Immune Grey Wolf Optimizer for Attribute Reduction: Application for Medical Systems

Badra Khellat Kihel¹,², Souad Khellat Kihel¹, Samira Chouraqui¹

¹Department of Informatics, Faculty of Mathematics and Informatics, Université des Sciences et de la Technologie d’Oran Mohamed-Boudiaf USTO-MB, B.P. 1505 EL M’NAOUER 31000 Oran, Algeria
²Department of Economics, Oran2 University Oran, Algeria

* Corresponding author’s Email: badra.khellat@univ-usto.dz

Abstract: The sizes of medical datasets are increased with an exponential rate. Extract useful knowledge from those data, analyzing and searching on this very vast amount of data is become a hard task for the scientists. Several metaheuristic algorithms for solving such problem were proposed for dimensionality reduction. In this work, a new stochastic search strategy inspired by the Grey Wolf optimization theory is proposed for feature subset selection. Grey Wolf Optimization algorithm (GWO) is a new metaheuristic optimization technique. Its principle is to reproduce the behavior of grey wolves in nature to hunt in a cooperative way. In order to improve the disadvantage of Grey Wolf Optimizer when solving the feature selection problem, three immune operators (selection, cloning and mutation) are embedded into the standard GWO and an Hybrid Immune Grey Wolf Algorithm (IGWO) is proposed. To ensure the diversity of the population of wolves and avoid premature convergence, we cloned and mutated Alpha, Beta and Delta wolves before selection of the new first, second, and third best solutions. The experiments were performed using ten medical datasets: Parkinson, Leukemia and heart diseases, Breast, Colon, and Prostate Cancers, and the results show that the proposed approach is able to explore different regions of a search space and avoid local optima reducing the dimensionality of datasets than can significantly improve the performances of the system for diseases recognition and decrease the required classification time.

Keywords: Grey wolf optimizer, Artificial immune system, Clonal selection, Medical system, Feature selection.

1. Introduction

In a computer aided diagnosis system, there is always a large amount of data which must be all saved and used in learning techniques to classify healthy and pathological patients. Achieving a successful diagnostic process must take into account several considerations related to knowledge extraction, compact representation and data analysis.

The problem of dimensionality reduction in medical system has been widely investigating due to its importance. Feature selection (FS) allows the reduction of feature space, which is crucial in reducing the training time and improving the prediction accuracy. This is achieved by removing irrelevant, redundant, and noisy features (i.e. selecting the subset of features that can achieve the best performance in terms of accuracy and computational time) [1, 2].

Many scholars have researched on feature selection methods. Two main methods for feature selection can be distinguished: Wrapper approaches include a learning-classification algorithm in the evaluation procedure, while filter approaches do not include such algorithm. Filter approaches are argued to be computationally less expensive and more general, while wrapper approaches can usually achieve better results [3].

Recently, an optimization technique called grey wolf optimizer algorithm has gained the interest of many researchers. This algorithm is a type of swarm intelligence algorithm based on the position change
of wolves according to the experience of leaders: Alpha, Beta and Delta [4, 5].

In this way, the wolves will gradually become the same as the leaders. To avoid premature convergence, we proposed a new Immune Grey Wolf Optimizer (IGWO), using immunological operators (cloning, mutation and selection) on Alpha, Beta and Delta wolves to update the wolves’ position, to select the best feature subset which include a small number of features and achieve a lower classification error rate than using all available features. The selected methods were evaluated using ten well-known medical datasets. The obtained results are promising and confirm the interest of using bio-inspired approaches for data mining in bioinformatics.

The rest of this paper is organized as follows: section 2 reviews existing works. Section 3 introduces bio-inspired algorithms (GWO and AIS). The proposed Immune grey wolf Optimizer Algorithm is clearly described in section 4. Section 5 illustrates and analyzes the experimental results. Finally, section 6 concludes the paper.

2. Related works

Recently, many FS based on new optimizers were proposed. Chen et al. proposed a novel rough set based method for feature selection using fish swarm algorithm [7]. The fish swarm algorithm is used for optimization through imitating the fish behaviours such as swarming, following and moving randomly. Mirjalili proposed a wrapper-feature selection approach that uses the binary dragonfly algorithm as a search strategy and the K-Nearest Neighborhood (KNN) classifier as an evaluator [8]. Another recent optimizer is Ant Lion that mimics the performances of ant lions in hunting prey is used as a wrapper feature selection method in [9, 10]. Firefly algorithm (FFA) was combined with artificial immune system (IFA) to reduce dimensionality in voluminous datasets in [11] to prevent the population from being trapped into local optimum. Grey Wolf Optimizer (GWO) is a popular swarm intelligence technique that has been successfully used for solving feature selection problems [4, 5]. In [4], GWO is used to find the best features in the Coronary artery disease identification dataset then Support Vector Machine classifier (SVM) for the evaluation of the GWO fitness function.

Grey wolf optimizer (GWO) has received widespread attention from scholars. The algorithm was inspired by grey wolves’ predation activity in nature. This algorithm has gained much interest in the hybrid metaheuristics field. In [13] authors proposed a multi-objective grey wolf optimizer in order to optimize problems with multiple objectives with a novel leader selection mechanism. An evolutionary GWO algorithm was proposed in [14] to resolve the distributed generation allocation problem using genetic algorithm. In [15], the authors proposed a hybrid GWO and Artificial Bee Colony (ABC) to improve the complex systems performance. Al-Tashi et al. proposed a binary version of hybrid particle swarm optimization PSO-GWO to solve feature selection problem [16]. In another study PSO was combined with GWO to outperform the convergence speed of the system [17]. Kohli et al. introduce the chaos theory into the GWO for the purpose of accelerating the global convergence speed [18]. These studies have revealed that the hybrid algorithms performed much better compared to standard metaheuristic methods.

In this work, we were particularly attracted by the use of bio-inspired methods and there hybridization for dimensionality reduction in medical systems. We have used two algorithms: Grey Wolf Optimizer and Artificial Immune System (AIS) to select the most relevant features in medical datasets, then we produce a new hybrid Immune Grey Wolf Optimizer (IGWO) in which we used GWO to select the three best wolves: Alpha, Beta and Delta then we used AIS in particular cloning, mutation and selection operators for the three best wolves to increase the global search space. The main aim of this hybridization is to study the influence of the initial population quality on the searching progress of GWO in feature selection task.

3. Methods

3.1 Grey Wolf Optimizer (GWO)

Grey Wolf Optimizer is a recent metaheuristic, which offers better performance in feature selection task. However, the new positions of wolves are mostly based on the experience of three solutions or leaders: Alpha, Beta and Delta in the position update [4].

In the feature selection problem, a representation for candidate feature subset must be chosen and encoded. In most studies, wolf is a binary string of length equal to the total number of features so that each bit encodes a single feature. A bit of ‘1’ (‘0’) implies the corresponding feature is selected (excluded).

For modeling the social hierarchy of wolves, the leaders of the park are considered as the alpha (α). The main responsibility of alpha is making decisions.
Beta (β) known to assist alpha in making decisions and the main responsibility of beta is the feedback suggestions. Delta (δ) performs as scouts, and controls omega (ω) wolves by obeying alpha and beta wolves. The omega wolves must obey every other wolf.

In GWO, α, β, and δ guides the hunting process and ω wolves follows them. The encircling behavior for the pack to hunt a prey can be expressed as

\[ X(t + 1) = X_p(t) - A.D \quad (1) \]

Where \( X_p \) is the position of prey, \( D \) is defined as:

\[ D = \left| C \cdot X_p(t) - X(t) \right| \quad (2) \]

Where \( X \) is the position of grey wolf, and \( t \) is the number of iterations. \( A \) and \( C \) are coefficient vectors calculated as follows:

\[ A = 2a \cdot r_1 - a \quad (3) \]

\[ C = 2 \cdot r_2 \quad (4) \]

Where \( r_1 \) and \( r_2 \) are independent random numbers between [0,1], and \( a \) is the encircling coefficient that is linearly decreased, from 2 to 0.

The leaders are guiding the omega wolves to move toward the optimal position. Therefore, we save the first three best solutions obtained and force the other search agents to update their positions according to the position of the best search agents.

\[ D_α = \left| C_1 \cdot X_α - X \right| \quad (5) \]

\[ D_β = \left| C_2 \cdot X_β - X \right| \quad (6) \]

\[ D_δ = \left| C_3 \cdot X_δ - X \right| \quad (7) \]

Mathematically, the new position of wolf is:

\[ X(t + 1) = \frac{X_1 + X_2 + X_3}{3} \quad (8) \]

Where \( X_1 \), \( X_2 \) and \( X_3 \) are the three best wolves in the swarm at a given iteration \( t \) calculated as follows:

\[ X_1 = \left| X_α - A_1 \cdot D_α \right| \quad (9) \]

\[ X_2 = \left| X_β - A_2 \cdot D_β \right| \quad (10) \]

\[ X_3 = \left| X_δ - A_3 \cdot D_δ \right| \quad (11) \]

Fig. 2 shows how a search agent updates its position according to alpha, beta, and delta in a search space.

3.2 Artificial Immune System (AIS)

The artificial immune systems work on three basic immunological principles which include clonal selection theory, negative selection principles and the mechanism of immune network [11, 20]. The main idea of the Clonal selection method is to multiply only the cells whose antibodies are able to recognize the antigens [21].

In order to clarify how an immune response is mounted when a nonself antigenic pattern is recognized by a B cell, the clonal selection theory has been developed. The clonal selection algorithm can be described as follows:

**Step 1:** Generate a set \( P \) of candidate solutions, composed of the subset of memory cells \( M \) added to the remaining population \( Pr \).

**Step 2:** Select the \( n \) best individuals of the population, based on an affinity measure.

\[ Affinity = \max \{ \text{accuracy} (\text{sub}) \} \quad (12) \]
Where accuracy (sub) evaluates the classification accuracy of the subset.

**Step 3:** Clone the n best individuals of the population to a temporary population of clones $C$ according to their affinity. The number of clones for each antibody $i$ is computed by

$$NC(i) = \text{round} \left( \frac{B \cdot \text{affinity}_i^2}{\sum_j \text{affinity}_j^2} \right) \quad (13)$$

Where $B$ is the cloning parameter, affinity is calculated in Eq. (12) and round () is the operator that rounds its argument to closest integer.

**Step 4:** A matures antibody population is generated $C^*$.

The mutation changes according to the representation of the data; in this direction we find various types of mutation in the case of real representation or binary presentation (which is our case).

In the first step we must calculate the number of bits to be inverted ($Nm$) using:

$$Nm = \text{round} \left( (L \quad - \quad \text{affinity}_C(i)) \cdot \text{rand}() \right) \quad (14)$$

Where $L$ represents the antibody size and $\text{affinity}_C$ is the vector of affinity calculated after cloning. Rand () is a mathematical function used to generate a random number that is greater than or equal to 0 and less than 1.

In binary valued individuals mutation is the flipping of variable values (inverse ‘0’ to ‘1’ and vice versa),

**Step 5:** Re-select the improved individuals from $C^*$ to compose the memory set $M$. Some members of $Ab$ can be replaced by other improved members of $C^*$.

**Step 6:** Replace $d$ antibodies by new ones (diversity introduction). The lower affinity cells have higher probabilities of being replaced [11, 20, 21].

4. **Hybrid system IGWO**

Several bio-inspired hybrid methods were proposed to perform FS problems. In this context, after a detailed study of bio-inspired approaches, we propose immunological and swarm intelligence methods for feature selection. Our study of immunological approaches oriented us to propose the Artificial Immune System (AIS) and our study of swarm intelligence methods guided us to opt Grey Wolf Optimizer (GWO) then their hybridization (IGWO).

In this work, we have used Grey Wolf Optimizer and Artificial Immune System to select the most relevant features with the use of k-nearest neighbors in each iteration to evaluate the error rate of the selected features.

The IGWO's basic idea is to increase the algorithm’s capability to exploit AIS with the ability to explore GWO to achieve good results. Initially, the population of grey wolves is randomly initialized (either bit 1 or 0). Afterward, the fitness of each wolf is evaluated using $Knn$. The best, second, and third best solutions defined as alpha, beta, and delta will be cloned then mutate using Eq.13 and Eq.14. Then, the position of wolf is updated by using Eq.8 on the selected three best wolves $X_\alpha, X_\beta$ and $X_\delta$ from the mutated population.

Next, the fitness of each wolf is evaluated and the positions of alpha, beta, and delta are updated. The algorithm is repeated until the stopping criterion is satisfied (number of iterations). Finally, the alpha solution is selected as the optimal feature subset.

5. **Experimental results**

5.1 Medical datasets used

The proposed system was evaluated on ten medical datasets: Parkinson, Leukemia and heart diseases, Breast, Colon, and Prostate Cancers.

5.2 Parameters tuning

In order to approve our system, few parameters tuning has to be made. Several parameters for GWO and AIS are required to be tuned to find the best value that can produce the optimum or near optimum.
Table 1. Datasets used

<table>
<thead>
<tr>
<th>Data set</th>
<th>Number of Features</th>
<th>Number of Classes</th>
<th>Number of Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spect Heart (Binary)</td>
<td>23</td>
<td>02</td>
<td>187</td>
</tr>
<tr>
<td>Spect Heart</td>
<td>44</td>
<td>02</td>
<td>187</td>
</tr>
<tr>
<td>Parkinson’s Disease</td>
<td>22</td>
<td>02</td>
<td>96</td>
</tr>
<tr>
<td>Parkinson 2</td>
<td>26</td>
<td>02</td>
<td>1040</td>
</tr>
<tr>
<td>Breast tissue</td>
<td>09</td>
<td>06</td>
<td>106</td>
</tr>
<tr>
<td>Wisconsin Breast Cancer</td>
<td>09</td>
<td>02</td>
<td>699</td>
</tr>
<tr>
<td>Wisconsin Breast Cancer Diagnostic</td>
<td>31</td>
<td>02</td>
<td>569</td>
</tr>
<tr>
<td>Colon</td>
<td>2000</td>
<td>02</td>
<td>62</td>
</tr>
<tr>
<td>Leukemia</td>
<td>7070</td>
<td>02</td>
<td>72</td>
</tr>
<tr>
<td>Prostate</td>
<td>5966</td>
<td>02</td>
<td>102</td>
</tr>
</tbody>
</table>

The population size of grey wolves is taken 10 and the maximum number of iterations is fixed at 100 iterations. The values of two vectors r1 and r2 are taken randomly in the range of (0, 1) and the controlling parameter a has linearly decreasing values from 2 to 0 over the course of iterations.

For AIS, the cloning parameter, b, is set at 0.1, the population size, N, and maximum number of iterations, T, are fixed at 10 and 100, respectively.

5.3 Results and discussion

In this section we present the experimental results obtained for the different approaches on datasets. Table 2 presents the results of the IGWO algorithm for feature selection in terms of number of selected features, classification error and CPU time. The IGWO is run 10 times on the ten medical databases and the statistics results of these 10 runs are provided Table 2.

In addition, the IGWO algorithm is compared with AIS and GWO algorithms in Table 3.

The results show that the proposed method is able to produce good performance on reducing the effects of the outliers and the noises and improve classification accuracy for disease recognition. The features selected can accomplish the goal of achieving higher accuracy with smaller size of features.

The results of the proposed algorithm show that the use of immunological operators for Alpha, Beta and Delta wolves influence the robustness and the algorithm convergence, by using the cloning, mutation and selection operators to calculate position that ensures diversity of the population.

Fig. 6 shows the Box-plot of error rates and number of selected features of the used feature selection methods for different datasets.

Metaheuristic methods have the limitation that, they often need tuning on the part of the user, to work correctly. Also, being random in nature, the solution produced is often nearly the global optimum. IGWO performs so well that, the results produced by it match the optimal by exhaustive search.

5.4 Comparative study

In this section, the results of the conducted experiments on the proposed systems are compared to some existed works in the literature. Table 4 depicts the error classification rate (Err) and the number of selected features (NF) from the proposed system as well as the existing swarm. The results showed a good performance of the proposed approach comparing to the existing algorithms. This is mainly due to the high exploration of the IGWO, because of the position for updating equations of GWO when combining the three best solutions Alpha, Beta and Delta.

6. Conclusion

In this work, a dimensionality reduction system has been evaluated on medical systems for the aim of disease recognition approving. Two metaheuristic algorithms have been studied such as Artificial
Immune System (AIS) and Grey Wolf Optimizer (GWO). Then, a new hybrid variant is proposed by using strengths of GWO and AIS.

The proposed algorithm was tested over ten different medical datasets: Parkinson, Leukemia and heart diseases, Breast, Colon, and Prostate Cancers. The obtained results after reduction are very encouraging and prove that hybrid approach is more reliable in giving best results as compared to some bio-inspired algorithms: AIS, GWO, ABC, ACO, FFA and IFA.

As future work, the proposed algorithm can be experimented with other classifiers such as Support Vector Machine (SVM) or Artificial Neural Network (ANN).

Table 2. Results for hybrid IGWO

<table>
<thead>
<tr>
<th>Data set</th>
<th>Number of Features (NF)</th>
<th>Error rate (Err)</th>
<th>CPU time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best</td>
<td>Worst</td>
<td>Average</td>
</tr>
<tr>
<td>Dataset1</td>
<td>11</td>
<td>14</td>
<td>12.2</td>
</tr>
<tr>
<td>Dataset2</td>
<td>12</td>
<td>20</td>
<td>15.6</td>
</tr>
<tr>
<td>Dataset3</td>
<td>8</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Dataset4</td>
<td>7</td>
<td>13</td>
<td>10.4</td>
</tr>
<tr>
<td>Dataset5</td>
<td>4</td>
<td>6</td>
<td>4.8</td>
</tr>
<tr>
<td>Dataset6</td>
<td>6</td>
<td>7</td>
<td>5.8</td>
</tr>
<tr>
<td>Dataset7</td>
<td>9</td>
<td>13</td>
<td>10.4</td>
</tr>
<tr>
<td>Dataset8</td>
<td>823</td>
<td>1011</td>
<td>880.6</td>
</tr>
<tr>
<td>Dataset9</td>
<td>2520</td>
<td>3132</td>
<td>2762.2</td>
</tr>
<tr>
<td>Dataset10</td>
<td>2124</td>
<td>2530</td>
<td>2359</td>
</tr>
</tbody>
</table>

Figure 5: Immune Grey Wolf Optimizer Algorithm
Table 3. Results for algorithms used (GWO, AIS, IGWO)

<table>
<thead>
<tr>
<th>Data set</th>
<th>GWO</th>
<th>AIS</th>
<th>IGWO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NF</td>
<td>Err</td>
<td>NF</td>
</tr>
<tr>
<td>Dataset1</td>
<td>10</td>
<td>0.325</td>
<td>9</td>
</tr>
<tr>
<td>Dataset2</td>
<td>16</td>
<td>0.150</td>
<td>21</td>
</tr>
<tr>
<td>Dataset3</td>
<td>11</td>
<td>0.020</td>
<td>12</td>
</tr>
<tr>
<td>Dataset4</td>
<td>14</td>
<td>0.341</td>
<td>13</td>
</tr>
<tr>
<td>Dataset5</td>
<td>04</td>
<td>0.264</td>
<td>05</td>
</tr>
<tr>
<td>Dataset6</td>
<td>06</td>
<td>0.025</td>
<td>07</td>
</tr>
<tr>
<td>Dataset7</td>
<td>07</td>
<td>0.042</td>
<td>16</td>
</tr>
<tr>
<td>Dataset8</td>
<td>840</td>
<td>0.145</td>
<td>1022</td>
</tr>
<tr>
<td>Dataset9</td>
<td>3150</td>
<td>0.069</td>
<td>3547</td>
</tr>
<tr>
<td>Dataset10</td>
<td>2552</td>
<td>0.078</td>
<td>2951</td>
</tr>
</tbody>
</table>

Figure 6 Box-plot for feature selection methods (a) Box-plot of error rate (b) Box-plot of number of selected features for small datasets and (c) for big datasets.

Figure 7 Number of selected features for comparative study

Figure 8 Error rates for comparative study

Table 4. Comparative study

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NF</td>
<td>Err</td>
<td>NF</td>
<td>Err</td>
<td>NF</td>
</tr>
<tr>
<td>Dataset1</td>
<td>9</td>
<td>0.102</td>
<td>19</td>
<td>0.111</td>
<td>15</td>
</tr>
<tr>
<td>Dataset2</td>
<td>34</td>
<td>0.062</td>
<td>34</td>
<td>0.080</td>
<td>19</td>
</tr>
<tr>
<td>Dataset3</td>
<td>15</td>
<td>0.022</td>
<td>14</td>
<td>0.023</td>
<td>13</td>
</tr>
<tr>
<td>Dataset4</td>
<td>16</td>
<td>0.169</td>
<td>20</td>
<td>0.170</td>
<td>17</td>
</tr>
<tr>
<td>Dataset5</td>
<td>7</td>
<td>0.332</td>
<td>9</td>
<td>0.510</td>
<td>7</td>
</tr>
<tr>
<td>Dataset6</td>
<td>7</td>
<td>0.017</td>
<td>8</td>
<td>0.019</td>
<td>7</td>
</tr>
<tr>
<td>Dataset7</td>
<td>21</td>
<td>0.010</td>
<td>26</td>
<td>0.010</td>
<td>18</td>
</tr>
</tbody>
</table>
Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, writing-original draft preparation, writing-review and editing, visualization, have been done by 1st and 2nd authors.

The supervision and project administration have been done by 3rd author.

References

