Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-019-01364-5

ORIGINAL RESEARCH

A novel hybrid wrapper–filter approach based on genetic algorithm, particle swarm optimization for feature subset selection

Fateme Moslehi[1] · **Abdorrahman Haeri**[1]

Received: 29 January 2019 / Accepted: 11 June 2019

Abstract
The classification is one of the main technique of machine learning science. In many problems, the data sets have a high dimensionality that the existence of all features is not important to the purpose of the problem, and this will decrease the accuracy and performance of the algorithm. In this situation, the feature selection will play a significant role, and by eliminating unrelated features, the efficiency of the algorithm will be increased. A hybrid filter-wrapper method is proposed in the present study for feature subset selection established with integration of evolutionary based genetic algorithms (GA) and particle swarm optimization (PSO). The presented method mainly aims to reduce the complication of calculation and the search time expended to achieve an optimum solution to the high dimensional datasets feature selection problem. The proposed method, named smart HGP-FS, utilizes artificial neural network (ANN) in the fitness function. The filter and wrapper methods are integrated in order to take the benefit of filter technique acceleration and the wrapper technique vigor for selection of dataset efficacious characteristics. Some dataset characteristics are eliminated through the filter phase, which in turn reduces complex computations and search time in the wrapper phase. Comparisons have been made for the effectiveness of the proposed hybrid algorithm with the usability of three hybrid filter-wrapper methods, two pure wrapper algorithms, two pure filter procedures, and two traditional wrapper feature selection techniques. The findings obtained over real-world datasets show the efficiency of the presented algorithm. The outcomes of algorithm examination on five datasets reveal that the developed method is able to obtain a more accurate classification and to remove unsuitable and unessential characteristics more effectively relative to the other approaches.

Keywords Feature selection · Hybrid (wrapper–filter) approach · Multi-objective optimization · Genetic algorithm · Particle swarm optimization (PSO)

1 Introduction

Data mining is a powerful new technology to extract hidden information from data warehouses. Data mining surveys data from different perspectives and finds useful patterns and knowledge from large volumes of raw data (Moslehi et al. 2019a).

Data mining methods such as clustering and classification are useful in many fields such as banking industry (Moslehi et al. 2019b) supply chain management, insurance industry, vehicular ad hoc networks (Zhang et al. 2018b) and Wireless Sensor Networks (Zhang et al. 2017a). Medical investigators have applied data mining to datasets for a multitude of patients to obtain basic understanding of genetic and environmental disease agents, and to create more effective diagnostic tools (Zhang et al. 2015a). Researchers are lent a profound insight into the environmental and genetic roots of diseases and more effective diagnostic methods are developed by mining large datasets of a large group of patients through data mining techniques (Zhang et al. 2017b; Zhang et al. 2012a). In recent years with rapid growth of internet (Zhang et al. 2016), Internet of things (IOT), Radio Frequency Identification (RFID) technology (Zhang et al. 2014) the volume of data stored has increased. It is a systemic necessity to save the information considering the growing body of information processed by application systems embedded into devices making them accessible on the

✉ Abdorrahman Haeri
ahaeri@iust.ac.ir

Fateme Moslehi
moslehi_fateme@ind.iust.ac.ir

[1] Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran

internet (Zhang and Zhang 2012). So the need for clearing and extracting appropriate information and methods such as feature selection has become increasingly important. Feature selection considerably reduces the running time by ruling out the irrelevant and redundant features, improving the classification accuracy, and/or simplifying the structure of the learned classifiers or the concept of models (Dash et al. 1997). However, with a large number of features, feature selection is hard to conduct (Narendra and Fukunaga 1977; Yang et al. 2008a, b).

Many features are involved in a complex classification problem. As a result, the classifier classifies the observations within a longer period of time. The first goal of feature selection is to minimize the dimensionality of the dataset, maximize classification accuracy and prevent over fitting (Wah et al. 2018) and can be improved the efficiency of the wireless sensor network and reduce energy consumption to extend the network lifetime as mentioned in (Zhang et al. 2012b).

1.1 Contribution

In feature selection problems the main challenge is how can discard part of the prepared and preprocessed data without undermining the quality. Numerous solutions and approaches have been developed for the feature selection problem. Some of these solutions and methods have a 30-year history, and the problem with some of these algorithms at the time of their introduction was their heavy computational load. As respects, this challenge was ironed out with the emergence of fast computers and big storage resources, but finding a quick algorithm to deliver this function is still important due to the big data sets used in new problems. Numerous researchers have recently proposed the compressive sensing (CS) technology to rule out data redundancy and control the number of nodes in the wireless sensor networks (WSNs) with the aim of reducing energy consumption in nodes. The aforesaid technology samples fewer points than the total number of points for signal collection and reconstruction with a high degree of probability (Zhang et al. 2018c; Zhang et al. 2013). Recently novel and efficient methods are applied for optimization problems like evolutionary game theory (Zhang et al. 2018d) Mobile learning (Zhang 2012) and graph theory (Zhang et al. 2018e) and heuristic methods (Zhang et al. 2015b).

Methods and algorithms of optimization are divided into two groups of the algorithm, exact algorithms, and approximate algorithms. Exact algorithms are able to obtain optimization response in an exact way but have no efficiency in robust optimization and their solution time increases exponentially. Approximate algorithms are able to find good responses (near optimization) for robust optimization problems in a short time solution. Approximate algorithms are divided into two groups of the algorithm, heuristic algorithms and metaheuristic algorithms (Chen and Ji 2007). Heuristic methods like Q-learning algorithm (Zhang et al. 2018f), identify limited search space and produce responses with good quality in acceptable computation times. Meanwhile, they can face various limitations in a real world easily and therefore they are used in vast commercial packages (Uçan and Altilar 2012). Evolutionary algorithms are an important solution in many areas such as the mobile ad hoc networks (Liu et al. 2019), quantitative association rules (Moslehi and Haeri 2019) and solving a traveling salesman problem (Haeri and Tavakkoli-Moghaddam 2012). A feature selection process starts with the exhaustive search through the subset of features, and discover the best feature among the main probable subclasses based on a certain assessment criterion. In case the feature set contains n features, the best subset has to be determined by evaluation of feature subsets through optimum feature selection process.

Since the evolutionary computation techniques have the global search option (Cervante et al. 2012), they have been employed as a strong solution and an alternative for the classic searching methods to solve these problems. Particle swarm optimization (PSO)) Omar and bin Othman 2013; Liu et al. 2011), genetic algorithms (GAs) (Frohlich et al. 2003; Huang and Wang 2006; Oreski and Oreski 2014), genetic programming (GP) (Guo et al. 2005; Muni et al. 2006; Lin et al. 2008)and ant colony optimization (ACO) (Kanan and Faez 2008; Sheikhan and Mohammadi 2012) have been widely used for selecting feature.

To obtain the tradeoff between searching exploration and exploitation, different strategies are used by the heuristic models. On the one hand, exploration allows finding clear-cut areas in the search space, and on the other hand, exploitation makes it possible to maintain better solutions through scanning the local search space. Among the aforementioned meta-heuristic search techniques, some employ the exploration process whereas others apply the exploitation process for better outcomes. Accordingly, the search algorithm performance can be progressed by applying hybrid methods. Hybridization combines positive properties of at least two methods, thereby boosting the yield of each technique. The present study aims to use GA and PSO as newly presented and effective meta-heuristic approaches to create an innovative hybrid method in order to boost the performance of general classification tasks (Zorarpaci and Özel 2016).

1.2 Roadmap

In this research, a hybrid model resulting from the combination of the filter and wrapper methods was proposed for the feature selection problem. This model mainly seeks to reduce the computational complexities of feature selection with high dimensional datasets using the information derived from the smaller datasets. As in wrapper process a feature subset is selected and the classifier runs on it in each iteration, after that the classification accuracy is computed from the obtained confusion matrix, because of that the wrapper approaches have high computational complexity (Hammami et al. 2018). And so the purpose of integrating the filter and wrapper approaches is to use the advantage of the speed of the filter techniques and the power of the wrapper techniques in selecting the relevant features of the dataset. To this end, a combination of the classification techniques, the genetic algorithm (GA), the particle swarm optimization (PSO) algorithm, and the artificial neural network model was used to extract the features influencing the dataset. The aim of integrating GA and PSO is to balance between exploration and exploitation to create an algorithm without the previous weakness by using their strengths that is the high convergence speed of particle swarm optimization algorithm and exploration ability of genetic algorithm. In this method, the features with a selection probability suiting the final subset are predicted align with classification, and the feature selection method is only applied to these features. Applying prediction methods is a powerful approach for obtaining good results which are considered by the authors such as (Zhang et al. 2018a, 2019). Hence, the computational complexity of the algorithm is reduced drastically and a subset with fewer features is selected. Accelerating the feature selection process with high dimensional datasets, increasing the classification accuracy, and reducing the number of selected features are the main advantages of the proposed approach. The attainment of these achievements in this paper is demonstrated by assessing and comparing this approach to similar algorithms and testing it on 5 high dimension datasets. Therefore, the present research goal is to answer the below questions.

- Will the involvement of the filter and wrapper approaches in the feature selection process improve the algorithm?
- Can the patterns extracted from smaller data sets are used to select the features of a larger data? Does this technique increase the efficiency and speed of the algorithm?
- Is it possible to predict the features to be included in the selected dataset before implementing the feature selection algorithm?
- Will the combination of the GA and PSO and the applying of the artificial neural network in the evaluation function of the proposed approach lead to the selection of a smaller dataset and improve the classification accuracy?

This paper is organized divided into eight sections. Section 2 reviews the proposed filter, wrapper and hybrid approaches. Section 3 gives a brief overview of feature selection and evolutionary algorithms. The proposed methodology is outlined in Sect. 4. Section 5 describes the experimental design. The experimental results are presented in Sect. 6 and discussed in Sect. 7. Section 8 draws the conclusions.

2 Related works

Several filter-based feature selection methods have been presented in the previous study by applying mutual information measure such as Bennasar et al. (2015) that have used joint mutual information maximization to select important features or Hoque et al. (2016) have developed a feature selection framework based on fuzzy mutual information. Wu et al. (2018) have introduced new methods for radiomics features extraction from glioma samples. This methods presented by adopting mutual information. Some other approaches have developed based on information gain while Lefkovits and Lefkovits (2017) used this measure in their paper and information gain in combination with the fuzzy rough set was applied by Chinnaswamy and Srinivasan (2017) to omit irrelevant genes in microarray datasets. The combination of these techniques has also been used. Osanaiye et al. (2016) have developed a novel ensemble-based feature selection algorithm that applies information gain, gain ratio, Chi-squared and ReliefF in a hybrid framework. The application of these techniques for ranking important features. In addition, Minimal-redundancy-maximal-relevance (mRMR) (Peng et al. 2005), Fisher Score (F-Score) (Gu et al. 2011), Sequential backward selection method for unsupervised data (SUD) (Dash et al. 1997) and Laplacian Score (L-Score) (He et al. 2006) are well-known filter feature selection methods.

Various wrapper approaches have been introduced to answer the feature selection problem based on evolutionary algorithms for instance: Huang et al. (2011) have applied ant colony optimization (ACO) algorithm for selecting useful features and eliminating the irrelevant ones in electromyography signals classification. Cervante et al. (2012) have developed two new filter feature selection algorithms by using binary particle swarm optimization (BPSO) and information theory. The mutual information of each pair of features was applied in the first algorithm that can choose a smaller feature subset and the entropy of each group of features was applied in another algorithm which can maximize predictive accuracy. In Xue et al. (2013b) a particle swarm optimization has been applied for extracting important features and elimination the inappropriate ones. A two-stage fitness function, which measured the error rate and the number of features, is applied in this approach. Hancer et al. (2015) have developed a novel hybrid feature selection algorithm with using binary artificial bee colony (ABC) algorithm. The results obtained over real-world datasets show the efficiency of the suggested method. The algorithm was examined on 10 benchmark datasets and the results outline the performance superiority of the proposed algorithm to previous algorithms. A new hybrid feature selection algorithm called HPSO-LS has developed by Moradi and Gholampour (2016). They have used particle swarm optimization in their proposed method have evaluated features subset by a local search strategy that is used in the fitness function. The correlation information of features is the measure for selecting relevant features. In the paper proposed by Shunmugapriya and Kanmani (2017) Ant colony optimization (ACO) and bee colony optimization (BCO) algorithms were proposed in a hybridized method to select the important features in the data set. Further analysis showed that the efficiency of the proposed method. Mafarja and Mirjalili (2017) have combined two evolutionary algorithms, whale optimization algorithm (WOA) and simulated annealing (SA) to select a small feature subset from a large number of features. The adaption of SA aimed to improve aspects of exploitation of research. Gu et al. (2018) have used the competitive swarm optimization algorithm to develop an efficient feature selection approach. In their analysis, the results achieved over real-world datasets showed the efficiency of the suggested method. Ibrahim et al. (2018) have introduced new approach with the aim of increasing the efficacy of the exploration and the exploitation steps. They used the slap swarm algorithm (SSA) with the PSO algorithm. A novel hybrid feature selection algorithm for heart disease classification have proposed by Jayaraman and Sultana (2019). They applied artificial gravitational cuckoo search algorithm combined with particle bee optimized associative memory neural network methods.

In recent years the hybrid feature selection methods have been one of the interesting research areas in the field of data mining. Therefore a large number of efficient algorithms have been proposed by researchers in this field. Shreem et al. (2012) introduced a new hybrid feature selection named R-m-GA in which the authors made use of both filter and wrapper algorithms. They further discovered selected features by the use of ReliefF algorithm and the MRMR technique to select effective features. Additionally, one of the hybrid methods is multi-filter multi-wrapper (MFMR) (Leung and Hung 2010), which aims to improve the prediction accuracy and the strength of selected features. Two filter-wrapper hybrid feature selection approaches have developed by Butler-Yeoman et al. (2015). They applied PSO algorithm in their frameworks. FastPSO and RapidPSO were compared to pure filter and wrapper PSO framework. In Apolloni et al. (2016), a two-stage wrapper–filter approach have presented. They have designed the filter stage by information gain and wrapper phase by binary DE method as a second stage. Chuang et al. (2016) have hybrid the filter and wrapper methods for microarray classification to propose a feature selection algorithm. Obtained results demonstrated that the proposed method simplified gene selection and the total number of parameters needed effectively, thereby gaining a higher classification accuracy compared to other feature selection approaches. In Bostani and Sheikhan (2017) two methods were used for omitting irrelevant features. The proposed algorithm used the binary gravitational search algorithm (BGSA) as wrapper phase and mutual information (MI) as the filtering stage. Jiang et al. (2017) have presented a multi-objective hybrid approach the mutual information as filter phase and the modified binary cuckoo search algorithm (MBCS) as wrapper phase. The gained results demonstrate that this algorithm can select a smaller subset of the feature with higher classification accuracy rather than similar approaches. Sahu (2018) has proposed a hybrid feature selection algorithm based on improved binary particle swarm optimization (IBPSO) and information gain for microarray gene. Hancer (2018) has used deferential evolution to propose a new hybrid wrapper and filter feature selection technique. This aim has been obtained by applying a fuzzy local search module simulating typical forward and backward selection technique to improve the *gbest* during the evolutionary process. Hammami et al. (2018) have presented a hybrid feature selection algorithm through the integration of the GA and local search. They have proposed FW-NSGA-II algorithm by applying a memetic framework.

In the area of hybrid feature selection approaches, which is the subject of this paper, various algorithms of the metaheuristic algorithms have been used in many papers. In these papers, only one basic evolutionary algorithm has been used. So far, no approach has been introduce based on the combination of two evolutionary algorithms. In the present study, the first combination of GA and PSO algorithms was used to select influence features and omit redundancy or irrelevance ones.

3 Preliminaries

The concepts used in this paper are explained as follow.

3.1 Feature selection

Feature reduction is one of the most applied methods which has been proposed in machine learning. The goal of feature reduction is eliminating unnecessary features and selecting the most appropriate features among primary features for the purpose of increasing the performance of learning algorithms (Zhao et al. 2010). In the area of machine learning (ML), feature construction and selection are the two main parameters. The two factors are typically very time-taking and intricate tasks since the attributes need manual crafting. The attributes undergo aggregating, combining or splitting to produce characteristics from raw data (Salguero et al. 2018). Typically doing an all-around search for the purpose of finding the most features in terms of computational cost is impossible. Therefore, feature reduction has become a major challenge in pattern identification and machine learning. This method enjoys high importance in many applications including categorization and regression, as in these applications, there are usually many features a lot of which are either ineffective or cause a decrease in learning accuracy. Eliminating these features lessens, in addition to increasing the accuracy, the computational complexities (Jović et al. 2015).

Feature selection methods are based on feature reduction. These methods try to reduce the dimensions of data via the selection of a subset of primary features. In these methods, the goal is to extract a subset with the minimum possible dimensions which is appropriate for the desired application. In most cases, the data analysis such as categorization on the reduced space works better than the main space (Jović et al. 2015).

Feature selection methods try to find the best subset from a set with N feature and 2^N subset. In all of these methods, based on the application and type of definition, a subset is selected as a response to being able to optimize an Evaluation Function. Although each method aims to find the most relevant features, regarding the vastness of possible responses, finding the optimized response is difficult and it is very costly in large and medium datasets (Jović et al. 2015).

Four categories namely filter; wrapper, embedded, and hybrid models can be used to classify the feature selection methods.

- *Filter approaches* A statistical analysis is applied to a feature set in the filter approach for solving the feature selection problem. (Moradi and Gholampour 2016). The filter methods evaluate and select the relevance of features by dint of a ranking method that rules out the low-scoring features. The filter methods have proven to be fast, scalable, computationally simple and classifier-independent. These methods are classified into two categories: univariate filter methods and multivariate filter methods. The univariate methods evaluate the features independently, thereby ignoring feature dependencies and creating poor feature subsets (Yongjun et al. 2012; Yusta 2009). Multivariate methods partly take into account feature dependencies and interaction with the classification algorithm unlike the univariate methods, which overlook these two parameters (Saeys et al. 2007).

Information gain is one of the commonly used criteria in this category for ranking the features. This measure is applied in this paper.

Information gain It is a measure developed based on the information theory of entropy. Entropy is a measure of disorderliness or noisiness, while information gain measures the decrease in entropy before and after inclusion of the features (Uğuz 2012; Yu and Liu 2004). A feature with a high information gain value is preferred to the other features. However, information gain does not remove the redundant features. The information gain of X provided by Y is calculated as follows:

$$IG(X|Y) = H(X) - H(X|Y), \tag{1}$$

where,

$$H(X) = -\sum_{i=1}^{k} P(x_i)\log_2(P(x_i)), \tag{2}$$

Is the entropy of the variable X, and

$$H(X|Y) = -\sum_{i=1}^{n} P(y_i)\sum_{j=1}^{k} P\left(x_j|y_i\right)\log_2\left(P\left(x_j|y_i\right)\right). \tag{3}$$

Is the entropy of X after observing another variable Y.

Continuous features need to be discretized when using entropy (Liu et al. 2002). Each feature is ranked based on its respective information gain value. In fact, a feature becomes more informative with an increase in the information gain.

- *Wrapper approaches* A set of methods which use Evaluation Function based on the error rate of learning method is called Wrapper method or Black Box. Each new subset of features in this method is created by Generation Function and this generation depends upon the search technique. Afterwards the generated subset is evaluated by machine learning method continuously. The number of Test-Suite errors, or error rate of learning method, specify the score of the subset (Tang et al. 2014). Wrapper approaches, apply a learning method to make a classifier in the assessment process. To generate a variety of feature subsets, they amplify or remove features followed by measuring the subsets according to the performance of the developed classifier. Generally speaking, the wrapper method has better performance than the filter method, but is more complex computationally (Purohit et al. 2010; Tran et al. 2014).
- *Hybrid approaches* These are the third category of approaches, which are intelligent as they try to take the advantages of both filter and wrapper approaches for feature selection. A trade-off is made by hybrid approaches between filter and wrapper procedures (Hsu et al. 2011; Ebrahimpour and Eftekhari 2017).
- *Embedded approach* These are the final approaches in this category that choose their features throughout the training stage. Furthermore, embedded approaches allow selecting the features by the classifier. As an example, Random Forest (Ye et al. 2013) represents one of the embedded approaches that work as described below. In the first step, some algorithms select the best data for training and subsequently, the relevant features are deter-

mined by the random forest (Chandrashekar and Sahin 2014; Ebrahimpour and Eftekhari 2017).

3.2 Particle swarm optimization

Particle swarm optimization (PSO) Algorithm is one of the strengths and well-known optimization tools for optimizing the features, which finds salient features through both local search and global search via repetition in features' space. Population in this algorithm consists of a set of accidental particles which move around the response space and optimized by repetition to find an optimized response. This procedure goes on until obtaining an appropriate convergence.

Here n stands for population's number of particles. The position of every ith particle of a c-dimension vector is represented as $X_i = \left(X_{i1}, X_{i2}, \dots, X_{ic}\right)$ in which c is the maximum number of dimensions of feature vector $(1 \leq j \leq c \textit{ and } 1 \leq i \leq n)$. Velocity vector of every ith particle is represented as $V_i = \left(V_{i1}, V_{i2}, \dots, V_{ic}\right)$.

Value of $Pbest_i\,(t)$: the best position of ith particle from the beginning till the repetition of t and $Gbest_i\,(t)$ is the best position of all particles of population till the repetition of t. velocity and position of particles are updated by the following Eqs. 4 and 5:

$$V_i(t+1) = w \times V_i(t) + c_1 \times u_1 \times \left(Pbest_i(t) - X_i(t)\right) + c_2 \times u_2 \times \left(Gbest_i(t) - X_i(t)\right), \tag{4}$$

$$X_i(t+1) = X_i(t) + V_i(t+1). \tag{5}$$

In the above equations u_1 and $u_2$are two incrementing numbers in (0, 1) and c_1 and $c_2$are individual and social impacts, respectively. There are three separate parts in the equation of updating velocity namely, inertia component, individual impact component, and social contact component, respectively. Weight parameter, w, in inertia component is a factor of controlling the balance in search algorithm. In the second component (individual cognition), updating of information is carried out based on the particle's local knowledge. Finally, in the third component updating is carried out based on the cooperation of particles. With 50 repetitions, this algorithm is selected as the most appropriate and effective feature (Seal et al. 2015).

3.3 Genetic algorithm

Genetic algorithm (GA) is a widely renowned evolutionary approach, which randomizes the generation of a population of strings. The best string in the population will access to an optimal solution by applying evolutionary operators, such as crossover, mutation, and reproduction (Hsieh et al. 2018). GA is a mathematical tool with a wide range of applications. It is very efficient in the optimization of the problems, especially when the respective objective functions are discontinuous and exhibit many local optima (Sadeghi et al. 2011). GA method can be simulated the search of the evolutionary phase, which is widely applied on feature selection problems. A greatly simplified and stylized simulation of the biological version of GA leads to the evolutionary process of GA. This process begins with a population of randomly generated individuals (chromosomes) specified by a probability distribution which is usually uniform and updates the population in stages that are referred to as generations. In each generation, the current population is employed to randomly choose multiple individuals according to the application of fitness function, bred using a crossover, and modified through mutation in order to generate a new population. The functions of genetic operators include:

- *Selection* Selection is concerned with the probabilistic survival of the fittest, i.e., more fit chromosomes are appointed to survive, where fitness represents a comparable measure that determines how well a chromosome can solve the current problem.
- *Crossover* In this operation, a random gene is chosen along the length of the chromosomes and all the genes after that point are swapped.
- *Mutation* This operation changes the new solutions to add stochastic in order to find better solutions. This is the probability that a bit within a chromosome will be flipped (Indira and Kanmani 2015).

Presumed on the evolutionary ideas of natural selection and genetic, genetic algorithms (GAs) are adaptive heuristic search algorithm. The concept of GAs is basically intended to mimic natural system processes essential for evolution. In the Genetic Algorithm (GA), the selection, crossover, mutation and other genetic operations are replaced to find the optimal solution to the problem. Starting with an initial feasible solutions set, the genetic algorithm carries out an efficient global search in the feasible domain using the objective function regardless of the other information. Thereafter, it converges to the global optimal solution with a probability of 1 (Zhang et al. 2018g, h). Thus, GAs manifests an intelligent utilization of a random search within a defined search space for problem-solving. There is little knowledge of the fact that GAs are among the best methods of solving a problem. As a very general algorithm, they will function properly in any search space. GA is a stochastic search algorithm, modeled on the natural selection process underlining biological evolution (Russell and Norvig 2008). Numerous search, optimization, and machine learning problems have successfully employed GA working iteratively by producing new populations of strings from old ones. Each string is the encoded binary, real, etc. version of a candidate solution. An evaluation function accompanies a fitness criterion to each string suggesting its fitness for the problem (Goldberg and Holland 1988).

4 Research methodology

In this paper, a novel hybrid feature selection approach has been proposed, which is the main application for the high-dimensional datasets. This approach is designed to decrease the search time and the computational complexity to extract the feature subset that yields the highest classification accuracy with the minimum number of features. To wit, the proposed algorithm reduces the search time spent to obtain the optimal solution to the feature selection problem of high-dimensional datasets. The idea is basically to use the information offered by the smaller datasets and use the information to rule out the features with a lower selection probability for being included in the final subset before implementing the feature section algorithm, so the feature selection algorithm is also applied to a smaller dataset to reduce the search time and the computational complexity of finding the optimal subset. To this end, the features with a lower selection probability for being included in the final subset are predicted and removed from the main dataset using a classification technique. In other words, the preprocessing of the features is carried out to evaluate their fitness using the statistical measure and the feature information prior to the implementation of the feature selection algorithm. As a result of this evaluation, some of the features are deleted in this phase. In fact, the proposed method is a combination of the wrapper and filter feature selection methods. The proposed approach also consists of two phases.

- In phase one, i.e. filter phase, some of the features are selected according to their competency. The competency of each criterion is measured using the statistical criteria and its selection probability for being included in the final subset is determined using the decision-tree technique.
- In phase two, i.e. wrapper phase, the final subset is selected using a hybrid approach developed by combining the genetic algorithm, PSO algorithm and the artificial neural network model. The features selected in phase one are the inputs of this phase and the proposed second phase approach will only be implemented on the selected features of the previous step.

The framework of the proposed approach is shown in Fig. 1 and each step is explained completely in the following.

4.1 Filter phase

As mentioned, this stage is in fact a pre-processing on the dataset's features and a filter to remove some features and reduce the size of the dataset. To this end, the information and patterns extracted from a smaller dataset are used to work with higher dimensional datasets. The features with a lower selection probability are predicted and removed from the features set using this information and the decision tree algorithm. Therefore, in the first step the hybrid algorithm, which is a combination of the genetic and PSO algorithms (HGP-FS), is implemented on the small dataset to select the effective features. In this research, the application of the algorithm to the selected dataset was iterated 40 times. The selected feature category is also find in each iteration. Afterwards, in step two the number of iterations of each feature is recorded in the selected subset. In the third step given the number of the presence of each feature in the selected subset, a frequency label (high frequency, medium frequency, low frequency, and the unseen feature) was attached to each feature. Table 1 explains this determination. In this case, frequency means the number of repetition. Afterward in step four, some characteristics and statistical criteria including the average, quartile, standard deviation, variance, and information gain are calculated per feature in the dataset to determine the utility and desirability of each feature. That is to state, a feature information table is created, wherein the columns present the average, quartile, standard deviation, variance, information gain, frequency rate values and frequency label and the rows introduce the dataset's feature. The "frequency label" field is also considered to be the target field. The same feature information table including the statistical criteria including information gain Eq. (3), average Eq. (7), variance Eq. (8), standard deviation Eq. (9), Q1 Eq. (10), Q2 Eq. (11), and Q3 Eq. (12), is created for the other bigger datasets.

Afterwards in the step five, the feature frequency label in a high dimensional dataset is predicted using a decision tree and the feature information table. Using decision is not as a measure for selecting features and the role of it is like a filter. Moreover, in this phase the accuracy of the decision tree is not important and has no effect in selecting features:

$$IG(X|Y) = H(X) - H(X|Y), \tag{1}$$

$$Avg\left(\bar{X}\right) = \frac{\sum_{i=1}^{n} X_i}{n}, \tag{7}$$

$$Var = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}, \tag{8}$$

$$Std = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}}, \tag{9}$$

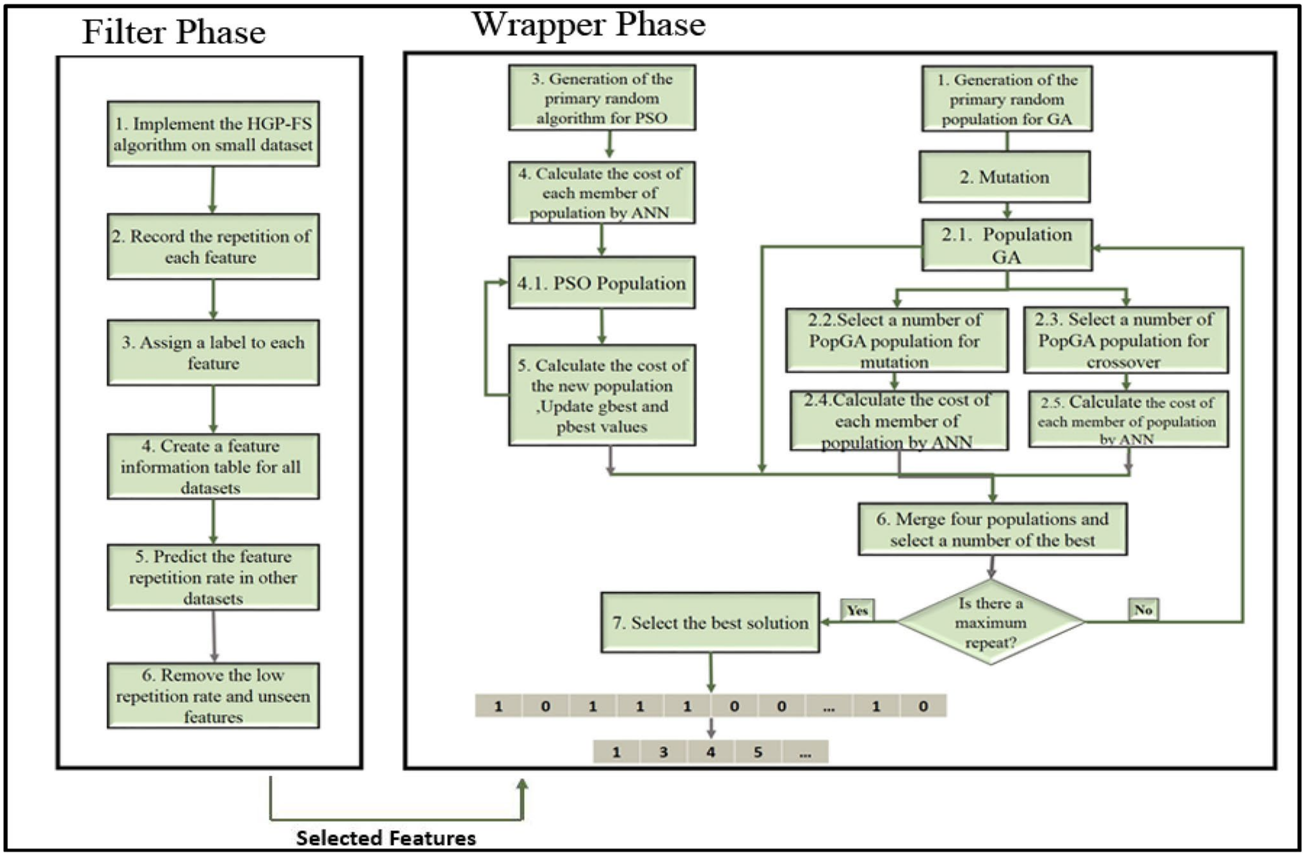

Fig. 1 the flowchart of the proposed algorithm

Table 1 Determining features' label

The number of feature repeat	Assigned label
25 and above	High frequency
16–24	Medium frequency
1–15	Low frequency
0	Unseen feature

$$Q1 = \frac{n+1}{4}, \tag{10}$$

$$Q2 = \frac{2(n+1)}{4}, \tag{11}$$

$$Q3 = \frac{3(n+1)}{4}. \tag{12}$$

The aim of creating the feature information table is to predict the feature frequency rate of the features in a higher dimensional dataset using a smaller dataset and a decision tree. In other words, the goal is to determine the characteristics of the features repeatedly included in the final subset of the small dataset. Table 1 explains the features' label.

Using the resulting information, the same of these features are selected in the other datasets and are passed to the next phase in step six. The other features are also removed. In fact, the main idea is to use the pattern identified in a smaller dataset for a bigger dataset. As a result, the features predicted to have a low iteration rate or no iteration are removed from the main dataset and are excluded from the rest of the algorithm process. Consequently, the computational complexity and implementation time decrease, improving the algorithm efficiency. In fact, prior to the implementation of the main algorithm (HGP-FS), a filter is defined for omitting some of the features to select the final subset using the information obtained from the smaller dataset.

4.2 Wrapper phase (HGP-FS) algorithm

In this phase, the features selected in the previous phase are searched to introduce the final subset. In this research, the main algorithm carries out the search for feature selection using a combination of the genetic algorithm and the particle swarm optimization (PSO) algorithm. There are explicit feature selection methods based on the mathematical criteria or the greedy algorithm. However, this method carries

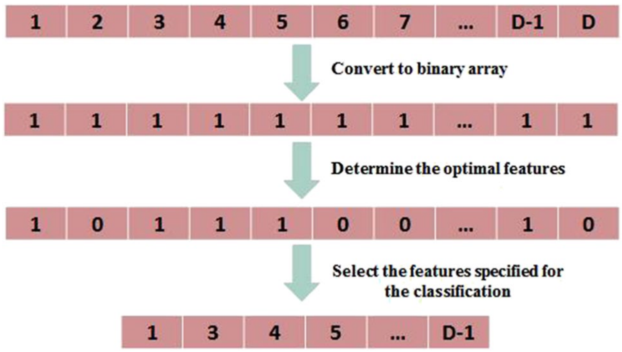

Fig. 2 Problem-making method for proposed algorithm

out the search using a combination of these two algorithms. Genetic algorithm and particle swarm optimization are evolution-based types of group computing technology (Sampath Kumar et al. 2010). The two approach somewhat possess their own features and benefits, as well as some faults and deficiencies (Su and Zhao 2017).

To this end, first the data are loaded with N records and D features. Therefore, the goal is to determine the feature selected from the D number of features as the appropriate feature to reduce the problem dimension provided that no harm is caused to the main problem (which is a classification problem in this research).

Therefore, it is important to select which one of these D features can maximize the classification accuracy. In general, when there are D features, the genetic problem is a D-dimensional problem. For the sake of simplicity, it is possible to imagine these D dimensions as a D-fold binary array, where 0 shows the non-selection of a feature and 1 shows the selection of a feature. Hence, it could be stated that the population members in the genetic algorithm and particle swarm optimization algorithms constitute a D-dimensional array containing binary values which is presented in Fig. 2.

The most important challenge in the meta-heuristic algorithms is the objective function or the cost function. In the proposed method, this function is known as the cost function and its input is an N × M matrix. Its output is also the classification error. The value of M equals the number of problem features (D) that will be used to measure the classification cost. It should be noted that $0 < M \leq D$. Besides, the value of M can never be zero because at least one feature must be selected for the classification.

The cost function is controlled through a neural network. That is to say, the selected data is put into a neural network, which classifies the data based on the corresponding labels. At last, the neural network returns an output as the classification error, which is assumed to be the cost (or "fitness") of each member of the population. The goal is to find the member that can impose the lowest cost on the problem.

As mentioned before, the population members must be D-dimensional. Hence, first two populations of certain sizes with D dimensions are generated stochastically. One population is used for the genetic algorithm and the other is used for the PSO algorithm. After generating the initial population and calculating the cost of each member, the method is implemented three times in parallel and it is repeated until the specified number of iterations is achieved. These three steps are taken for the genetic algorithm and the PSO algorithm. In the genetic algorithm, two steps are taken simultaneously, while the PSO algorithm involves only one step. The steps of the genetic algorithm are illustrated in Fig. 3.

Step one (population generation through crossover): In this procedure, first a certain number of members are selected using the roulette wheel algorithm. Afterwards, the crossover is performed for every two members of the population (which are called the parents in this paper). Crossover is carried out in the following three states: single-point, double-point, and uniform. One state is selected stochastically in the crossover process.

Hence, in the step one the offspring are generated within crossovers.

In step two the other offspring are generated through mutation. In mutation, one member of the population is selected and mutation is carried out for all values using a given probability value. It is noteworthy that some of the values might change depending on the probability values. For instance, for a member with 60 values and a mutation probability of 0.2, only two members are subjected to mutation and the values are reversed.

Therefore, the second step generates some other offspring by offspring of mutation and the cost of each population member is calculated using the cost function.

In step three, other offspring are generated using the particle swarm optimization (PSO) algorithm, which is designed based on the collective intelligence of animal species. In addition, animals that collectively search for food are constantly moving and searching for food based on two factors: the best search accomplished by each member of the population and the best search done in the entire population. Over time, the population moves in the direction that leads to the best solution.

This method is implemented in parallel to the genetic algorithm (GA), and in each movement, the population moves in the direction that leads to the new solutions that are better than the previous solutions. In step four when the population moves in a certain direction, the cost of each population member is computed by the cost function. Afterwards, in step five the g_{best} and p_{best} variables are updated according to Eqs. (4) and (5) for each member considering the new costs. Here, g_{best} is the best solution seen by the population so far, and p_{best} is the best solution observed by

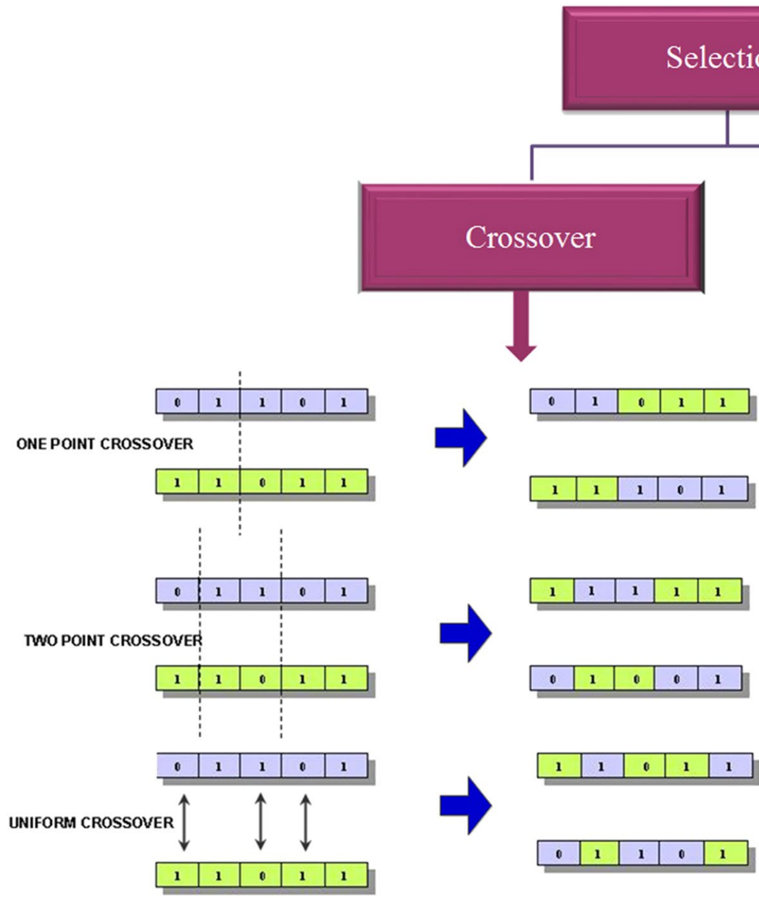

Fig. 3 Schema of the crossover and mutation operations

each member so far. Next, the same solutions are passed to the next step of the particle swarm optimization algorithm.

$$V_i(t+1) = w \times V_i(t) + c_1 \times u_1 \times \left(Pbest_i(t) - X_i(t)\right) + c_2 \times u_2 \times \left(Gbest_i(t) - X_i(t)\right), \tag{4}$$

$$X_i(t+1) = X_i(t) + V_i(t+1). \tag{5}$$

After the same movement is done in these three steps, four populations are obtained as follows:

1. The initial genetic population.
2. The crossover population.
3. The mutation population.
4. The PSO population.

At the end of each movement (generation) in step six, these four generations are merged and a particular number of them is selected for the next generation of the genetic algorithm (GA). This selection is based on the cost and fitness of the population, while the PSO population does not change.

At the final step after reaching the specified number of iterations, the algorithm ends. It finds the best solution in the last resulting population and introduces it as the final solution, which determines the feature to be selected to ensure the satisfactory classification accuracy.

All of the algorithm steps are illustrated in the Algorithm 1 and the pseudocode of proposed algorithm is explained in Fig. 4.

Algorithm1:

Step of algorithm
Filter phase:
1. Implement the HGP-FS algorithm on the dataset with fewer fields for effective feature selection
2. Record the number of repetition of each feature in the selected dataset
3. Assign a label (high frequency, average frequency, low frequency, and the unseen feature) to each feature
4. Create a feature information table, wherein the columns show the average, quartile, standard deviation, variance, information gain, and frequency rate values and the rows show the features of the initial dataset
5. Create a feature information table for the other bigger datasets
6. Predict the feature frequency rate in the datasets using the feature information table of the smaller data and the decision tree
7. Remove the features with low frequency and the unseen feature and send the other features to phase two
Wrapper phase (HGP-FS):
8. Generation of the primary random population for genetic algorithm (PopGA) and calculation of the cost of each part of the population
9. Stochastically generate an initial population for the particle swarm optimization (PSO) algorithm (PopPSO) and calculate the cost of each population member
10. Do Steps 11 to 19
11. Select a number of PopGA members for the crossover (PopC)
12. Perform the crossover on the selected PopC population and then calculate the population cost
13. Select a number of PopGA members for the mutation operation (PopM)
14. Perform mutation on the PopM population and then calculate the population cost
15. Move the PopPSO population and then calculate the new population cost
16. Update the g_{best} and p_{best} values in the particle swarm optimization algorithm
17. Merge populations PopGA, PopC, PopM, and PopPSO
18. Sort the merged population by cost
19. Select the best members of the merged population for the number of the PopGA members to create the new PopGA population
20. Go to step 11 if the counter does not reach the specified limit
21. Find the best solution from the final merged population
22. Extract the specified features from the data based on the final solution and use it to train the classifier

```
Start:
    Filter phase:
      Fs = Features of a small dataset
      Fb = Features of bigger datasets
      Selected features<-- empty set
      Run the HGP-FS algorithm on the dataset with fewer fields
          For each feature fd ∈ Fs
                Frequency rate = the number of repetition of feature in the selected dataset
                Assign a label according to Frequency rate
                Calculate the Avg, Q1, Q2, Q3, Std, Var, Info gain
          End For
      Create a feature information table for the initial dataset
      For each feature fd ∈ Fb
          Calculate the Avg, Q1, Q2, Q3, Std, Var, Info gain
      End For
      Create a feature information table for the other bigger datasets
      Predict the feature frequency rate in the bigger datasets
      For each feature fd ∈ Fb
          If the assigned label is low frequency or unseen Then
              Remove the feature
          Else
                Selected features <-- the feature
      End If
      End For
    End Filter phase
    Wrapper phase (HGP-FS):
      PopGA <-- Generate the random initial population for GA algorithm
      For each population member
              Calculate the population cost
      End For
      PopPSO <-- Generate an initial population for the PSO algorithm
      For each population member
              Calculate the population cost
      End For
      While (the counter does not reach the specified limit) do
          PopC <-- Select a number of PopGA members
          Crossover with a crossover probability crossover the parents
          For each population member
                Calculate the population cost
          End For
          PopM <-- Select a number of PopGA members
          Mutation with a mutation probability mutate
          For each population member
                Calculate the population cost
          End For
          For each Particle do
                Update velocity                     according Eq. (4)
                Update particle-position            according Eq. (5)
                Evaluate Fitness Functions
                Update Pbest
                Update Gbest
          End For
          Merge populations PopGA, PopC, PopM, and PopPSO
          Sort the merged population by cost
          Select the best members of the merged population for the number of the PopGA
          Create the new PopGA population
      End While
      Find the best solution from the final merged population
      Extract the specified features from the data based on the final solution
      Use the specified features to train the classifier
End.
```

Fig. 4 The pseudocode of proposed algorithm

Table 2 Definition of the datasets

Dataset	No. of features	No. of classes	No. of examples
Lung	56	3	32
Hill-Valley	100	2	606
Gas 6	128	6	476
Musk 1	166	2	476
Madelon	500	2	2000
Isolet 5	617	26	1559

5 Experiment design

This paper has investigated several experiments to prove the effectiveness of the proposed algorithm. The steps for implementing the proposed algorithm on the real-world datasets are explained in this section.

5.1 Datasets

The proposed algorithm (Smart HGP-FS) was implemented on six large-scale datasets received from the UCI to evaluate the potency of this approach. Table 2 introduced these datasets. The datasets include a different number of variables, classes and examples that preparing a general and wide survey of the suggested and employed approaches.

5.2 Implementation of the first phase (filter)

As explained in Sect. 4, the HGP-FS algorithm is first executed with a number of 40 iterations on the "Lung" dataset with 52 features. The selected subset of features is selected from each iteration; and then, the number of repetitions for each feature in these 40 iterations is recorded. A label is assigned to each feature based on its number of repetitions (high repetition, medium repetition, low repetition, and no repetition). In the next step, an information table is formed for this data set. In this information table, a number of statistical criteria –including the average, variance, standard deviation, quartiles, and information gain– are calculated and recorded for each feature. Moreover, the repetition rate and the label of each feature are also added to the table as the target field. An overview of the first five features in Lung data set is plotted in Table 3.

This information table of the features—which includes the statistical measure of the average, variance, standard deviation, quartiles, and information gain—, can also be arranged for other data sets with higher dimensions (Hill-Valley, Gas 6, Musk 1, Madelon, and Isolet 5). Following that, the repetition rates of the features in the mentioned datasets are predicted using a decision tree (ID3) and the information table of the features. Features with high and medium predicted repetition rates are chosen in this step and transferred to the next phase (wrapper phase). Other features with low repetition or no repetition labels are eliminated in this step. The execution results for this phase are recorded in set in this step is recorded in Table 4. In this table, the number of features selected from each dataset in this step is recorded.

5.3 Implementation of the second phase (Wrapper)

In this step, the implemented algorithm (HGP-FS) is executed on the Madelon, Gas 6, Hill-Valley, Musk 1, and Isolet datasets in order to search for the optimal subset. It is worth mentioning that the search is not conducted on all the features in these data sets. In this step, the HGP-FS algorithm only examines the selected features from the previous step, and the final subset will be selected from those features.

6 Experimental results

The proposed method was compared with similar methods from two perspectives:

1. From the analytical perspective and performance of the algorithm.

The proposed algorithm is a wrapper-filter combination. So far, the proposed combined methods have been implemented using an evolutionary algorithm, but the method proposed in this paper has been presented using the combination of two powerful algorithms of GA and PSO. In the filter phase of this method, features which are less likely to be selected as appropriate features are removed using the decision tree. In this method, the existing data in the smaller datasets are used to remove the inappropriate features in the larger datasets. This idea was not presented in previous methods. The process of omitting unimportant features in high dimensional large datasets is time-consuming. In addition to being a combined method, the proposed method tries to reduce time, something which has been neglected in the previous methods. In the similar methods presented so far, only the increase of prediction accuracy was considered, but the reduction of the searching time and computational complexity are among the objectives of the proposed algorithm.

2. From the perspective of the result and efficiency of the algorithm.

Until obtain more precise analysis, the proposed algorithm has been executed 10 times on every dataset, and the average results are recorded in Tables 5, 6, 7 and 8. In the

Table 3 An overview of the first five features in Lung dataset

Feature	Avg	Var	Std	Q1	Q2	Q3	Info gain	Fre-quency rate	Frequency label
F1	0.031	0.030	0.174	0	0	0	0.042	29	High
F2	0.68	0.074	0.272	0.5	0.5	1	0.207	20	Medium
F3	0.68	0.114	0.338	0.42	0.67	1	0.129	16	Low
F4	0.67	0.072	0.268	0.5	0.5	1	0.145	33	High
F5	0.28	0.202	0.449	0	0	1	0.005	11	Low

Table 4 Result of first phase

Dataset	No. of the selected features in first phase
Hill-Valley	53
Gas 6	100
Musk 1	115
Madelon	320
Isolet 5	378

Table 5 Result of smart HGP-FS vs. HGP-FS

Dataset	Criterion	HGP-FS	Smart HGP-FS
Hill-Valley	Ave-Size	43.85	**24.2**
	Ave-Acc	55.08	**58.02**
	Best-Acc	56.83	**60.66**
	Ave-Time (min)	360	**60**
Gas 6	Ave-Size	64	**44.75**
	Ave-Acc	100	**99.82**
	Best-Acc	100	**100**
	Ave-Time (min)	250	**30**
Musk 1	Ave-Size	84.5	**80.33**
	Ave-Acc	84.23	**86.11**
	Best-Acc	85.81	**89.58**
	Ave-Time (min)	390	**90**
Madelon	Ave-Size	483.5	**226.5**
	Ave-Acc	70.46	**59.12**
	Best-Acc	71.25	**60.75**
	Ave-Time (min)	1000	**420**
Isolet 5	Ave-Size	351.33	**183**
	Ave-Acc	76.31	**77.62**
	Best-Acc	77.69	**78.51**
	Ave-Time (min)	900	**360**

experiments, all of the instances in each dataset are randomly divided into two sets: 70% as the training set and 30% as the test set. The best result in the comparisons is shown in bold format for each data. The classification performance of a selected feature subset is assessed by 10-fold cross-validation on the test set. The most important performance factor in classification algorithms is the prediction accuracy of unobserved data. Feature selection techniques have two main goals: increasing the prediction accuracy and reducing the numbers of the selected features. Therefore, these two important criteria (the prediction accuracy and the size of the selected subset) are examined and compared in this research in order to evaluate the performance of feature selection techniques. The prediction accuracy can be calculated by Eq. (6). Where TP and TN are true positives and negatives, and FP and FN are false positives and negatives:

$$Accuracy\ Rate = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

The comparisons performed between this method and other techniques are arranged in four sections. This hybrid method (Smart HGP-FS) is first compared to the HGP-FS wrapper algorithm, and the results of this comparison are recorded in Table 5. Since the proposed method in this research is a hybrid method, it is compared with variant kinds of techniques (hybrid, filter, wrapper, and traditional methods) in order to prove its efficiency. The results are plotted in the Tables 6, 7, and 8.

The k-NN method is one of major and powerful classification algorithms and basis of this method is on a similarity measure. To assess the efficiency of the presented feature selection approach, the five nearest neighbors (5-NN) is applied. Experimental results on five datasets are reported in Table 5, 6, 7 and 8. "Ave-Size" shows the mean size of the feature subsets selected by proposed algorithm in 10 implementations. In experimental results the mean test accuracy of the feature subsets selected by each algorithm is represented by "Ave-Acc", the best test accuracy is showed by "Best-Acc".

6.1 Competitor algorithms

To measure the performance of the Smart HGP-FS approach, three hybrid methods: FW-NSGA-II (Hammami et al. 2018), FastPSO and RapidPSO (Butler-Yeoman et al. 2015), two pure filter algorithm: GA based framework called FilterAlgo (Xue et al. 2013a) and FilterPSO (Butler-Yeoman et al. 2015), two pure wrapper approaches: GA based algorithm called WrapperAlgo (Mukhopadhyay and Maulik 2013) and WrapperPSO (Butler-Yeoman et al. 2015) and two traditional wrapper feature selection techniques: LFS (Gutlein

Table 6 Result of smart HGP-FS vs. hybrid approaches

Dataset	Criterion	5-nn All	FW-NSGA-II	FastPSO	RapidPSO	Smart HGP-FS
Hill-Valley	Ave-Size	100	52.6	56.1	61.0	**24.2**
	Ave-Acc	55.45	56.80	54.13	54.00	**58.20**
	Best-Acc	55.45	60.20	56.44	56.68	**60.66**
Gas 6	Ave-Size	128	54.8	55.4	63.2	**45.8**
	Ave-Acc	99.82	98.01	99.88	**99.91**	99.82
	Best-Acc	99.82	99.07	**100**	**100**	**100**
Musk 1	Ave-Size	166	82.1	91.1	94.4	**80.33**
	Ave-Acc	76.10	78.31	78.99	78.41	**86.11**
	Best-Acc	76.10	78.76	82.39	83.02	**89.58**
Madelon	Ave-Size	500	250.2	258.5	268.6	**226.5**
	Ave-Acc	52.60	56.08	54.41	54.00	**59.12**
	Best-Acc	52.60	61.06	58.13	57.55	**60.75**
Isolet 5	Ave-Size	617	310.5	316.1	326.8	**186.75**
	Ave-Acc	77.12	**79.60**	77.71	77.33	**79.35**
	Best-Acc	77.12	**82.75**	79.62	78.65	79.74

Table 7 Result of smart HGP-FS vs. filter and wrapper approaches

Dataset	Criterion	FilterPSO	FilterAlgo	WrapperPSO	WrapperAlgo	Smart HGP-FS
Hill-Valley	Ave-Size	48.9	62.9	50.4	49.7	**24.2**
	Ave-Acc	54.53	50.28	54.12	52.78	**58.20**
	Best-Acc	57.43	55.33	58.17	57.77	**60.66**
Gas 6	Ave-Size	66.5	69.15	42.0	**39.89**	45.8
	Ave-Acc	**99.87**	89.40	**99.86**	98.38	**99.86**
	Best-Acc	**100**	97.23	**100**	**100**	**100**
Musk 1	Ave-Size	82.9	80.91	81.9	82.6	**80.33**
	Ave-Acc	78.05	71.93	79.41	79.89	**86.11**
	Best-Acc	85.53	81.06	85.53	86.20	**89.58**
Madelon	Ave-Size	307.5	310.0	245.5	244.2	**226.5**
	Ave-Acc	52.48	51.58	54.08	55.28	**59.12**
	Best-Acc	55.71	54.11	57.32	59.99	**60.75**
Isolet 5	Ave-Size	369.9	365.8	306.6	306.0	**186.75**
	Ave-Acc	75.87	69.87	78.01	77.98	**79.35**
	Best-Acc	78.27	77.54	**79.81**	78.98	79.74

Table 8 Result of smart HGP-FS vs. traditional approaches

Dataset	Criterion	LFS	GSBS	Smart HGP-FS
Hill-Valley	Size	9	95	24.2
	Acc	55.49	54.40	**58.20**
Gas 6	Size	4	8	45.8
	Acc	99.82	99.12	**99.86**
Musk 1	Size	12	124	80.33
	Acc	80.71	82.86	**86.11**
Madelon	Size	7	250	226.5
	Acc	71.03	**74.88**	59.12
Isolet 5	Size	27	585	186.75
	Acc	76.28	**80.77**	**79.35**

et al. 2009) GSBS (Caruana and Freitag 1994) are applied as benchmark approaches in the experiments.

6.2 Performance comparison with HGP-FS algorithm

The HGP-FS method is an independent hybrid approach applying the GA and PSO evolutionary algorithms for the feature selection problem. In this research, a pre-processing phase is added to this method until to increase the speed and decrease the search time. The purpose of this section of the comparison is to demonstrate the effect of the first phase of the smart HGP-FS algorithm (the filter phase) on the overall results. In this table, the average execution time of one iteration for each data set in the method is presented in minutes.

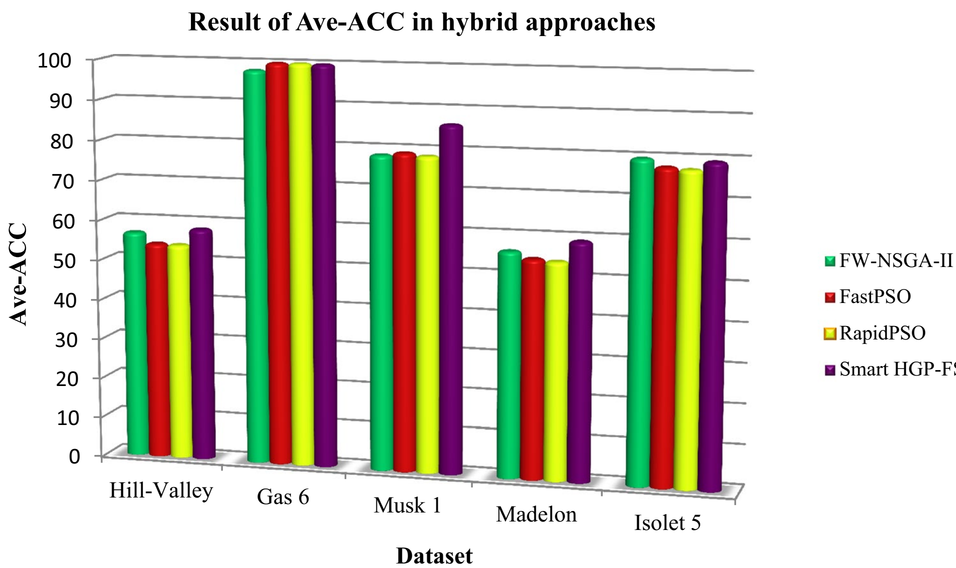

Fig. 5 Result of Ave-Acc in hybrid approaches

Another aim of this section is a comparison between search times with and without the filter phase.

As evident in, the results of the smart HGP-FS approach are meaningfully better than HGP-FS in points of increasing prediction accuracy and decreasing the number of selected features; which implies that combining the filter method with the HGP-FS approach significantly increases the efficiency of feature selection process. In aspect of search time, it is evident that using the filter phase decreases the search time and somewhat resolves the problem of time complexity in wrapper methods. In general, the results of the smart HGP-FS algorithm have improved with respect to HGP-FS in aspect of accuracy, the size of the selected subset, and search time. These criteria distinctly prove the superiority of the smart HGP-FS method.

6.3 Performance comparison with hybrid approaches

In this section, the gained results of the Smart HGP-FS approach and three hybrid filter-wrapper methods are surveyed in order to compare the efficiency of the proposed algorithm with other hybrid methods. The first column in Table 6 lists the results of the 5-NN with all variables and no feature selection. The FastPSO and RapidPSO methods are implemented using the PSO algorithm, and the FW-NSGA-II method is executed using the GA. The results of the Table 6 presents the full superiority of the method proposed by this research in all datasets. The performance comparison is shown in Figs. 5 and 6. Using the Smart HGP-FS algorithm, the classification accuracy is higher than the other hybrid methods in all data sets except Isolet 5. This method also demonstrates much higher performance in point of the size of the selected subset. As can be observed, the number of selected features using the Smart HGP-FS algorithm is much smaller than the other methods; which demonstrates the significant improvement of the proposed method regarding this criterion. For example in the developed approach for Madelon dataset, the average number of selected features is 226.5 that is lower than the other approaches and the average obtained classification accuracy is 59.12 where is higher than the other techniques. In the Isolet 5 dataset the classification accuracy by applying Smart HGP-FS method is only around 0.25% lower than the FW-NSGA-II technique. Since the number of selected features in the Smart HGP-FS method is almost half the number of features in the FW-NSGA-II method, it can be stated that the overall performance of the Smart HGP-FS method is higher in this dataset as well.

The classification accuracy in the Smart HGP-FS algorithm leads to the selection of a smaller subset, which in turn increases the prediction accuracy. Therefore, this method is also more successful than the other methods in eliminating the irrelevant features according to the results and comparisons. In general, it can be expressed that the proposed algorithm has a better efficiency in comparison with the other hybrid approaches.

6.4 Performance comparison with filter and wrapper approaches

In this section, the result of the proposed method is surveyed in contrast to the pure filter and pure wrapper approaches. The FilterAlgo and WrapperAlgo methods are developed based on the GA and the FilterPSO and WrapperPSO are developed using the PSO algorithm. The comparison in this section is conducted with two objectives: first, to demonstrate the supremacy of the combination of the filter and wrapper

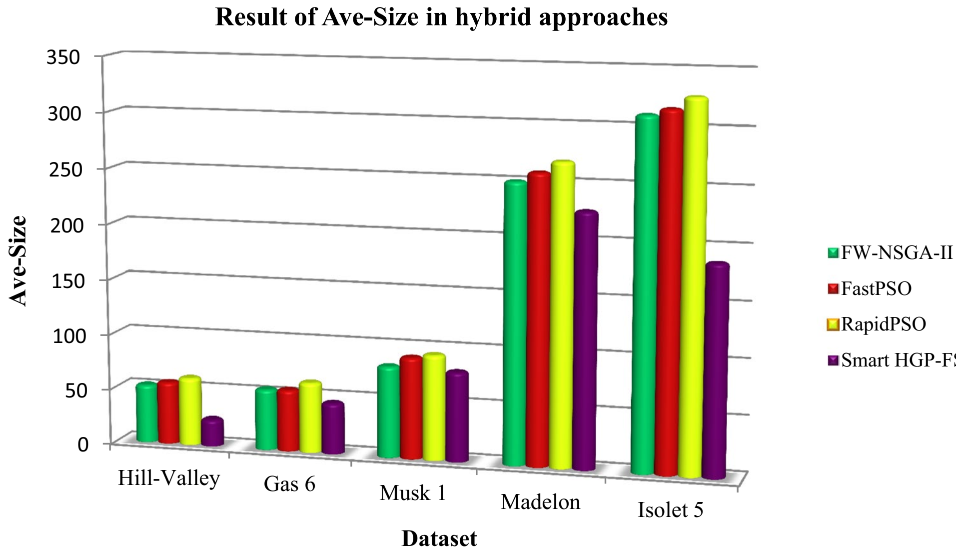

Fig. 6 Result of Ave-Size in hybrid approaches

approaches to non-hybrid approaches; and second, to present the effect of combining the GA and PSO evolutionary algorithms in comparison with basic evolutionary algorithms. The recorded results in Table 7 demonstrate the better performance of the hybrid Smart HGP-FS method with respect to the pure filter and pure wrapper methods in terms of classification accuracy and the size of the selected subset. The performance comparison is shown in Figs. 7 and 8. The proposed hybrid method demonstrates higher performance and better results with respect to the filter methods in all datasets. The performance of the hybrid smart HGP-FS method is higher than the wrapper methods in all states and datasets except for the Gas 6 dataset (in terms of the size of the selected subset). As mentioned in the below table for Hill-Valley dataset the number of selected features in the Smart HGP-FS method is half the other methods with the higher classification accuracy. In the proposed algorithm for Isolet 5 dataset, the gained prediction accuracy is similar to the other approaches but this accuracy is achieved by just 186.75 features where 50% is lower than other techniques.

6.5 Performance comparison with traditional approaches

Moreover, the proposed method is compared the deterministic approaches including the linear forward selection (LFS) (Gutlein et al. 2009) that uses forward search to add selected features and greedy step wise based selection (GSBS) (Caruana and Freitag 1994) that removes features by applying backward search approaches derived from the sequential forward (SFS), sequential floating forward (SFFS) and sequential backward (SBS) feature selection approaches are selected. The comparison between proposed approach and the deterministic approaches are mentioned in Table 8.

According to the indicated results in Table 8, the proposed HGP-FS algorithm obtained better results compared

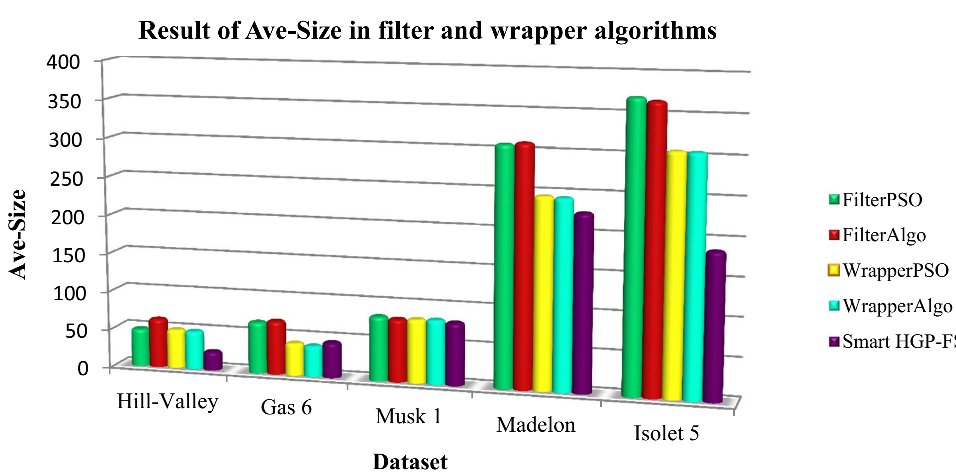

Fig. 7 Result of Ave-Size in filter and wrapper algorithms

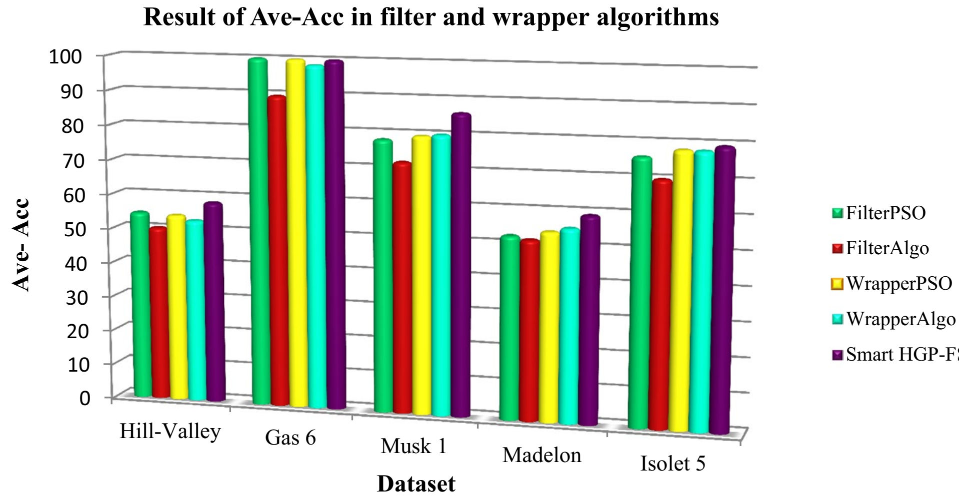

Fig. 8 Result of Ave-Acc in filter and wrapper algorithms

Table 9 The summary of results

Dataset	Criterion	Average of results				Smart HGP-FS superiority than other approaches		
		Smart HGP-FS	Filter approaches	Wrapper approaches	Hybrid approaches	Filter approaches (%)	Wrapper approaches (%)	Hybrid approaches (%)
Hill-Valley	Size	24.2	55.9	50.05	56.57	56.71	51.65	57.22
	Acc	58.20	52.40	53.45	54.98	9.96	8.16	5.54
Gas 6	Size	45.8	67.82	40.94	57.80	32.47	− 11.86	20.76
	Acc	99.82	94.63	99.12	99.27	5.23	0.74	0.59
Musk 1	Size	80.33	80.91	82.25	89.20	1.92	2.33	9.94
	Acc	86.11	74.99	79.65	78.57	12.91	7.50	8.76
Madelon	Size	226.5	308.75	244.8	259.10	26.64	7.49	12.58
	Acc	59.12	52.03	54.68	54.83	11.99	7.51	7.26
Isolet 5	Size	186.75	367.85	54.68	317.80	49.23	39.03	41.24
	Acc	79.35	72.87	77.99	78.21	8.17	1.71	1.43
Total	Size					33.40	17.73	28.35
	Acc					9.65	5.12	4.72

to traditional and deterministic algorithms. In three out of five aspects, the present method obtained higher accuracy of classification compared to other procedures. Although LFFS and LFS procedures have selected those datasets with fewer features, the accuracy of classification resulted from those datasets was proved to be low. GSBS algorithm performs aggressively with the whole initial datasets the process of feature selection. Further, this method selects a subset with the higher numbers of features and did not reveal an improvement in the accuracy of classification. Therefore, the proposed procedure showed a higher efficiency compared to deterministic methods and extracts a subset that resulted in the gained higher accuracy of the classification and a remarkable improvement in this area.

7 Discussion

In this paper, a hybrid (filter- wrapper) feature selection is presented. The principal purpose of this method is to reduce the computational complexity and search time in datasets with high dimensions. The comparisons are performed between five datasets with high dimensions. In the first step of the comparisons, this method is compared with HGP-FS; which is a pure wrapper method without the filter phase. The goal of this comparison is to evaluate the effect of the filter phase on the overall results of the algorithm. The results demonstrated that using the filter phase and eliminating some of the features in this phase has a significant effect on the improvement of the algorithm's results. Adopting the

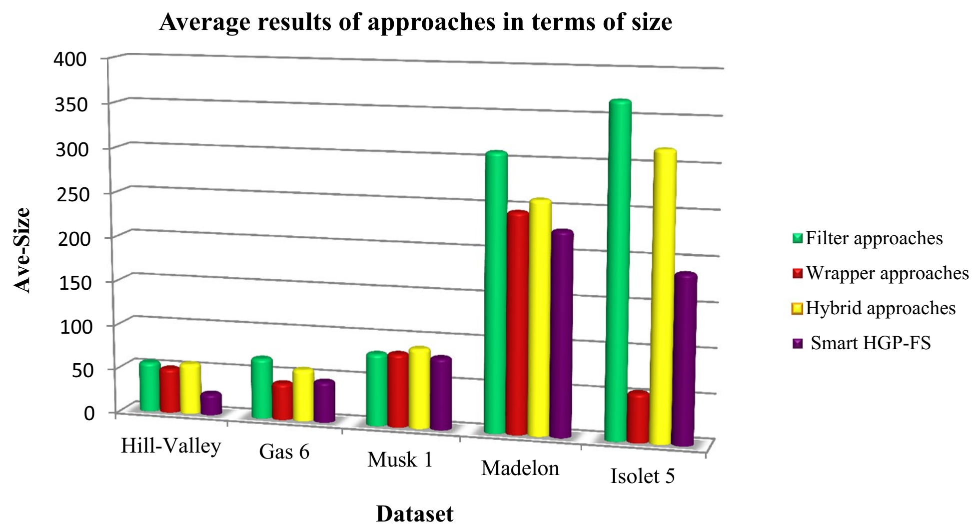

Fig. 9 Average results of approaches in terms of size

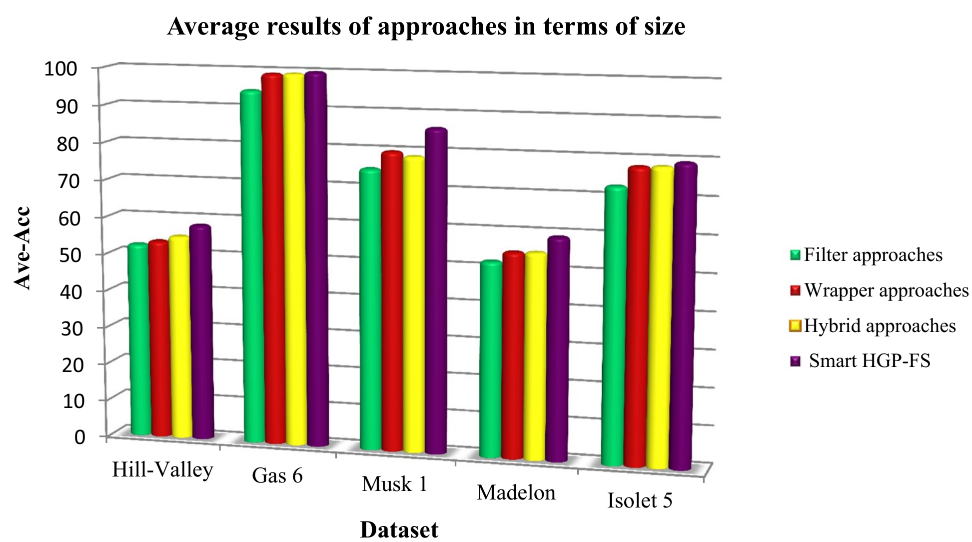

Fig. 10 Average results of approaches in terms of prediction accuracy

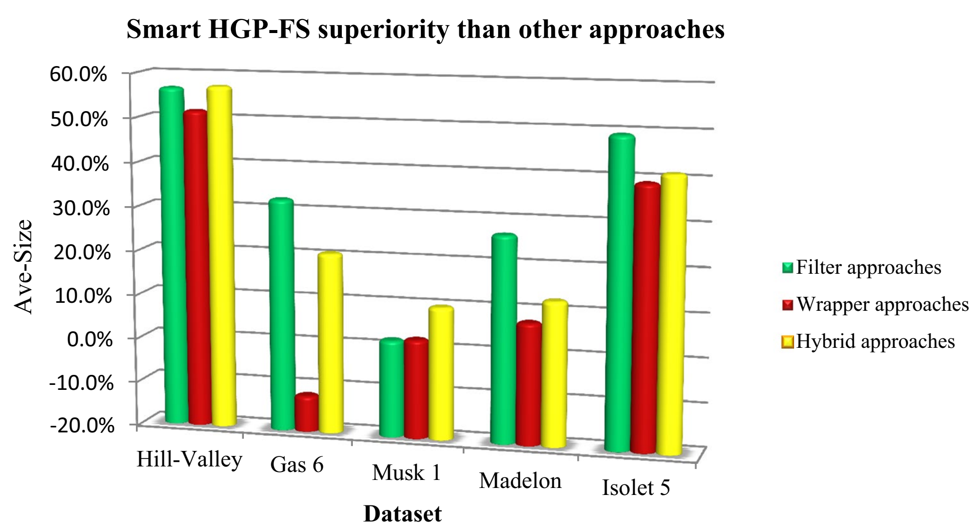

Fig. 11 Smart HGP-FS superiority than other approaches in terms of the number of selected features

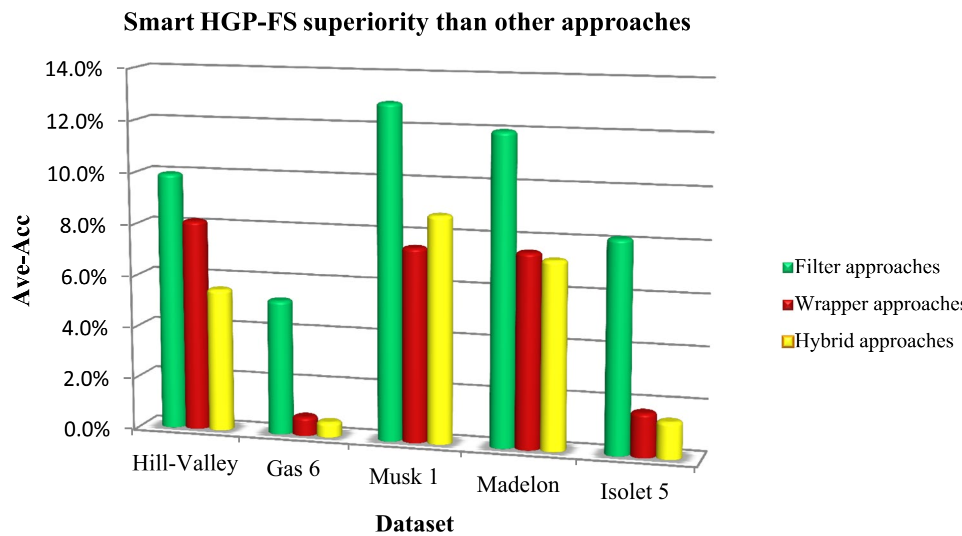

Fig. 12 Smart HGP-FS superiority than other approaches in terms of the prediction accuracy

filter phase leads to the selection of a smaller number of features in a shorter period of time; resulting in an increase in the accuracy of the classification through all datasets.

Following that, the proposed algorithm is compared to three hybrid methods (a GA-based method, and two PSO-based methods), two filter methods (a GA-based method and a PSO-based method), and two wrapper methods (a GA-based method and a PSO-based method). The comparison with other hybrid methods proves that the combination of GA and PSO algorithm improves the results of the hybrid methods. The results of the comparisons with pure filter and pure wrapper algorithms indicate that the hybridization of the filter and wrapper approaches—as well as the combination of the GA and the PSO algorithm—increases the efficiency of the feature selection algorithm. The summary of the conducted comparisons is presented in Table 9. In this table, the obtained results of the Smart HGP-FS algorithm are compared separately with three filter, hybrid, and wrapper approaches. For each approach, the first column (Ave) shows the average results of the compared algorithms, and the second column (Superiority Ratio) proves the superiority percentage of the Smart HGP-FS method with respect to this average value. As expressed in the table, the proposed algorithm has a 28.58% improvement ratio in terms of the subset size, and 4.27% in point of classification accuracy with respect to the hybrid approaches. Also, the superiority percentage of this method with respect to the wrapper methods is 17.97% higher in terms of the number of selected features, and 4.68% higher in terms of prediction accuracy. The performance of smart HGP-FS is 33.60% better than the filter methods in terms of the size of the selected subset, and 9.24% better in terms of classification accuracy. This method is significantly more superior to all the other methods in terms of selected subset size due to the pre-processing and filters phases. The results of this section are compared in Figs. 9, 10, 11 and 12.

In general, it can be stated that the proposed approach—with respect to the other algorithms—can extract a subset with a smaller number of features and higher classification accuracy and performs better in eliminating the irrelevant variables of the dataset.

8 Conclusions

The main objective of this paper is to present a novel and integrated technique to select relevant features in the dataset and remove the non-relevant features. For this purpose, a method was presented in the present work by integrating three powerful algorithms named PSO, GA and ANN having all the advantages and strengths together. The presented method mainly intends to adopt the information drawn out from smaller datasets to decrease the complex computation of feature selection with large datasets. In the main, integration of GA and PSO algorithms aims to compromise between examination and utilization leading to an optimal solution. The mentioned method, called Smart HGP-FS, used the ANN algorithm in the objective function. The current study has tried to develop an algorithm with no earlier faults through integration of the above algorithms and the use of their merits, i.e., the high convergence speed of PSO algorithm, and exploring the capability of GA. Since the proposed method in this research is a hybrid method, it is compared with different kinds of techniques (hybrid, filter, wrapper, and traditional methods) in order to prove its efficiency. To do this, the five nearest neighbors (5-NN) is applied as a classifier. The findings obtained from comparing the developed algorithm and related algorithms demonstrated a superior efficiency of the introduced algorithm to remove unimportant features and improve the precision of classifying the investigated datasets confirming the stability of the developed approach. Therefore, this algorithm will be beneficial as a reliable and effective procedure in solving the feature selection problems.

References

Apolloni J, Leguizamón G, Alba E (2016) Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments. Appl Soft Comput 38:922–932

Bennasar M, Hicks Y, Setchi R (2015) Feature selection using joint mutual information maximisation. Expert Syst Appl 42(22):8520–8532

Bostani H, Sheikhan M (2017) Hybrid of binary gravitational search algorithm and mutual information for feature selection in intrusion detection systems. Soft Comput 21(9):2307–2324

Butler-Yeoman T, Xue B, Zhang M (2015) Particle swarm optimisation for feature selection: a hybrid filter-wrapper approach. In: CEC, pp 2428–2435

Cervante L, Xue B, Zhang M, Shang L (2012) Binary particle swarm optimisation for feature selection: a filter based approach. In: Evolutionary computation (CEC), 2012 IEEE congress. IEEE, pp 1–8

Chandrashekar G, Sahin F (2014) A survey on feature selection methods. Comput Electr Eng 40(1):16–28

Chen KJ, Ji P (2007) A genetic algorithm for dynamic advanced planning and scheduling (DAPS) with a frozen interval. Expert Syst Appl 33(4):1004–1010

Chinnaswamy A, Srinivasan R (2017) Hybrid information gain based fuzzy roughset feature selection in cancer microarray data. In: Power and advanced computing technologies (i-PACT), 2017 Innovations. IEEE, pp 1–6

Chuang LY, Ke CH, Yang CH (2016) A hybrid both filter and wrapper feature selection method for microarray classification. arXiv:1612.08669

Dash M, Liu H, Yao J (1997) Dimensionality reduction of unsupervised data. In: Tools with artificial intelligence, 1997. Proceedings, 9th IEEE international conference. IEEE, pp 532–539

Ebrahimpour MK, Eftekhari M (2017) Ensemble of feature selection methods: a hesitant fuzzy sets approach. Appl Soft Comput 50:300–312

Caruana R, Freitag, D (1994) Greedy attribute selection. In: Proceedings of the 11th international conference on machine learning, New Brunswick

Frohlich H, Chapelle O, Scholkopf B (2003) Feature selection for support vector machines by means of genetic algorithm. In: Tools with artificial intelligence, 2003. Proceedings. 15th IEEE international conference. IEEE, pp 142–148

Goldberg DE, Holland JH (1988) Genetic algorithms and machine learning. Mach Learn 3(2):95–99

Gu Q, Li Z, Han J (2011) Generalized fisher score for feature selection. In: Proceedings of the international conference on uncertainty in artificial intelligence

Gu S, Cheng R, Jin Y (2018) Feature selection for high-dimensional classification using a competitive swarm optimizer. Soft Comput 22(3):811–822

Guo H, Jack LB, Nandi AK (2005) Feature generation using genetic programming with application to fault classification. IEEE Trans Syst Man Cybern Part B (Cybernetics) 35(1):89–99

Gutlein M, Frank E, Hall M, Karwath A (2009) Large-scale attribute selection using wrappers. In: Computational intelligence and data mining, 2009. CIDM'09. IEEE symposium. IEEE, pp 332–339

Haeri A, Tavakkoli-Moghaddam R (2012) Developing a hybrid data mining approach based on multi-objective particle swarm optimization for solving a traveling salesman problem. J Bus Econ Manag 13(5):951–967

Hammami M, Bechikh S, Hung CC, Said LB (2018) A multi-objective hybrid filter-wrapper evolutionary approach for feature selection. Memet Comput 11(2):193–208

Hancer E (2018) Differential evolution for feature selection: a fuzzy wrapper–filter approach. Soft Comput. https://doi.org/10.1007/s00500-018-3545-7

Hancer E, Xue B, Karaboga D, Zhang M (2015) A binary ABC algorithm based on advanced similarity scheme for feature selection. Appl Soft Comput 36:334–348

He X, Cai D, Niyogi P (2006) Laplacian score for feature selection. In: Advances in neural information processing systems, pp 507–514

Hoque N, Ahmed HA, Bhattacharyya DK, Kalita JK (2016) A fuzzy mutual information-based feature selection method for classification. Fuzzy Inf Eng 8(3):355–384

Hsieh Y, Lee P, You P (2018) Immune based evolutionary algorithm for determining the optimal sequence of multiple disinfection operations. Sci Iran. https://doi.org/10.24200/sci.2018.20324

Hsu HH, Hsieh CW, Lu MD (2011) Hybrid feature selection by combining filters and wrappers. Expert Syst Appl 38(7):8144–8150

Huang CL, Wang CJ (2006) A GA-based feature selection and parameters optimization for support vector machines. Expert Syst Appl 31(2):231–240

Huang H, Xie HB, Guo JY, Chen HJ (2011) Ant colony optimization–based feature selection for surface electromyography signals classification. Comput Biol Med 42(1):30–38

Ibrahim RA, Ewees AA, Oliva D, Elaziz MA, Lu S (2018) Improved salp swarm algorithm based on particle swarm optimization for feature selection. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-018-1031-9

Indira K, Kanmani S (2015) Mining association rules using hybrid genetic algorithm and particle swarm optimisation algorithm. Int J Data Anal Tech Strateg 7(1):59–76

Jayaraman V, Sultana HP (2019) Artificial gravitational cuckoo search algorithm along with particle bee optimized associative memory neural network for feature selection in heart disease classification. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-019-01193-6

Jiang Y, Liu X, Yan G, Xiao J (2017) Modified binary cuckoo search for feature selection: a hybrid filter–wrapper approach. In: 2017 13th International conference on computational intelligence and security (CIS). IEEE, pp 488–491

Jović A, Brkić K, Bogunović N (2015) A review of feature selection methods with applications. In: Information and communication technology, electronics and microelectronics (MIPRO) 2015 38th international convention. IEEE, pp 1200–1205

Kanan HR, Faez K (2008) An improved feature selection method based on ant colony optimization (ACO) evaluated on face recognition system. Appl Math Comput 205(2):716–725

Lefkovits S, Lefkovits L (2017) Gabor feature selection based on information gain. Proc Eng 181:892–898

Leung Y, Hung Y (2010) A multiple-filter-multiple-wrapper approach to gene selection and microarray data classification. IEEE/ACM Trans Comput Biol Bioinf 7(1):108–117

Lin JY, Ke HR, Chien BC, Yang WP (2008) Classifier design with feature selection and feature extraction using layered genetic programming. Expert Syst Appl 34(2):1384–1393

Liu H, Hussain F, Tan CL, Dash M (2002) Discretization: an enabling technique. Data Min Knowl Discov 6(4):393–423

Liu Y, Wang G, Chen H, Dong H, Zhu X, Wang S (2011) An improved particle swarm optimization for feature selection. J Bionic Eng 8(2):191–200
Liu S, Zhang DG, Liu XH, Zhang T, Gao JX, Cui YY (2019) Dynamic analysis for the average shortest path length of mobile ad hoc networks under random failure scenarios. IEEE Access 7:21343–21358
Mafarja MM, Mirjalili S (2017) Hybrid whale optimization algorithm with simulated annealing for feature selection. Neurocomputing 260:302–312
Moradi P, Gholampour M (2016) A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy. Appl Soft Comput 43:117–130
Moslehi F, Haeri A (2019) A genetic algorithm based framework for mining quantitative association rules without specifying minimum support and minimum confidence. Sci Iran. https://doi.org/10.24200/SCI.2019.51030.1969
Moslehi F, Haeri A, Gholamian M (2019a) A novel selective clustering framework for appropriate labeling of the clusters based on K-means algorithm. Sci Iran. https://doi.org/10.24200/SCI.2019.51110.2010
Moslehi F, Haeri A, Gholamian M (2019b) Investigation of effective factors in expanding electronic payment in Iran using datamining techniques. J Ind Syst Eng 12(2):61–94
Mukhopadhyay A, Maulik U (2013) An SVM-wrapped multiobjective evolutionary feature selection approach for identifying cancer-microRNA markers. IEEE Trans Nanobiosci 12(4):275–281
Muni DP, Pal NR, Das J (2006) Genetic programming for simultaneous feature selection and classifier design. IEEE Trans Syst Man Cybern Part B (Cybernetics) 36(1):106–117
Narendra PM, Fukunaga K (1977) A branch and bound algorithm for feature subset selection. IEEE Trans Comput 9:917–922
Omar N, bin Othman MS (2013) Particle swarm optimization feature selection for classification of survival analysis in cancer. Int J Innov Comput 2(1):1–7
Oreski S, Oreski G (2014) Genetic algorithm-based heuristic for feature selection in credit risk assessment. Expert Syst Appl 41(4):2052–2064
Osanaiye O, Cai H, Choo KKR, Dehghantanha A, Xu Z, Dlodlo M (2016) Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing. EURASIP J Wirel Commun Netw 2016(1):130
Peng H, Long F, Ding C (2005) Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 27(8):1226–1238
Purohit A, Chaudhari N S, Tiwari A (2010) Construction of classifier with feature selection based on genetic programming. In: Evolutionary computation (CEC), 2010 IEEE congress. IEEE, pp 1–5
Russell SJ, Norvig P (2008) Artificial intelligence: a modern approach
Sadeghi H, Zolfaghari M, Heydarizade M (2011) Estimation of electricity demand in residential sector using genetic algorithm approach. J Ind Eng Prod Res 22(1):43–50
Saeys Y, Inza I, Larranaga P (2007) A review of feature selection techniques in bioinformatics. Bioinformatics 23(19):2507–2517
Sahu B (2018) A combo feature selection method (filter +wrapper) for microarray gene classification. Int J Pure Appl Math 118(16):389–401
Salguero AG, Medina J, Delatorre P, Espinilla M (2018) Methodology for improving classification accuracy using ontologies: application in the recognition of activities of daily living. J Ambient Intell Humaniz Comput 10(6):2125–2142
Sampath kumar R, Alagumurthi N, Ramesh R (2010) Integrated total cost and tolerance optimization with genetic algorithm. Int J Comput Intell Syst 3(3):325–333
Seal A, Ganguly S, Bhattacharjee D, Nasipuri M, Gonzalo-Martin C (2015) Feature selection using particle swarm optimization for thermal face recognition. In: Applied computation and security systems. Springer, New Delhi, pp 25–35
Sheikhan M, Mohammadi N (2012) Neural-based electricity load forecasting using hybrid of GA and ACO for feature selection. Neural Comput Appl 21(8):1961–1970
Shreem SS, Abdullah S, Nazri MZA, Alzaqebah M (2012) Hybridizing reliefF, MRMR filters and GA wrapper approaches for gene selection. Theor Appl Inf Technol 46(2):1034–1039
Shunmugapriya P, Kanmani S (2017) A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid). Swarm Evolut Comput 36:27–36
Su S, Zhao S (2017) A hierarchical hybrid of genetic algorithm and particle swarm optimization for distributed clustering in large-scale wireless sensor networks. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-017-0619-9
Tang J, Alelyani S, Liu H (2014) Feature selection for classification: a review. In: Data classification: algorithms and applications, p 37
Tran B, Xue B, Zhang M (2014) Overview of particle swarm optimisation for feature selection in classification. In: Asia-Pacific conference on simulated evolution and learning. Springer, Cham, pp 605–617
Uçan F, Altilar DT (2012) Using genetic algorithms for navigation planning in dynamic environments. Appl Comput Intell Soft Comput. https://doi.org/10.1155/2012/560184
Uğuz H (2012) A hybrid system based on information gain and principal component analysis for the classification of transcranial Doppler signals. Comput Methods Programs Biomed 107(3):598–609
Wah YB, Ibrahim N, Hamid HA, Abdul-Rahman S, Fong S (2018) Feature selection methods: case of filter and wrapper approaches for maximising classification accuracy. Pertanika J Sci Technol 26(1):329–340
Wu Y, Liu B, Wu W, Lin Y, Yang C, Wang M (2018) Grading glioma by radiomics with feature selection based on mutual information. J Ambient Intell Humaniz Comput 9(5):1671–1682
Xue B, Cervante L, Shang L, Browne WN, Zhang M (2013a) Multi-objective evolutionary algorithms for filter based feature selection in classification. Int J Artif Intell Tools 22(04):1350024
Xue B, Zhang M, Browne WN (2013b) Particle swarm optimization for feature selection in classification: a multi-objective approach. IEEE Trans Cybern 43(6):1656–1671
Yang CS, Chuang LY, Ke CH, Yang CH (2008a) Boolean binary particle swarm optimization for feature selection. In: Evolutionary computation, 2008. CEC 2008. (IEEE world congress on computational intelligence). IEEE congress. IEEE, pp 2093–2098
Yang CS, Chuang LY, Li JC, Yang CH (2008b) Chaotic maps in binary particle swarm optimization for feature selection. In: Soft computing in industrial applications, 2008. SMCia'08. IEEE conference. IEEE, pp 107–112
Ye Y, Wu Q, Huang JZ, Ng MK, Li X (2013) Stratified sampling for feature subspace selection in random forests for high dimensional data. Pattern Recogn 46(3):769–787
Yongjun P, Minghao P, Kiejung P, Keun HY (2012) An ensemble correlation-based gene selection algorithm for cancer classification with gene expression data. Bioinformatics 28(24):3306–3315
Yu L, Liu H (2004) Efficient feature selection via analysis of relevance and redundancy. J Mach Learn Res 5:1205–1224
Yusta SC (2009) Different metaheuristic strategies to solve the feature selection problem. Pattern Recognit Lett 30(5):525–534
Zhang DG (2012) A new approach and system for attentive mobile learning based on seamless migration. Appl Intell 36(1):75–89
Zhang DG, Zhang XD (2012) Design and implementation of embedded un-interruptible power supply system (EUPSS) for web-based mobile application. Enterp Inf Syst 6(4):473–489

Zhang DG, Zhu YN, Zhao CP, Dai WB (2012a) A new constructing approach for a weighted topology of wireless sensor networks based on local-world theory for the internet of things (IOT). Comput Math Appl 64(5):1044–1055. https://doi.org/10.1016/j.camwa.2012.03.023

Zhang D, Zhao CP, Liang YP, Liu ZJ (2012b) A new medium access control protocol based on perceived data reliability and spatial correlation in wireless sensor network. Comput Electr Eng 38(3):694–702. https://doi.org/10.1016/j.compeleceng.2012.02.016

Zhang D, Li G, Zheng K, Ming X, Pan ZH (2013) An energy-balanced routing method based on forward-aware factor for wireless sensor networks. IEEE Trans Ind Inf 10(1):766–773

Zhang D, Wang X, Song X, Zhao D (2014) A novel approach to mapped correlation of ID for RFID anti-collision. IEEE Trans Serv Comput 7(4):741–748

Zhang DG, Wang X, Song XD (2015a) New medical image fusion approach with coding based on SCD in wireless sensor network. J Electr Eng Technol 10(6):2384–2392. https://doi.org/10.5370/JEET.2015.10.6.2384

Zhang DG, Zheng K, Zhang T, Wang X (2015b) A novel multicast routing method with minimum transmission for WSN of cloud computing service. Soft Comput 19(7):1817–1827

Zhang DG, Zheng K, Zhao DX, Song XD, Wang X (2016) Novel quick start (QS) method for optimization of TCP. Wirel Netw 22(1):211–222

Zhang DG, Liu S, Zhang T, Liang Z (2017a) Novel unequal clustering routing protocol considering energy balancing based on network partition & distance for mobile education. J Netw Comput Appl 88:1–9

Zhang DG, Niu HL, Liu S (2017b) Novel PEECR-based clustering routing approach. Soft Comput 21(24):7313–7323

Zhang DG, Zhou S, Tang YM (2018a) A low duty cycle efficient MAC protocol based on self-adaption and predictive strategy. Mob Netw Appl 23(4):828–839. https://doi.org/10.1007/s11036-017-0878-x

Zhang D, Ge H, Zhang T, Cui YY, Liu X, Mao G (2018b) New multi-hop clustering algorithm for vehicular ad hoc networks. IEEE Trans Intell Transport Syst 99:1–14

Zhang DG, Zhang T, Zhang J, Dong Y, Zhang XD (2018c) A kind of effective data aggregating method based on compressive sensing for wireless sensor network. EURASIP J Wirel Commun Netw 2018(1):159. https://doi.org/10.1007/s11036-018-1123-y

Zhang DG, Chen C, Cui YY, Zhang T (2018d) New method of energy efficient subcarrier allocation based on evolutionary game theory. Mob Netw Appl. https://doi.org/10.1007/s11036-018-1123-y

Zhang DG, Tang YM, Cui YY, Gao JX, Liu XH, Zhang T (2018e) Novel reliable routing method for engineering of internet of vehicles based on graph theory. Eng Comput 36(1):226–247

Zhang D, Zhang T, Liu X (2018f) Novel self-adaptive routing service algorithm for application in VANET. Appl Intell 49(5):1866–1879. https://doi.org/10.1007/s10489-018-1368-y

Zhang DG, Zhang T, Dong Y, Liu XH, Cui YY, Zhao DX (2018g) Novel optimized link state routing protocol based on quantum genetic strategy for mobile learning. J Netw Comput Appl 122:37–49

Zhang DG, Liu S, Liu XH, Zhang T, Cui YY (2018h) Novel dynamic source routing protocol (DSR) based on genetic algorithm-bacterial foraging optimization (GA-BFO). Int J Commun Syst 31(18):e3824

Zhang DG, Gao JX, Liu XH, Zhang T, Zhao DX (2019) Novel approach of distributed & adaptive trust metrics for MANET. Wirel Netw 25(16):3587–3603. https://doi.org/10.1007/s11276-019-01955-2

Zhao Z, Morstatter F, Sharma S, Alelyani S, Anand A, Liu H (2010) Advancing feature selection research. In: ASU feature selection repository 1–28

Zorarpacı E, Özel SA (2016) A hybrid approach of differential evolution and artificial bee colony for feature selection. Expert Syst Appl 62:91–103

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.