagnosis of diabetes disease, using Particle Swarm Optimization (PSO), Firefly (FF) and Homogeneity-Based Algorithm (HBA) metaheuristics in conjunction with fuzzy system. Here, the fitness function in the optimization process is the total misclassification cost that is in term of false positive, false negative and unclassifiable rates. The results prove that HBA approach achieves better results than the other metaheuristics. With execution time, FF was faster than the PSO and HBA methods.

Keywords: Metaheuristic · PSO · Firefly · HBA · Fuzzy system · PIMA dataset

1 Introduction

Diabetes is a complex and chronic disease, which is characterized as an illness that may occur when the pancreas is unable to produce insulin or when this hormone is not effectively used by the human body. Diabetes includes two major classes: type 1, which results in lack of insulin, due to beta-cells destroyment by the human immune system, while type 2 diabetes, results from insulin resistance. Nowadays, diabetes has become a worldwide epidemic that affects a big number of people [14].

Diabetes recognition is an important and difficult task that requires a reliable algorithm in order to reduce the probability of classification error. Recently, there has been several works, interested in the development of automatic tools for healthcare provision. Most of these researches have investigated the development of CAD (Computer Aided-Diagnosis) that provides a second opinion for physicians and helps them in their diagnosis tasks.

Thus, in order to identify the diabetes diagnosis of patient, in this work, we propose the employment of three well-known population-based metaheuristics, namely: PSO (Particle Swarm Optimization), FF (Firefly) and HBA (Homogeneity-Based Algorithm), in conjunction with FIS (Fuzzy Inference System), to evaluate their effectiveness. Note that, in this study, the main goal is to minimize the fitness function value,

that represents the total misclassification cost of FIS model, which is in term of FP (false positive), FN (false negative), UC (unclassifiable) errors. The experiments are made on the PIMA diabetes data set, taken from the UCI repository [10]. In the medical field, a (Uc) case means a patient that cannot be diagnosed by the prediction system. This is because of scanty and inadequate information about the patient. In the medical applications, the Uc error is very required, because it may lead to additional examinations that help doctors to make the right diagnosis.

The reminder of this paper is organized as follows: Sect. 2 presents some related works on diabetes diagnosis. In Sect. 3, the PSO, FF and HBA metaheuristics and the FIS model are defined. In Sect. 4, we describe our proposed methodology and we discuss the obtained results, based on PIMA diabetes data set. Finally, in Sect. 5, we conclude the paper.

2   Related Works

In the literature, several metaheuristic approaches have been successfully applied in medical diagnosis, such as the PIMA diabetes dataset. In [1], authors developed a Homogeneity-base Algorithm (HBA), for the classification of PIMA diabetes dataset. Authors introduced the concept of Homogeneity Degree (HD), to achieve a simultaneously balance between the fitting and the generalization of the inferred classification models. The HD value calculates the density of learning points in a given homogenous hypersphere. The proposed HBA, compared to standalone data mining approaches such as SVM (Support Vector Machine) and DT (Decision Tree), presented high performance accuracy, but less efficiency due to the HBA's computational time.

In another study [2], Pham and Triantaphyllou proposed a Convexity-Based Algorithm (CBA) approach. In this work, a new concept of convex density (CD) was introduced to optimize the total misclassification cost value, based on defragmenting convex regions. The obtained results, tested on PIMA dataset, present significant improvement compared to other well-known data mining techniques.

Beloufa and Chikh [3] presented a novel approach based on ABC (Artificial Bee Colony) metaheuristic for automatic recognition of diabetes dataset from the UCI repository. In this work, authors modify the original ABC approach by adding a blended genetic operator for better intensification and diversification of the search space. This operator serves mainly to automatically update the Fuzzy Inference System membership functions and rules. Experimental results prove that the proposed ABC metaheuristic, found minimal number of rules, while improving the final classification performance.

Al-Muhaideb and Menai [4] introduced a two-stage metaheuristic optimization method, called HColonies (Hybrid ant-bee colonies), which is a hybrid system between ACO (Ant Colony Optimization) and ABC (Artificial Bee Colony) approaches. The main idea of HColonies is to use ACO approach to create initial population solutions for the ABC metaheuristic. This is done to accelerate the search and obtain good performance results. In the first stage, the Ant-Miner+ is adopted to generate a population of food sources. In the second stage, a modified ABC method, based on new operators is employed, to fit the appropriate problem. Results obtained by the

HColonies method on the PIMA diabetes, illustrate the effectiveness and robustness of the proposed metaheuristic, toward change in its parameters.

3 Overview of FIS and the Used Metaheuristics

3.1 Particle Swarm Optimization

The first metaheuristic used in this work, is a population-based metaheuristic, called PSO: particle swarm optimization (Kennedy and Eberhart in 1995) [5]. PSO mimics the cooperative behavior concepts of natural organisms such as birds and fish. The PSO algorithm starts with a random swarm that constitutes a set of particles. Each particle 'i' is characterized by its corresponding velocity Vi and its position Xi. Then, at each iteration, particles modify their velocity and their position using formulas (1) and (2).

$$Xi(t) = Xi(t-1) + Vi(t) \tag{1}$$

$$V_i(t) = V_i(t-1) + c_1 r_1 (P_{besti}(t-1) + X_i(t-1)) + c_2 r_2 (G_{besti}(t-1) + X_i(t-1)) \tag{2}$$

where (c_1, c_2) are two factors that represent the cognitive attraction and the social attraction respectively. (r_1, r_2) are two random numbers uniformly distributed between [0, 1]. (P_{besti}, G_{besti}) define the best position obtained by a given particle i and the best position ever found in the entire swarm respectively. The different steps of the PSO algorithm are given in Algorithm 1 as follows:

```
Algorithm 1: Particle Swarm Optimization
- Particles initialization.
- While (stopping criteria not met){
     -Evaluate f (Xi):fitness function of each particle i
     -For all particles i do {
          - Compute Velocities Vi using formula(2).
          - Compute the new position Xi using formula(1).
             - If ( f (Xi) <  f (P_besti)) Then P_besti = Xi.
             - If ( f (Xi) <  f (G_besti)) Then G_besti = Xi.
             - Update (Xi, Vi).
                              }
                               }
```

3.2 Firefly Metaheuristic

The second used metaheuristic is Firefly (FF) algorithm [6]. FF is a nature inspired population based swarm method, that imitates the flashing behaviors of fireflies. Each firefly is considered as a candidate solution in the search space. There are three fundamental key issues, regarding the FF metaheuristic:

- r_{ij}: that denote the distance between fireflies i and j.
- Attractiveness ($\beta(r_{ij})$): calculated by the backdrop of the fitness function as follows:

$$\beta(r_{ij}) = \beta_0 \, e^{-\gamma r_{ij}^2} \tag{3}$$

Where β_0 means the attractiveness at r = 0, and γ is the absorption coefficient.
- Movement: is determined by the Eq. (4):

$$Xi = Xi + \beta_0 \, e^{-\gamma rij2}(Xj - Xi) + \alpha(\text{rand} - 1/2) \tag{4}$$

Where α is a random number, uniformly distributed between [0, 1].

The main steps of FF metaheuristic are summarized in Algorithm 2.

```
Algorithm 2: Firefly metaheuristic
- Fireflies initialization.
- Define the light intensity I.
- Define the absorption coefficient γ.
- Define f(x): fitness function of x=(x1,x2,..,xd).
- While (Max-generations not met){
   - For each firefly i do{
    - For (j=1 to i) do {
       - If ( I (j) >  I (i)) Then {
         Move firefly i toward firefly j using eq(4)
                                   }
       -Attractiveness varies with distance r via e-γr2
       -Compute new solutions.
       -Update I(i).
                      }
                      }
   - Rank fireflies and find the best one.

                              }
```

3.3 Homogeneity-Based Algorithm

In [8], Pham and Triantaphyllou introduced a new metaheuristic called Homogeneity-based Algorithm (HBA). The main objective of HBA metaheuristic, is to achieve a simultaneously balance between the fitting and the generalization, using the concept of homogenous set and homogeneity degree (HD) [1, 7–9]. HBA approach may be applied in conjunction with data mining techniques such as SVM (Support Vector Machine), to minimize the fitness function formula (5):

$$TC = \min(CFP * RateFP + CFN * RateFN + CUC * RateUC) \tag{5}$$

Where: CFP, CFN, CUC are the unit penalty cost for the false positive, false negative and unclassifiable rates respectively. TC represents the total misclassification cost value. From the two inferred models, obtained using a data mining approach, HBA breaks each set into hyperspheres, covering decision regions. Next the Homogeneity Degree (HD) value, corresponding to each hypersphere is calculated, using the Eq. (6):

$$HD(S) = \ln(n_s)/h \qquad (6)$$

Where S is a given homogenous set, n_s is the number of samples in S and h is the minimal most frequent distance in a set S [1, 8, 9]. After that, HBA adopts four thresholds: (β^-, β^+, α^-, α^+), to expand or break down each homogenous set, using their corresponding HD values. Please note that, (α^-, α^+) are used to expand the negative and the positive homogenous sets respectively. On the other hand, (β^-, β^+) are used to fragment the negative and the positive homogenous sets respectively [8]. The HBA metaheuristic iterates until all the homogenous sets are processed. Within the scope of HBA, the GA approach is employed, to adjust the (β^-, β^+, α^-, α^+) factors values.

3.4 Fuzzy Inference System

Fuzzy Inference System (FIS) [15] is a well-known artificial intelligence approach, that is based on the theory of fuzzy sets and fuzzy logic to extend the classical crisp sets theory. In the literature, the FIS model has been widely employed in the medical field [16–19].

The basic architecture of the FIS model as shown in Fig. 1, consists of three main phases:

- Fuzzification: transform the crisp input into linguistic variables (fuzzy input).
- Inference Engine (IE): uses the fuzzy input and the rules defined in the knowledge base module, to derive fuzzy sets for each variables.
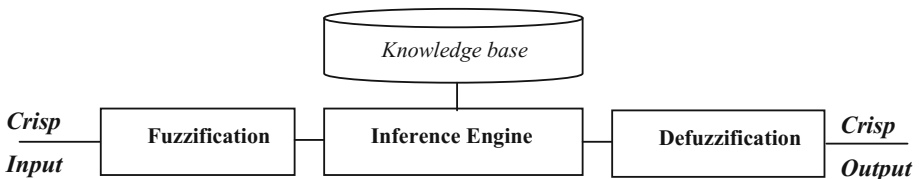- Defuzzification: transform the obtained fuzzy output by the IE into crisp output

Fig. 1. The main structure of the FIS model

The FIS model used in this paper is the Takagi-Sugeno fuzzy model.

4   Computational Results

4.1   Description of Data

In this study, we employ the PIMA diabetes dataset, obtained from the UCI repository [10]. The main description of PIMA dataset is depicted in Table 1. The PIMA dataset is composed of 768 instances, where 268 samples belong to the positive class, and 500 samples belong to the negative class.

Table 1. PIMA diabetes dataset description

Attribute	Description
Npreg	Number of times pregnants
Glu	Plasma glucose concentration
Bp	Diastolic blood pressure
Skin	Triceps skin fold thickness
Insulin	2-Hour serum insulin
BMI	Body mass index
PED	Diabetes pedigree function
Age	Age of the patient

4.2   Experimental Design

In this work, the procedure of conducting the experiments is based on the fuzzy inference system (FIS) [15] and metaheuristics techniques (PSO, FF, HBA). The individual system was designed and developed with their default parameters values. We developed the FIS approach, using the concept of membership functions and rules, that denote the relationship between input (x_i) and output (o_i). The use of FIS, permit to increase the transparency of the classifier. The main architecture of the proposed F-Metaheuristic is depicted in Fig. 2.
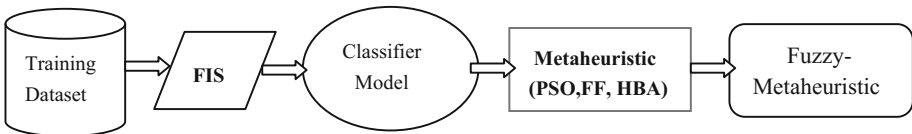

Fig. 2. Architecture of the proposed F-Metaheuristic approach

For the development of (PSO, FF, HBA) methods, the objective was to minimize the fitness function values that define the total misclassification cost (TC) of FIS approach. In this work, we propose three different configurations of TC formulas, depicted in Fig. 3.
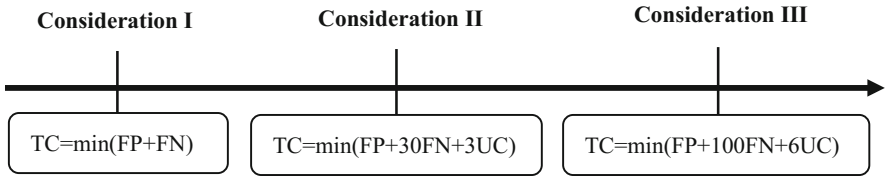
Consideration I	Consideration II	Consideration III
TC=min(FP+FN)	TC=min(FP+30FN+3UC)	TC=min(FP+100FN+6UC)

Fig. 3. The proposed three scenarios of TC

In the first consideration, we assume that unclassifiable (UC) cost is equal to 0, while FN and FP penalty costs are set to (1, 1) respectively. This is the case on the majority of works in the literature. Therefore, TC can be defined as:

$$TC = \min(1 * \text{RateFP} + 1 * \text{RateFN}) \tag{7}$$

In the second consideration, we suppose that UC is taken into account, while the FN rate is penalized more than (UC, FP) and the UC is penalized more than the FP rate. Thus, TC is defined as follows:

$$TC = \min(1 * \text{RateFP} + 30 * \text{RateFN} + 3 * \text{RateUC}) \tag{8}$$

In the last consideration, we assume the case where the FN rate is penalized much more than the (FP, UC) costs, while UC still penalized more than FP rate. Therefore, the TC value is described as follows:

$$TC = \min(1 * \text{RateFP} + 100 * \text{RateFN} + 6 * \text{RateUC}) \tag{9}$$

It is surprised that despite the various works reported in the literature, on the PIMA dataset, the majority of researchers still neglects the evaluation of UC cases. It is to be emphasized that the UC error is very essential in medical diagnosis problem. This is because, this may help the physician to make the right decision, based on more examinations. In this work, we compute the UC rates that represent the number of patients which cannot be diagnosed by the classifier. A reason for that is that the information gathered about the patient are scanty or inadequate. The defaults parameters of the three used metaheuristics are given in Table 2.

4.3   Measure for Performance Evaluation

In order to conduct a valid experiment, in this work, we propose the use of three different scenarios for the fitness function (TC), as it is described in Sect. 4.2. Please note that TC is in term of three type of errors (FP, FN, UC) that measure mistakes made by the FIS or the F-metaheuristic, during the validation phase. The sensitivity (Se), and the specificity (Sp) are two other well-known metrics that are used in this study. (Se, Sp) measure the true positive and the true negative rates respectively, defined as follows:

Table 2. Summary of default parameters used in the implementation of the model

Approach	Parameters values
Fuzzy	Membership number: 2 for each attribute
	Membership type: Generalized Bell
	FIS type: Takagi-Sugeno
	Rules number: 256
PSO	Swarm size: 40
	Generations number: 10–50
	Cognitive attraction (c1) = 0.5
	Social attraction (c2) = 1.25
FF	Population size: 25
	Generations number: 10–100
	Light absorption coefficient = 1
	$\beta 0 = 2$
HBA	Initial ($\beta-$, $\beta+$, $\alpha-$, $\alpha+$): 0
	Final ($\beta-$, $\beta+$, $\alpha-$, $\alpha+$): updated using Genetic Algorithm
	Iterations number: 10–50
	$\gamma = 0.01$, is a parameter, used to judge whether a region is homogenous or not

$$Se = TP/TP + FN \tag{10}$$

$$Sp = TN/TN + FP \tag{11}$$

Where (TP, TN) are the true positive and the true negative numbers respectively, while (FP, FN) are false positive and false negative numbers respectively. In this study, we also calculate two other metrics that are:

- Solicitation degree (SD) of each rule, generated by the FIS (F-metaheuristic) approach. SD defines the rules activation degree between 50% and 100%.
- Accuracy: that defines the correct classification rate.

4.4   Results and Discussion

We conducted different experiments to juxtapose the three used metaheuristics (PSO, FF, HBA), for the given default parameters, depicted in Table 2. As discussed in Sect. 4.1, we adopt the PIMA diabetes dataset. The experimental methodology ran as presented in Sect. 4.2. The experiments were made in Matlab, version 2012a. As this work presents a comparative study of metaheuristics, the comparative results related to the fitness function value, are presented in Table 3. This table presents the (FP, FN, UC) rates and the obtained TC values. Note that F-PSO means: the FIS system is used in conjunction with PSO metaheuristic. Same definitions are valid for F-FF and F-HBA. The improvement column defines the improvement rate of the FIS-metaheuristic compared to the FIS approach.

Table 3. Results obtained under the three proposed considerations

Nbr. consideration	Approach	FP (%)	FN (%)	UC (%)	TC (%)	Improvement (%)
Consideration I	FIS	67.18	0.0	32.18	67.18	
	F-PSO	21.35	19.8	58.85	41.15	67.18
	F-FF	16.14	19.27	11.45	35.41	47.29
	F-HBA	0.0	8.33	55.2	8.33	87.60
Consideration II	FIS	67.18	0.0	32.18	163.72	
	F-PSO	21.35	19.8	58.85	791.9	No.improv
	F-FF	16.14	19.27	11.45	628.6	No.improv
	F-HBA	0.0	8.33	55.2	415.5	No.improv
Consideration III	FIS	67.18	0.0	32.18	260.26	
	F-PSO	21.35	19.8	58.85	2354	No.improv
	F-FF	16.14	19.27	11.45	2011	No.improv
	F-HBA	0.0	8.33	55.2	1164	No.improv

According to Table 3, the average value of TC on PIMA dataset were 28.3%, under the first consideration. This value of TC was less than the TC average value by about 67.35%. In the second consideration, the TC values are equal to (791, 628, 415) for (F-PSO, F-FF, F-HBA) respectively. This value of TC was not optimal than the FIS TC value. The average value of TC is equal to 612%, on the PIMA dataset. The obtained results for the last consideration are also summarized in Table 3. In this case, the obtained average TC value is equal to 1843. The (F-PSO, F-FF, F-HBA) have not yielded optimal TC on the PIMA dataset, compared to the FIS TC value. In addition, the obtained computational results presented in this table, confirm that the three (C_{FP}, C_{FN}, C_{UC}) values affect the final TC values.

To make the results clearer, we plotted the (Accuracy, Se, Sp) in Fig. 4. This figure presents also the average improvement of the FIS-metaheuristics, compared to the standards FIS. The experiments illustrate that F-HBA metaheuristic achieved better results in term of (Accuracy, Se, Sp), which were successful for small number of generations. FF was slightly worse than PSO metaheuristic. It clearly appears that the HBA metaheuristic achieves minimal TC error value and better accuracy for small generations number. Also, $(\alpha+, \alpha-, \beta+, \beta-)$ values, played a great role in method performance. However, the F-HBA computational time is higher than (F-PSO, F-FF, Fuzzy) approaches. This is due to the compute of homogenous sets and homogeneity degree. F-FF was slightly faster the F-PSO metaheuristic in term of execution time.

Tables 4 and 5 present the solicitation degree (SD) of some rules, obtained by the F-HBA model, for the TN and UC cases. Rules cited in those tables are:

R97: if (Npreg is sm)& (Glu is bg)& (Bp is bg)& (Skin is sm)& (Insulin is sm)& (BMI is sm)& (PED is sm)& (Age is sm) then Class 97

R101: if (Npreg is sm)& (Glu is bg)& (Bp is bg)& (Skin is sm)& (Insulin is sm)& (BMI is bg)& (PED is sm)& (Age is sm) then Class101

R109: if (Npreg is sm)& (Glu is bg)& (Bp is bg)& (Skin is sm)& (Insulin is bg)& (BMI is bg)& (PED is sm)& (Age is sm) then Class109
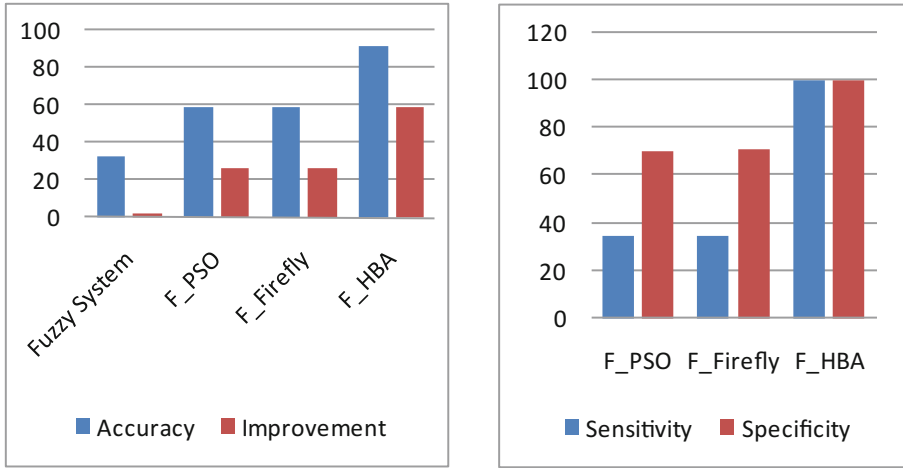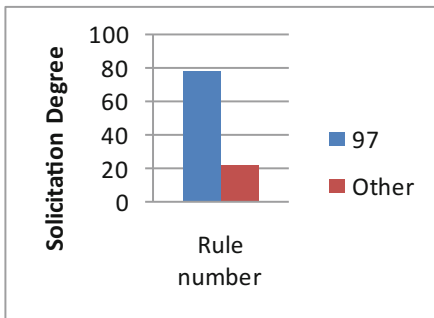
Fig. 4. Accuracy, Se, Sp comparisons for PIMA dataset

As described in Table 2, we choose two membership functions for each descriptor (sm: small, bg: big). Figure 5 presents a TN case that activates R97 with SD = 78.22. This example has been misclassified by the FIS, while correctly classified by F-HBA.

Table 4. SD of some rules for TN case

Rule	Solicitation degree
[101]	0.008
[97]	0.02

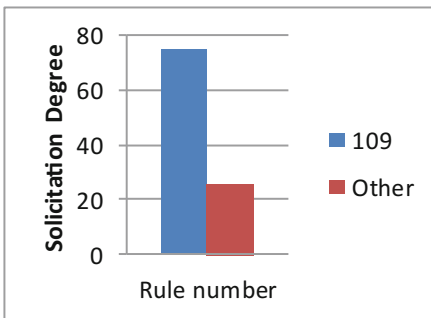| Parameter	Value
Npreg	6.77
Glu	18.84
BP	16.58
Skin	0
Insulin	0
BMI	9.32
Ped	8.99
Age	8.61

Fig. 5. Example of TN case

For the UC cases, Table 5 illustrates the SD of rules 101 and 97. According to this table, R101 presents the higher SD value. The example depicted in Fig. 6 presents a patient that was correctly classified by the F-HBA approach, while misclassified by the

FIS. In fact, this patient has a Glu value greater than 1.4 g/l and a body mass index greater than 40 kg/m^2, which indicates a morbid obesity.

Table 5. SD of some rules for UC case

Rule	Solicitation degree
[97]	0.008
[101]	0.06
[109]	0.008

Parameter	Value
Npreg	0
Glu	17.77
BP	16.95
Skin	20.92
Insulin	110.96
BMI	21.28
Ped	11.78
Age	9.00

Fig. 6. Example of UC case

This example has activated R109 with SD equal to 74.45. Experiments proves that the HBA metaheuristic can be of great interest for diabetes disease diagnosis.

5   Conclusion

This study investigates the problem of medical data classification that involves an optimization phase, and so the employment of metaheuristics approaches is very recommended. Various works dedicated in the diagnosis of diabetes have been carried out over the last decades. In the present work, three metaheuristic algorithms were developed to identify whether a patient is a subject of diabetes or not. The performance of PSO (Particle Swarm Optimization), FF (Firefly), HBA (Homogeneity-Based Algorithm) has been compared for the minimization of FP (False Positive), FN (False Negative), UC (Unclassifiable) rates of the FIS (Fuzzy Inference System) model. Computational experiments based on PIMA dataset from the UCI repository, illustrate that the HBA metaheuristic obtained better performance among the other used methods. In the near future, we aim to pay more attention to employ other metaheuristics such as Krill Herd [11], Dragonfly [12] and Whale [13] metaheuristics.

References

1. Pham, H.N.A., Triantaphyllou, E.: Prediction of diabetes by employing a new data mining approach which balances fitting and generalization. In: Yin, L.R. (ed.) Computer and Information Science. SCI, vol. 131, pp. 11–26. Springer, Berlin (2008). Chapter 2

2. Pham, H.N.A., Triantaphyllou, E.: A meta-heuristic approach for improving the accuracy in some classification algorithms. Comput. Oper. Res. **38**, 174–189 (2011)

3. Beloufa, F., Chikh, M.A.: Design of fuzzy classifier for diabetes disease using modified artificial bee colony algorithm. Comput. Methods Programs Biomed. **112**, 92–103 (2013)

4. Al-Muhaideb, S., Menai, M.E.-B.: HColonies: a new hybrid metaheuristic for medical data classification. Appl. Intell. **41**(1), 282–298 (2014)

5. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: Proceedings of IEEE International Conference on Neural Networks IV, pp. 1942–1948. IEEE, Piscataway (1995)

6. Yang, X.S.: Nature-Inspired Metaheuristic Algorithms. Luniver Press, Bristol (2008)

7. Bekaddour, F., Amine, C.M.: IHBA: an improved homogeneity-based algorithm for data classification. In: Amine, A., Bellatreche, L., Elberrichi, Z., Neuhold, Erich J., Wrembel, R. (eds.) CIIA 2015. IAICT, vol. 456, pp. 129–140. Springer, Cham (2015). doi:10.1007/978-3-319-19578-0_11

8. Pham, H.N.A., Triantaphyllou, E.: The impact of overfitting and overgeneralization on the classification accuracy in data mining. In: Maimon, O., Rokach, L. (eds.) Soft Computing for Knowledge Discovery and Data Mining, Part 4, pp. 391–431. Springer, New York (2007). Chapter 5

9. Pham, H.N.A., Triantaphyllou, E.: An application of a new meta-heuristic for optimizing the classification accuracy when analyzing some medical datasets. Expert Syst. Appl. **36**(5), 9240–9249 (2009)

10. UCI Repository of Machine Learning Databases, Department of Computer Science, University of California at Irvine. https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes. Accessed 2017

11. Gandomi, A.H., Alavi, A.H.: Krill herd: a new bio-inspired optimization algorithm. Commun. Non Linear Sci. Numer. Simul. **17**(12), 4831–4845 (2012). doi:10.1016/j.cnsns.2012.05.010

12. Mirjalili, S.: Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. Neural Comput. Appl. **27**(4), 1053–1073 (2016)

13. Ebrahimi, A., Khamehchi, E.: Sperm whale algorithm: an effective metaheuristic algorithm for production optimization problems. J. Nat. Gas Sci. Eng. **29**, 211–222 (2016)

14. Al-Rubeaan, K., Youssef, A.M., Ibrahim, H.M., Al-Sharqawi, A.H., AlQumaidi, H., AlNaqeb, D., Aburisheh, K.H.: All-cause mortality and its risk factors among type 1 and type 2 diabetes mellitus in a country facing diabetes epidemic. Diabetes Res. Clin. Pract. **118**, 130–139 (2016)

15. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**(3), 338–353 (1965)

16. Miranda, G.H.B., Felipe, J.C.: Computer-aided diagnosis system based on fuzzy logic for breast cancer categorization. Comput. Biol. Med. **64**, 334–346 (2015). http://dx.doi.org/10.1016/j.compbiomed.2014.10.006

17. Nilashi, M., Ibrahim, O., Ahmadi, H., Shahmorad, L.: A knowledge-based system for breast cancer classification using fuzzy logic method. Telemat. Inform. **34**(4), 133–144 (2017)

18. Muthukaruppan, S., Er, M.J.: A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease. Expert Syst. Appl. **39**, 11657–11665 (2012)

19. Ghazavi, S.N., Liao, T.W.: Medical data mining by fuzzy modeling with selected features. Artif. Intell. Med. **43**(3), 195–206 (2008)