Machine Learning with Applications 5 (2021) 100054

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

Ensemble Based Filter Feature Selection with Harmonize Particle Swarm Optimization and Support Vector Machine for Optimal Cancer Classification

Tengku Mazlin Tengku Ab Hamid *, Roselina Sallehuddin, Zuriahati Mohd Yunos, Aida Ali

Faculty of Engineering, School of Computing, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia

A B S T R A C T

Explosive increase of dataset features may intensify the complexity of medical data analysis in deciding necessary treatment for the patient. In most cases, the accuracy of diagnosis system is vitally impacted by the data dimensionality and classifier parameters. Since these two processes are dependent, conducting them independently could deteriorate the accuracy performance. Filter algorithm is used to eliminate irrelevant features based on ranking. However, independent filter still incapable to consider features dependency and resulting in imbalance selection of significant features which consequently degrade the classification performance. In order to mitigate this problem, ensemble of multi filters algorithm such as Information Gain (IG), Gain Ratio (GR), Chi-squared (CS) and Relief-F (RF) are utilized as it can considers the intercorrelation between features. The proper kernel parameters settings may also influence the classification performance. Hence, a harmonize classification technique using Particle Swarm Optimization (PSO) and Support Vector Machine (SVM) is employed to optimize the searching of optimal significant features and kernel parameters synchronously without degrading the accuracy. Therefore, an ensemble filter feature selection with harmonize classification of PSO and SVM (Ensemble-PSO-SVM) are proposed in this research. The effectiveness of the proposed method is examined on standard Breast Cancer and Lymphography datasets. Experimental results showed that the proposed method successfully signify the classifier accuracy performance with optimal significant features compared to other existing methods such as PSO-SVM and classical SVM. Hence, the proposed method can be used as an alternative method for determining the optimal solution in handling high dimensional data.

1. Introduction

In machine learning, redundant data are unfavourable as it constitutes numerous problems. Generally, redundant data occurs when an information of the same entry being duplicated inconsistently on the same datasets and resulting wasteful data redundancies. Data redundancy might also exist in the datasets due to coding inefficiency or overcomplex data storing processes which may leads to many issues caused by accessing the wrong redundant datasets. Consequently, the explosive increase of redundant data might consume the storage of memory space over time and raise the dimensionality issues known as the "curse of dimensionality" in handling high dimensional data analysis.

Several typical issues in data analysis raised by redundant data are loss of accuracy due to explosive increase of irrelevant features and bias due to ambiguity of the feature's distribution. Some classifiers for decision making tools such as SVM and other classification techniques could not perform accurately when too many irrelevant features were included in the classification tasks. On the other hand, the issues of data redundancy in medical dataset raised when creating diagnosis from case files. Redundant data cause much greater affect as the disease diagnosis will influence the proper treatment a patient should receive especially in cancer prediction, as it is crucial to discover the cancer recurrence and the diagnosis involved in the treatment of a particular patient. Consequently, the analysis of cancer prediction could be misclassified due to the biased of redundant features during the classification process.

Accuracy is a major issue in handling high dimensional dataset with redundant features as it may influence the reliability of the data analysis results (Hamouda et al., 2017; Xue et al., 2021). The diagnosis accuracy of breast cancer, heart disease, diabetes, and other diseases are highly depending on the quality of input data used by medical experts. Most medical datasets consist of numerous features related to the patient's medical information. The escalating number of features has increased the computational complexity which tremendously degrade the accuracy of classification models. Besides, redundant features may contain irrelevant information and less beneficial to be used in disease prediction. Hence, feature selection is required for eliminating redundancy and irrelevant features effectively to improve the accuracy of classifier performance.

Feature selection refers to the process of selecting features from a dataset to describe information of a particular data (Miao & Niu, 2016; Xue et al., 2021). Feature selection can be categorized into filter, wrapper, and embedded techniques. In filter techniques, features are evaluated using simple ranking criteria based on dependency, distance measures, entropy value or feature score which ranked according to the intrinsic values (Raj & Mohanasundaram, 2020). Filter techniques are independent as it does not require any classification algorithm employment which particularly efficient in dealing with the

* Corresponding author.
E-mail addresses: tgmazlynn@gmail.com (T.M.T.A. Hamid), roselina@utm.my (R. Sallehuddin), zuriahati@utm.my (Z.M. Yunos), aida@utm.my (A. Ali).

https://doi.org/10.1016/j.mlwa.2021.100054
Received 19 December 2020; Received in revised form 20 May 2021; Accepted 30 May 2021
Available online 6 June 2021

high dimensional data. In contrast, wrapper and embedded techniques requires certain classification algorithm in evaluating the significance of features to produce more accurate prediction (Zhang et al., 2019). However, the involvement of classification algorithm in wrapper and embedded techniques tend to provide fewer effective results due to high consumption of storage space and computational time. Consequently, filter techniques are much preferred due to the flexibility and simplicity of ranking based evaluation.

In recent years, filter techniques such as Information Gain (IG), Gain Ratio (GR), Chi-squared (CS) and Relief-F (RF) were recommended as the most eminent and convenience filter algorithm for handling high dimensional data due to its simplest ranking strategies and (Bommert et al., 2020). IG, GR, CS, and RF produce better classification accuracy when the irrelevant features are eliminated from the dataset using a statistical ranking score evaluation and a set of threshold values. Features with the highest ranking score that exceed the threshold values are selected as significant whereas features that does not exceed the threshold values are not included into the classification tasks (Alirezanejad et al., 2019). However, the accuracy of classifier performance such as SVM still severely affected when the total number of selected features from each filter algorithm are too large and too small. The cause of this imbalance is because the independent filter algorithm only focuses on evaluating features individually instead of considering the interactions or dependencies between features which made them still unable to produce the optimum number of features relevant for classification and cause the classifier to perform poorly (Ali et al., 2019).

Consequently, ensemble feature selection using multi filters algorithm were proposed to handle the imbalance selection of features in datasets. Dongare et al. considered Correlation based Feature Selection (CFS) to derive the initial subset of features and then further analysed the best selected features using filter ranking such as IG, CS, RF and Symmetrical Uncertainty (SU). The ensemble feature selection shows an improvement of classification in terms of dimensionality reduction by reducing the irrelevant and redundant features while obtaining the topmost significant features using multi filters algorithm ranking approaches (Dongare et al., 2018). Singh et al. considered CFS and SU to identify the significant features by partitioning the ranked features and evaluate the symmetrical uncertainties and correlation of each feature in the ranked list simultaneously (Singh et al., 2014). The results produce a scalable and effective performance as it synchronized tuples of feature simultaneously which allow more evaluation of potential features in high dimensional dataset. This proved that assemble of multi filters algorithm highly improves classification accuracy performance such that the advantage of ranking based techniques allows redundant features can be eliminated, features with highest correlation can be determined, and the independently weak but strong in group features can be identified (Canedo et al., 2012). Therefore, in this study, an ensemble based multi filters feature selection will be utilized before the classification process.

One of the most reliable classification algorithms that widely adopted for various range of problems is SVM. As this study focuses on medical data, SVM has shown to provide satisfying performance in classifying various types of diseases especially in cancer data (Huang et al., 2018; Lee, 2019). An optimal SVM classifier performance can be distinguished dynamically even when handling lower dimensional data. Although SVM provide a reliable classification performance, it is quite sensitive towards the kernel parameter settings. For instance, Radial Basis Function (RBF) is the common kernel of SVM that consists of cost function parameter (C) and kernel function parameter (y). An improper selection of C and y parameters could consequently influence the selection of optimal features and negatively affect the accuracy performance. The sensitivity of kernel parameters settings can be reduced by employing a search technique that capable to optimize the optimum values of C and y. Since the values of classification parameters could also influence the selection of features, a separate optimization process among these two aspects could restrain the aim of achieving the optimal solution of datasets (Rani & Ramyachitra, 2018; Xue et al., 2019). Hence, a harmonize optimization process is essential to optimize the searching of C and y parameters while determining the optimal number of significant features synchronously without degrading the accuracy performance.

PSO is an outperformed search algorithm which known to produce an efficient optimization using less parameters and faster rate of convergence in the searching process (Huang & Dun, 2008; Xue et al., 2019). This has been proven in several studies that have applied PSO technique in various classifier including SVM. Improvement of PSO and SVM synchronization is due to the reason that PSO is easily employed for parallel processing where the searching of optimal significant features and SVM parameters values can be tuned synchronously to determine the set of optimal features with highest accuracy performance. The less parameters usage in PSO algorithm provide less sensitive impact towards SVM solutions compared to other heuristic algorithms (Xue et al., 2016). Besides, PSO is capable to generate robust solutions towards SVM classification with a lot faster training time and stable convergence rates than classical SVM. Based on the aforementioned advantages, PSO and SVM is employed synchronously for harmonize classification in this study.

In this study, we proposed an ensemble based multi filters feature selection by IG, GR, CS and RF with harmonize classification by PSO and SVM to improve the classification in high dimensional medical data. It is believed that the acceptable results presented by the proposed method can be used as a possible tool to assist the medical experts for proper disease diagnosis and better decision making. Other main contribution of this paper includes (1) an optimum top significant feature to be used in determining the disease prediction can be identified based on ensemble filters feature selection, (2) an enhanced SVM classifier that harmonized with PSO can significantly improve the classification performance by optimizing feature selection with SVM parameters synchronously, (3) the appropriate combination of C and y can be determined using CCD method for optimal solution.

The rest of this paper is organized in the following manner. Section 2 will describe the literature review and related works. Section 3 explains the implementation of the proposed method. Section 4 discusses the experimental data and presents the analysis results. Finally, the summary and conclusions are provided in Section 5.

2. Related Works

In this section, discussions on the literature of IG, GR, CS and RF and the implementation of ensemble IG, GR, CS and RF for ranking and assemble of relevant features based on feature's occurrence rate are described. Next, SVM which applied as classifier are explained. Finally, the harmonize classification of PSO and SVM which employed in this study is also discussed.

2.1. IG, GR, CS, and RF as Ensemble Filters

IG determine the relevance of features by calculating the information gain between the features and class labels to measure their level of dependence (Fahrudin et al., 2016). To obtain the ranking score, IG must evaluate the measure of entropy value in each attribute as its relevance score. The highest information gain is equivalent to the smallest entropy value in which a feature is considered relevant if it obtains high information gain. This means the decrease of entropy value indicates the information is gained based on the new added information. However, IG is biased towards features with large number of distinct values. This proved that IG may lead to overfitting issues due to its inability to handle redundant features since the feature are selected in univariate way.

GR is utilized to improve the biased of IG towards features with large distinct values (Dai & Xu, 2013). GR determine the relevance of features by adapting branch mechanism to evaluate the significance of

information using the measure of entropy value. The size and number of branches are considered in identifying the significant features. Based on the branch mechanism, an evenly distributed information belongs to multiple branches will produce higher gain ratio value whereas an uneven information belongs to a single branch will produce smaller gain ratio value.

CS is utilized to test the independence of data between two features by measuring how these two features deviates each other (Lee et al., 2011). The aims of CS are to identify features which are highly dependent. This means when the two features are independent, smaller CS value will be obtained whereas the higher CS value indicates that the features are highly dependent which can be selected for training tasks.

RF calculates the relevance of features using continuous testing to evaluate the difference of features weight in the similar class (nearest Hit) and different class (nearest Miss). The significant feature is selected based on its ability to separate instance from different classes. This means that feature with higher score is indicated by the higher feature's weight in the similar class (high nearest Hit) and lower score feature is indicates by the higher feature's weight in the different class (high nearest Miss). RF is highly capable in handling an incomplete and multi class data (Urbanowicz, Meeker et al., 2018).

IG, GR, CS and RF are univariate filters algorithm which independently calculates the rank of features without including any classifiers where its corresponding scores are determined by the specific ranking evaluation in each filter. Due to the computational efficiency and simple ranking interpretation, IG, GR, CS and RF are mostly recommended for feature selection. However, these independent filter algorithms have major disadvantage in which they do not considers the influence of the selected features subset on the performance of training algorithm and resulting an imbalance number of selected features. This may lead to the problem of finding the optimal feature subset.

Therefore, combining multi filters algorithm is highly suggested as it capable of handling redundant features and balance out the selection of features before classification tasks (Hancer et al., 2018; Pardo et al., 2019). This research utilizes multi filters algorithm feature selection to rank the features according to its relevancy and combine the ranked features output by considering the rate of features occurrence across each filter, thus increasing the classification accuracy by identifying the top significant features subset significant for classification.

2.2. Support Vector Machine (SVM)

SVM classification model is a supervised learning that can analyse and recognize the patterns of data. The performance of SVM is highly influenced by the values of kernel parameter and types of kernel function that are selected in the training task. The purpose of kernel function is to constructs a nonlinear hyperplane in an input space to perform the classification (Huang et al., 2018).

In this study, RBF kernel function which consists of regularization parameter (C) and kernel function parameter (y) is employed. C also known as cost penalty parameter identifies the trade-off cost between reducing the training error and complexity of the model, whereas y determines the mapping of nonlinear hyperplane from the input space into high dimensional feature space. The optimal parameter values of these parameters are determined using cross-validation method.

2.3. Particle Swarm Optimization (PSO)

PSO is one of the most common metaheuristics searching algorithm developed by Kennedy and Eberhart (Assarzadeh & Nilchi, 2015). PSO algorithm emulates the interactions of swarm behaviour in nature such as bird flocking, fish schooling and ant colony to share food information. PSO has been applied to various research areas in optimization and combination with other existing algorithms. PSO algorithm performs the search of the optimal solution by agents known as particles, which direction are adapted by two positive acceleration constants such as cognitive and social learning factor (C_1 and C_2) and two random parameters (r_1 and r_2) which set within (0, 1), and inertia weight (w) (Harb & Desuky, 2014). In general, the particles are scattered randomly to stimulate the search in all possible locations. An optimal position of each particle is influenced by its best individual position and the best group position but tends to move in random direction. Each particle is defined by two vectors named position and velocity, where each particle changes its position corresponding to the new velocity.

The primary advantage of PSO is that it requires smaller number of parameters to be tuned and constraints tolerance compared to other methods such as Genetic Algorithm (GA) and Ant Colony Optimization (ACO) (Assarzadeh & Nilchi, 2015; Xue et al., 2021). Previous studies have illustrated that PSO is much preferred for optimization due to its strong exploration ability in searching the optimal solution using simple mathematical operators (Prasad et al., 2018; Zeng et al., 2018). Due to the memory and knowledge of the solution reserved by all particles, the strong exploration ability is produced and provide the exchange of information behaviours between particles in solving optimization problems. In terms of memory space, the computational time of PSO does not easily influenced by the number of features, thus producing lower computational cost (Sakri et al., 2018).

Meanwhile, the computational time for GA in searching the optimal solution are highly influenced by the number of features in the searching process which consequently reduce the computation efficiency (Moslehi & Haeri, 2019). Even though GA is quite effective for rapid searching in less recognized spaces, the usage of complex operators such as crossover and mutation tends to produce poor optimal features (Dankolo et al., 2017). In addition, the absence of memory in GA may cause the knowledge of particle to be easily destructed by the changes in population. In comparison to GA, ACO also consumed higher searching time in the optimization process which may increase high tendency in arriving at local minima (Ghimatgar et al., 2018). Besides, as the optimization process of ACO can stop in single area, it is incapable to achieve the optimal solution. Due to these reasons, PSO is proven to be more reliable and qualified the requirements for optimum searching solution based on its capability to perform synchronous optimization process with any classifier particularly SVM using swarm intelligence.

2.4. Synchronous Optimization of PSO and SVM

Based on literature studies, hybrid implementation of PSO have been adapted to solve the problems of multi objective optimization in various studies. For examples, Xue et al. (2019) utilized PSO to tune the SVM parameters (C, y) for gene selection in large microarray data. Similarly, Zeng et al. (2018) employed PSO in determining C and y for improving the classification of high dimensional Alzheimer's disease data. The proposed PSO-SVM approach outperform better accuracy performance in disease prediction compared to GA and standard SVM. This is because a single PSO searching might be inefficient for large scale problems due to time complexity and risk of falling to local minima that might require multi objective optimization.

In terms of computational complexity, incorporating PSO may also consists of several limitations regarding the particle's velocities dimension that determine how large the search space is permitted for each particle can take (Raj et al., 2016). According to previous studies, they found that the particles may not explore further than local optimum areas appropriately when the velocity is too small, whereas if the velocity is too high, the particles could pass a higher optimal solution (Xue et al., 2019). However, since PSO can be implemented easily for parallel processing with any classifier particularly SVM, the searching of optimal SVM parameters and significant features can be conducted simultaneously without degrading the accuracy using swarm intelligence.

Generally, the harmonize classification of PSO and SVM works by initializing the position and velocity of the particles. Then, the parameters C and y were entered, and the fitness of each particle was

assessed using the fitness function. The optimum significant feature in the data is determined by evaluating the fitness of training accuracy of the particles. The best individual and global positions with the highest fitness values were updated. Next, the new position and velocity of each particle were updated. The searching process stops when the maximum number of iterations (100) is achieved or when the improvement of fitness function between two consecutive iterations is lower than the specified minimum amount of improvement even though as slightest as 0.0001.

Based on the related works, several research gaps were highlighted. Firstly, it is observed that most independent filter algorithm only focused on evaluating the intrinsic characteristics of features and ignoring the features interactions which indicates that an independent filter algorithm still lacks considering features dependencies (Ali et al., 2019). Consequently, imbalance number of irrelevant features that may contribute to poor classification accuracy were produced and resulting in difficulty to identify features that accurately significant for classification. Secondly, the tuning of parameters C and y using grid search method are computationally prohibitive due to the requirements of high parameters range which could led to impossibility in achieving the optimal accuracy when the optimization and classification processes are conducted separately (Wang & Chen, 2020). Therefore, the main advantage of proposed ensemble multi filters feature selection based on the benefit of IG, GR, CS and RF is to effectively eliminate irrelevant features prior to classification with the consideration of features occurrence. Then, the implementation of harmonize classification method based on the reliability of PSO and SVM using CCD search method were conducted synchronously to attain the optimal solution.

3. Proposed Method

The proposed method consists of Phase 1, the ensemble multi filters feature selection process and Phase 2, the harmonize classification process. In the first phase, four filters algorithm are utilized for ensemble feature selection through combination of IG, GR, CS, RF algorithms using occurrence rate evaluation and SVM classification. In the second phase, the harmonize classification of PSO and SVM were carried out on Phase 1 output for synchronous optimization. Fig. 1 shows the workflow of the proposed method.

3.1. Phase 1: Ensemble Multi Filters Feature Selection

In Phase 1, there are 3 stages involved to develop the ensemble multi filters feature selection process. Firstly, Stage 1 describes the utilization of multi filters algorithm ranking where IG, GR, CS and RF perform features ranking to identify the initial top ranked features from the dataset based on each entropy value, gain ratio value, feature score and feature weight, respectively. Secondly, Stage 2 describes the assemble of ranking outputs and occurrence rate evaluation to obtain a set of ensemble features with highest occurrence. Stage 3 describes the assemble selection process and SVM classification to evaluate the improvement of ensemble features towards the accuracy performance.

3.1.1. Stage 1: Multi (N) Filters Algorithm Ranking

In the first stage, the ranking scores of IG, GR, CS and RF are utilized on each feature to identify their significance level based on their scores in entropy value, gain ratio value, feature score and feature weight, respectively. The higher the ranking score, the higher the significance level. In contrast, lower ranking scores signify lower significance level. The most relevant feature is indicated at the top of the rank, while the least relevant feature is indicated at the bottom of the rank. Following the suggestions in Hamid, Sallehuddin and Yunos (2019), threshold value of 0.05 are used in each filter to select the ranked features. The ranking score that achieved higher than threshold value will be selected whereas, the ranking score that obtained lower than the threshold value will be eliminated from the dataset. The process involved in Stage 1 is shown in Fig. 2.

3.1.2. Stage 2: Ensemble Ranking Outputs and Occurrence Rate

In the second stage, the feature outputs that has been ranked by IG, GR, CS and RF are assembled and the occurrence rate of each ranked feature across the four filters algorithm are computed. The purpose of this process is to determine the optimum number of top ranked features that are significant to be included in classification tasks. Fig. 3 shows the process involved in Stage 2.

3.1.3. Stage 3: Assemble Selection and SVM Classification

In the third stage, the assemble selection process is performed by evaluating the rate of features occurrence across each filter algorithm. Fig. 4 shows the process involved in Stage 3. The highest occurring features signify its high significance as the features are frequently selected among all filters. In contrast, the less occurring features indicates its low significance. The maximum occurrence rate to assemble features is set according to the number of filters algorithm utilized (Hamid, Sallehuddin, Yunos and Ali, 2019). Since four filter algorithms are utilized in this study, the maximum occurrence rate is set to 4. The ensemble process continues until the maximum occurrence rate is obtained. Then, a set of optimum top ranked features is obtained and used as input for classification task.

In order to evaluate the performance of selected features towards classification accuracy, SVM is employed as the classifier. Via 10-fold cross validation, the datasets are divided into ten partitions where in each partition, nine subsets are used for training task and one subset is used for testing task to avoid overfitting of data that may influence the training performance. The output from SVM is used as input for harmonize classification by PSO and SVM which will be optimized synchronously in the next phase.

3.2. Phase 2: Harmonize PSO and SVM Classification

In Phase 2, the reduced dataset with ensemble features is used as input for harmonize classification by PSO and SVM to determine the optimal SVM parameters and optimal features synchronously. The process involved in harmonize classification phase is shown in Fig. 5. All required parameters to perform PSO and SVM were initialized across all experimental datasets based on the values presented in Table 1.

For PSO parameters, two main learning factors such as social learning factor (C_1) and cognitive learning factor (C_2) are set based on a published paper by Brezočnik (2017) and Rahman et al. (2009) which suggested that the best value for C_1 and C_2 is 2, such that (C_1, C_2) = (2, 2). The value of C_1 and C_2 can be range between 1 to any numbers as long as the total of C_1 and C_2 must not exceed more than 4 (Harb & Desuky, 2014). The maximum iteration and the fitness function of PSO are set to 100 and 0.95 respectively to avoid overtraining (Brezočnik, 2017). Meanwhile, the number of population size are set using different values from 10, 25, 40, 55, 70, 85 and 100, so that the robustness of the proposed method as the population size increases can be observed.

For SVM parameters, C and y are two main parameters in RBF kernel function that need to be properly selected. These parameters play a big role where C sets the trade-off cost between the reduction of training error and model complexity while y controls the mapping of nonlinear hyperplane in high dimensional feature space of SVM. Since there are no specific values for C and y, these parameters were initialized using Centre Composite Design (CCD) method as suggested by Srisukkham et al. (2017) that requires less parameter combinations compared to grid search method. Following the CCD method, nine parameters combination based on three-level full factorial design (3^k factorial) are employed to evaluate the optimum values of C and y in each population size. The nine parameters combination are (2, 2), (2, 4), (2, 8), (4, 2), (4, 4), (4, 8), (8, 2), (8, 4) and (8, 8). These parameters are evaluated through trial and error for optimal searching process. SVM via 10-fold cross validation is applied in order to validate the accuracy of harmonize classification method.

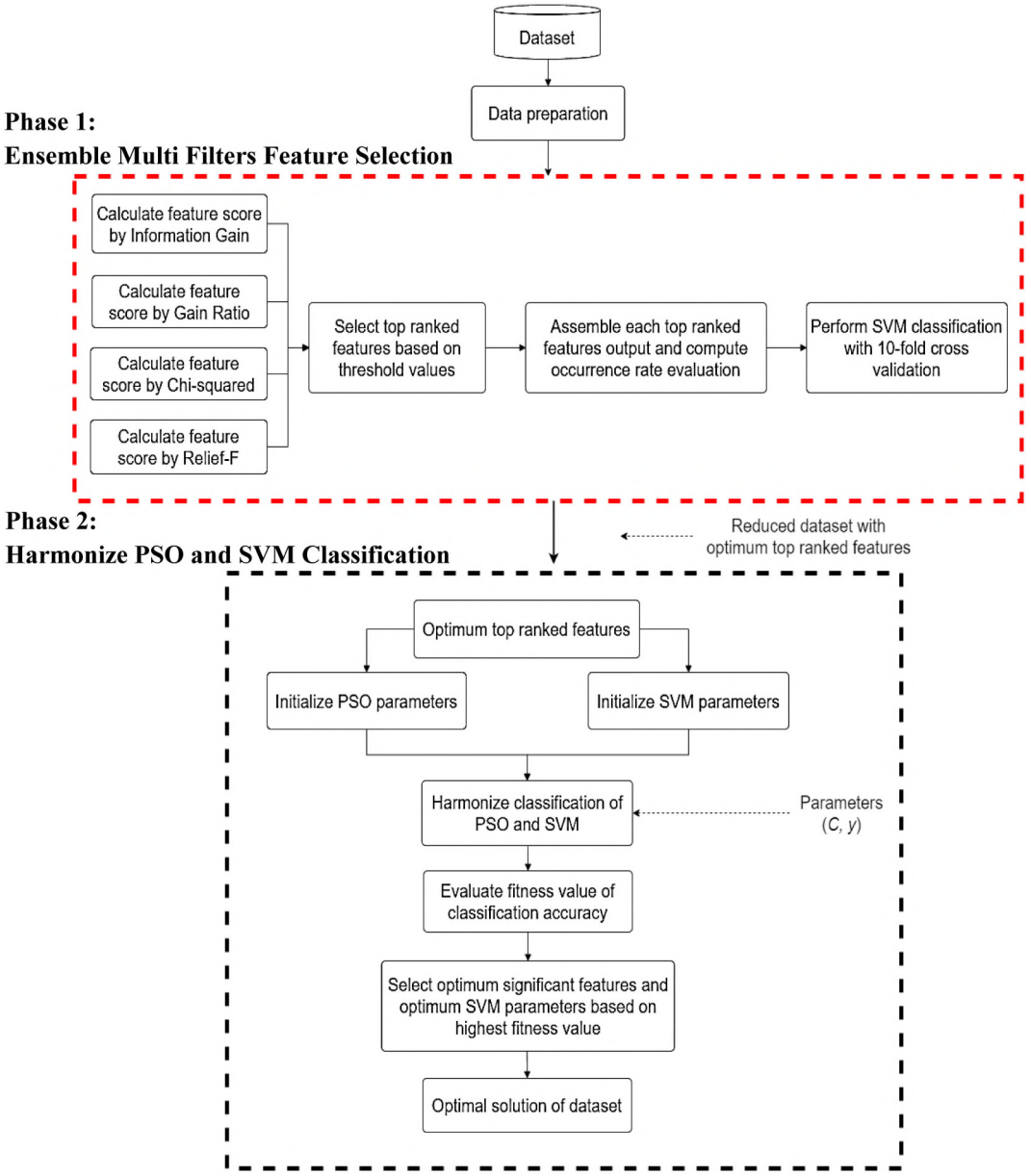

Fig. 1. Flowchart of proposed method (Ensemble-PSO-SVM).

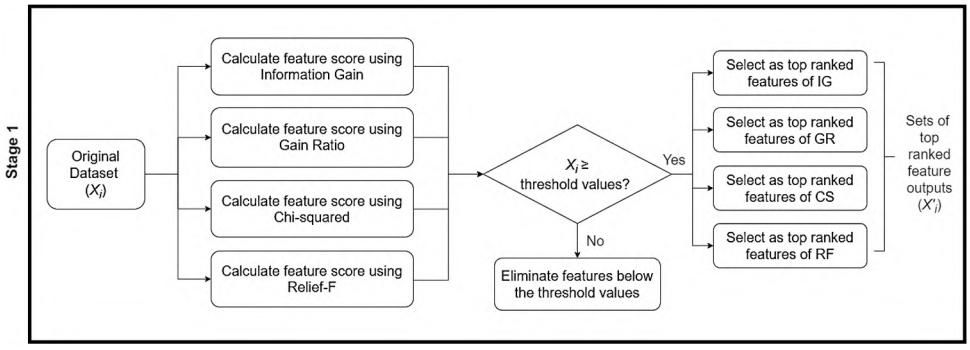

Fig. 2. Utilization of multi (N) filters algorithm ranking in Stage 1.

During harmonize searching process, the CCD method utilizes all nine parameter combinations of C and y to evaluate each particle. Parameter combinations that achieve the best global position for specific particle is used to generate the fitness value of that particle. The best pair of parameter settings identified for the final optimal solution is used to train the SVM using the whole training set. In order to ensure

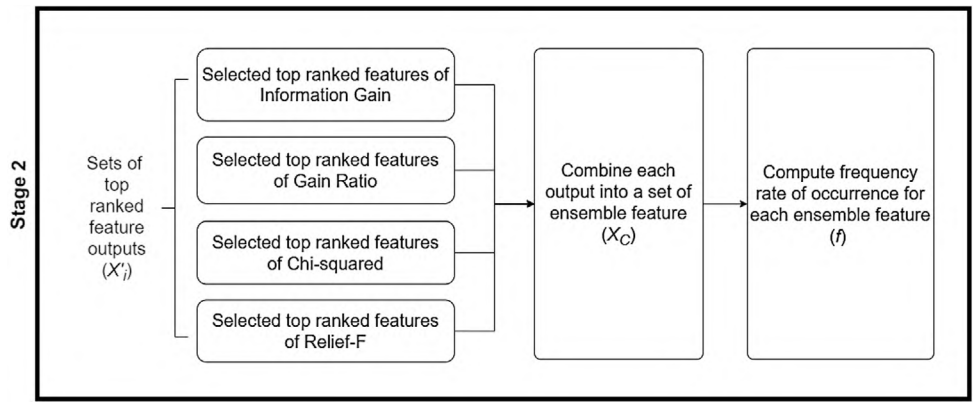

Fig. 3. Ensemble of filters ranking outputs and occurrence rate evaluation in Stage 2.

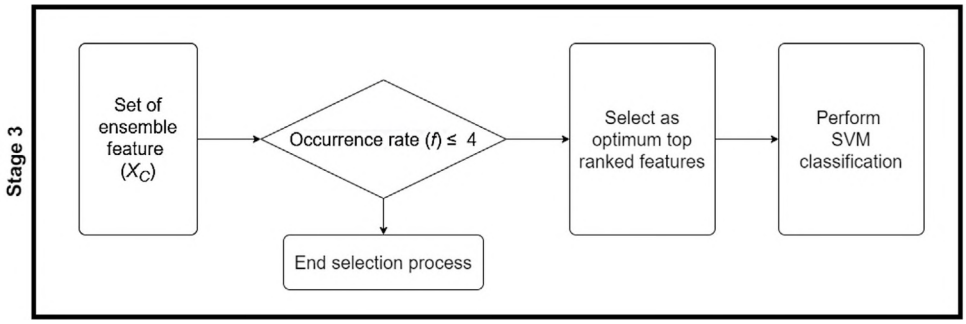

Fig. 4. Assemble selection and SVM classification in Stage 3.

Table 1
Settings of PSO and SVM parameters on all experimental datasets.

PSO Population Size (S)	10	25	40	55	70	85	100
SVM Parameters (C, y)				(2, 2) (2, 4) (2, 8) (4, 2) (4, 4) (4, 8) (8, 2) (8, 4) (8, 8)			
PSO Learning Factors (C_1, C_2)				(2, 2)			
PSO Maximum Iteration				100			

the optimum accuracy of harmonize classification has been obtained, measurement of performance validation is conducted. The result is then validated via 10-fold cross validation to calculate the average performance. The final output from harmonize classification of PSO and SVM will be a reduced dataset with optimal significant features and classification parameters with highest accuracy performance.

4. Results and Discussions

In this section, discussion on the result is divided into three sections. The first section will discuss the result obtained from filter ranking by IG, GR, CS and RF, while the second section will discuss on the ensemble of filters ranking outputs and the classification result of the selected features using SVM classifier. The third section will discuss on the harmonize classification phase. This section begins with description of the experimental dataset used and performance evaluation employed in this study.

4.1. Experimental Dataset and Performance Evaluation

The effectiveness of the proposed model is verified using two publicly available dataset such as Breast Cancer dataset and Lymphography dataset obtained from UCI Machine Learning Repository, which can be retrieved from https://archive.ics.uci.edu/ml/datasets.php. All experimental dataset features are shown in Table 2.

The first dataset, UCI Breast Cancer dataset contains 286 instances represented by nine attributes and two predictive classes which are class recurrence event and class no-recurrence event. Class recurrence event represents malignant cases of breast cancer while class no-recurrence event represents benign cases of breast cancer. Out of 286 instances, 85 instances are in class recurrence event and 201 instances are in class no-recurrence event. The second dataset, UCI Lymphography dataset contains 148 instances represented by eighteen attributes and four predictive classes such as normal, metastases, malignant and fibrosis. Out of 148 instances, 2 instances are in class normal, 81 instances are in class metastases, 61 instances are in class malignant and 4 instances are in class fibrosis.

In order to validate these datasets, the performance evaluations were separated into two categories which are the ensemble filters feature selection phase and harmonize classification phase. The performance measures that are used in evaluating the performance of ensemble filters feature selection phase are accuracy, sensitivity, and specificity. On the other hand, the fitness values evaluation is used to measure the performance of harmonize classification phase. To further validate the significance of the proposed method, SVM classification via 10-fold cross validation and comparisons study with previous research are conducted. Finally, the best result obtained from each dataset are highlighted.

4.2. Results on UCI Breast Cancer Dataset

In this section, all experimental results of proposed method obtained by UCI Breast Cancer dataset is presented and discussed.

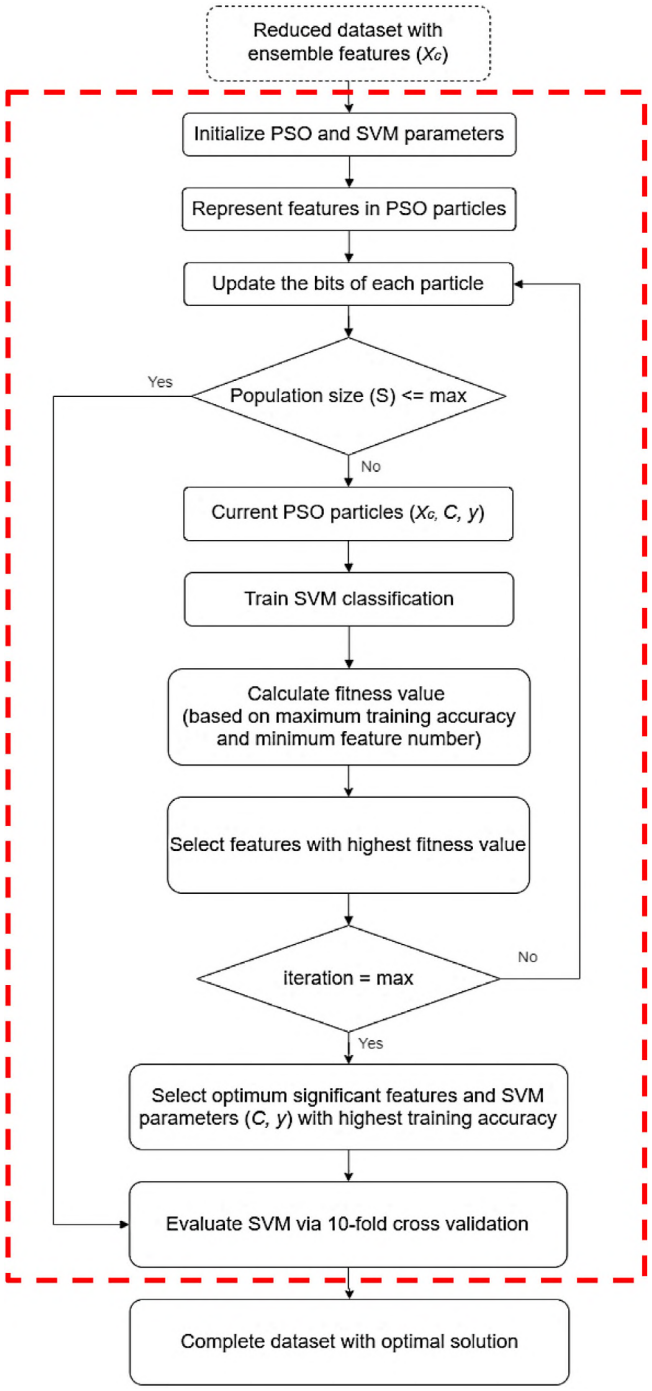

Fig. 5. Harmonize classification process by PSO and SVM.

4.2.1. Filters Ranking by IG, GR, CS and RF

The ranking score based on the entropy value, gain ratio value, feature score and feature weight given for the experimental dataset is utilized independently in order to demonstrate the differences in features ranking order and number of features when ranking score are considered. The higher the value of ranking scores, the higher the significance of the feature compared to the other features. The ranked features are sorted in descending order where the top ranked feature indicate the most significant feature, while the last ranked feature indicated the least significant feature. The bold features indicate significant feature that achieved the threshold value, whereas the red highlighted features indicate irrelevant feature that achieved below the threshold value which can be eliminated from the dataset.

Table 2
Attributes of features in experimental datasets.

Datasets	Features Name	
UCI Breast Cancer	Age	A1
	Menopause	A2
	Tumour Size	A3
	INV-nodes	A4
	Node Capsule	A5
	Degree of Malignancy	A6
	Breast	A7
	Breast Quadrant	A8
	Irradiation	A9
UCI Lymphography	Lymphatics	C1
	Block of Afferent	C2
	Block of Lymph Capsule	C3
	Block of Lymph Subcapsular	C4
	Bypass	C5
	Extravasates	C6
	Regeneration	C7
	Early Uptake	C8
	Diminish of Lymph Nodes	C9
	Enlargement of Lymph Nodes	C10
	Changes in Lymph Nodes	C11
	Defect in Nodes	C12
	Changes in Nodes	C13
	Changes in Structure	C14
	Special Forms	C15
	Dislocation of Nodes	C16
	Exclusion of Nodes	C17
	Number of Nodes	C18

Table 3
Features ranking and ranking scores for UCI Breast Cancer dataset.

Filter Algorithms	IG ($t = 0.05$)		GR ($t = 0.05$)		CS ($t = 0.05$)		RF ($t = 0.05$)	
Ranking Score	**A6**	0.078	**A5**	0.071	**A6**	28.875	**A6**	0.093
	A4	0.071	**A4**	0.054	**A4**	26.594	**A8**	0.062
	A3	0.061	**A6**	0.051	**A5**	19.731	**A2**	0.057
	A5	0.051	A9	0.033	**A3**	17.039	**A1**	0.051
	A9	0.026	A3	0.02	**A9**	9.792	A7[a]	0.048
	A1	0.012	A1	0.006	**A1**	3.956	A3	0.05
	A8	0.01	A8	0.005	A8	3.462	A9	0.033
	A7[a]	0.003	A7[a]	0.002	A7[a]	0.887	A5	0.026
	A2	0.003	A2	0.003	A2	0.94	A4	0.018
Total Selected Features	4		3		6		4	

[a]Eliminated features.

Each feature is ranked by IG, GR, CS, and RF individually and the results of filters ranking are shown in Table 3. Based on the table, several similar features are selected by IG, GR, CS and RF despite the different ranking strategy in each filter. For UCI Breast Cancer dataset, IG, CS and RF identifies A6 as the top dominant ranked features, while the top dominant ranked features identified by GR is A5. In contrast, A7 can be ignored and eliminated since it does not achieve the required threshold value in each filter. In terms of total selected features, both IG and RF selected 4 features, GR selected 3 features and CS selected 6 features out of nine dataset features.

4.2.2. Ensemble Ranking Outputs and Assemble Selection

Next, the ensemble of filters ranking outputs based on occurrence rate evaluation is performed. The occurrence rate of each selected features in IG, GR, CS and RF are computed to demonstrate the difference in number of features being selected across each filter. The intercept point is evaluated until the maximum occurrence rate is achieved. Since four filters are utilized in this study, the maximum occurrence rate of selected features is counted from 1 until 4 rates of occurrence. The highest number of tick mark indicates the frequent occurring features while the less occurring features have fewer number of tick mark. For assemble selection process, the measure of accuracy is used. The

Table 4
Occurrence rate of selected features for UCI Breast Cancer dataset.

Ranks	Selected Features				Top Dominant Features	Rate of Occurrence			
	IG	GR	CS	RF		1	2	3	4
1	**A6**	**A5**	**A6**	**A6**	A6	✓	✓	✓	✓
2	**A4**	**A4**	**A4**	**A8**	A5	✓	✓	✓	–
3	**A3**	**A6**	**A5**	**A2**	A4	✓	✓	✓	–
4	**A5**	A9	**A3**	**A1**	A8	✓	–	–	–
5	A9	A3	**A9**	A7	A3	✓	✓	–	–
6	A1	A1	**A1**	A3	A2	✓	–	–	–
7	A8	A8	A8	A9	A1	✓	✓	–	–
8	A7	A7	A7	A5	A9	✓	–	–	–
9	A2	A2	A2	A4	A7	–	–	–	–

Table 5
Optimum ensemble features for UCI Breast Cancer dataset.

Rate of Occurrence	1	2	3	4
Optimum Ensemble Features	1. A6 2. A5 3. A4 4. A8 5. A3 6. A2 7. A1 8. A9	**1. A6** **2. A5** **3. A4** **4. A3** **5. A1**	1. A6 2. A5 3. A4	1. A6
Total Selected Features	8	**5**	3	1
SVM Accuracy (%)	70.63	**72.91**	69.58	69.93

higher the accuracy of occurrence rate, the higher the significance of the ensemble features.

The computed occurrence rate for each of the ranked features in each IG, GR, CS and RF for UCI Breast Cancer dataset are shown in Table 4. From the table, A6 is identified as the frequent occurring features across each filter and achieved the intercept level. This means A6 is the most frequently selected by IG, GR, CS and RF which then followed by A5, A4, A3 and A1. This indicates A6, A5, A4, A3 and A1 are the optimum selected features compared to A8, A2 and A9 that only selected once among each filter. Based on the result of assemble selection in Table 5, it is observed that 5 features successfully signify the highest classification accuracy by 72.91% through second rate of occurrence which are A6, A5, A4, A3, and A1, compared to other features in different rate of occurrence with lower accuracy. Even though A6 is the highest occurring features across each filter, the classification accuracy is still unable to increase using this single information. This indicates that A6, A5, A4, A3, and A1 are optimum ensemble features relevant for the next harmonize classification as they have the ability to improve the accuracy of classifier with variety of significant information.

4.2.3. Classification Performance of Ensemble Features

In this section, the classification performance of independent IG-SVM, GR-SVM, CS-SVM, RF-SVM, Ensemble-SVM and SVM is presented. The classification performance of the original dataset by SVM is also highlighted to benchmark the SVM performance with the existence of irrelevant features. Classification accuracy is used to measure the performance of the selected features from each method.

The accuracy performance of independent IG-SVM, GR-SVM, CS-SVM, RF-SVM, Ensemble-SVM and SVM for UCI Breast Cancer dataset for are shown in Table 6. From the results, the classification accuracy increases by reducing the data dimensionality using Ensemble-SVM in comparison with the classification using independent IG-SVM, GR-SVM, CS-SVM, RF-SVM and SVM. Ensemble-SVM achieved the highest classification accuracy by 72.91% using five optimum top ranked features compared to using full dataset features. It is a noticeable increase in the classification accuracy by the Ensemble-SVM when the rates of features occurrence are considered compared to independent IG-SVM, GR-SVM, CS-SVM, RF-SVM and SVM. This increment is due to ability

Table 6
Accuracy improvement in Ensemble-SVM, IG-SVM, GR-SVM, CS-SVM, RF-SVM and SVM for UCI Breast Cancer dataset.

Method	Accuracy (%)	Selected Features
Breast Cancer $_{SVM}$	56.69	All
Breast Cancer $_{IG\text{-}SVM}$	66.20	A6>A4>A3>A5
Breast Cancer $_{GR\text{-}SVM}$	69.10	A5>A4>A6
Breast Cancer $_{CS\text{-}SVM}$	67.60	A6>A4>A5>A3>A9>A1
Breast Cancer $_{RF\text{-}SVM}$	68.90	A6>A8>A2>A1
Breast Cancer $_{ENSEMBLE\text{-}SVM}$	**72.91**	**A6>A5>A4>A3>A1**

Table 7
Fitness values based on different (C, y) for UCI Breast Cancer dataset.

Population size	y	C			Best pair (C, y)
		2	4	8	
10–25	2	0.7413	0.7413	**0.9541**	(8, 2)
	4	0.7832	0.7657	0.9511	
	8	0.8636	0.8636	0.9231	
40–70	2	0.7587	0.8322	**0.9601**	(8, 2)
	4	0.7832	0.8532	0.9511	
	8	0.8147	0.8951	0.8986	
85–100	2	0.7692	0.8077	**0.9615**	(8, 2)
	4	0.7867	0.8217	0.9511	
	8	0.7902	0.8427	0.8951	

of Ensemble-SVM to eliminate irrelevant features that are plagued with redundancy which may reduce the classifier performance.

In contrast, independent IG-SVM, GR-SVM, CS-SVM, RF-SVM and SVM obtain lower classification accuracy. This is probably due to the difference in ranking of A2, A8 and A9 obtained by the occurrence rate evaluation where Ensemble-SVM identified A2, A8 and A9 as the least significant features, while other independent filter such as RF-SVM identifies these features as significant. This shows that Ensemble-SVM have the ability to improve the accuracy using appropriate number of significant features which relevant to be included for training task.

4.2.4. Optimum Parameters (C, y) for UCI Breast Cancer Dataset

In this section, the effects of different C and y parameters towards the fitness performance of harmonize PSO and SVM classification for UCI Breast Cancer dataset are determined. As shown in Table 7, the highest fitness value of ensemble features is obtained by parameters $C = 8$ and $y = 2$ at every population size in comparison with other parameter combinations. As the size of population increases, the higher training accuracy of fitness value is achieved. For instance, the fitness value increase and decrease inconsistently from population 10 until population 40 when $C = 8$ and $y = 2$ (population 10 = 0.9541, population 25 = 0.8986, population 40 = 0.9601). However, the fitness value has become constant starting from population 55 until population 100 by achieving the optimum training accuracy of 0.9615. This indicates that the optimum parameters (C, y) which successfully signify the classification performance of UCI Breast Cancer dataset is (8, 2). For this case, lower kernel function y has the ability to improve the inconsistency of training accuracy by increasing the significance of classifying each training samples correctly using larger value of C.

4.2.5. Optimum Features for UCI Breast Cancer Dataset

In this section, the effects of different C and y parameters towards number of selected features from harmonize PSO and SVM classification for UCI Breast Cancer Dataset are determined. As shown in Table 8, a consistent accuracy performance is produced despite different population size of harmonize classification. The increase in population size may cause the selection of optimum features to vary. In searching for optimal solution, combination of SVM parameters and different population size influence the selection of optimum features by fastening or delaying the searching process. For example, the optimal solution of UCI Breast Cancer dataset identified four optimum features

Table 8
Optimum features based on different (C, y) for UCI Breast Cancer dataset.

Population Size	(C, y)	Fitness Value	Optimum Features
10–25	(2, 2)	0.7413	A4, A6
	(2, 4)	0.7832	A4, A3, A6
	(2, 8)	0.8636	A4, A3, A5
	(4, 2)	0.7413	A4, A6
	(4, 4)	0.7657	A3, A4, A1
	(4, 8)	0.8636	A4, A3, A5, A6, A1
	(8, 2)	**0.9541**	**A1, A3, A4, A6, A5**
	(8, 4)	0.9511	A1, A3, A6, A5, A4
	(8, 8)	0.9231	A1, A3, A5, A6, A4
40–70	(2, 2)	0.7587	A1, A3, A6
	(2, 4)	0.7832	A1, A3, A5, A6
	(2, 8)	0.8147	A1, A3, A4, A5, A6
	(4, 2)	0.8322	A1, A3, A4, A5
	(4, 4)	0.8532	A1, A4, A5, A6, A3
	(4, 8)	0.8951	A1, A3, A4, A5, A6
	(8, 2)	**0.9601**	**A1, A3, A4, A6, A5**
	(8, 4)	0.9511	A1, A3, A6, A4, A5
	(8, 8)	0.8986	A1, A3, A4, A6, A5
85–100	(2, 2)	0.7692	A1, A5, A6
	(2, 4)	0.7867	A1, A4, A6
	(2, 8)	0.7902	A1, A3, A6, A5
	(4, 2)	0.8077	A1, A3, A5
	(4, 4)	0.8217	A1, A4, A5, A6
	(4, 8)	0.8427	A1, A4, A5, A6, A3
	(8, 2)	**0.9615**	**A1, A3, A4, A6, A5**
	(8, 4)	0.9511	A1, A3, A6, A4, A5
	(8, 8)	0.8951	A1, A3, A4, A5, A6

with 0.7832 fitness value at 100 population using $C = 2$ and $y = 2$, while the optimal solution at 100 population using $C = 8$ and $y = 2$ identified five optimum features with 0.9615 higher fitness value. This proved that the searching of optimal features is highly influenced by proper parameters settings of C and y. As the optimal solution for UCI Breast Cancer dataset, (8, 2) is the appropriate SVM parameters (C, y) to trade off a correct classification of the training samples without affecting the classification accuracy and A1, A3, A4, A6, and A5 are the most optimum significant features of the dataset.

4.3. Results on UCI Lymphography Dataset

In this section, all experimental results of proposed method obtained by UCI Lymphography dataset is presented and discussed.

4.3.1. Filters Ranking by IG, GR, CS and RF

The same process of features ranking is performed on UCI Lymphography dataset based on entropy value, gain ratio value, CS statistical value and feature score and the results are presented in Table 9. From the table, C13 ("*Changes in Nodes*") is identified as the most significant feature by IG and RF, while GR and CS identifies C13 as the third significant feature. IG, GR and CS identifies C6 ("*Extravasates*") as the least significant feature, while RF identifies C6 as the third least significant feature. In simple terms, only IG and RF identifies "*Changes in Nodes*" as top dominant ranked features and "*Extravasates*" as least dominant ranked features. Thus, C6 can be ignored and eliminated from the dataset. In terms of total selected features, IG selected 15 features, GR selected 17 features, CS selected 16 features and RF selected 13 features out of nineteen dataset features.

4.3.2. Ensemble Ranking Outputs and Assemble Selection

Next, the ensemble of filters ranking outputs based on occurrence rate evaluation for UCI Lymphography dataset is performed and the results is discussed. From Table 10, 13 features were identified as the highest occurring features across IG, GR, CS and RF which are C13, C9, C18, C7, C2, C10, C15, C12, C8, C11, C17, C5 and C16. In contrast, 4 features that are less selected by each filter are C1, C14, C4 and C3 which can be considered as less significance. Based on

Table 9
Features ranking and ranking scores for UCI Lymphography dataset.

Filter Algorithms	IG ($t = 0.05$)		GR ($t = 0.05$)		CS ($t = 0.05$)		RF ($t = 0.05$)	
Ranking Score	**C13**	0.406	**C9**	0.578	**C1**	143.019	**C13**	0.286
	C18	0.328	**C7**	0.383	**C9**	111.175	**C2**	0.241
	C10	0.212	**C13**	0.249	**C13**	103.866	**C15**	0.182
	C14	0.188	**C2**	0.177	**C12**	98.492	**C10**	0.139
	C15	0.186	**C8**	0.156	**C18**	66.609	**C18**	0.135
	C2	0.175	**C4**	0.153	**C11**	64.56	**C7**	0.086
	C9	0.161	**C18**	0.136	**C14**	64.336	**C8**	0.088
	C1	0.158	**C15**	0.128	**C7**	53.466	**C5**	0.076
	C11	0.149	**C11**	0.125	**C10**	49.094	**C17**	0.075
	C12	0.151	**C10**	0.123	**C15**	34.183	**C11**	0.067
	C8	0.137	**C1**	0.098	**C2**	30.616	**C9**	0.063
	C7	0.137	**C5**	0.092	**C8**	22.803	**C12**	0.055
	C5	0.074	**C17**	0.091	**C4**	18.365	**C16**	0.057
	C17	0.067	**C12**	0.089	**C5**	14.085	C14	0.039
	C16	0.066	**C14**	0.074	**C17**	12.842	C3	0.027
	C4	0.042	**C16**	0.071	**C16**	11.327	C6[a]	0.026
	C3	0.035	**C3**	0.053	C3	6.694	C1	0.016
	C6[a]	0.031	C6[a]	0.031	C6[a]	4.267	C4	0.009
Total Selected Features	15		17		16		13	

[a]Eliminated features.

Table 10
Occurrence rate of selected features for UCI Lymphography dataset.

Ranks	Selected Features				Top Dominant Features	Rate of Occurrence			
	IG	GR	CS	RF		1	2	3	4
1	**C13**	**C9**	**C1**	**C13**	C13	✓	✓	✓	✓
2	**C18**	**C7**	**C9**	**C2**	C9	✓	✓	✓	✓
3	**C10**	**C13**	**C13**	**C15**	C1	✓	✓	✓	–
4	**C14**	**C2**	**C12**	**C10**	C18	✓	✓	✓	✓
5	**C15**	**C8**	**C18**	**C18**	C7	✓	✓	✓	✓
6	**C2**	**C4**	**C11**	**C7**	C2	✓	✓	✓	✓
7	**C9**	**C18**	**C14**	**C8**	C10	✓	✓	✓	✓
8	**C1**	**C15**	**C7**	**C5**	C15	✓	✓	✓	✓
9	**C11**	**C11**	**C10**	**C17**	C14	✓	✓	✓	–
10	**C12**	**C10**	**C15**	**C11**	C12	✓	✓	✓	✓
11	**C8**	**C1**	**C2**	**C9**	C8	✓	✓	✓	✓
12	**C7**	**C5**	**C8**	**C12**	C11	✓	✓	✓	✓
13	**C5**	**C17**	**C4**	**C16**	C4	✓	✓	–	–
14	**C17**	**C12**	**C5**	C14	C17	✓	✓	✓	✓
15	**C16**	**C14**	**C17**	C3	C5	✓	✓	✓	✓
16	C4	**C16**	**C16**	C6	C16	✓	✓	✓	✓
17	C3	**C3**	C3	C1	C3	✓	–	–	–
18	C6	C6	C6	C4	C6	–	–	–	–

assemble selection in Table 11, C13, C9, C18, C7, C2, C10, C15, C12, C8, C11, C17, C5 and C16 obtained the highest classification accuracy by 85.31%. Even though the difference between the number of features is small in each rate occurrence, the highest accuracy is achieved by the smallest number of features compared to the larger number of features. This shows that C13, C9, C18, C7, C2, C10, C15, C12, C8, C11, C17, C5 and C16 are the optimum ensemble features for the next harmonize classification phase.

.

4.3.3. Classification Performance of Ensemble Features

In this section, the accuracy performance between independent IG-SVM, GR-SVM, CS-SVM, RF-SVM, Ensemble-SVM and SVM for UCI Lymphography dataset is presented. As shown in Table 12, the dataset achieved the highest classification accuracy at 10 cross-validations with 85.31%. The result shows that the classification accuracy of IG-SVM, GR-SVM, CS-SVM and RF-SVM are higher than SVM. However, there is an increased in the accuracy when ensemble of IG, GR, CS and RF is processed. The improvement in accuracy performance is because Ensemble-SVM identifies an optimum number of top ranked features relevant for classification. Out of 19 features, IG-SVM selects 15 features, GR-SVM selects 17 features, CS-SVM selects 16 features and RF-SVM selects 13 features as significant features. Similar with

Table 11
Optimum ensemble features for UCI Lymphography dataset.

Rate of Occurrence	1	2	3	4
Optimum Ensemble Features	1. C13 2. C9 3. C1 4. C18 5. C7 6. C2 7. C10 8. C15 9. C14 10. C12 11. C8 12. C11 13. C4 14. C17 15. C5 16. C16 17. C3	1. C13 2. C9 3. C1 4. C18 5. C7 6. C2 7. C10 8. C15 9. C14 10. C12 11. C8 12. C11 13. C4 14. C17 15. C5 16. C16	1. C13 2. C9 3. C1 4. C18 5. C7 6. C2 7. C10 8. C15 9. C14 10. C12 11. C8 12. C11 13. C17 14. C5 15. C16	**1. C13** **2. C9** **3. C18** **4. C7** **5. C2** **6. C10** **7. C15** **8. C12** **9. C8** **10. C11** **11. C17** **12. C5** **13. C16**
Total Selected Features	17	16	15	**13**
SVM Accuracy (%)	81.76	81.96	83.42	**85.31**

Table 12
Accuracy improvement between Ensemble-SVM, IG-SVM, GR-SVM, CS-SVM, RF-SVM and SVM in UCI Lymphography dataset.

Method	Accuracy (%)	Selected Features
Lymphoma $_{SVM}$	73.80	All
Lymphoma $_{IG\text{-}SVM}$	82.43	C13>C18>C10>C14>C15>C2>C9> C1>C11>C12>C8>C7>C5>C17>C16
Lymphoma $_{GR\text{-}SVM}$	81.76	C9>C7>C13>C2>C8>C4>C18>C15>C11 >1>C5>C17>C12>C14>C16>C3
Lymphoma $_{CS\text{-}SVM}$	82.43	C1>C9>C13>C12>C18>C11>C14>C7> C10>C15>C2>C8>C4>C5>C17>C16
Lymphoma $_{RF\text{-}SVM}$	84.45	C13>C2>C15>C10>C18>C7>C8> C5>C17>C11>C9>C12>C16
Lymphoma$_{ENSEMBLE\text{-}SVM}$	**85.31**	**C13>C9>C18>C7>C2>C10>C15> C12>C8>C11>C17>C5>C16**

Table 13
Fitness values based on different (C, y) for UCI Lymphography dataset.

Population Size	y	C			Best pair (C, y)
		2	4	8	
10–25	2	0.6014	0.8041	0.9189	(8, 8)
	4	0.6554	0.8378	0.9324	
	8	0.7838	0.8784	**0.9662**	
40–70	2	0.7838	0.8378	0.9257	(8, 8)
	4	0.8041	0.8783	0.9527	
	8	0.8176	0.8987	**0.9662**	
85–100	2	0.7432	0.8176	0.9324	(8, 8)
	4	0.7635	0.8716	0.9594	
	8	0.8041	0.8987	**0.9662**	

RF-SVM, Ensemble-SVM also selects 13 features as the most optimum informative features. However, the ranking of features is different where RF-SVM unable to identify C9 as the top ranked features and ranked it as lowest significance among other features which caused the SVM accuracy to decrease. On the other hand, Ensemble-SVM identifies C9 among the top significant features thus contributes to improving the classification accuracy of the dataset. This shows that a proper ranking of features highly influenced the improvement of classification accuracy.

4.3.4. Optimum Parameters (C, y) for UCI Lymphography Dataset

In this section, the effects of different C and y parameters towards the fitness performance of harmonize PSO and SVM classification for UCI Lymphography dataset are determined. As shown in Table 13, the highest fitness value is achieved by 0.9662 using $C = 8$ and $y = 8$ compared to other parameter values. This fitness value is found constant and maintains using $C = 8$ and $y = 8$ across all population size even though other fitness value with the same C value ($C = 8$) shows an improvement. This indicates that the optimum parameters (C, y) which signify the classification performance with optimum training accuracy of UCI Lymphography is (8, 8). For this case, high fitness value is produced using equal value of C and y. This shows that implementing a larger margin for regularization of parameters is not essential as RBF kernel alone acts as a reliable regularization (Huang et al., 2018). However, an optimum value of C is necessary to simplify the decision function to determine the optimum number of significant features without degrading the training accuracy.

Table 14
Optimum features based on different (C, y) for UCI Lymphography dataset.

Population Size	(C, y)	Fitness Value	Optimum Features
10–25	(2, 2)	0.6014	C1, C9
	(2, 4)	0.6554	C4, C12, C17
	(2, 8)	0.7838	C10, C13
	(4, 2)	0.8041	C7, C8, C13
	(4, 4)	0.8378	C2, C4, C8, C11, C14
	(4, 8)	0.8784	C2, C11, C12, C17, C18
	(8, 2)	0.9189	C1, C2, C3, C4, C5, C7, C9, C10, C14
	(8, 4)	0.9324	C2, C3, C8, C11, C13, C15
	(8, 8)	**0.9662**	**C2, C8, C10, C12, C13, C14, C17, C18**
40–70	(2, 2)	0.7838	C10, C13
	(2, 4)	0.8041	C7, C8, C13
	(2, 8)	0.8176	C1, C3, C5, C8, C9, C15
	(4, 2)	0.8378	C2, C4, C8, C11, C14
	(4, 4)	0.8783	C2, C11, C12, C17, C18
	(4, 8)	0.9054	C1, C2, C13, C14, 16
	(8, 2)	0.9189	C1, C2, C3, C4, C5, C7, C9, C10, C14
	(8, 4)	0.9594	C1, C2, C4, C7, C10, C12, C14
	(8, 8)	**0.9662**	**C2, C8, C10, C12, C13, C14, C17, C18**
85–100	(2, 2)	0.7432	C8, C12, C14
	(2, 4)	0.7635	C3, C4, C7, C18
	(2, 8)	0.8041	C1, C2, C3, C4, C15
	(4, 2)	0.8176	C3, C7, C9, C10, C15
	(4, 4)	0.8716	C2, C3, C5, C7, C13, C16, C17
	(4, 8)	0.8987	C2, C5, C10, C16, C17, C18
	(8, 2)	0.9324	C2, C10, C12, C13, C18
	(8, 4)	0.9594	C2, C13, C14, C17, C18
	(8, 8)	**0.9662**	**C2, C8, C10, C12, C13, C14, C17, C18**

4.3.5. Optimum Features for UCI Lymphography Dataset

In this section, the effects of different pairs of C and y parameters towards the number of selected features from harmonize PSO and SVM classification for UCI Lymphography dataset are listed. As shown in Table 14, the optimal solution of UCI Lymphography dataset at 100 population using $C = 8$ and $y = 8$ identified eight optimum features with highest fitness value of 0.9662. In contrast, the optimal solution for 100 population using $C = 2$ or $C = 4$ produces smaller number of features with lower fitness value. As the optimal solution for UCI Lymphography dataset, (8, 8) is the proper SVM parameters (C, y) that trades off a correct classification of the training samples and C2, C8, C10, C12, C13, C14, C17, and C18 are the most optimum significant features of the dataset.

4.4. Performance Evaluation and Validation

The overall performance of proposed method on all datasets were evaluated using performance metrics such as accuracy, sensitivity, specificity, and AUC. The average harmonize classification performance were validated via 10-fold cross validation to obtain the comprehensive result. Comparisons study between the proposed method with independent IG-SVM, GR-SVM, CS-SVM, RF-SVM and Ensemble-SVM was conducted to observe the effectiveness of harmonize classification method towards the classification accuracy.

Table 15
Summary of selected features by Ensemble-PSO-SVM and Ensemble-SVM for all datasets.

Datasets	UCI Breast Cancer		UCI Lymphography	
Methods	Ensemble-SVM	Ensemble-PSO-SVM	Ensemble-SVM	Ensemble-PSO-SVM
Original features	9	9	18	18
Selected features	5	5	13	8
Features name	A6, A5, A4, A3, A1	A1, A3, A4, A6, A5	C13, C9, C18, C7, C2, C10, C15, C12, C8, C11, C17, C5, C16	C2, C8, C10, C12, C13, C14, C17, C18
(C, y)	(1, 0)	(8, 2)	(1, 0)	(8, 8)
Accuracy (%)	72.91	**96.15**	85.31	**96.62**

Table 16
10-fold cross validation results of proposed Ensemble-PSO-SVM method on UCI Breast Cancer dataset.

Fold	Benign	Malignant	Total Correctly Classified	Total Incorrectly Classified	Average Accuracy (%)
1	271	15	194	92	94.50
2	276	10	195	91	96.00
3	275	13	200	86	96.50
4	274	12	193	93	95.00
5	272	14	197	89	95.50
6	271	15	198	88	95.50
7	271	15	198	88	95.50
8	273	13	196	90	95.50
9	271	15	198	88	95.50
10	**208**	**78**	**275**	**11**	**96.15**

4.4.1. Overall Classification Performance of Proposed Method

In this section, the overall classification performance of proposed method, Ensemble-PSO-SVM in comparisons with Ensemble-SVM, independent IG-SVM, GR-SVM, CS-SVM and RF-SVM on all experimental datasets are presented.

As illustrated in Fig. 6, Ensemble-PSO-SVM produces higher accuracy performance in both experimental datasets compared to Ensemble-SVM and independent IG-SVM, GR-SVM, CS-SVM and RF-SVM. The implementation of harmonize classification proves the capability of swarm intelligence to determine the optimum number of features by evaluating the optimal position of features while synchronously considers appropriate values of SVM parameters that are able to influence the training accuracy of PSO. By using Ensemble-PSO-SVM, UCI Breast Cancer dataset achieves 96.15% accuracy using 5 optimum features compared to Ensemble-SVM which also selected 5 optimum features but with only 72.91% accuracy. This is due to the selection of proper kernel parameters that has caused the accuracy of Ensemble-PSO-SVM to increase significantly. UCI Lymphography dataset also achieves highest classification performance by Ensemble-PSO-SVM with 96.62% accuracy using 8 optimum features compared to independent IG-SVM, GR-SVM, CS-SVM, RF-SVM and Ensemble-SVM with lower accuracy performance.

In terms of sensitivity, Ensemble-PSO-SVM produce higher sensitivity performance for both experimental datasets. This shows the capability of Ensemble-PSO-SVM in classifying true positive data correctly for each dataset. By the increasing accuracy performance of Ensemble-PSO-SVM, both datasets produced highest sensitivity results with 99.00% for UCI Breast Cancer dataset and 99.81% for UCI Lymphography dataset, respectively. This shows the essential of obtaining slightest improvement in medical diagnosis in order to derive the necessary treatment a patient should receive.

In terms of selected features, a higher classification accuracy does not necessarily can be obtained using too small number of features, while too large number of features does not guarantee a lower classification accuracy. Table 15 shows the summary of total selected features obtained by Ensemble-PSO-SVM for all datasets. For UCI Breast Cancer dataset, an optimal solution is achieved by Ensemble-PSO-SVM when appropriate value of C and y are considered. UCI Breast Cancer dataset significantly achieves 96.15% classification accuracy using Ensemble-PSO-SVM using 5 optimum features. These features were also selected by Ensemble-SVM but with different ranking by achieving only 72.91% accuracy performance. This shows that a proper selection of kernel parameters values has caused the accuracy of Ensemble-PSO-SVM to increase significantly compared to Ensemble-SVM. UCI Lymphography dataset also achieves the highest classification performance by 96.62% of accuracy compared to Ensemble-SVM with 85.31% of accuracy. 8 out of 18 features are selected as the optimum significant features for predicting lymphoma cancer. Based on the results, the accuracy obtained by Ensemble-PSO-SVM using optimal number of features is higher compared to Ensemble-SVM. This means even though smaller number of features may reduce the computational complexity of classification, independent filter algorithms still unable to collect more information to optimally evaluate the significance of features individually. Thus, this indicates that Ensemble-PSO-SVM perceives the ability to determine an optimum number of significant features and appropriate classification parameters for SVM, where more information can be collected through harmonize classification to produce an optimal solution.

4.4.2. Performance Validation on Proposed Method

For performance validation, the results obtained from Ensemble-PSO-SVM is evaluated using 10-fold cross validation. The complete dataset is randomly partitioned into 10 equal size subsets where one part of the subsets is used as testing data and the remaining nine subsets are used as training data (Huang et al., 2018). These training and test sets are run 10 times to estimate the average highest classification performance. The results of proposed method via 10-fold cross validation for all experimental datasets is presented.

As shown in Table 16, Ensemble-PSO-SVM maintains high accuracy value in majority number of folds which indicates the proposed method have successfully provides a good and acceptable prediction. UCI Breast Cancer dataset performs highest accuracy in 2-fold, 3-fold, and 10-fold cross validation. Meanwhile in Table 17, UCI Lymphography dataset performs highest accuracy value at 9-fold and 10-fold cross validation. The average of cross validation results shows that 10-fold cross validation is effective since the independent test sets could validate for performance evaluation. This signify the importance of ensemble ilters feature selection and harmonize classification in increasing the predictive performance of SVM.

4.5. Comparisons Study

In this section, the performance of Ensemble-PSO-SVM is further validated based on the comparative analysis with several published methods that used the similar UCI Breast Cancer dataset.

As shown in Table 18, Ensemble-PSO-SVM performs better than the previous published method of PSO-CFS, PSO-RBF, PSO-KNN, PSO-NB, PSO-DT and PSO-Bayes (Harb & Desuky, 2014). The higher the accuracy value, the better the performance of the methods. This indicates

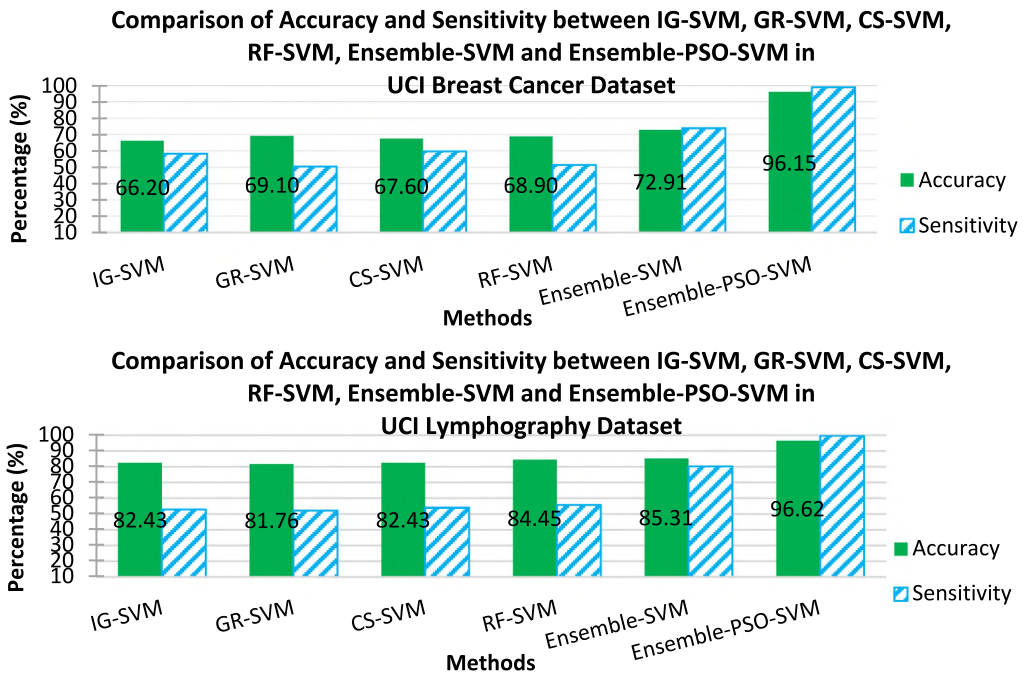

Fig. 6. Comparison of accuracy and sensitivity in IG-SVM, GR-SVM, CS-SVM, RF-SVM, Ensemble-SVM and Ensemble-PSO-SVM for all datasets.

Table 17
10-fold cross validation results of proposed Ensemble-PSO-SVM method on UCI Lymphography dataset.

Fold	Malignant Lymph	Metastases	Fibrosis	Normal	Total Correctly Classified	Total Incorrectly Classified	Average Accuracy (%)
1	56	89	3	1	120	28	81.08
2	56	89	3	1	120	28	81.08
3	61	84	2	1	117	31	79.05
4	53	92	3	0	122	26	82.43
5	54	88	3	0	122	26	82.43
6	62	84	2	0	121	27	81.76
7	59	86	2	1	124	24	83.78
8	56	89	3	0	124	24	83.78
9	60	82	4	2	141	7	95.27
10	**60**	**82**	**4**	**2**	**141**	**7**	**95.27**

that utilizing ensemble multi filters feature selection with occurrence rate evaluation can increase the accuracy of harmonize classification significantly. Irrelevant and imbalance number of features may result in bias thus degrading the accuracy of harmonize classification. By using Ensemble-PSO-SVM, the irrelevant features are eliminated, the optimum number of significant features and classification parameters that provides highest accuracy performance are maintained. From Table 18 Ensemble-PSO-SVM achieved the highest accuracy performance compared to other seven methods. This indicates that the feature selection using ensemble multi filters algorithm with harmonize classification of PSO and SVM successfully improves the accuracy performance while producing an optimum number of features. Ensemble-PSO-SVM has proven that "*Age*", "*Tumour Size*", "*INV-nodes*", "*Node Capsule*" and "*Degree of Malignancy*" are the most optimum features significant to predict malignant cases from benign cases of UCI Breast Cancer dataset. Hence, these features should be given more attention in practical medical diagnosis due to the primary information contained in these features.

Table 18
Comparative analysis of the proposed method with previous methods (Harb & Desuky, 2014).

Method	Accuracy (%)	Total Selected Features
Ensemble-PSO-SVM	**96.15**	**5**
PSO-CFS	72.03	5
PSO-RBF	76.22	4
PSO-KNN	76.22	5
PSO-NB	75.52	4
PSO-DT	74.13	5
PSO-Bayes	73.08	3
SVM	56.69	9

5. Conclusion

In a nutshell, irrelevant or redundant data in medical dataset may increase the dimensionality and accuracy issues when performing a diagnosis or decision making from a case file. Thus, the requirement to develop a reliable classification method that can identify the relevant data with high accuracy is essential. However, the improper settings of classification parameters may also influence the effectiveness of the classification method as improper selection of kernel parameter will degrade the accuracy of the relevant data. Improper classification parameters can also influence the selection of features that may misguide the classier due to overfitting of data. The common approaches to eliminate irrelevant features while producing optimal solution of classification is by using feature selection and optimization techniques.

In this paper, an ensemble multi filter feature selection with harmonize classification method, Ensemble-PSO-SVM has been proposed. Here, the ensemble multi filters feature selection method utilized are Information Gain, Gain Ratio, Chi-square, and Relief-F where they identify and combine relevant features in dataset with the consideration of features occurrences across each algorithm. Support Vector Machine is then used to evaluate the classification performance of the selected

features. An advantage of ensemble IG, GR, CS and RF is that the occurrence of the selected features subset on the training algorithm can be considered, and the optimum number of significant features can be obtained even after the data dimensionality is reduced. After the ensemble features are obtained, it is then optimized with the classification parameters synchronously by Particle Swarm Optimization and Support Vector Machine.

Experimental results achieved by UCI Breast Cancer dataset shows that Ensemble-SVM has produced considerable improvement in classification performance of the dataset compared to independent IG, GR, CS, RF and SVM in terms of accuracy. Consideration of features occurrence in ensemble IG, GR, CS and RF has aided in increasing the capability of IG, GR, CS and RF to determine the optimum number of significant features accurately. This is due to the capability of Ensemble-SVM to analyse attributes as irrelevant by evaluating the dependencies between features in the feature space of each filter using different approach from independent IG, GR, CS and RF. Therefore, Ensemble-SVM can contribute better learning and generalization ability in SVM classifier. By eliminating the irrelevant features and identifying the optimum number of significant features, the harmonize classification performance of PSO and SVM achieves well in terms of accuracy, sensitivity, and specificity when compared to classical SVM. The increase in harmonize classification of PSO-SVM is due to the appropriate values of kernel parameters used and the optimal number of significant features included into training task have significantly enhance the accuracy performance.

It is believed that the encouraging results signified by Ensemble-PSO-SVM can be implemented to assist medical experts in the healthcare centre for more effective and accurate diagnosis. For future work, this research will focus on analysing the proposed method on other dataset of different domains. In addition, other searching methods for optimizing the parameters of the classifier may also be analysed for appropriate parameter settings.

CRediT authorship contribution statement

Tengku Mazlin Tengku Ab Hamid: Conceived and designed the analysis, Conceptualization, Methodology, Collected the data, Investigation, Data curation, Contributed data or analysis tools, Software, Resources, Performed the analysis, Formal analysis, Wrote the paper, Writing - original draft, Visualization. **Roselina Sallehuddin:** Conceived and designed the analysis, Supervision, Performed the analysis, Validation, Wrote the paper, Writing - review & editing, Project administration, Funding acquisition. **Zuriahati Mohd Yunos:** Conceived and designed the analysis, Supervision, Performed the analysis, Validation, Wrote the paper, Writing - review & editing, Funding acquisition. **Aida Ali:** Conceived and designed the analysis, Supervision, Wrote the paper, Writing - review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study is supported by the Fundamental Research Grant Scheme (FRGS vot:5F154) which sponsored by Ministry of Higher Education (MOHE), Malaysia. Authors would like to thank Research Management Centre (RMC) Universiti Teknologi Malaysia and Applied Industrial Analytics Research Group (ALIAS) for the support and motivation in research activities.

References

Ali, H., Salleh, M. N. M., Saedudin, R., Hussain, K., & Mushtaq, M. F. (2019). Imbalance class problems in data mining: A review. *Indonesian Journal of Electrical Engineering and Computer Science*, *14*(3), 1560–1571. http://dx.doi.org/10.11591/ijeecs.v14.i3.pp1552-1563.

Alirezanejad, M., Enayatifar, R., Motameni, H., & Nematzadeh, H. (2019). Heuristic filter feature selection methods for medical datasets. *Genomics*, *112*(2), 1173–1181. http://dx.doi.org/10.1016/j.ygeno.2019.07.002.

Assarzadeh, Z., & Nilchi, A. R. N. (2015). Chaotic particle swarm optimization with mutation for classification. *Journal of Medical Signals and Sensors*, *5*(1), 12–20. http://dx.doi.org/10.4103/2228-7477.150380.

Bommert, A., Sun, X., Bischl, B., Rahnenfuhrer, J., & Lang, M. (2020). Benchmark for filter methods for feature selection in high-dimensional classification data. *Computational Statistics and Data Analysis*, *143*, Article 106839. http://dx.doi.org/10.1016/j.csda.2019.106839.

Brezočnik, L. (2017). Feature selection for classification using particle swarm optimization. In *IEEE EUROCON 2017 17th International Conference on Smart Technologies* (pp. 966–971). http://dx.doi.org/10.1109/EUROCON.2017.8011255.

Canedo, V. B., Marono, N. S., & Betanzos, A. A. (2012). An ensemble of filters and classifiers for microarray data classification. *Pattern Recognition*, *45*(1), 531–539. http://dx.doi.org/10.1016/j.patcog.2011.06.006.

Dai, J., & Xu, Q. (2013). Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification. *Applied Soft Computing*, *13*, 211–221. http://dx.doi.org/10.1016/j.asoc.2012.07.029.

Dankolo, N. H. M., Sallehuddin, R., & Mustaffa, N. H. (2017). A study of metaheuristic algorithms for high dimensional feature selection on microarray data. *AAIP Conference Proceedings*, *1905*(1), 1–6. http://dx.doi.org/10.1063/1.5012198.

Dongare, S. A., Ande, V. K., & Tirandasu, R. K. (2018). A feature selection approach for enhancing the cardiotocography classification performance. *International Journal of Engineering and Techniques*, *4*(2), 222–226. http://dx.doi.org/10.29126/23951303/IJET-V4I2P33.

Fahrudin, T. M., Syarif, I., & Barakbah, A. R. (2016). Feature selection algorithm using information gain based clustering for supporting the treatment process of breast cancer. In *IEEE International Conference on Informatics and Computing* (pp. 6–11). http://dx.doi.org/10.1109/IAC.2016.7905680.

Ghimatgar, H., Kazemi, K., Helfroush, M. S., & Aarabi, A. (2018). An improved feature selection algorithm based on graph clustering and ant colony optimization. *Knowledge-Based Systems*, *159*, 270–285. http://dx.doi.org/10.1016/j.knosys.2018.06.025.

Hamid, T. M. T. A., Sallehuddin, R., & Yunos, Z. M. (2019). Utilization of filter feature selection with support vector machine for tumours classification. *IOP Conference Series: Materials Science and Engineering*, *551*(1), 1–5. http://dx.doi.org/10.1088/1757-899X/551/1/012062.

Hamid, T. M. T. A., Sallehuddin, R., Yunos, Z. M., & Ali, A. (2019). Ensemble based multi filters algorithm for tumor classification in high dimensional microarray dataset. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(1.6), 116–123. http://dx.doi.org/10.30534/ijatcse/2019/1881.62019.

Hamouda, S. K. M., Abo El-Ezz, H. R., & Wahed, M. E. (2017). Intelligent system for predicting, diagnosis and treatment of breast cancer. *International Journal of Biomedical Data Mining*, *6*, 128. http://dx.doi.org/10.4172/2090-4924.1000128.

Hancer, E., Xue, B., & Zhang, M. (2018). Differential evolution for filter feature selection based on information theory and feature ranking. *Knowledge-Based Systems*, *140*, 103–119. http://dx.doi.org/10.1016/j.knosys.2017.10.028.

Harb, H. M., & Desuky, A. S. (2014). Feature selection on classification of medical datasets based on particle swarm optimization. *International Journal of Computer Applications*, *104*(5), 14–17. http://dx.doi.org/10.5120/18197-9118.

Huang, S., Cai, N., Pacheco, P. P., Narandes, S., Wang, Y., & Xu, W. (2018). Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics and Proteomics*, *15*, 41–51. http://dx.doi.org/10.21873/cgp.20063.

Huang, C. L., & Dun, J. F. (2008). A distributed PSO-SVM hybrid system with feature selection and parameter optimization. *Applied Soft Computing*, *8*, 1381–1391. http://dx.doi.org/10.1016/j.asoc.2007.10.007.

Lee, W. M. (2019). Supervised learning-classification using support vector machines. In W. M. Lee (Ed.), *Python® Machine Learning*. http://dx.doi.org/10.1002/9781119557500.ch8.

Lee, I. H., Lushington, G. H., & Visvanathan, M. (2011). A filter-based feature selection approach for identifying potential biomarkers for lung cancer. *Journal of Clinical Bioinformatics*, *1*(11), 1–8. http://dx.doi.org/10.1186/2043-9113-1-11.

Miao, J., & Niu, L. (2016). A survey on feature selection. *Procedia Computer Science*, *91*, 919–926. http://dx.doi.org/10.1016/j.procs.2016.07.111.

Moslehi, F., & Haeri, A. (2019). An evolutionary computation-based approach for feature selection. *Journal of Ambient Intelligence and Humanized Computing*, *11*, 3757–3769. http://dx.doi.org/10.1007/s12652-019-01570-1.

Pardo, B. S., Canedo, V. B., & Betanzos, A. A. (2019). On developing an automatic threshold applied to feature selection ensembles. *Information Fusion*, *45*, 227–245. http://dx.doi.org/10.1016/j.inffus.2018.02.007.

Prasad, Y., Biswas, K. K., & Hanmandlu, M. (2018). A recursive PSO scheme for gene selection in microarray data. *Applied Soft Computing*, *71*, 213–225. http://dx.doi.org/10.1016/j.asoc.2018.06.019.

Rahman, S. A., Bakar, A. A., & Hussein, Z. A. M. (2009). Filter-wrapper approach to feature selection using RST-DPSO for mining protein function. In *IEEE 2nd Conference on Data Mining and Optimization* (pp. 71–78). http://dx.doi.org/10.1109/DMO.2009.5341906.

Raj, D. M. D., & Mohanasundaram, R. (2020). An efficient filter-based feature selection model to identify significant features from high-dimensional microarray data. *Arabian Journal for Science and Engineering*, *45*, 2619–2630. http://dx.doi.org/10.1007/s13369-020-04380-2.

Raj, S., Ray, K. C., & Shankar, O. (2016). Cardiac arrhythmia beat classification using DOST and PSO tuned SVM. *Computer Methods and Programs in Biomedicine*, *136*, 163–177. http://dx.doi.org/10.1016/j.cmpb.2016.08.016.

Rani, R. R., & Ramyachitra, D. (2018). Microarray cancer gene feature selection using spider monkey optimization algorithm and cancer classification using SVM. *Procedia Computer Science*, *143*, 108–116. http://dx.doi.org/10.1016/j.procs.2018.10.358.

Sakri, S., Rashid, N., & Zain, Z. (2018). Particle swarm optimization feature selection for breast cancer recurrence prediction. *IEEE Access*, 29637–29647. http://dx.doi.org/10.1109/ACCESS.2018.2843443.

Singh, B., Kushwaha, N., & Vyas, O. P. (2014). A feature subset selection technique for high dimensional data using symmetric uncertainty. *Journal of Data Analysis and Information Processing*, *2*, 95–105. http://dx.doi.org/10.4236/jdaip.2014.24012.

Srisukkham, W., Zhang, L., Neoh, S. C., Todryk, S., & Lim, C. P. (2017). Intelligent leukaemia diagnosis with bare-bones PSO based feature optimization. *Applied Soft Computing*, *56*, 405–419. http://dx.doi.org/10.1016/j.asoc.2017.03.024.

Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. *Journal of Biomedical Informatics*, *85*, 189–203. http://dx.doi.org/10.1016/j.jbi.2018.07.014.

Wang, M., & Chen, H. (2020). Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis. *Applied Soft Computing*, *88*, Article 105946. http://dx.doi.org/10.1016/j.asoc.2019.105946.

Xue, Y., Tang, Y., Xu, X., Liang, J., & Neri, F. (2021). Multi-objective feature selection with missing data in classification. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 1–10. http://dx.doi.org/10.1109/TETCI.2021.3074147.

Xue, Y., Xue, B., & Zhang, M. (2019). Self-adaptive particle swarm optimization for large-scale feature selection in classification. *ACM Transactions on Knowledge Discovery from Data*, *13*(5), 1–27. http://dx.doi.org/10.1145/3340848.

Xue, B., Zhang, M., Browne, W. N., & Yao, X. (2016). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, *20*(4), 606–626. http://dx.doi.org/10.1109/TEVC.2015.2504420.

Zeng, N., Qiu, H., Wang, Z., Liu, W., Zhang, H., & Li, Y. (2018). A new switching-delayed-PSO- based optimized SVM algorithm for diagnosis of Alzheimer's disease. *Neurocomputing*, *320*, 195–202. http://dx.doi.org/10.1016/j.neucom.2018.09.001.

Zhang, J., Xiong, Y., & Min, S. (2019). A new hybrid filter/wrapper algorithm for feature selection in classification. *Analytica Chimica Acta*, *1080*, 43–54. http://dx.doi.org/10.1016/j.aca.2019.06.054.