International Arab Conference on Information Technology (ACIT'2016)

A comparative study of the liver disorders prediction based on Neuro-Fuzzy and Metaheuristics approaches

Fatima Bekaddour[1], Chikhi Salim[1], Bekaddour Okkacha[2]

[1] MISC Laboratory, Mentouri University, Canstantine, Algeria

[2] Computer Science Department , Tlemcen University, Algeria

Abstract: *In this work, we propose the application of some well-known metaheuristics to enhance the medical classifier performance. IHBA (Improved Homogeneity-Based Algorithm), Simulated Annealing (SA), Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) metaheuristics have been applied in conjunction with the neuro-fuzzy system to minimize the value of an objective function proposed in our previous work. We validate our computational results, based on the liver disorders dataset obtained from the UCI repository. Results show that the IHBA approach found the best performances. Both SA and PSO outperform the GA metaheuristic and the standard neuro-fuzzy model.*

Keywords: *Metaheuristics, Neuro-Fuzzy, IHBA, HBA, SA, PSO, GA, Bupa Liver Disorders, Medical Informatics.*

---

1. Introduction

The liver is a biggest internal organ that performs many functions including blood clotting and making proteins, to prevent significant damage to the human body. Several diseases may occur in the liver such as hepatitis, cancers, cirrhosis. Liver disorders referred to any disturbance of liver function that makes this organ unable to accomplish its functions. The most common liver diseases are caused by one of the following factors: alcohol consumption, toxins, medications or drug, obesity, diabetics, genetic disorders...

Metaheuristics have been widely used for diagnosis in medicine. The efficiency and the effectiveness of metaheuristics optimization as tools that help doctors to make decisions in the medical sector, has been investigated by any researchers in the literature. Computational results prove that metaheuristics are suitable in solving challenging classification problems. As for now, various metaheuristics approaches have been proposed such as: Cuckoo Search Algorithm [2], Krill Herd Algorithm [3], Artificial Chemical Reaction Algorithm [4] and Golden Ball metaheuristic [5].

In this work and in order to confirm the absence or the presence of a liver disorders for a patient, we suggest to employ four metaheuristics: Simulated Annealing (SA) [6], Particle Swarm Optimization (PSO) [7], Genetic Algorithm (GA) [8], and IHBA metaheuristics (Improved Homogeneity-Based Algorithm) [1]. The experiments were made on the liver disorders (LD) dataset obtained from the UCI repository [16].

This paper is organized as follows: section 2 introduces an overview of some related works. Then, the theory of SA (Simulated Annealing), PSO (Particle Swarm Optimization), GA (Genetic Algorithm), and IHBA (Improved Homogeneity-based Algorithm) metaheuristics are presented in section 3. Next, the adopted methodology and the computational results are illustrated in section 4. Finally, we conclude the paper.

2. Related works

- Lee and Mangasarian (2001) [9] proposed a Smooth Support Vector Machine (SSVM) approach. In this work, a new formulation of the support vector machine method was introduced, based on a smooth unconstrained optimization and a fast Newton–Armijo algorithm to makes the SSVM converges globally and quadratically. The obtained results tested on BUPA liver disease show that the proposed approaches present better generalization ability. The SSVM yielded 70.33 % accuracy.
- Polat et al. (2007) [10] developed a Fuzzy Artificial Immune Recognition System (FAIRS) for the automatic classification of BUPA liver disorders disease. Authors investigated the use of a fuzzy resource allocation mechanism based on fuzzy logic concepts to enhance the performance system. Results were compared with other approaches for the same dataset. The FAIRS

achieved the best classification accuracy of 83.7%.

- Alia and Betar (2011) [11] developed a two stage approach for the prediction of several standard medical benchmarks including BUPA liver disorders, Diabetes, Iris, etc. First, they applied a Harmony Search (HS) metaheuristic for exploring the search space to achieve (near)-optimal clusters centers. Next, the best cluster centers obtained in the first stage are used as initial clusters for the c-means algorithm. Results from the HS-based algorithm proves better performances when compared to the HCM (Hard c-means) and the FCM (Fuzzy c-means) approaches.
- Ivanoe (2013) [12] presented a novel method based on Differential Evolution (DE) for the automatic recognition of items in several medical benchmarks such as haberman, thyroid, liver disorders, diabetes... In this work, the author compared different versions of DE algorithms and obtained 65.85 ±2.18% accuracy by using a best/2/bin version. Additionally, the proposed approach proves good classification performance when compared with fifteen classifiers systems widely used in the literature.

3. Optimization Metaheuristics

This section describes the Simulated Annealing (SA), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and the Improved Homogeneity-Based Algorithm metaheuristics which were used in this paper.

3.1. Simulated Annealing (SA)

Simulated Annealing (Kirkpatrick S. et al. 1983) [6] is one of the most popular single-based search metaheuristic. The SA mimics the annealing process's cooling schedule. The SA starts from a high temperature value (Tmax) randomly initialized. Then, at each iteration a new solution (S1) is selected in the neighborhood of the current solution (S0). Neighbor that improves the current solution is accepted depending on the probability value ($P^{-\Delta E/T}$), where $\Delta E = f(S1) - f(S0)$. During the search process, the temperature is decreased gradually to avoid "metastable" structures (Local minima of energy).

Input: A given Problem
Output: (sub) optimal solution.
- Initial Temperature T.
- Generate randomly an initial solution s0.
- While (T< 0) {
  - Generate a new solution s'.
  - Compute $\Delta E = f(s') - f(s0)$.
  - If ($\Delta E \leq 0$) Then
    s→s'.
    else
    accept with probability $P^{-\Delta E/T}$ }

3.2. Particle Swarm Optimization (PSO)

The second metaheuristic used in this work is the particle swarm optimization (Kennedy and Eberhart in 1995) [7] algorithm. PSO is a population based solution approach that simulates the social and the cooperative behavior concepts of natural organisms (birds, fish …). The PSO algorithm starts with a swarm that constitutes a set of particles. Each particle is characterized by its position Xi and its velocity Vi. At each cycle, particles change their velocity and their position using the following equations (1) and (2).

$$X_i(t) = X_i(t-1) + V_i(t) \quad (1)$$

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

The two factors (c1,c2) represent the cognitive attraction and the social attraction respectively. (r1, r2) are two random numbers uniformly distributed between [0,1]. (Pbesti ,Gbesti) are the best position obtained by the particle i and the best position ever found in the entire population respectively. The pseudo-code of the PSO metaheuristic is given bellow:

- Particles random initialization.
- While (stopping criteria not met){
  - Evaluate f (Xi) of each particle
  - For all particles i do{
    - Calculate velocities Vi using formula (2).
    - Calculate the new position Xi using formula (1).
    - If (f (Xi) < f (Pbesti)) Then Pbesti = Xi.
    - If (f (Xi) < f (Gbesti)) Then Gbesti = Xi.
    - Update (Xi, Vi).
    }
  }

3.3. Genetic Algorithm (GA)

Genetic Algorithm (Holland J. H., 1975) [8] is a population-based search metaheuristic that imitates the Darwinian evolution theory. The GA starts from a set of chromosomes (called population) randomly initialized, then at each generation, some pairs of solution from the current population are chosen to make them arising using recombination operation (Crossover / Mutation). This phase is called Selection operation. After that, the replacement phase that determines which chromosomes will survive to construct the next generation is applied. The GA metaheuristic is repeated until a stopping criterion is reached (maximum number of generation, maximum number of the fitness function evaluation...). The different steps of the GA metaheuristic are represented as follows:

-Generate randomly a population of individuals
-Evaluate each individual.
-Repeat {
-Select chromosomes based on their fitness values.
-Chromosomes recombination (Crossover, Mutation)
-Replace current population by the new population.
}
Until (stopping criteria are reached)

3.4. Improved Homogeneity-Based Algorithm (IHBA)

In 2007, Pham and Triantaphyllou proposed a novel metaheuristic called Homogeneity-based Algorithm (HBA) [13]. The main idea of HBA, is to achieve a simultaneously balance between the fitting and the generalization [1, 13, 14, 15], using the concept of homogenous set and homogeneity degree (HD). The new metaheuristic can be used in conjunction with standard data mining approaches such as SVM (Support Vector Machine), FS (Fuzzy Logic System), to reduce the following total misclassification cost (TC) equation:

$$TC = \min(CFP * RateFP + CFN * RateFN + CUC * RateUC) \quad (3)$$

CFP, CFN, CUC are the unit penalty cost for the false positive, false negative and unclassifiable rates respectively. HBA starts by applying a classification approach on a given training dataset to infer the positive and the negative classification models. After that, the two inferred models are broken into hyperspheres and their Homogeneity Degree (HD) are calculated using the following formula:

$$HD(S) = ln(ns)/h \quad (4)$$

Where S is a given homogenous set, ns is the number of samples in S and h is the minimal most frequent distance in a set S [13, 14, 15].

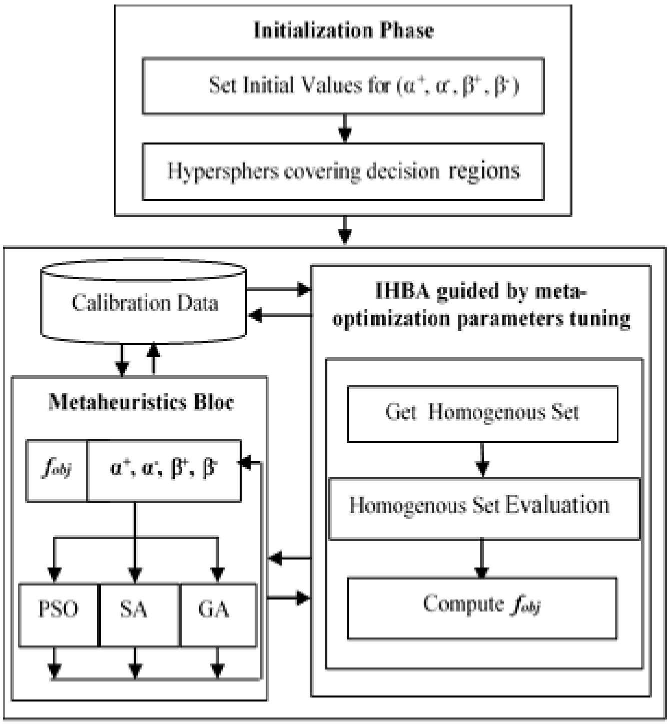

Figure1. Architecture of the IHBA metaheuristic [1]

Next, the (β-,β+,α-,α+) factors are used to expand or break down each homogenous set, depending on their 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. The HBA metaheuristic iterates until all the homogenous sets are processed. The HBA metaheuristic uses the GA approach to adjust the (β-,β+,α-,α+) thresholds values.

In [13, 14, 15], Pham and Triantaphyllou apply the HBA metaheuristic to classify several medical datasets, including: Diabetics, Appendicitis, Breast Cancer and Liver Disorders diseases. Results from the Homogeneity-based algorithm proves better performances when compared to the standalone approaches. Nevertheless, the HBA metaheuristic has three main problems. The first is the neglect of the structural complexity of the used classifier, due to the way objective function is described [1]. The second obstacle found among the deployment of the HBA is the use of the GA (Genetic Algorithm) approach only to adjust the (β-,β+,α-,α+) parameters control values. This may reduce the freedom degree of this metaheuristic. The large amount of data dimensionality that increase the HBA's temporal complexity is the third complication that arise from the employment of the HBA metaheuristic. In this regard, in this work, we adopt the IHBA (Improved Homogeneity-Based Algorithm) [1] approach.

The IHBA modify the HBA objective function presented in formula (5) as follows:

$$fobj = Penality * \frac{\alpha 1 * TC_{Lear} + \alpha 2 * TC_{Test}}{\alpha 1 + \alpha 2} \qquad (5)$$

Where ($\alpha 1$, $\alpha 2$) >0 ϵ $\Re$, are two factors for weighting the training and the test errors respectively. Penalty is defined as follow:

$$Penality = 5 * 10^{-8} * e^{f(x)} + 5 * 10^{-5} * y + 1$$

Where: y is the number of epochs necessary for training the classifier. f(x) is the structural complexity of the neuro-fuzzy system. Here, f(x) represents the number of rules of the neuro-fuzzy system. (TC_{Lear} and TC_{test}) are the misclassification rates for the learning and the validation phases respectively. Their formulas are described in equations (7), (8), (10), (11), (13), (14), (16), (17).

The IHBA algorithm starts by using a datamining approach to infer classifications models, and then homogenous sets are computed. Next, Homogeneity Degree (HD) of each homogenous set are calculated, and their f_{eval} are also evaluated using equations (6), (9), (12), (15). The IHBA iterates to obtains optimal parameters values (α+, α- ,β+ ,β-), using the metaheuristics bloc. The main architecture of the IHBA metaheuristic is depicted in Figure 1.

4. Computational Results

4.1. Dataset Description

In this work, we used the Liver disorders LD dataset, obtained from the UCI data set repository [16]. This dataset provided by the British United Provident Association (BUPA), has been used by many researchers in the literature for analyzing their works.

Attribut	Description
MCV	Mean Corpuscular Volume
Alkphos	Alkaline Phosphotase
SGPT	Alamine Aminotransferase
SGOT	Aspartate Aminotransferase
GammaGT	Gamma-Glutamyl Transpeptidase
Drinks	Number of half-pint equivalents of alcoholic beverages drunk per day

TABLE1. CARACTERISTICS OF THE LD DATASET

The main characteristics of this dataset is presented in Table 1. The LD dataset is composed of 345 samples and six continuous attributes. There are 145 instances that belong to the positive diagnosis (class 1) and 200 instances belong to the negative diagnosis (class 2).

4.2. Experimental methodology

The procedure of conducting the experiments in this work is presented as follow:

- **Phase 1:** The value for f_{eval1} was evaluated by applying the training dataset T on the Neuro-fuzzy approach and then testing the classification systems using the validation benchmark based on equations (6), (9), (12), (15) as criteria of performance.
- **Phase 2:** The value for f_{eval2} was obtained by applying the training data set T on the Neuro-Fuzzy algorithm in conjunction with SA, GA, PSO and IHBA metaheuristics and then by using the test dataset as in phase 1.
- **Phase 3:** The two values for f_{eval1} and f_{eval2} were compared with each other and all the obtained results were discussed.

We hope that the value for f_{eval2} would be less (or equal) to the one achieved by $f_{eval\ 1}$. The experiments were conducted into four metaheuristics via Matlab optimization Toolbox

4.2. Results and Discussion

This section shows in detail the obtained experimental results for the prediction of the liver disorders disease. We supposed four cases for the objective function values. Results are presented in Tables 2, 4, 5, 6. Those tables show the values of the false positive, false negative and the unclassifiable errors. The values of the total misclassification cost for the learning (TC_{tr}) and the generalization (TC_{test}) phases are also computed.

Additionally, we calculate the value of f_{eval} as described below in formulas (6), (9), (12) and (15). The last colon is the improvement rate that presents any improvement of the (SA, PSO, GA, IHBA) metaheuristics when compared to the Neuro-Fuzzy system. Results show that the choice of the three cost errors (False Positive, False Negative, Unclassifiable) and the thresholds values (α-, α+, β-, β+) have a great influence on the final objective function value.

Metaheu	Training Results				Validation Results				f_{eval}	Imp (%)
	FP	*FN*	*UC*	TC_{tr}	*FP*	*FN*	*UC*	TC_{tesr}		
Neuro-F	7.7	0.3	91.5	99.5	52	0.0	45	97	98.2	
SA	13	5.8	0.0	18.8	13	9.4	74.4	96.8	60.7	38.2
PSO	22	0.7	0.0	22.7	10	6.9	76.7	93.6	61.0	37.8
GA	17	6.1	0.0	23.1	23	1.1	65.1	89.2	63.7	35.1
IHBA	30	1.9	0.0	31.9	0.0	0.0	77.9	77.9	54.9	44.1

- **Scenario 1:** we study the case where the FP, FN, UC cases are all penalized by one unit. The results of this scenario are presented in Table 2.

TABLE2. RESULTS FOR MINIMIZING $F_{EVAL} = P * \frac{1 * TC_{Lear} + 1 * TC_{Test}}{2}$

Also, we fix the values of (1,1) for (α1, α2) respectively. In this case, we assume that ability to train the classifier and the ability to generalize have the same importance.

$$fobj = penalty * [(1 * TCtr) + (1 * TCTest)] / 2 \quad (6)$$

Where:

$$TCtr = Rate\ FPtr + Rate\ FNtr + Rate\ UCtr \quad (7)$$

$$TCtest = Rate\ FPtest + Rate\ FNtest + Rate\ UCtest \quad (8)$$

Table 3 below shows values for (α-, α+, β-, β+) factors when the IHBA metaheuristic found the optimal values of f_{eval}. Please, note that initial values of (α-, α+, β-, β+) were equal to 0.

TABLE3. INITIAL & FINAL PARAMETERS VALUES

Table 2 presents that the used metaheuristics, partically the IHBA metaheuristic, found the minimal value of f_{eval} equal to 60.7, 61.00, 63.7, 54.9 for the SA, PSO, GA and IHBA metaheuristics respectively. In addition, the average value of f_{eval} on the liver disorders dataset were 60.07% . This value of f_{eval} were optimal than the average value obtained by the neuro-fuzzy

Metaheu	Training Results				Validation Results				f_{eval}	Imp
	FP	FN	UC	TC_{tr}	FP	FN	UC	TC_{tesr}		(%)
Neuro-F	7.7	0.3	91.5	295	52	0.0	45	238	290	
SA	11	7.3	0.0	168	27	0.0	69.7	263	92.4	68.1
PSO	22	0.7	0.0	58	10	6.9	76.7	388	92.4	68.1
GA	22	3.5	0.0	144	32	0.0	67.4	266	134	53.8
IHBA	7.7	0.0	0.0	15.4	10	0.0	65.2	215	33.6	88.4

system by about 38.8%.

- ***Scenario2*** **:** In this second scenario, we assume the consideration where the false positive rate would be

Metaheu	Training Results				Validation Results				f_{eval}	Imp
	FP	FN	UC	TC_{tr}	FP	FN	UC	TC_{tesr}		(%)
Neuro-F	7.7	0.3	91.5	47.1	52	0.0	45	313	288	
SA	13	6.5	0.0	97.5	12	8.1	75.6	96.3	101	64.9
PSO	22	0.7	0.0	134	10	6.9	76.7	80.7	89.8	65.7
GA	17	6.1	0.0	120	23	1.1	65.1	141	146	49.3
IHBA	30	1.9	0.0	185	0.0	1.4	71.0	4.32	21.8	92.4

more penalized than the false negative cost. Also, we ignore the penalization of the unclassifiable case.

$$fobj = penalty * [(0.1 * TCtr) + (1 * TCTest)] / 1.1 \quad (9)$$

Where:

$$TCtr = 6 * Rate\ FPtr + 3 * Rate\ FNtr \quad (10)$$

$$TCtest = 6 * Rate\ FPtest + 3 * Rate\ FNtest \quad (11)$$

TABLE 4. RESULTS FOR MINIMIZING $F_{EVAL} = P * \frac{0.1 * TC_{Lear} + 1 * TC_{Test}}{1.1}$

(α1, α2) values were set to (0.1, 1) respectively. This consideration seems to be realistic, due the fact that the classifier learning ability is less relevant than the generalization ability. The obtained results under this scenario are summarized in Table 4. This table presents that the obtained f_{eval} average value was equal to 89.65%. Moreover, the (SA, PSO, GA, and IHBA) metaheuristics reached optimal values of f_{eval} equal to 69.6%, when compared to the f_{eval} found by the neuro-fuzzy system.

Parameter	Before Optimization	After Optimization
α-	0	19.26
α+	0	11.9
β-	0	0.21
β+	0	0.93

- ***Scenario 3***: Now, we suppose that the (FP, FN, UC) rates are penalized by two, twenty and tree units respectively. In this case, we assume that the FN rate is more penalized than the false positive and the unclassifiable cases.

$$fobj = penalty * [(1 * TCtr) + (0.1 * TCTest)] / 1.1 \quad (12)$$

Where:

$$TCtr = 2Rate\ FPtr + 20Rate\ FNtr + 3Rate\ UCtr \quad (13)$$

$$TCtest = 2Rate\ FPtest + 20Rate\ FNtest + 3Rate\ UCtest \quad (14)$$

TABLE 5. RESULTS FOR MINIMIZING $F_{EVAL} = P * \frac{1 * TC_{Lear} + 0.1 * TC_{Test}}{1.1}$

Also, we fix the values of (1, 0.1) for (α1, α2) factors, to study the case where the learning ability is much more important than the generalization ability. The obtained results under this scenario are presented in Table 5. This table presents that the average value of the f_{eval} on the LD dataset were 88,1. This value of f_{eval} were less than the average value of f_{eval} , obtained by the neuro-fuzzy system by about 69.6%.

- ***Scenario 4***: In this last scenario, we study the case where the unclassifiable cost is much more penalized than the others costs errors. The (FP, FN) costs are penalized by (2, 3) units respectively. The results of this scenario are presented in Table 6. Here, we also fix the values of (1, 1) for (α1, α2) respectively.

$$fobj = penalty * [(1 * TCtr) + (1 * TCTest)] / 2 \quad (15)$$

Where:

$$TCtr = 2Rate\,FPtr + 3\ Rate\,FNtr + 20Rate\,UCtr \quad (16)$$

$$TCtest = 2Rate\,FPtest + 3Rate\,FNtest + 20Rate\,UCtest (17)$$

The results under this scenario are summarized in Table 6. This table proves that the obtained f_{eval} average value was set to 769. Moreover, the (SA, PSO, GA, and IHBA) metaheuristics found optimal values of f_{eval} equal to 48.6%, when compared to neuro-fuzzy objective function value

TABLE6. RESULTS FOR MINIMIZING $F_{EVAL} = P * \frac{1*TC_{Learn} + 1*TC_{Test}}{2}$

Metaheu	Training Results				Validation Results				f_{eval}	Imp
	FP	*FN*	*UC*	TC_{tr}	*FP*	*FN*	*UC*	TC_{tesr}		(%)
Neuro-F	7.7	0.3	91.5	1846	52	0.0	45	1004	1496	
SA	13	5.8	0.0	43.4	13	9.4	74.4	1542	832	44.4
PSO	22	0.7	0.0	24.1	10	6.9	76.7	1574	839	43.9
GA	17	6.1	0.0	52.3	23	1.1	65.1	1351	736	50.8
IHBA	7.7	0.0	0.0	15.4	10	0.0	65.2	1324	669	55.3

According to tables 2-4-5-6, it clearly appears that the IHBA f_{eval} values outperform the other metaheuristics (SA, PSO, GA) and the neuro- fuzzy inference system. However, the IHBA and the PSO metaheuristics show higher temporal complexity than SA and GA approaches. Moreover, SA was slightly better than the PSO metaheuristic. GA was slightly worse than the SA and the PSO metaheuristics. This proves that the IHBA approach can work well in diagnosis of the liver disorders disease.

5. Conclusion

This work aims to prove how metaheuristics optimization approaches, constitutes a good tool for enhancing traditional data mining classifiers. Particularly, the neuro-fuzzy system was used in conjunction with some well-known metaheuristics on the prediction of the liver disorders disease. We have tested Simulated Annealing, Genetic Algorithm, Particle Swarm Optimization approaches and our previous proposed metaheuristic IHBA [1] on the LD benchmark obtained from the UCI repository. The obtained computational results strongly show that Neuro-Fuzzy classifier in conjunction with the used metaheuristics can be of great interest to doctors for their diagnosis task. Moreover, the results show the advantage of the IHBA over GA, PSO and SA algorithms. The future works will pay more attention to test other metaheuristics optimization approaches in other medical datasets, in particular, big data.

References

[1] Fatima Bekaddour, Mohammed Amine Chikh. IHBA: An Improved Homogeneity-Based Algorithm for Data Classification. Computer Science and Its Applications. Vol. 456 of the series IFIP Advances in Information and Communication Technology , pp 129-140, 2015.

[2] X.S.Yang, S.Deb. Cuckoo search via levy flights, in: World Congresson Nature & Biologically Inspired Computing (NaBIC2009). IEEE Publications, pp 210–214, 2009.

[3] A.H.Gandomi, A.H.Alavi. Krill herd: a new bio-inspired optimization algorithm. Communications in Nonlinear Science and Numerical Simulation 17(12), pp. 4831–4845, 2012 http://dx.doi.org/10.1016/j.cnsns.2012.05.010

[4] Alatas, B. A novel chemistry based metaheuristic optimization method for mining of classification rules. Expert Syst. Appl. 39, pp. 11080-11088, 2012.

[5] Osaba, E., Diaz, F., Onieva, E.: Golden ball: a novel meta-heuristic to solve combinatorial optimization problems based on soccer concepts. Applied Intelligence 41, pp. 145-166 , 2014.

[6] Kirkpatrick, S., Gelatt, C.D., Jr., and Vecchi,. Optimization by simulated annealing. Science 220 , pp. 671-680, 1983.

[7] P J. Kennedy and R. C. Eberhart. Swarm Intelligence. Morgan Kaufmann, San Francisco, CA, 2001.

[8] J.H. Holland (1975). Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, Michigan; re-issued by MIT Press 1992.

[9] Lee, Y. J., Mangasarian, O. L. RSVM: Reduced support vector machines. In Proceedings of the first SIAM international conference on data mining, Chicago, IL, USA, 2001.

[10] Kemal Polat, Seral Sahan, Halife Kodaz, Salih Günes. Breast cancer and liver disorders classification using artificial immune recognition system (AIRS) with performance evaluation by fuzzy resource allocation mechanism. Expert Syst. Appl. 32(1): pp. 172-183 , 2007.

[11] Alia, O.M., Al-Betar, M.A., Mandava, R., Khader, A.T. Data Clustering Using Harmony Search Algorithm. In: Panigrahi, B.K., Suganthan, P .N., Das, S., Satapathy, S.C. (eds.) SEMCCO 2011, Part II. LNCS, Vol. 7077, pp. 79–88. Springer, Heidelberg 2011 .

[12] Ivanoe De Falco. Differential Evolution for automatic rule extraction from medical databases. Applied Soft Computing, Vol. 13, pp. 1265–1283, 2013.

[13] Pham HNA, Triantaphyllou E. The impact of overfiting and overgeneralization on the classificaton accuracy in data mining. In: Maimon O,Rokach L, editors. Soft computing for knowledge discovery and data mining, pp. 391–431, part 4, chapter 5, New York, NY, USA: Springer, 2007.

[14] Pham HNA, Triantaphyllou E. Prediction of diabetes by employing a new data mining approach which balances fitting and generalization. In: Yin Lee R, editor. Studies in computa-tion intelligence, Vol. 131, pp. 11–26, chapter 2, Berlin, Germany: Springer, 2008.

[15] Pham HNA, Triantaphyllou E. An application of a new meta-heuristic for optimizing the classification accuracy when analyzing some medical datasets. Expert Systems with Applications, Vol. 36, number 5, pp. 9240–9249, 2009.

[16] UCI repository of machine learning databases. University of California at Irvine, Department of Computer Science, http://archive.ics.uci.edu/ml/datasets/Liver+Disorders, last accessed (2016).