**TELKOMNIKA Telecommunication, Computing, Electronics and Control**
Vol. 18, No. 5, October 2020, pp. 2421~2431
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v18i5.15141 

# Integrated bio-search approaches with multi-objective algorithms for optimization and classification problem

**Mohammad Aizat Basir[1], Mohamed Saifullah Hussin[2], Yuhanis Yusof[3]**
[1,2]Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, Malaysia
[3]School of Computing, Universiti Utara Malaysia, Malaysia

**Article Info**

***Article history:***

Received Jan 2, 2020
Revised May 1, 2020
Accepted May 14, 2020

***Keywords:***

Bio-inspired
Classification
ENORA
Feature selection
NSGA-II

**ABSTRACT**

Optimal selection of features is very difficult and crucial to achieve, particularly for the task of classification. It is due to the traditional method of selecting features that function independently and generated the collection of irrelevant features, which therefore affects the quality of the accuracy of the classification. The goal of this paper is to leverage the potential of bio-inspired search algorithms, together with wrapper, in optimizing multi-objective algorithms, namely ENORA and NSGA-II to generate an optimal set of features. The main steps are to idealize the combination of ENORA and NSGA-II with suitable bio-search algorithms where multiple subset generation has been implemented. The next step is to validate the optimum feature set by conducting a subset evaluation. Eight (8) comparison datasets of various sizes have been deliberately selected to be checked. Results shown that the ideal combination of multi-objective algorithms, namely ENORA and NSGA-II, with the selected bio-inspired search algorithm is promising to achieve a better optimal solution (i.e. a best features with higher classification accuracy) for the selected datasets. This discovery implies that the ability of bio-inspired wrapper/filtered system algorithms will boost the efficiency of ENORA and NSGA-II for the task of selecting and classifying features.



***Corresponding Author:***

Mohammad Aizat Basir,
Faculty of Ocean Engineering Technology and Informatics,
Universiti Malaysia Terengganu (UMT),
21030 Kuala Nerus, Terengganu, Malaysia.
Email: aizat@umt.edu.my

## 1. INTRODUCTION

Enormous dataset normally consists of a large number of attributes. These attributes are repetitive/irrelevant on a regular basis and influence the data mining model. In cases where the rule has so many constraints, with a wide number of characteristics, the rule becomes more complicated and difficult to understand. By understanding this problem, it is important to the the number of features to be used in the creation of information mining models. In realistic situations, it is proposed that the obsolete and redundant measurements should be removed in order to minimize processing time and labor costs. In [1] claimed that a dataset with a large number of attributes is known as a dataset with a high dimensionality. This condition would lead to the curse of the dimensionality theorem, where the time of measurement is the exponential function of the number of dimensions. In addition, the high dimension of space searching leads to the redundancy of features in the model. The ultimate solution is to reduce the search dimension while preventing the loss of vital information in the results. Large number of attributes in each potential rule can create ambiguous representation, making it difficult to understand, use, and exercise. The complexity of the attribute can then be minimized by reducing the number of attributes and removing irrelevant attributes that will increase processing time and boost storage performance.

Feature selection (FS) is defined in [2] as the process of removing features from the database that are irrelevant to the task to be performed. Feature selection promotes data comprehension, reduces calculation and storage requirements, reduces computational process time, and reduces the size of the data collection, making model learning easier. FS has become increasingly popular in applications in genomics, health sciences, economics, banking, among others [3-5] as well as in psychology and social sciences [6, 7].

Feature selection algorithms categorized into 2 main group: supervised, unsupervised and semi-supervised; this relies upon whether the training set is, or not, labelled. Feature selection models are also categorized into filter, wrapper and embedded models. The first ones apply statistical measures to assign a score to each feature; features are ranked by their score, and either selected to be kept or removed from the data set. Filter models do not interact with learning algorithms, and they can be univariate (when features are evaluated one by one) or multivariate (when they are evaluated in subsets). Wrapper methods define the selection of a set of features as a search problem, where different combinations are prepared, evaluated and compared to other combinations. Finally, the underlying idea of embedded models is learning which features best contribute to th accuracy of the model while the model is being created.

Feature selection consists of four stages, typically referred to as subset creation, subset evaluation, stop criterion, and result validation.During the phase of subset evaluation the goodness of a subset produced by a given subset generation procedure is measured. Examples of subset evaluation measures for multivariate filter methods are the distance [8], the uncertainty [9], the dependence [10], and the consistency [4], while wrapper methods mostly use the accuracy [11]. The stopping criterion establishes when the feature selection process must finish; it can be defined as a control procedure that ensures that no further addition or deletion of features does produce a better subset, or it can be as simple as a counter of iterations. Finally, in the phase of result validation the validity of the selected subset is tested.

A recent overview, categorization and comparison of existing methods for selecting features is shown in [12]. A significant downside to these techniques is that they only consider a single criterion when looking for a subset, and do not seek to limit the number to attributes chosen; they can then be referred to as single-objective feature selection methods. However, the single mechanisms do not suffice when the number of features is particularly high, and a separate feature selection process does improve the performances of the learned model.

Evolutionary (or genetic) computation uses a simple evolutionary metaphor. The problem, according to this metaphor, plays the function of an atmosphere in which a population of individuals resides, each representing a potential solution to the problem. The degree of adaptation of each person to his or her environment is expressed by a measure of adequacy known as fitness function. Unlike evolution in nature, evolutionary algorithms have the ability to slowly evolve solutions to the problem. Algorithms begin with an initial population of random solutions and, in each iteration, the best individuals are selected and combined using variation operators, such as crossovers and mutations, to create the next generation. The cycle is repeated until each of the stop criteria is met. Some problems involve multi-objective optimization (MO) in particular where there is an implicit tension between two or more problem objectives; the selection function, in which one must optimize the accuracy of the classifier and reduce the number of features, is an example of such a problem.

Multi-objective evolutionary algorithms [13, 14] have proven to be very successful in finding optimal solutions to multiple objective problems. Multi-objective evolutionary algorithms are especially suitable for multi-objective optimization because they look for multiple optimal solutions in parallel and are able to find a set of optimal solutions in their final population in a single sprint. When an optimal solution set is available, the most suitable solution can be chosen by applying a preference criterion. The goal of a multi-objective search algorithm, therefore, is to discover a family of solutions that are a good approximation to the Pareto front. In the case of multi-objective feature selection, each front-end solution may represent a subset of features with an related trade-off between, for example, accuracy and model complexity.

In multi-objective feature selection methods, two common methods are known as ENORA and NSGA-II. ENORA (evolutionary non-dominated radial slots based algorithm) is one of the multi-objective evolutionary algorithm selection techniques for random search [15, 16] with the following two objectives: minimizing the number of selected features and minimizing the root mean squared error (RMSE) of the Random Forest (RF) model, a well-known regression model learning algorithm [17]. In addition, the multi-objective evolutionary algorithm known as the NSGA-II (non-dominated sorted genetic algorithm) [18] is considered a norm in the multi-objective evolutionary computing community, both in terms of the hypervolume statistics of the last population and in terms of the RMSE of the chosen person. The NSGA-II wrapper solution is introduced for the identification of designated persons in [19].

A change in the dominant relationship is implemented in [20] to consider an arbitrary large number of goals and is used in a combination of NSGA-II, logistic regression, and naive Bayes with Laplace correction as classification algorithms. In [8], the selection of a multi-objective function is applied to a diagnostic issue in the medicine. For an application in engineering, a multi-objective algorithm that minimizes the error identification rate, undetected identification rate and the number of selected features is proposed in [9]. In [21] a multi-objective Bayesian artificial immune system is used for the selection of features in classification problems, with the goal of reducing both the classification error and the cardinality of the subset of features. In [10] a wrapper approach is proposed to optimize the data mining algorithm error rate and the model size of the learning algorithm using NSGA and NSGA-II. A multi-objective estimation of the distribution algorithm is proposed in [11] for the selection of a function subset based on a common modeling of objectives and variables. Figure 1 shows the complete flow of ENORA/NSGAII adapted from [22].

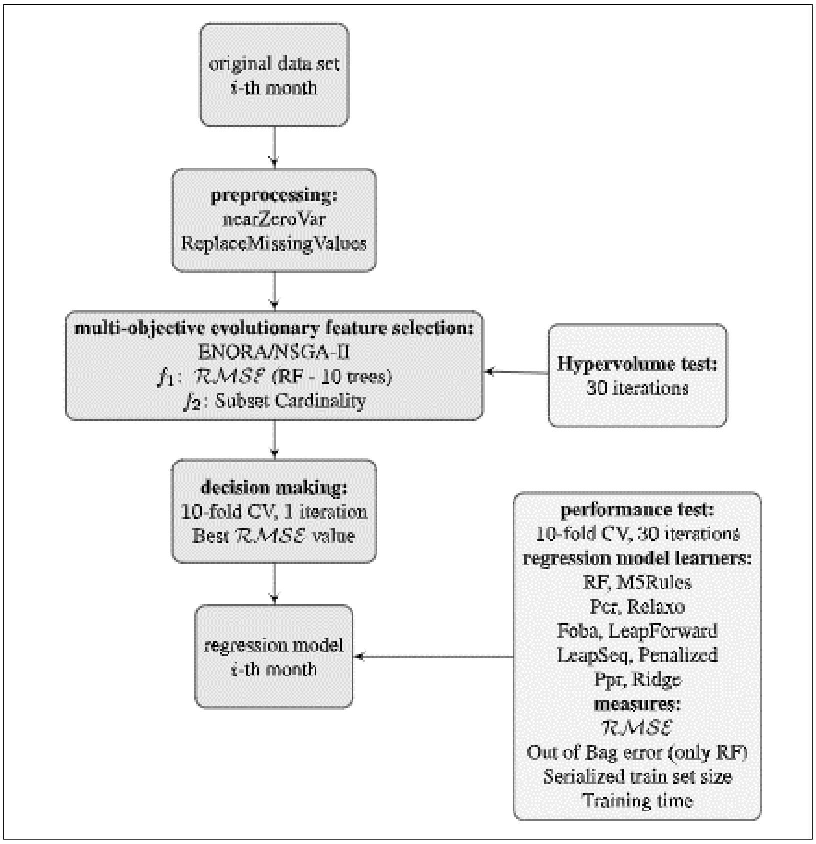


Figure 1. Flow chart of an ENORA/NSGA-II adapted from [22]

A multi-objective approach to the collection of function subsets using ACO and fuzzy has been proposed [23]. ACO was used in research to effectively solve the fuzzy multi-objective problem. Their work shows that the proposed approach can produce better subsets and achieve higher classification accuracy. ACO was also used with a genetic algorithm to pick a function for pattern recognition in [24]. The method consists of two interesting models, the visibility density model (VMBACO) and the pheromone density model (PMBACO) for the optimal solution for selecting and de-selecting features. Promising results have been obtained where the proposed approach demonstrates robustness and adaptive efficiency relative to other approaches. Similarly, the ant colony optimization (ACO) algorithm was used in the medical field to identify important features for the diagnosis of Raman-based breast cancer [25]. Experimental results demonstrated that ACO has the capability to boost the diagnostic accuracy of Raman-based diagnostic models. Similarly, ACO was used in the area of network security to detect intrusion [26]. Figure 2 presents basic pseudo-code of an ant algorithm.

New meta-heuristic algorithm artificial bee colony (ABC) [27] has been used for the collection of features in computed tomography (CT Scan) images of cervical cancer that help to recognize existing cancers.

For the handling of high dimensional problems, [28] suggested a new method of selection of features based on ABC with gradient-boosting decision tree. The research result has shown that the proposed method effectively reduces the size of the dataset and achieves superior classification accuracy by using the selected features. Similarly, the hybrid approach [29] used the ABC algorithm with a differential evolution algorithm to address the high dimensional problem. The developed hybrid approach demonstrates the ability to pick good features for the classification tasks and thus increases the run-time efficiency and accuracy of the classifier. A multi-objective artificial bee colony (MOABC) model has been developed [30]. The developed algorithm was incorporated with a fuzzy approach to evaluating the relevance of the function subsets. Experimental findings indicate a substantial contribution to seeking a successful subset of features. Figure 3 demontrates basic pseudo-code of bee algorithm.

```
Input: Instance x ϵ I of Π_opt
        Set algorithm parameters ()
        i, j ← 0
        for j= 1 to colonies do
            Ant s_0 ← Create sub-colony and release
             agent
             while not-termination conditions
                    on sub-colony do
                        i=i+1
                        Manage_ants activity ()
                        Manage_Pheromone ()
                        Manage_Demon Action ()
                        Selection Procedure ()
                        Compute solution Quality ()
             end while
             j=j+1
             S_best←candidate to be optimal solution
             Update pheromone on arc ()
         end for
Output: S_best"candidate" to be the best found
                solution x ϵ I
```

Figure 2. The basic pseudo-code of an ant algorithm

```
Input: Instance x ϵ I of Π_opt
        set algorithm parameters ()
        Initialize Pop_0←Initial population () with
        random solutions.
        Evaluate fitness of the population(Pop_0).
        while not termination condition do
                Select sites for neighbourhood search
                N(s)
                Recruit bees for selected sites (more
                bees for best e sites) and evaluate
                fitnesses.
                Select the fittest bee from each patch.
                Assign remaining bees to search
                randomly and evaluate their fitnesses.
        end while
         Generate a new population of scout bees
S_best←optimal solution Pop_i
Output: S_best "candidate" to be the best found solution
x ϵ I
```

Figure 3. The pseudo-code of a bee algorithm

Bat algorithm has been used effectively in engineering [31]. Multi-objective binary bat algorithm (MBBA) proposed by [32] modified bat position update strategy that works better with binary problems and also implemented mutation operator to boost local search capability and support the diversity of algorithms. The experimental results show that the proposed MBBA is a competitive multi-objective algorithm that outperforms NSGA-II. Bat algorithm has also been used in the area of renewable energy [33], which has great potential for application of the proposed algorithm to the wind power network. Similarly, in the medical sector, a modified bat algorithm (MBA) for feature selection developed by [34] performed significantly well to remove unwanted and repetitive data on breast cancer prior to diagnosis. In [35], the hybrid binary bat enhanced particle swarm optimization algorithm (HBBEPSO) was developed and claimed to have the ability to scan the feature space for appropriate combinations of features. Figure 4 outline the basic pseudo-code of bat algorithm.

A multi-objective algorithm based on a cuckoo search algorithm has been applied to the optimization problem [36-38]. In the dimensional reduction problem, a new multi-objective cuckoo search algorithm [39] has been developed to search the space attribute with minimal correlation between the selected attributes. Experimental findings have shown that the proposed multi-objective CS method has successfully outperformed particle swarm optimization (PSO) and genetic algorithm (GA) optimization algorithms. For example, a hybrid rough set based on a modified cuckoo search algorithm has been proposed [39]. The algorithm developed demonstrates the ability to reduce the number of features in the reduction set without losing the accuracy of the classification. In [40] also proposed a prediction algorithm for heart disease based on the cuckoo search system. Two algorithms, namely cuckoo search algorithm (CSA) and cuckoo optimization algorithm (COA), have been used for subset generation and the results show that both algorithms have achieved better predictive accuracy on selected datasets. Figure 5 summarise the general pseudo-code of Cuckoo algorithm.

Firefly algorithm has been invented by Yang [41] and has been used in many areas, especially in the selection of apps. New firefly algorithm based on the Ada-boost method has recently been developed in the medical field [42] to diagnose liver cancer. The developed hybrid method used by firefly algorithm to improve the resulting Ada-boost algorithm can help physicians recognize and classify safe and unhealthful individuals. It can also be used in medical centers to improve accuracy and speed and reduce costs. In addition, [43] proposes the collection of features in the Arabic text classification based on firefly algorithm. The proposed algorithm has been successfully applied to various combinatorial problems and has achieved high precision in the development of the Arabic text classification. In the multi-objective question, the firefly algorithm was successfully applied to the scheduling problem field, such as in [44-46]. Figure 6 presents the basic pseudo-code of firefly algorithm.

Objective function $f(x), x = (x^1, \ldots, x^n)$.

Initialize the bat population $x_i$ and $v_i$, $i = 1, 2, \ldots, m$.
**For** each bat
  Define pulse frequency $f_i$, loudness $A_i$ and pulse rates $r_i$
**EndFor**
**While** t<T
  **For** each bat $x_i$
    Generate new solutions through Eqs.(1-3);
    **If** rand> $r_i$
      Select a solution among the best solutions;
      Generate a local solution around the best solutions by Eq.(6).
    **EndIf**
    **If** rand< $A_i$ && $f(x_i) < f(\hat{x})$
      Accept the new solutions;
      Increase $r_i$ and reduce $A_i$ through Eqs.(4-5).
    **EndIf**
  **End For**
**EndWhile**

Figure 4. The pseudo-code of a bat algorithm

**Begin**
The objective function f(x); $x = (x_1; x_2; \ldots; x_d)^T$;
Create opening populace of n host nests $x_i$ (i = 1; 2; ...; n);
**Set:** $S_{best} = S_0$;
$\gamma_{best}$ = eval ($S_0$; D; M);
**While** (t < Max Generation) or (Halt condition)
  **Begin**
  Get a cuckoo arbitrarily by means of levy flight;
  S = generate(D);
  $\gamma_{best}$ = eval ($S_0$; D; M);
  if ($\gamma$ greater than best)
    $\gamma_{best} = \gamma$;
    $S_{best} = S$;
  Estimate its superiority/suitability $F_i$;
  Select a nest amongst n (say, j) arbitrarily;
**If** ($F_i > F_j$)
  Substitute j by means of the new-fangled solutions;
**End If**
A portion (pa) of inferior quality nest are uncontrolled and fresh ones are made;
Retain the finest solutions (or the nest with excellence solutions);
Ranked the solutions and discover the recent finest one;
**End While;**
**Return** $S_{best}$;
Post process outcomes along with visualizations;
**End.**

Figure 5. The pseudo-code of a cuckoo search algorithm

```
input: A firefly is a document of words x = (x_1, x_2, ..., x_d)
       Generate randomly a swarm of fireflies
       The dimension of the firefly' s position : d = document size
       Discretize the fireflies positions (Eq.1 & 2)
       Formulate light intensity I = score (Eq. 3)
       Define absorption coefficient γ = [0.01, 100] (from [12])
       Define randomization parameter α = 1 (from [12])
       Define attractiveness value β_0 = 1 (from [12])
       Set a maximum number of iterations: MaxGeneration
Int t = 0;
while (t ≤ MaxGeneration) do
    for (i = 1 : n) do
        for (k = 1 : n) do
            if (I_k > I_i) then
                /*Move firefly i towards firefly k by doing the following*/
                Calculate the distance r (Eq.5);
                Calculate the attractiveness β (Eq.4);
                Calculate the new position x_i of the firefly i (Eq.6);
                Evaluate the firefly by updating the light intensity (score: Eq.3);
            end
        end
        Rank fireflies and find the current best;
        Give the position of the current best firefly (in real numbers);
        Discretize the current best real position (Eq.1 & 2);
    end
    Return the intensity with the discrete position;
end
```

Figure 6. The pseudo-code of a firefly algorithm

In this paper, we suggest an optimal combination of a selection mechanism based on evolutionary subset generation. Wrapper and filtered approaches have been used. Bio-search algorithms have been combined with ENORA and NSGA-II to perform the optimum collection of apps. Inspired by the ability of bio-search algorithms to select features, the purpose of this paper is to present optimized ENORA and NSGA-II algorithms by deploying bio-search algorithms to obtain an optimum number of attributes for selected datasets. The key concept is to incorporate integrated algorithms by numerous reductions between multi-objective algorithms and bio-search algorithms for the collection of features. Description of the execution steps are listed in the next section.

## 2. RESEARCH METHOD

Methodology of this paper is represented in Figure 7 has been presented in the form of the workflow. It consists of series of steps and mention in details through out this section.

- Step 1

Data collection: datasets were selected from UCI Machine Learning Repository [47] (refer Table 1 for profile of the selected datasets). These datasets consist of various sizes and mix domains in order to examine the capability of algorithms to perform attribute selection.

- Step 2

Data handling: missing values in the dataset has been pre-processed to be ready for experimentation. Dataset that has missing value (symbolized as '?' in original dataset) should be replaced either with 0 or mean value. Both methods have been tested and a result indicates insignificant difference in terms of performance. This research decided using value of "0" to be replaced for missing values.

- Step 3

Load clean datasets: all datasets have been trained and tested using WEKA software. WEKA also has been used to do the data pre-processing in step 2. In WEKA software, the detailed parameter setting for all algorithms has been set up to be further experimented in step (4) and step (5) as shown in Table 2.

- Step 4

Subset generation (1): in this step, two (2) reduction processes which are ENORA and NSGA-II algorithms with filtered method have been executed. The output of this first subset generation considered not an optimal subset and need to be furthered reduced. The extended reduction is needed to get an optimal reduction which been done in step (5).

- Step 5

Subset generation (2): in this step, the output in step (4) will be furthered reduced with five (5) bio-search methods (ant, bat, bee, cuckoo and firefly) + wrapper used in order to search for the optimal attributes. This experiment process reflects research done in [48] which claimed that balance of exploitation and exploration need to be accomplished for efficient space searching. This second generation of the subset considered an optimal subset.

– Step 6

Subset evaluation: in this step, the output of subset generation (1) and subset generation (2) will be evaluated through classification performance. This step is to confirm the performance of subset generation with good classification accuracy in order to produce an optimal feature selection model.

– Step 7

Production of optimal feature selection model: In this final step, various combinations of bio-search methods and reduction algorithms were carefully selected to perform a feature selection model. Optimal numbers of reductions with good classification accuracy are the criteria for choosing the best selected list.

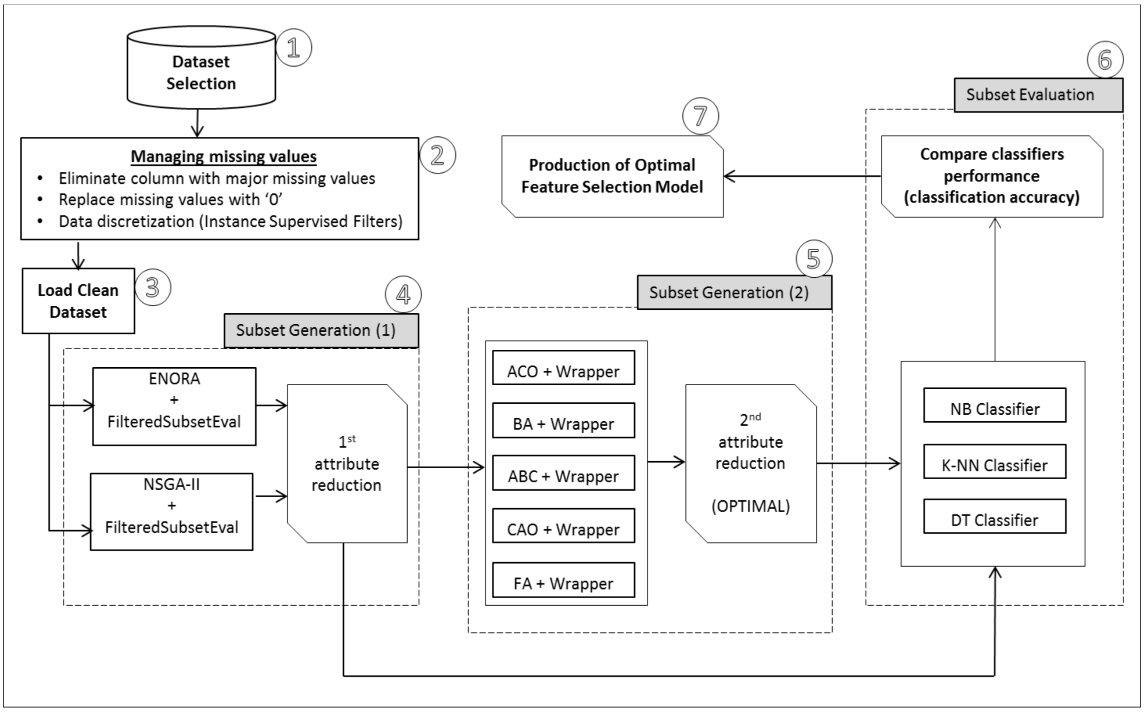


Figure 7. Methodology of the research

Table 1. Profile of the selected datasets

| Size | Dataset | #Attr | #Inst | #Class |
|---|---|---|---|---|
| Small | Breastcancer | 9 | 367 | 2 |
| Small | Parkinson | 22 | 197 | 2 |
| Small | Ozone | 72 | 2536 | 2 |
| Medium | Clean1 | 166 | 476 | 2 |
| Medium | Semeion | 265 | 1593 | 2 |
| Large | Emails | 4702 | 64 | 2 |
| Large | Gisette | 5000 | 13500 | 2 |
| Large | Arcene | 10000 | 900 | 2 |

Table 2. Details parameter setting

| Searc Algo | Population Size | Specific setting |
|---|---|---|
| Ant | 20 | Evaporation rate: 0.9 \| Pheromone rate: 2.0 \| Heuristic rate: 0.7 |
| Bat | 20 | Frequency: 0.5 \| Loudness: 0.5 |
| Bee | 30 | Radius Damp: 0.98 \| Radius Mutation: 0.80 |
| Cuckoo | 20 | Pa rate: 0.25 \| Sigma rate: 0.70 |
| Firefly | 20 | Beta zero: 0.33 \| Absorption Coefficient: 0.001 |
| ENORA | 100 | Generation: 10 |
| NSGA-II | 100 | Generation: 10 |

*Fixed setting for all bio-search algorithms: Iteration: 20, Mutation Probability: 0.01

## 3. RESULTS AND DISCUSSION

Table 3 shows the comparison of reduction performance between ENORA vs NSGA-II in the first subset generation and second subset generation. It can be seen that ENORA+filtered method managed to reduce the attributes for seven (7) datasets (Ozone, Parkinson, Clean1, Semeion, Emails, Gisette, Arcene) except for Breastcancer datasets where the original attributes remained. Semeion, Emails, Gisette and Arcene datasets achieved more than 95% reduction. Similar situation with NSGA-II where the first subset generation achieved more attribute reduction than ENORA. Emails dan Gisette datasets have reached almost 100% reduction which is extreme cases to be considered in the first subset generation. However, the massive reduction using ENORA and NSGA-II of these attributes with filtered approach still does not approve the optimal selection. Even though the performance of NSGA-II better than ENORA in term of much less selected attributes in first reduction, this condition still not promising to get the optimal set of attributes.The second subset generation need to be executed to obtain absolute optimal reduction set. Extended experiment has been conducted to optimize the ENORA and NSGA-II algorithms with five (5) bio-search algorithm and wrapper method. A result shows more reduction happened for all datasets. Extreme case has been discovered by Ozone dataset where twelve (12) attributes in the first reduction with ENORA have been reduced to only one (1) attribute in the second reduction. Same result also been achieved with NSGA-II. Further experiment been conducted to optimize the ENORA and NSGA-II algorithms with five (5) bio-search algorithm and wrapper method. Results shows superior reduction for all datasets for ENORA and NSGA-II. Ozone dataset maintain the same result as all searching space has been fully explored. Overall, all bio-search algorithms succeeded to acquire near-optimal solutions (optimal features) in second subset generations. This result confirmed the adaptive behavior of bio-search algorithm with wrapper methods to perform optimal features selection for ENORA and NSGA-II algorithms. Also, the ability of random search function that exists in the bio-search algorithms gives more advantages to select the best optimum features. For reduction purposes, it can be concluded that bio-search algorithms with wrapper method can be used to reduce attributes from all sizes of data.

Table 3. Comparison of attribute reduction: ENORA vs ENORA + Bio-Search

| Dataset | #Attr Ori | Subset Generation (1) | | Subset Generation (2) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | #Attr | | # Attr [ENORA + (Wrapper + Bio Search)] | | | | | # Attr [NSGA-II + (Wrapper + Bio Search)] | | | | |
| | | ENORA + Filtered | NSGA-II + Filtered | Ant | Bat | Bee | Cuc | Fly | Ant | Bat | Bee | Cuc | Fly |
| Breastcancer | 9 | 9 (0.0%)* | 9 (0.0%)* | 7 | 7 | 7 | 6 | 7 | 7 | 7 | 7 | 6 | 7 |
| Parkinson | 22 | 9 (59.1%)* | 7 (68.2%)* | 5 | 6 | 6 | 7 | 6 | 5 | 3 | 5 | 5 | 5 |
| Ozone | 72 | 12 (85.7%)* | 6 (91.7%)* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Clean1 | 166 | 22 (86.7%)* | 19 (88.6%)* | 14 | 13 | 14 | 14 | 14 | 15 | 17 | 15 | 15 | 14 |
| Semeion | 265 | 5 (98.1%)* | 7 (97.4%)* | 4 | 4 | 4 | 4 | 4 | 4 | 6 | 6 | 6 | 4 |
| Emails | 4702 | 79 (98.3%)* | 40 (99.1%)* | 18 | 24 | 11 | 13 | 34 | 8 | 11 | 4 | 7 | 14 |
| Gisette | 5000 | 66 (98.7%)* | 49 (99.0%)* | 23 | 28 | 18 | 15 | 31 | 14 | 20 | 13 | 18 | 20 |
| Arcene | 10000 | 391(96.1%)* | 216(97.8%)* | 101 | 37 | 36 | 37 | 133 | 93 | 86 | 56 | 80 | 84 |

* % of reduction from original attributes

Table subset generation (2) shows the comparison of classification accuracy of ENORA with various classifiers for classification perfomance. Surprisingly that attributes selected from all datasets by ENORA in the first reduction does not improve the classification accuracy which maintained the same accuracy results of the original datasets (refer to Table 4). Clearly, attributes selected in second subset generation by ENORA and bio-search algorithms with wrapper method successfully increased the classification accuracy. All five (5) algorithms (ant, bat, bee, cuckoo and firefly) proved to have good classification accuracy for all datasets except Gisette dataset. But it is still considered acceptable since the percentage of reduction achieved more than 50% (refer to Table 4) then still maintaining good classification accuracy for Gisette dataset. Generally, it can be seen all bio-search algorithm performed well to achieve better classification accuracy with various classifiers. The highlighted column in Table 5 shows the selected best performance of classification results which reflects the model to be developed (refer to Table 6).

Table 7 shows the comparison of classification accuracy of NSGA-II with various classifiers for classification perfomance. Interestingly to highlight that attributes selected from all datasets by NSGA-II in the first subset generation show inconsistent results which improved the accuracy for the half of the datasets (refer to Table 4). Another half shows decrement of classification accuracy. Obviously, the first subset generation results by NSGA-II algorithm need to be optimized in order to get better classification accuracy. In second subset generation, NSGA-II and bio-search algorithms with wrapper method shows significant increment for all datasets. The highlighted column in Table 5 shows the selected best performance of classification results which reflects the model to be developed (refer to Table 6). Table 6 shows the IDEAL feature selection model on various sizes of datasets. This model which consist of combination list of algorithms can be a guideline for searching optimal number of attributes based on dataset size.

Table 4. Subset generation (1) classification accuracy: original data, ENORA, NSGA-II using DT, NB and k-NN

| Dataset | No Reduction Accuracy (%) | | | 1st Reduction [ENORA + Filtered] Accuracy (%) | | | 1st Reduction [NSGA-II + Filtered] Accuracy (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| | DT | NB | k-NN | DT | NB | k-NN | DT | NB | k-NN |
| Breastcancer | 96.2 | 96.2 | 95.8 | 96.2 | 96.2 | 95.4 | 96.2 | 96.2 | 95.4 |
| Parkinson | 84.8 | 83.3 | 92.4 | 89.4 | 90.9 | 92.4 | 87.9 | 90.9 | 87.9 |
| Ozone | 93.3 | 71.9 | 92.4 | 93.7 | 80.5 | 92.2 | 93.9 | 89.4 | 93.5 |
| Clean1 | 85.8 | 85.2 | 83.3 | 75.9 | 82.1 | 80.2 | 82.1 | 86.4 | 86.4 |
| Semeion | 94.5 | 93.0 | 97.6 | 92.4 | 91.3 | 92.4 | 93.4 | 90.6 | 93.4 |
| Emails | 72.7 | 86.4 | 72.7 | 77.3 | 86.4 | 86.4 | 77.3 | 77.3 | 77.3 |
| Gisette | 91.5 | 91.5 | 92.9 | 88.2 | 85.0 | 85.6 | 86.8 | 86.2 | 85.9 |
| Arcene | 70.6 | 76.5 | 91.2 | 85.3 | 79.4 | 88.2 | 85.3 | 79.4 | 88.2 |

Table 5. Subset generation (2) classification accuracy: ENORA using DT, NB and k-NN

| Dataset | 2ndReduction [ENORA + (Wrapper + Bio Search)] Accuracy (%) | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Decision Tree (DT) | | | | | Naïve Bayes (NB) | | | | | k-Nearest Neighbour (k-NN) | | | | |
| | Ant | Bat | Bee | Cuc | Fly | Ant | Bat | Bee | Cuc | Fly | Ant | Bat | Bee | Cuc | Fly |
| Breastcancer | **96.2** | 96.2 | 96.2 | 96.2 | 96.2 | 96.6 | 96.6 | 96.6 | **96.6** | 96.6 | 95.4 | 95.4 | 95.4 | **96.2** | 95.4 |
| Parkinson | **87.9** | 89.4 | 89.4 | 89.4 | 89.4 | 90.9 | 90.9 | 90.9 | **93.9** | 90.9 | 90.9 | 92.4 | 92.4 | **92.4** | 92.4 |
| Ozone | **93.9** | **93.9** | **93.9** | **93.9** | **93.9** | **93.9** | **93.9** | **93.9** | **93.9** | **93.9** | **93.9** | **93.9** | **93.9** | **93.9** | **93.9** |
| Clean1 | **80.9** | **80.9** | **80.9** | **80.9** | **80.9** | **84.0** | **84.0** | 83.3 | **84.0** | **84.0** | 82.1 | **82.7** | **82.7** | 82.1 | 82.1 |
| Semeion | **92.4** | **92.4** | **92.4** | **92.4** | **92.4** | **91.3** | **91.3** | 91.3 | **91.3** | **91.3** | 92.4 | **92.4** | **92.4** | 92.4 | 92.4 |
| Emails | 77.3 | 77.3 | 77.3 | 72.7 | **77.3** | 86.4 | 86.4 | **86.4** | 86.4 | 86.4 | 86.4 | 86.4 | 86.4 | 86.4 | **86.4** |
| Gisette | 84.4 | 87.1 | 86.2 | 83.5 | **87.6** | 85.3 | 85.3 | **86.2** | 81.8 | 85.9 | 83.2 | 86.2 | 86.5 | 82.9 | **86.8** |
| Arcene | 88.2 | 88.2 | 76.5 | 88.2 | **94.1** | 88.2 | 88.2 | **91.2** | 88.2 | 88.2 | 85.3 | 85.3 | 82.4 | 88.2 | **91.2** |

Table 6. Ideal feature selection model

| List | Dataset size | Multi-objective algo | Reduction algo | Bio-search algo | Classifier |
|---|---|---|---|---|---|
| 1. | Small | ENORA | Wrapper | Cuckoo | NB |
| 2. | | NSGA-II | | Ant, Cuckoo, Firefly | NB |
| 3. | Medium | ENORA | | Bee, Bat | k-NN |
| 4. | | NSGA-II | | Ant, Cuckoo | NB |
| 5. | Large | ENORA | | Firefly | k-NN |
| 6. | | NSGA-II | | Bat | DT |

Table 7. Subset generation (2) classification accuracy: NSGA-II using DT, NB and k-NN

| Dataset | 2ndReduction [NSGA-II + (Wrapper + Bio Search)] Accuracy (%) | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Decision Tree (DT) | | | | | Naïve Bayes (NB) | | | | | k-Nearest Neighbour (k-NN) | | | | |
| | Ant | Bat | Bee | Cuc | Fly | Ant | Bat | Bee | Cuc | Fly | Ant | Bat | Bee | Cuc | Fly |
| Breastcancer | **96.2** | 96.2 | **96.2** | **96.2** | **96.2** | **96.6** | 96.6 | 96.6 | **96.6** | **96.6** | 95.4 | 95.4 | 95.4 | **96.2** | 95.4 |
| Parkinson | **87.9** | 84.8 | **87.9** | **87.9** | **87.9** | **93.9** | 86.4 | 90.9 | **93.9** | **93.9** | 90.9 | 86.3 | 90.9 | **90.9** | 90.9 |
| Ozone | **93.9** | 93.9 | **93.9** | **93.9** | **93.9** | **93.9** | 93.9 | 93.9 | **93.9** | **93.9** | 93.9 | 93.9 | 93.9 | **93.9** | 93.9 |
| Clean1 | **84.0** | 83.3 | 83.3 | 83.3 | 83.3 | **87.0** | 84.6 | 85.8 | **87.0** | 84.6 | **85.2** | 82.1 | 84.6 | 81.5 | 84.0 |
| Semeion | **92.6** | 93.4 | 93.4 | 93.4 | 92.6 | **93.2** | 93.2 | 93.2 | **93.2** | 93.2 | **93.2** | 93.2 | 93.2 | 93.2 | 93.2 |
| Emails | 77.3 | **77.3** | 77.3 | 77.3 | 77.3 | 77.3 | 77.3 | **77.3** | 77.3 | 77.3 | 77.3 | 77.3 | 77.3 | 77.3 | **77.3** |
| Gisette | 88.2 | **88.2** | 88.2 | 88.8 | 88.2 | 86.5 | 86.2 | **86.5** | 86.5 | 86.5 | 85.9 | 85.3 | 85.9 | 85.3 | **85.9** |
| Arcene | 85.3 | **91.2** | 76.5 | 76.5 | 82.4 | 85.3 | 82.4 | **88.2** | 85.3 | 85.3 | 85.3 | 88.2 | 88.2 | 91.2 | **91.2** |

## 4. CONCLUSION AND FUTURE WORK

In summary, the impact of this paper on data mining can be seen as leading in particular to alternative optimization techniques. This alternative technique provides a better understanding of the implementation of various bio-inspired algorithms in the exploration and utilization of the search space, in particular for the optimization of multi-objective algorithms. This paper explores a new ideal feature selection model that has been compared and evaluated on eight (8) datasets. The ideal lists for the selection of features have been determined on the basis of the produced good classification accuracy with the relevant features. However, the limitation of algorithms needs to be addressed. One of the limitations is the computational cost (longer computation time), and it takes time to discover the formulation of the list. The next research work to be explored would therefore be the study on different bio-search algorithms and the formulation of the correct setting of parameters for new optimization techniques.

## ACKNOWLEDGEMENTS
The authors would like to recognize Universiti Malaysia Terengganu (UMT), Universiti Utara Malaysia (UUM) and Kementerian Pendidikan Malaysia (KPM) for the support of services and facilites. This research was sponsored by the UMT and Ministry of Education Malaysia (MOE) under TAPE-RG research grant (Vot No. 55133).

## REFERENCES
[1] R. Jensen and Q. Shen, "Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches," *Wiley-IEEE Press,* 2008.
[2] H. Liu and H. Motoda, "Feature Selection for Knowledge Discovery and Data Mining," *Kluwer Academic Press,* 1998.
[3] R. Caruana and D. Freitag, "Greedy Attribute Selection," *Machine Learning Proceedings*, 1994.
[4] A. Arauzo-Azofra, J. M. Benitez, and J. L. Castro, "Consistency measures for feature selection," *J. Intell. Inf. Syst.*, vol. 30, pp. 273-292,2008.
[5] X. Zhang, Y. Hu, K. Xie, S. Wang, E. W. T. Ngai, and M. Liu, "A causal feature selection algorithm for stock prediction modeling," *Neurocomputing*, vol. 142, pp. 48-59, 2014.
[6] B. A. Blesser, T. T. Kuklinski, and R. J. Shillman, "Empirical tests for feature selection based on a psychological theory of character recognition," *Pattern Recognition,* vol. 8, no. 2, pp. 77-85, 1976.
[7] J. Tang and H. Liu, "Feature selection for social media data," *ACM Transactions on Knowledge Discovery from Data*, vol. 8, no. 4, pp. 1-27, 2014.
[8] C. H. Chen, "On information and distance measures, error bounds, and feature selection," *Information Sciences*, vol. 10, no. 2, pp. 159-173, 1976.
[9] L. Sun, J. Xu, and Y. Tian, "Feature selection using rough entropy-based uncertainty measures in incomplete decision systems," *Knowledge-Based Syst.*, vol. 36, pp. 206-216, 2012.
[10] S. K. das Subrata, "Feature Selection with a Linear Dependence Measure," *IEEE Transactions on Computers*, vol. C-20, no. 9, pp. 1106-1109, 1971.
[11] R. Kohavi and G. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1, pp. 273-324, 1997.
[12] V. Kumar, "Feature Selection: A literature Review," *Smart Comput. Rev.*, vol. 4, no. 3, pp. 211-229, 2014.
[13] K. Deb, "Multi-objective Optimisation Using Evolutionary Algorithms: An Introduction,"*Jhon Wiley and Sons Ltd*, 2011.
[14] Coello Coello, Carlos, Lamont, Gary B., van Veldhuizen, David A., "Evolutionary Algorithms for Solving Multi-Objective Problems," *Genetic and Evolutionary Computation,Springer,* 2007.
[15] P. M. Narendra and K. Fukunaga, "A Branch and Bound Algorithm for Feature Subset Selection," *IEEE Transactions on Computers,* vol. C-26 , no. 9*,* pp. 917-922, 1977.
[16] P. Gupta, D. Doermann, and D. DeMenthon, "Beam search for feature selection in automatic SVM defect classification," *Proc. - Int. Conf. Pattern Recognit.*, 2002.
[17] H. Vafaie and K. De Jong, "Genetic algorithms as a tool for feature selection in machine learning," *Proceedings International Conference on Tools with Artificial Intelligence, ICTAI*, 1992.
[18] J. Langford *et al.*, "Evolutionary Feature Selection," *Encyclopedia of Machine Learning*, Springer US, pp. 353-353, 2011.
[19] L. Cervante, B. Xue, M. Zhang, and L. Shang, "Binary particle swarm optimisation for feature selection: A filter based approach," *IEEE Congress on Evolutionary Computation,* 2012.
[20] Z. Yong, G. Dun-wei, and Z. Wan-qiu, "Feature selection of unreliable data using an improved multi-objective PSO algorithm," *Neurocomputing*, vol. 171, pp. 1281-1290, 2016.
[21] A. Al-Ani and M. Deriche, "Feature selection using a mutual information based measure," *Object recognition supported by user interaction for service robots*, vol. 4, pp. 82-85, 2002.
[22] F. Jiménez, G. Sánchez, J. M. García, G. Sciavicco, and L. Miralles, "Multi-objective evolutionary feature selection for online sales forecasting," *Neurocomputing*, vol. 234, pp. 72-92, 2017.
[23] H. Falaghi, M. R. Haghifam, and C. Singh, "Ant colony optimization-based method for placement of sectionalizing switches in distribution networks using a fuzzy multiobjective approach," *IEEE Trans. Power Deliv.*, vol. 24, no. 1, pp. 268-277, 2008.
[24] Y. Wan, M. Wang, Z. Ye, and X. Lai, "A feature selection method based on modified binary coded ant colony optimization algorithm," *Appl. Soft Comput. J.*, vol. 49, pp. 248-258, 2016.
[25] O. Fallahzadeh, Z. Dehghani-Bidgoli, and M. Assarian, "Raman spectral feature selection using ant colony optimization for breast cancer diagnosis," *Lasers Med. Sci.*, vol. 333, no. 8, pp. 1799-1806, 2018.
[26] T. Mehmod and H. B. M. Rais, "Ant colony optimization and feature selection for intrusion detection," *Advances in Machine Learning and Signal Processing,* pp. 305-312, 2016.
[27] V. Agrawal and S. Chandra, "Feature selection using Artificial Bee Colony algorithm for medical image classification," *2015 8th International Conference on Contemporary Computing,* 2015.
[28] H. Rao et al., "Feature selection based on artificial bee colony and gradient boosting decision tree," *Appl. Soft Comput. J.*, vol. 74, pp. 34-42, 2019.

[29] E. ZorarpacI and S. A. Özel, "A hybrid approach of differential evolution and artificial bee colony for feature selection," *Expert Syst. Appl.*, vol. 62, pp. 91-103, 2016.
[30] E. Hancer, B. Xue, M. Zhang, D. Karaboga, and B. Akay, "A multi-objective artificial bee colony approach to feature selection using fuzzy mutual information," *2015 IEEE Congress on Evolutionary Computation,* 2015.
[31] X.-S. Yang, "Bat algorithm for multi-objective optimisation," *Int. J. Bio-Inspired Comput.*, vol. 3, no. 5, pp. 267-274, 2011.
[32] L. M. Amine and K. Nadjet, "A multi-objective binary bat algorithm," *ACM International Conference Proceeding Series*, 2015.
[33] T. Niu, J. Wang, K. Zhang, and P. Du, "Multi-step-ahead wind speed forecasting based on optimal feature selection and a modified bat algorithm with the cognition strategy," *Renew. Energy*, vol. 118, pp. 213-229, 2018.
[34] S. Jeyasingh and M. Veluchamy, "Modified bat algorithm for feature selection with the Wisconsin Diagnosis Breast Cancer (WDBC) dataset," *Asian Pacific J. Cancer Prev.*, vol. 18, no. 5, pp. 1257-1264, 2017.
[35] M. A. Tawhid and K. B. Dsouza, "Hybrid Binary Bat Enhanced Particle Swarm Optimization Algorithm for solving feature selection problems," *Appl. Comput. Informatics*, 2018.
[36] M. Akbari and H. Rashidi, "A multi-objectives scheduling algorithm based on cuckoo optimization for task allocation problem at compile time in heterogeneous systems," *Expert Syst. Appl.*, vol. 60, 2016.
[37] Q. Wang, S. Liu, H. Wang, and D. A. Savić, "Multi-objective cuckoo search for the optimal design of Water Distribution Systems," *Proceedings of the 2012 International Conference on Civil Engineering and Urban Planning*, 2012.
[38] K. Chandrasekaran and S. P. Simon, "Multi-objective scheduling problem: Hybrid approach using fuzzy assisted cuckoo search algorithm," *Swarm Evol. Comput.*, vol. 5, pp. 1-16, 2012.
[39] W. Yamany, N. El-Bendary, A. E. Hassanien, and E. Emary, "Multi-Objective Cuckoo Search Optimization for Dimensionality Reduction," *Procedia Computer Science*, 2016.
[40] A. M. Usman, U. K. Yusof, S. Naim, and S. Naim, "Cuckoo inspired algorithms for feature selection in heart disease prediction," *Int. J. Adv. Intell. Informatics*, vol. 4, no. 2, 2018.
[41] X. S. Yang, "Firefly Algorithms," *Nature-Inspired Optimization Algorithms*, pp. 1-10, 2014.
[42] S. Ardam and F. Soleimanian Gharehchopogh, "Diagnosing Liver Disease using Firefly Algorithm based on Adaboost," *Journal of Health Administration,* vol. 22, no. 1, pp. 61-77, 2019.
[43] S. Larabi Marie-Sainte and N. Alalyani, "Firefly Algorithm based Feature Selection for Arabic Text Classification," *Journal of King Saud University - Computer and Information Sciences*, vol. 32, no. 3, pp. 320-328, 2018.
[44] S. Karthikeyan, P. Asokan, and S. Nickolas, "A hybrid discrete firefly algorithm for multi-objective flexible job shop scheduling problem with limited resource constraints," *Int. J. Adv. Manuf. Technol.*, vol. 72, pp. 1567-1579, 2014.
[45] S. Karthikeyan, P. Asokan, S. Nickolas, and T. Page, "A hybrid discrete firefly algorithm for solving multi-objective flexible job shop scheduling problems," *Int. J. Bio-Inspired Comput.*, vol. 7, no. 6, pp. 386-401, 2015.
[46] H. Wang et al., "A hybrid multi-objective firefly algorithm for big data optimization," *Appl. Soft Comput. J.*, vol. 69, pp. 806-815, 2018.
[47] A. Asuncion and D. J. Newman, "UCI machine learning repository," *Irvine, CA: University of California, School of Information and Computer Science*, 2017. [Online]. Available: http://archive.ics.uci.edu/ml.
[48] M. Montazeri, M. Montazeri, H. R. Naji, and A. Faraahi, "A novel memetic feature selection algorithm," *The 5th Conference on Information and Knowledge Technology*, pp. 295-300, 2013.