Contents lists available at https://citedness.com/index.php/jdsi

Data Science Insights

Journal Page is available to https://citedness.com/index.php/jdsi

Research article

A Review on Support Vector Machine Problems and Solutions

[1] *Hiba Basim Alwan, University of Technology – Iraq, College of Computer Science, Department of Artificial Intelligence, Baghdad, Iraq*
[2] *Ku Ruhana Ku-Mahamud, University Utara Malaysia, School of Computing, Sintok, Kedah, Malaysia*
*email: [1, *] 110154@uotechnology.edu.iq,[2] ruhana@uum.edu.my*
** Correspondence*

ARTICLE INFO

Article history:
Received 26 October 2025
Revised 12 January 2026
Accepted 23 February, 2026
Available online 28 February 2026

Keywords:
Support Vector Machine
Model Selection
Feature Subset Selection
Non-Simultaneous Approach
Simultaneous Approach

Please cite this article in IEEE style as:
H. Al-Dulaimi and K. R. Ku-Mahamud, "A Review on Support Vector Machine Problems and Solutions", Data Science Insights, vol. 4, no. 1, pp. 7–20, Feb. 2026.

ABSTRACT

In this paper, a review is presented particularly on Support Vector Machine (SVM) problems, so as to understand these problems and to identify the approaches to solve them. The aim is to organize the main SVM problems in a manner that provides a clear view for the readers. The approaches for solving SVM problems were classified into non-simultaneous and simultaneous approaches based on constraints considered in solving the problems. Algorithms for model selection and feature subset selection are classified into heuristic and non-heuristic approaches. Very promising result can be obtained if the bio-inspired algorithms are simultaneously applied with SVM for classification problem.

Correspondence:
Hiba Basim Alwan, University of Technology – Iraq, College of Computer Science, Department of Artificial Intelligence, Baghdad, Iraq

Data Science Insights is an open access under the with CC By 4.0 license.

1. Introduction

One of the most popular data mining and machine learning algorithms is Support Vector Machine (SVM) [1], [2], [3], [4], which is presented by [5] as a classifier technique [6], [7]. However, SVM still captures researchers' attention even after about more than twenties years of research. This is due to its excellent classification performance, which was built on a statistical learning approach, and its problems that are still not completely solved. These problems are related with how to set SVM parameters as well as how to select the best feature subset [8].

The main concept of SVM is to obtain the Optimal Separating Hyperplane (OSH) between the positive and negative samples. This can be completed through maximizing the margin between two parallel hyperplanes. By finding this plane, SVM can then forecast the classification of unlabelled samples to be asked on which side of the separating plan the sample lies [9], [10], [11], [12], [13], [14], [15], [16], [17], [18]. SVM in fact is also named as maximum margin classifier [19]. Margin means the distance of the closest training sample to the hyperplane [20], [21], [22], [23], [24], [25], [26] [27].

SVM was improved based on the concept of structural risk minimization [17], [22], [28], [29], [30], [31], [32], [33], [34], [35], which means, it decreases the trade-off between the experimental mistake and the difficulty that arises from the classification of estimating functions to prevent over fitting [21], [36]. The concept of structural risk minimization plans the data into high dimensional domains via kernel functions by using kernel tricks [37], [38], [39], [40], [41], [42], [43].

SVM has an extremely good generalization capability and strong theoretical foundation [34], [43]. Generalization capability can be defined as the ability of SVM to classify unknown data examples correctly through constructed SVM. This is achieved by learning SVM from training examples, which is also known as SVM performance [44]. SVM manipulates the "curse of dimensionality", which means the computational complexity for the SVM training or testing is not affected by the feature space dimensionality [34], [37].

The work of the SVM algorithm will be as follows: given a training dataset, where each sample has been assigned to belong to one of two groups, an SVM training algorithm constructs a model that assigns new examples into one group or another [45].

There are two types of SVM: binary and multi-class. Binary SVM is the core of SVM. It is capable of distinguishing between two classes. Multi-class SVM expands binary SVM by being able to classify three or more classes. The main algorithm involved in multi-class SVM is to divide the classification problem to many binary problems with its own classifier [46]. Two main methods are related to SVM multi-class classification: One-Against-All (OAA) and One-Against-One (OAO). In OAA, a group of binary classifiers are trained to isolate each class from all others. Based on the largest decision value, each data object would be classified to the class. The main advantage of this method is its fast approach; unfortunately, it is prone to errors obtained from the marginally unstable training groups. The other method, OAO, is a sequence of classifiers that is implemented on a pair of classes with the popular calculated class preserved in each object. Subsequently, a max-min operator is applied to calculate to which class the object will finally be allocated. This method is considered as being more accurate, but it is slower than OAO [47].

Nevertheless, there are two main problems in the SVM classifier that influence the classification accuracy; these are: **(a)** tuning SVM parameters to be used in the SVM classifier, which is also known as Model Selection (MS); and **(b)** selecting the optimal feature subset. These two problems are essential because they affect each other [48], [49], [50], [51], [52], [53], [54]. Other SVM problems include selecting suitable kernel functions [15].

The objective of this review is to summarize and record a comprehensive survey on the two most important problems in SVM. By providing the information related to the two main problems, this study aims to inspire further research in exploring the new least studied area or any potential in the area of SVM classifier problem with better insight into potential algorithms that can be applied to solve these problems. This review has concentrated on those publications that are considered most significant, within the authors' knowledge on SVM classifier problems from 2008 to 2013.

The paper is organized as follows: the main problems of SVM are presented in Section 2. Additionally, this section is divided into two subsections; the first one, Section A, discusses the Model Selection (MS) problem, while the second part, Section B, is related to the problem of selecting suitable feature subset to be introduced to the SVM classifier. Next, in Section 3, the solution algorithms designed to solve the SVM classifier problems are discussed and classified into non-simultaneous approach and simultaneous approach, which in turn, are divided into other subsections that deal with solving MS problem and feature subset selection problem. Further, the datasets used to classify the SVM classifier from the related literature are also discussed. Finally, the summary on the SVM classifier problems with new potential research directions are presented in Section 4.

2. Support Vactor Machine Problem

The classification work of SVM strongly depends on selecting the appropriate kernel parameter and suitable value for the regularization parameter C [40], [44], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66]. Additionally, selecting small numbers and suitable feature subset has an effect on the performance of SVM [34], [42], [55], [58], [67], [68], [69], [70], [71]. The following two subsections discuss these two main problems.

2.1. Model Selection

SVM has recently been seen as a stronger approach for resolving difficulties in pattern classification; however, its execution depends, especially, on the variables chosen for it [72], [73]. Variables chosen for SVM are very complex and difficult to resolve via conventional optimization approaches [74], [75], [76], [77]. The difficulty in selecting the values of SVM variables that decrease the prediction of test error is called the MS problem [64], [75], [78], [79], [80], [81]. However, researchers use previous skills in choosing SVM variables. In order to seek suitable variable groups, some implementation conditions of the prototype need to be selected. This point, at which a minimum training group mistake or validation group mistake is achieved, is chosen as the variable group to apply training and thereafter forecasting [82].

The variables involved in optimization, C and kernel parameters, are necessary in constructing a successful and highly executable SVM prototype. When utilizing SVM, it is difficult to choose parameters in order to obtain a high generalization execution. Therefore, investigating better approaches to choose optimal variables is often carried out. The purpose of optimizing variables for SVM is to obtain variables that decrease generalization mistakes [83], [84]. It is seen as difficult because it requires either exhaustive searching through the space of variables or optimizing approaches that discover only a bounded subgroup of the potential values [85]. There is also no regular methodology that accepts advance approximation of their optimal values. In fact, in the present classification work, obtaining good values for these parameters is not an easy MS problem that requires either an exhaustive search through the space of hyper variables, or an optimization approach that searches only a bounded subgroup of the potential values [44]. Below is a discussion on these SVM parameters.

2.1.1. Regularization Parameter *C*

The regularization parameter C monitors the trade-off between the complexity of the machine and the amount of non-separable points (training error). In other words, the C parameter monitors the trade-off between margin maximization and error minimization [44], [86], [87], [88]. If C is too large, the assessment of precise percentage

is very high in the training stage, but very low in the testing stage. If C is too small, the assessment precision is unsatisfied, and the model is useless [89].

2.1.2. Kernel Function Parameter γ

The kernel function parameter γ indirectly defines non-linear planning from the input domain to a high dimensional feature (attribute) domain, where the maximal margin hyperplane is obtained [44], [88]. An extremely large value for kernel parameter can be due to over fitting; otherwise, it is due to under fitting. Kernel parameter has much greater impact on assessment than C [89].

2.2. Feature Subset Selection

One of the important components of machine learning is feature analysis, which includes two main operations, namely feature extraction and Feature Selection (FS) [90], [91]. FS is the process of determining a subset of fields in the database, minimizing the number of fields that appear during data classification [48], [92], [93]. The main idea behind FS is to select a subset of input variables by deleting features that contain less or no information [94]. FS aims to decrease the dimension of the initial feature group by determining the unauthentic features, which would eventually supply the best performance under a certain classification dataset [95], [96], and to delete unrelated, unneeded, or noisy features while preserving the richness of the instructive ones [97], [98]. FS may be considered as an optimization problem that seeks potential feature subsets, which ultimately determines the optimal one [97]. Exhaustive search is the easiest way to identify the best subsets. A FS problem is a hard combinatorial process in machine learning, but it has high practical importance [99].
SVM is not able to avoid the influence of huge numbers of unrelated or redundant features on the classification results. Thus, selecting a few numbers of suitable features would result in obtaining good classification accuracy [34].

2.3. Kernel Function Selection

There are many types of kernel functions, such as polynomial, Radial Base Function (RBF), linear and sigmoid kernel function. RBF is the most popular kernel function because of its capability to manage high dimensional data [100], good performance in major cases [101], and it only needs to use one parameter, which is kernel parameter gamma (γ) [48]. However, the disadvantage of RBF is that it will fail if it deals with large numbers of features [51].

3. Solutions to SVM Problems

There are two main strategies to classify the patterns using SVM: non-simultaneous and simultaneous. Below is a full discussion on these two approaches.

3.1. Non-Simultaneous Approach

The non-simultaneous strategy may be divided into two approaches: optimizing only the value for the SVM parameters, or optimizing only the feature subset selection. One of the main downsides to the non-simultaneous strategy is that selecting the feature subset and tuning the SVM parameters are completed separately and this would greatly influence the classification process as there might be a loss of relevant information during this process [42]. The non-simultaneous algorithms can also be divided according to the type of algorithms that have been used, i.e., non-heuristics search algorithms and heuristics search algorithms. The subsections below discuss these search algorithms.

3.1.1. Non-Heuristics Search Algorithm for Model Selection Problem

The non-simultaneous strategy may be divided into two approaches: optimizing only the value for the SVM parameters, or optimizing only the feature subset selection. One of the main downsides to the non-simultaneous strategy is that selecting the feature subset and tuning the SVM parameters are completed separately and this would greatly influence the classification process as there might be a loss of relevant information during this process [42]. The non-simultaneous algorithms can also be divided according to the type of algorithms that have been used, i.e., non-heuristics search algorithms and heuristics search algorithms. The subsections below discuss these search algorithms.

Table 1: Summary of non-optimization solutions for SVM MS problem

Ref.	SVM Class	MS Algorithm	Kernel	Dataset
[37]	Binary	GS	ARD	4 UCI
[70]	Binary	GS	RBF	Three UCI datasets
[86]	Binary	k-NN, nearest enemy, redundant fast clustering and 10-fold CV	Gaussian	6 UCI datasets
[88]	Binary	New Gaussian kernel parameter selection method is proposed and then integrated with a one-dimension CV seek mechanism	Gaussian	8 UCI
[100]	Multi	GS	Gaussian	12 real-world problems
[101]	Binary	Manual	Gaussian	Monk's problem UCI
[102]	Binary	GS	RBF	154 Patients
[103]	Multi	Manual	Linear	Liver disease metabolism
[104]	Binary	5-fold CV	Gaussian	Electro Cardio Gram signal and toy dataset
[105]	Multi	Manual	RBF	6 UCI datasets
[106]	Multi	Kernel matrix approximation	Gaussian	5 UCI, Statlog and Delve dataset
[107]	Binary	GS	RBF	Wisconsin breast cancer UCI
[108]	Binary	$\Sigma = 100$ and $C = 1000$ computed by trial and error	RBF	3 medical UCI
[109]	Binary	10-fold CV	Gaussian, polynomial	4 datasets
[110]	Binary	GS-based ACO technique	RBF	Banana dataset from International Development Association repository

From Table 1, it can be seen that binary SVM classifier was the most used class as compared to multi-class SVM classifier. Furthermore, GS, CV, and manual approaches were the most commonly used algorithms to tune SVM parameters. Besides that, RBF, Gaussian, and UCI datasets were the favourite kernel function and repository datasets that have been used by the authors.

3.1.2. Non-Heuristics Search Algorithm for Feature Subset Selection Problem

In order to select the feature subset, it can be seen as three groups, depending on how the classification model is built using machine learning and FS search, namely wrapper, filter, and embedded (also known as the hybrid approach) [100], [104], [106], [111], [112], [113], [114]. Filter and wrapper approaches may be further classified into five main methods: forward selection, backward elimination, forward/backward combination, random choice and, finally, instance-based method [115], [116].

A filtering method can be used in the pre-processing phase to minimize space dimensionality, over fitting, and independency on the classifier. The main idea behind using a filtering approach is to look for related features and delete the non-related features through using statistical measurements [100], [104], [111], [106], [112], [117], [118], [119]. There are two steps in using a filter approach: in the first step, criteria such as information, distance, dependence, or stability are used for FS without the classifier; in the second step, with the selected features, a classifier would learn from the training data and be examined on the test data [100], [111], [119].

A wrapper approach searches through the area of all potential feature subsets by applying the forecasted accuracy based on classification algorithms. These approaches are exhaustively calculated and rely on the classification algorithm. However, they also consider feature dependencies [39], [100], [104], [106], [112], [117], [118], 120]. A wrapper approach is seen as useful, but is very slow to run because the learning approach is called repeatedly. Hence, wrappers do not perform well on large datasets that have many features [104], [106], [117], [118], [121]. The quality of features and feature subset size are often used as the terminating condition in wrapper techniques [100]. The wrapper approach involves two steps: (i) choosing the good feature subset by utilizing the accuracy of the classifier (on the training data) as a condition to learn and understand; and (ii) a learning classifier is taken from the training data with the best feature subset, and would be examined on the test data [119].

In the hybrid or embedded approach, the advantage of both feature and wrapper techniques is utilized. The optimal feature subset is constructed into the classifier when it is being built. The interaction with classification prototype enhances the calculation complexity and considers feature dependencies [111], [112], [118].

If a comparison is made between filter and wrapper approaches, it is obvious that the filter approach is able to perform at a much faster rate as compared to the wrapper approach. Therefore, it can be used on large datasets with many features. The generic existence allows them to work with any learner. This is not the same when it

comes to the wrapper approach because it needs to be re-executed when changing from one learning approach to another. The filter technique is more efficient than the wrapper technique based on calculation. However, a major disadvantage is that an optimal selection of features may not be treated separately from the inductive and representational influences of a learning approach, which is used to build the classifier. The wrapper technique, on the other hand, includes extra calculation, which measures the nominated feature subset by running a provided learning approach on the database using each feature subset under restriction [37], [104], [106], [117], [118], [122]. In literatures, techniques like Balance index, Recursive Feature Elimination (RFE), Embedded RFE, Improved F-score Sequential Forward Search (IFSFS), Term Frequency-Inverse Document Frequency (TF-IDF), minimum Redundancy-Maximum-Relevance (mRMR), Kernal F-score FS (KFFS), Mutual Information (MI), Particle Swarm Optimization (PSO), filter and wrapper techniques have been used in order to select SVM feature subset and datasets like Electro Encephalo Graphy (EEG) signals, Electro Cardio Gram (ECG) signal and UCI datasets were used to test SVM performance. Studies related to the FS algorithm and dataset for testing are summarized in Table 2.

Table 2: Non-optimization solutions for SVM feature subset selection problem

Ref.	FS Algorithm	Dataset
[37]	Kernel polarization	Four UCI datasets
[67]	Gaussian marginal densities for a Bayesian prototype, wrapper approach	Toy and ten alternation commercial Tunisian banks during a period from 2000 to 2006
[70]	Improved F-score (filter approach) and IFSFS (wrapper approach)	Three UCI datasets
[100]	Fuzzy complementary criterion (forward filter technique)	Twelve real-world classification problems
[102]	Balance Index	154 Patients
[103]	Embedded RFE	Liver disease metabolism
[104]	Embedded	ECG signal and toy datasets
[106]	MI (filter approach) and discrete PSO (wrapper approach)	Six UCI datasets
[107]	F-score	Wisconsin breast cancer UCI dataset
[108]	KFFS	Three medical UCI datasets
[109]	Wrapper technique based on sequential backward selection	Diabetes and Wisconsin Breast Cancer datasets from UCI Colorectal Microarray and Lymphoma Microarray datasets from DNA microarray
[123]	RFE	EEG signals
[124]	1) a method that gathers Kullback-Leibler MI and distance information; (2) a text mining method, TF-IDF; (3) mRMR	100 bacterial genomes

From Table 2, it can be seen that many different approaches have been used for the feature subset selection approach. Similar to non-optimization solutions for the SVM MS problem, UCI datasets were also the most commonly used repository in the abovementioned studies.

3.1.3. Heuristics Search Algorithm for Model Selection Problem

The variables chosen for SVM are very complex and difficult to resolve via conventional optimization approaches [74], [75], [76], [77]. Therefore, investigating better approaches to choose optimal variables is often carried out. The purpose of optimizing variables for SVM is to obtain variables that decrease generalization mistakes [84], [124]. In fact, in the present classification work, obtaining good values for these parameters is not an easy MS problem that requires an optimization approach that searches only a bounded subgroup of the potential values [44]. Many researchers choose SVM variables experimentally by applying a bounded number of values and preserving the values that provide the minimum test mistake. Manual selection is avoided because it is not accurate and does not provide any assurance on the quality of the result [85]. Evolutionary algorithms such as Genetic Algorithm (GA), PSO, Simulated Annealing (SA), Ant Colony Optimization (ACO), Differential Evolution (DE), and Bare Bones Differential Evolution (BBDE) as well as Multi-Layer Perceptron (MLP) can simplify the choice of optimal value processes for SVM parameters [30], [44], [78]. Table 3 summarizes the literature that used optimization approaches to solve the SVM MS problem.

Table 3: Summary on optimization solutions for SVM MS problem

Ref.	MS Algorithm	Kernel Function	Dataset
[15]	GA	RBF	Eight UCI datasets
[30]	BBDE hybrid with barebones PSO and DE	---	Iris, Breast, Sonar and pima UCI data set
[44]	ACO and GS	RBF	Breast cancer, diabetes, heart, thyroid and titanic UCI datasets
[57]	DE	Linear, polynomial, RBF, sigmoid	Target Miner and two UCI datasets
[59]	Standard PSO and real GA	RBF	Letter recognition database and speech dataset for speaker recognition processes
[63]	GA with a gradient descent method	RBF	banana, breast cancer, diabetes, flare-solar, German, heart, image, ring norm, splice, thyroid, titanic, two norm, and waveform UCI datasets
[64]	PSO and GA	---	---
[110]	GS-based ACO	RBF	One dataset
[125]	GA	RBF	2,300 mobile phone reviews data
[126]	PSO	RBF	Erythematous-squamous disease UCI dataset
[127]	GA, PSO, feed forward MLP	Linear, quadratic, Gaussian, polynomial	---
[128]	Exponent weighted ACO	---	Chinese text provided by Sogou Labs
[129]	Modified PSO	Gauss, wavelet	---
[130]	SA	---	Hepatitis disease, two-spiral dataset and a chain link dataset
[131]	GA	RBF	Five datasets, three from UCI and two found in LibSVM
[132]	ACO	---	---
[133]	ACO	Gaussian	Chinese bank credit risk

From Table 3, it can be seen that PSO, GA, SA, DE, and ACO were the most used optimization algorithms to tune SVM parameters. RBF, Gaussian, and UCI datasets were the most applied kernel function and dataset repository that have been used in the studies.

3.1.4. Heuristics Search Algorithm for Feature Subset Selection Problem

Feature subset selection is considered as a problem that needs to be optimized. Researches related to FS constantly deal with the issue of choosing suitable input features to produce good classification results. Therefore, it has been noticed that successful application using datasets of large numbers of features would be difficult to manipulate [39], [70], [134], [135], [136], [137]. In order to find the optimal feature subset for SVM, optimization techniques such as Empirical model decomposition (EMD), Association Rules (ARs), Kernel Condition Random Field (KCRF), Principle Component Analysis (PCA), Tabu Search (TS), SA, GA, PSO, ACO, Ant System (AS), Elitist Ant System (EAS) and Ant Colony System (ACS) can be used to select the suitable features [115], [116]. Table 4 summarizes the literature that used optimization approaches to solve the SVM feature subset selection problem.

Table 4: Summary of literature on optimization solutions for SVM feature subset selection problem

Ref.	FS Algorithm	Dataset
[52]	ACO and GA via a wrapper approach	Lab plant image database of a cotton area
[55]	ACO	15 non-image UCI datasets
[89]	EMD method to extract features, distance evaluation method to select features subset	Vibration signals from rolling element bearings
[126]	Ars	Erythematous-squamous disease UCI dataset
[128]	KCRF to segment words, extract features and representation, PCA and noise elimination threshold to select feature	Chinese text provided by Sogou Labs
[129]	Dynamic weighted ACO	Eight UCI datasets
[133]	Average linkage method and then computed t-tests for each feature. After that, rough sets are utilized to select suitable feature subsets	Chinese bank credit risk
[138]	ACO-Selection (ACO-S)	Five microarray datasets

[140]	AS and EAS	1,445 online Arabic newspaper archives, taken from Al-Jazeera, Al-Nahar, Al-Hayat, Al-Ahram, and Al-Dostor
[148]	Integrating ACS via a wrapper approach	Ten binary datasets taken from statistical and UCI datasets

It can be seen from the summary that ACO is the most used optimization algorithm to optimize feature subset selection for SVM classifier. Moreover, UCI was the most used dataset repository in these studies.

3.2. Simultaneous Approach

SVM has been used to solve classification problems with acceptable accuracy while simultaneously optimizing both feature subset selection and SVM parameters. The current research direction has moved towards simultaneously optimizing both feature subset selection and tuning SVM parameters by using optimization algorithms. This approach will increase the classification accuracy because selecting the suitable feature subset and values for SVM parameters influence each other, and in turn, will influence the classification accuracy [39], [48], [49], [51], [53], [54], [58], [141], [142], [143], [144], [145], [146]. Techniques such as Mixed-Variable ACO with continuous ACO (ACOR) to optimize continuous parameters ($ACO_{MV\text{-}R}$), PSO, ACO, GA, Immune Clonal Algorithm (ICA), Cat Swarm Optimization (CSO), Clonal Selection Algorithm (CSA), Real-valued Gravitational Search Algorithm (RGSA), Binary (discrete) GSA (BGSA), and Adaptive Cohort Intelligence (SACI) are used to simultaneously optimize feature subset selection and value for SVM parameters.

A total of fourteen similar works [48], [49], [50], [51], [53], [54], [93], [141], [143], [146], [147], [148] suggested using hybrid systems to enhance classification accuracy by using limited, suitable feature subsets. All fourteen works optimized feature subset and SVM parameters, which are C and γ RBF kernel variables, simultaneously. Ultimately, SVM is used to measure the quality of the solution for all hybrid systems. However, the difference is what the hybrid system is based on. The summarization of these works is illustrated in Table 5.

Table 5: Studies on simultaneous optimization of model and feature subset selections

Ref.	MS and FS Algorithm	Kernel Function	Dataset
[48]	GA	RBF	UCI
[49]	Bees algorithm	RBF	UCI
[50]	Discrete PSO	RBF	UCI
[51]	Mixed discrete and continuous PSO	RBF	Distributed parallel
[53]	CSA	RBF	UCI
[54]	GA	RBF	UCI
[93]	ACO	RBF	UCI
[94]	Modified Binary PSO (MBPSO)	RBF	UCI
[141]	$ACO_{MV\text{-}R}$-SVM	RBF	UCI
[142]	SACI	RBF	---
[143]	RGSA to optimize MS and BGSA to optimize feature subset selection	RBF	UCI
[145]	CSO	RBF	UCI
[146]	Continuous Hide-and-Seek SA to optimize MS and discrete SA to optimize feature subset selection	RBF	UCI
[147]	PSO	RBF	UCI
[148]	ICA	RBF	UCI
[149]	Dual formula SVM involving a penalization function based on 0-norm approximation and modifying Gaussian kernel utilizing gradient descent approximation for kernel optimization and feature deletion	Linea Polynomial Gaussian	---

From the summary, it can be seen that the UCI datasets were the most used data repository and RBF is the most applied kernel function. $ACO_{MV\text{-}R}$, ACO, GA, PSO ICA, SA, Bees, CSA, CSO, and RGSA were the optimization algorithms that have been used to simultaneously optimize SVM MS and feature subset selection.

4. Conclusion

Selecting SVM parameters' value and feature subset as an optimization problem is crucial in solving the classification problem. Many approaches have been applied to classify patterns with high classification accuracy and SVM is considered as an excellent approach to classify patterns. SVM suffers from dealing with huge numbers of features and poor settings to its parameters, which influence its performance. Simultaneously optimizing SVM parameters and feature subset selection has been shown to improve the performance of SVM. The simultaneous

operation is produced by hybridizing ACO and other optimization algorithms with SVM. It is worth pointing out that, the novelty of this study lies in the review that is specially made to focus on pattern classification using SVM classifier.

Based on the many articles that have been reviewed, the simultaneous solution approach is the top trend to solve SVM problems. It is important to select the right optimization algorithm to be simultaneously applied with SVM so that an optimal solution will be achieved. From this review, it is found that a very promising result can be obtained if the bio-inspired algorithms are simultaneously applied with SVM.

As discussed in this review, SVM problems will remain an endless study domain in the future as its performance in classification can be mapped into real-world classification problems. It is also found that, several proposed solution approaches that have been discussed are significant to be used and have been adapted in several real-world classification problems.

Acknowledgement
None

Conflicts of Interest
The authors declare that they do not have any conflicts of interest.

Authors' contributions
The authors' contributions are as follows: Hiba Basim Alwan: conceptualization, methodology, analysis, and writing. Ku Ruhana Ku-Mahamud: review and editing.

Data Availability
No dataset has been used in this study.

References

[1] K. Michał and N.-S. Ewa, "Support vector machine for malware analysis and classification," in *Proc. of the IEEE/WIC/ACM Int. Joint Conf. on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, Warsaw, Poland, pp. 415–420, 11–14 Aug. 2014.

[2] L. Dalian, T. Yingjie, B. Rongfang, and S. Yong, "Self-universum support vector machine," *Pers Ubiquit Comput*, *18*, pp. 1813–1819, 2014.

[3] V. L. Martin, H. Jos, W. Lei., J. Hailong, and P.D. G. Jose, "Reconfigurable support vector machine classifier with approximate computing," *IEEE Computer Society Annual Symposium on VLSI*, pp. 13–18, 2017.

[4] A. F. Emmanuel, Y. O. Philip, and A. Johnson, "Prediction of transient stability status using walsh hadamard transform and support vector machine," *IEEE PES-IAS Power Africa*, pp. 301–306, 2017.

[5] V. Vapnik, "The nature of statistic learning theory," NY-USA: Springer, 1995.

[6] C. Amira, T. Haifa, and S. Hichem, "SVM-based indoor localization in wireless sensor networks," in *Proc. of the 13th Int. Wireless Communications and Mobile Computing Conf. (IWCMC)*, Valencia, Spain, 26–30 June 2017, pp.1144–1149.

[7] L. Xingxin, Z. Youwen, W. Jian, L. Zhe, L. Yining, and Z. Mingwu, "On the soundness and security of privacy-preserving SVM for outsourcing data classification," *IEEE Transactions on Dependable and Secure Computing*. 2017.

[8] Z. Si-Da, M. Yuan-Chen, L. Li, K. Jie, M. Zhi, and Y. Lei, "Output-only modal parameter estimator of linear time-varying structural systems based on vector TAR model and least squares support vector machine," *Journal of Mechanical Systems and Signal Processing*, 98(2018), 722–755.

[9] Shubham S., S. N. Garsha, Somsrsa C. (2016). Characterization of classifier performance on left and right limb motor imagery using support vector machine, classification of EEG signal for left and right limb movement. *Proceedings of the 5th International Conference on Wireless Networks and Embedded Systems (WECON)*, pp. 1–4, 14–16 Oct. 2016, Chitkara University, Rajpura, India

[10] Vapnik, V. & Vashist, A. (2009). A new learning paradigm: Learning using privileged information. *Neural Networks: The Official Journal of The International Neural Network Society*, *22*(5–6), 544–557.

[11] Qi, Z., Tian, Y., & Shi, Y. (2013). Robust twin support vector machine for pattern classification. *Pattern Recognition*, *46*(1), 305–316.

[12] Fu, Z., Lu, G., Ting, K. M., & Zhang, D. (2013). Efficient nonlinear classification via low-rank regularised least squares. *Neural Computing & Applications*, *22*(7–8), 1279–1289.

[13] Khalid, S. & Razzaq, S. (2012). Frameworks for multivariate m-mediods based modelling and classification in Euclidean and general feature spaces. *Pattern Recognition*, *45*(3), 1092–1103.

[14] Khashei, M., Zeinal Hamadani, A., & Bijari, M. (2012). A novel hybrid classification model of artificial neural networks and multiple linear regression models. *Expert Systems with Applications*, *39*(3), 2606–2620.

[15] Tsai, C. W., Liao, M. Y., Yang, C. S., & Chiang, M. C. (2011). Classification algorithms for interactive multimedia services: A review. *Multimedia Tools and Applications*, 1–29.

[16] Li, S. & Tan, M. (2010). Tuning SVM parameters by using a hybrid CLPSO-BFGS algorithm. *Neurocomputing*, *73*(10–12), 2089–2096.

[17] Al-Naami, B., Al-Nabulsi, J., Amasha, H., & Torry, J. (2010). Utilizing wavelet transform and support vector machine for detection of the paradoxical splitting in the second heart sound. *Medical and Biological Engineering and Computing*, *48*(2), 177–184.

[18] Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.

[19] Kuan-Chou L., Tse.-Yen Liu, Po.-Han. C., & Chih.-Ting. L. (2017). Use Support Vector Machine (SVM) to estimate gas concentration in mixture condition. *Proceedings of the 2017 IEEE International Conference on Applied System Innovation*, pp. 744–746.

[20] Xiaolin H., Lei S., & Johan A. K. S. (2017). Solution path for pin-SVM classifierswith positive and negative t value. *IEEE Transactions on Neural Networks and Learning System*, *28*(7), pp. 1584–1593.

[21] Bajla, I., Rublik, F., Arendacka, B., Farkas, I., Hornisova, K., & Stole, S. (2009). Segmentation and supervised classification of image objects in EPO doping-control. *Machine Vision and Applications*, *20*(4), 243–259.

[22] Zhang, H. & Mao, H. (2009). A PSO-based framework for dynamic SVM model selection. *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*, 8–12 July 2009, Montréal Québec, Canada, 1227–1234.

[23] Liu, J. & Yuan , X. (2009). Obscure bleeding detection in endoscopy images using support vector machines. *Optimization Engineering*, *10*(2), 289–299.

[24] Wang, Z. & Chen S. (2007). New least squares vector machines based on matrix patterns. *Neural Processing Letters*, *26*(1), 41–56.

[25] Zhang, B., Pham, T., & Zhang, Y. (2007). Bagging Support Vector Machines for classification of SELDI-TOF mass spectra of ovarian cancer serum samples. In M. Orgun & J. Thornton (Eds.), *Advances in artificial intelligence* (pp. 820–826). Berlin Heidelberg: Springer.

[26] Choi, Y. & Noh, J. (2004). Relevance feedback for content-based image retrieval using proximal support vector machine. In A. Lagana, M. Gavrilova, V. Kumar, Y. Mun, C. J. Tan, & O. Gervasi (Eds.), *Computational Science and its Applications*, (pp. 942–951). Berlin Heidelberg: Springer.

[27] Ren, J., Shen, Y., Ma, S., & Guo, L. (2004). Applying multi-class SVMs into scene image classification. In B. Orchard, C. Yang & M. Ali (Eds.), *Innovations in applied artificial intelligence*, (pp. 924–934). Berlin Heidelberg: Springer.

[28] Yuxia, H. & Hongtao, Z. (2012). Chaos optimization method of SVM parameters selection for chaotic time series forecasting. *Physics Procedia*, *25*, 588–594.

[29] Li, N. & Zhou, R. (2011). A new approach to fault pattern classification of gasoline engine vibration. *Proceedings of the 8th International Conference on Fuzzy systems and Knowledge Discovery*, 26–28 July 2011, Shanghai, 1892–1896.

[30] Qiu, D., Li, Y., Zhang, X., & Gu, B. (2011). Support Vector Machine with parameter optimization by bare bones differential evolution. *Proceedings of the 7th International Conference of the Natural Computation*, 26–28 July 2011, Shanghai, 263–266.

[31] Jin, Z., Chen, Y., & Ma, X. (2010). Model selection for SVM based on similarity margin of inner product. *Proceedings of the Conference on Control and Decision Conference*, 26–28 May 2010, Xuzhou, 1653–1657.

[32] Kapp, M., Sabourin, R., & Maupin, P. (2009). A PSO-based framework for dynamic SVM model selection. *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*, 8–12 July 2009, Montréal Québec, Canada, 1227–1234.

[33] Lu, N., Zhou, J., He, Y., & Liu, Y. (2009). Particle swarm optimization for parameter optimization of support vector machine model. *Proceedings of the 2nd International Conference on Intelligent Computation Technology and Automation*, 10–11 October 2009, Changsha, Hunan, 283–286.

[34] Liu, W. & Zhang, D. (2009). Feature subset selection based on improved discrete particle swarm and support vector machine algorithm. *Proceedings of The Information Engineering and Computer Science*, 19–20 December 2009, Wuhan, 1–4.

[35] Luo, Z.-Y., Zhang, W.-F., Ye, B.-Y., & Cai, L.-Q. (2008). Wavelet SVM ensemble for pattern classification with quantum-inspired evolutionary algorithm. *Proceedings of The International Conference on Wavelet Analysis and Pattern Recognition*, 30–31 August 2008, Hong Kong, 485–490.

[36] Ju, W., Shan, J., Yan, C. & Cheng, H. (2009). Discrimination of disease-related non-synonymous single nucleotide polymorphisms using multi-scale RBF kernel fuzzy support vector machine. *Pattern Recognition Letters*, *30*(4), 391–396.

[37] Wang, T., Huang H., Tian S., & Xu J. (2010). Feature selection for SVM via optimization of kernel polarization with gaussian ARD kernels. *Expert Systems with Applications*, *37*(9), 6663–6668.

[38] Ye, Y., Chen, L., Wang, D., Li, T., Jiang, Q., & Zhao, M. (2009). SBMDS: An interpretable string based malware detection system using SVM ensemble with bagging. *Journal in Computer Virology*, *5*(4), 283–293.

[39] Shieh, M. & Yang, C. (2008). Multiclass SVM-RFE for product from feature selection. *Expert Systems with Applications*, *35*(1–2), 531–-541.

[40] Lessmann, S., Stahlbock, R., & Crone, S. (2006). Genetically constructed kernels for support vector machines. In Hans H., Herbert K. & Jorn S. (Eds.), *Operations research*, (pp. 257–262). Berlin Heidelberg: Springer.

[41] Kadoury, S. & Levine, M. (2006). Finding faces in gray scale images using locally linear embeddings. In H. Sundaram, M. Naphade, J. Smith & Y. Rui (Eds.), *Image and video retrieval*, (pp. 221–230). Berlin Heidelberg: Springer.

[42] Nguyen, H., Ohn, S., Chae, S., Song, D., & Lee, I. (2006). Optimizing weighted kernel function for support vector machine by genetic algorithm. In A. Gelbukh & C. Garcia (Eds.), *Advances in artificial intelligence*, (pp. 583–592). Berlin Heidelberg: Springer.

[43] Adriana-Cristina E. & Valentin S. (2014). Enhanced intrusion detection system based on bat algorithm-support vector machine. *Proceedings in the 11th International Conference on Security and Cryptography (SECRYPT)*, 28–30 Aug. 2014, Vienna, Austria.

[44] Zhang, X., Chen, X., & He, Z. (2010). An ACO-based algorithm for parameter optimization of support vector machines. *Expert System with Applications*, *37*(9), 6618–6628.

[45] Eriko N., Asit M., & Debbie M. (2015). A sparse matrix vector multiply accelerator for support vector machine. *Proceedings of the International Conference on Compilers, Architecture and Synthesis for Embedded Systems (CASES)*, Amsterdam, 4–9 Oct. 2015, pp. 109–116.

[46] Harvey, J. (2009). GPI acceleration object classification algorithm using NVIDIA CUDA. (Master Thesis, Rochester Institute of Technology, 2009). Retrieved from http://ritdml.rit.edu/bitstream/handle/1850/10894/35445_pdf_00B0B24A-DFD8-11DE-9A30-D21AD352ABB1.pdf?sequence=1.

[47] Tzotsos, A. & Argialas, D. (2008). Support Vector Machine classification for object-based image analysis. In T. Huang, V. Kecman & I. Kopriva (Eds.), *Kernal based algorithms for mining huge data sets supervised, semi-supervised, and unsupervised learning* (pp. 663–677). Berlin Heidelberg: Springer.

[48] Huang, C. & Wang, C. (2006). A GA-based feature selection and parameters optimization for support vector machines. *Expert Systems with Applications*, *31*(2), 231–240.

[49] Pham, D., Muhamad, Z., Mahmuddin, M., Ghanbarzadeh, A., Koc, E., & Otri, S. (2007). Using the bees algorithm to optimise a support vector machine for wood defect classification. *Conference on Innovative Production Machines and Systems Virtual*.

[50] Lin, S., Lee, Z.-J., Chen, S.-C., & Tseng, T.-Y. (2008). Parameter determination of support vector machine and feature selection using simulated annealing approach. *Applied Soft Computing*, *8*(4), 1505–1512.

[51] Huang, C. & Dun, J. (2008). A distributed PSO-SVM hybrid system with feature selection and parameter optimization. *Applied Soft Computing*, *8*(4), 1381–1391.

[52] Li, X. & Chen, Z. (2010). Weed identification based on shape features and ant colony optimization algorithm. *Proceedings of the International Conference on Computer Application and System Modeling*, 22–24 October 2010, Taiyuan, 384–387.

[53] Ding, S. & Li, S. (2009). Clonal selection algorithm for feature selection and parameters optimization of support vector machines. *Proceedings of the 2nd International Symposium on Knowledge Acquisition and Modeling*, 30 November–1 December 2009, Wuhan, 17–20.

[54] Zhao, M., Fu, C., Ji, L., Tang, K., & Zhou, M. (2011). Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. *Expert System with Applications*, *38*(5), 5197–5204.

[55] Chen, H., Sheng-Han Z., Yue X., & Wen-Bing C. (2016). The study on hard landing prediction model with optimized parameter SVM method. *Proceedings of the 35th Chinese Control Conference*, pp. 4283–4287, July 27–29, 2016, Chengdu, China.

[56] Yunxiang L. & Jiongjun D. (2015). Parameter optimization of the SVM for big data. *Proceedings of the 8th International Symposium on Computational Intelligence and Design*, 12–13 Dec. 2015, pp. 341–344, Hangzhou, Zhejiang, China.

[57] Bhadra, T., Bandyopadhyay, S., & Maulik, U. (2012). Differential evolution based optimization of SVM parameters for meta classifier design. *Procedia Technology*, *4*, 50–57.

[58] Zavar, M., Rahati, S., Akbarzadeh-T, M.-R., & Ghasemifard, H. (2011). Evolutionary model selection in a wavelet-based Support Vector Machine for automated seizure detection. *Expert Systems with Applications*, *38*(9), 10751–10758.

[59] Hric, M., Chmulik, M., & Jarina, R. (2011). Model parameters selection for SVM classification using Particle Swarm Optimization. *Proceedings of The 21st International Conference of the Radioelektronika*, 19–20 April 2011, Brno, 1–4.

[60] Jin, Y., Hu, Y., Huang, J., & Zhang, J. (2011). Study on influences of model parameters on the performance of SVM. *Proceedings of The International Conference on Electrical and Control Engineering*, 16–18 September 2011, Yichang, 3667–3670.

[61] Liao, S., Yang, C., & Ding, L. (2011). Approximate parameter tuning of support vector machines. *Proceedings of the 3rd International Workshop on Intelligent System and Applications*, 28–29 May 2011, Wuhan, 1–4.

[62] Sun, J., Zheng, C., Li, X., & Zhou, Y. (2010). Analysis of the distance between two classes for tuning SVM hyperparameters. *Transaction on Neural Networks*, *21*(2), 305–318.

[63] Zhou, X. & Xu, J. (2009). A SVM model selection method based on hybrid genetic algorithm and empirical error minimization criterion. In H. Wang, Y. Shen, T. Huang, & Z. Zeng (Eds.), *The Sixth International Symposium on Neural Networks*, (pp. 245–253). Berlin Heidelberg: Springer.

[64] Zhang, Y. (2008). Evolutionary computation based automatic SVM model selection. *Proceedings of the 4th International Conference on Natural Computation*, 18–20 October 2008, Jinan, 66–70.

[65] Acir, N., Ozdamar, O. & Guzelis, C. (2006). Automatic classification of auditory brainstem responses using SVM-based selection algorithm for threshold detection. *Engineering Applications of Artificial Intelligence*, *19*(2), pp. 209-218.

[66] Wu, Z. & Li, C. (2006). Feature selection with transductive support vector machines. In I. Guyon, M. Nikravesh, S. Gunn & L. Zadeh (Eds), *Feature extraction foundation and applications* (pp. 325–341). Berlin Heidelberg: Springer.

[67] Feki, A., Ishak, A., & Feki, S. (2012). Feature selection using Bayesian and multiclass support vector machines approaches: application to bank risk prediction. *Expert Systems with Applications*, *39*(3), 3087–3099.

[68] Tan, J., Zhang, Z., Zhen, L., Zhang, C., & Deng, N. (2012). Adaptive feature selection via a new version of support vector machine. *Neural Computing & Applications*, *23*(3–4), 937–945.

[69] Hu, Q., Pan, W., Song, Y., & Yu, D. (2012). Large-margin feature selection for monotonic classification. *Knowledge-Based Systems*, *31*, 8–18.

[70] Xie, J. & Wang, C. (2011). Using Support Vector Machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases. *Expert Systems with Applications*, *38*(5), 5809–5815.

[71] Dunbar, M., Murray, J., Cysique, L., Brew, B. ,& Jeyakumar, V. (2010). Simultaneous classification and feature selection via convex quadratic programming with application to HIV-associated neurocognitive disorder assessment. *European Journal of Operational Research*, *206*(2), 470– 478.

[72] Malek S. & Abdellatif E. A. (2016). Feature selection and parameter optimization of support vector regression for electric load forecasting. *2nd International Conference on Electrical and Information Technologies ICEIT '2016*, pp. 288–293, 4–7 May 2016, Morocco.

[73] Wei, C. & Yuan H.-M. (2014). An improved GA-SVM algorithm. *Proceedings of the 9th IEEE Conference on Industrial Electronics and Applications*, 9–11 June 2014, pp. 2137–2141, Hangzhou, China.

[74] Reif, M., Shafait, F., & Dengel, A. (2012). Meta-learning for evolutionary parameter optimization of classifiers. *Machine Learning*, *87*(3), 357–380.

[75] Anguita, D. & Ghio, A. (2011). In-sample model selection for support vector machines. *Proceedings of The International Joint Conference on Neural Networks*, 31 July–5 August 2011, San Jose, CA, 1154–1161.

[76] Xusheng, G., Wei, C., & Yongxiang, L. (2010). Parameter selection for SVM using niches genetic algorithm in determined range. *Proceedings of the 2nd International Conference on Information Management and Engineering*, 16–18 April 2010, Chengdu, 492–496.

[77] Zong, Q., Liu, W., & Dou, L. (2006). Parameters selection for SVR based on PSO. *Proceedings of The 6th World Congress on Intelligent Control*, 21–23 July 2006, Dalian, China, 2811–2814.

[78] Kapp, M. N., Sabourin, R., & Maupin, P. (2012). A dynamic model selection strategy for support vector machine classifiers. *Applied Soft Computing*, *12*(8), 2550–2565.

[79] Glasmachers, T. & Igel, C. (2010). Maximum likelihood model selection for 1-norm soft margin SVMs with multiple parameters. *Transactions on Pattern Analysis and Machine Intelligence*, *32*(8), 1522–1528.

[80] Chapelle, O. & Vapnik, V. (2000). *Model selection for Support Vector Machines*. Retrieved January 1, 2011, from http://olivier.chapelle.cc/pub/nips99.pdf.

[81] Vapnik, V. & Chapelle, O. (2000). Bounds on error expectation for support vector machines. *Neural Computation*, *12*(9), 2013–2036.

[82] Saini, L., Aggarwal, S., & Kumar, A. (2010). Parameter optimization using genetic algorithm for support vector machine-based price-forecasting model in national electricity market. *IET Generation, Transmission & Distribution*, *4*(1), 36–49.

[83] Zhang, X., Chen, X., Zhang, Z., & He, Z. (2008). A grid–based ACO algorithm for parameters optimization in Support Vector Machines. *Proceedings of The International Conference on General Computing*, 26–28 August 2008, Hangzhou, 805–808.

[84] Dong, Y., Xia, Z., Tu, M., & Xing, G. (2007). An optimization method for selecting parameters in support vector machines. *Proceedings of the 6th International Conference on Machine Learning and Applications*, 13–15 December 2007, Cincinnati, OH, 1–6.

[85] Imbault, F. & Lebart, K. (2004). A stochastic optimization approach for parameter tuning of support vector machines. *Proceedings of the 17th International Conference on Pattern Recognition*, 23–26 August 2004, 597–600.

[86] Lázaro-Gredilla, M., Gómez-Verdejo, V., & Parrado-Hernández. Low-cost model selection for SVMs using local features. (2012). *Engineering Application of Artificial Intelligence*, *25*(6), 1203–1211.

[87] Liu, Q., Chen, C., Zhang, Y., & Hu, Z. (2011). Feature selection for support vector machines with RBF kernel. *Artificial Intelligence Review*, *36*(2), 99–115.

[88] Tang, Y., Guo, W., & Gao, J. (2009). Efficient model selection for support vector machine with Gaussian kernel function. *Proceedings of The Symposium on Computational Intelligence and Data Mining*, 30 March–2 April 2009, Nashville, TN, 40–45.

[89] Liu, A., Cao, H., Chen, X., He, Z., & Shen, Z. (2013). Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings. *Neurocomputing*, *99*, 399–410.

[90] Bing X., Mengjie Z., Will N. B., & Xin Y. (2015). *A survey on evolutionary computation approaches to feature selection*. Accepted for Publication in IEEE Transactions on Evolutionary Computation Journal.

[91] Amit K. S. & Vimal K. D. (2015). A survey on feature selection algorithms, *International Journal on Recent and Innovation Trends in Computing and Communication*, *3*(4), pp.1895–1899.

[92] Sivagaminathan, R. & Ramakrishnan, S. (2007). A hybrid approach for feature subset selection using neural networks and ant colony optimization. *Expert Systems with Applications*, *33*(1), 49–60.

[93] Huang, C. (2009). ACO-based hybrid classification system with feature subset selection and model parameters optimization. *Neurocomputing*, *73*(1–3), 438–448.

[94] Vieira, S., Sousa, J., & Runkler, T. (2007). Ant colony optimization applied to feature selection in fuzzy classifiers. In P. Melin, O. Castillo, L. Aguilar, J. Kacprzyk & W. Pedrycz (Eds.), *Foundations of fuzzy logic and soft computing* (pp. 778–788). Berlin Heidelberg: Springer.

[95] Kabir, M., Shahjahan, M., & Murase, K. (2009). An efficient feature selection using ant colony optimization algorithm. In C. Leung & J. Chan (Eds.), *Neural information processing*, (pp. 242–252). Berlin Heidelberg: Springer.

[96] Kabir, M., Shahjahan, M., & Murase, K. (2012). A new hybrid ant colony optimization algorithm for feature selection. *Expert Systems with Applications*, *39*(3), pp. 3747–3763.

[97] Abd-Alsabour, N. & Randall, M. (2010). Feature selection for classification using an ant colony system. *Proceedings of the 6th International Conference on e-Science Workshops*, 7–10 December 2010, Brisbane, QLD, 86–91.

[98] Sanz, S., Cumplido, M., Cruz, F., & Calzon, C. (2002). Feature selection via genetic optimization. In J. Dorronsoro (Ed.), *Artificial Neural Networks*, (pp. 547–552). Berlin Heidelberg: Springer.

[99] Frohlich, H., Chapelle, O., & Scholkopf, B. (2003). Feature selection for support vector machines by means of genetic algorithms. *Proceedings of the 15th International Conference on Tools with Artificial Intelligence*, 3–5 November 2003, 142–148.

[100] Moustakidis, S. & Theocharis, J. (2010). SVM-FuzCoC: A novel SVM-based feature selection method using a fuzzy complementary criterion. *Pattern Recognition*, *43*(11), 3712–3729.

[101] Zhang, H., Xiang, M., Ma, C., Huang, Q., Li, W., Xie, W., Wei, Y., & Yang, S. (2009). Three-class classification models of LogS and LogP derived by using GA-CG-SVM approach. *Molecular Diversity*, *13*(2), 261–268.

[102] Garde, A., Voss, A., Caminal, P., & Benito, S. (2013). SVM-based feature selection to optimize sensitivity–specificity balance applied to weaning. *Computers in Biology and Medicine*, *43*(5), 533–540.

[103] Lin, X., Yang, F., Zhou, L., Yin, P., Kong, H. Xing, W., Li, X., Jia, L., Wang, Q., & Xu, G. (2012). A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information. *Journal of Chromatography B*, *910*, 149–155.

[104] Atienza, F., Alvarez, J., Munoz, A., Vinagre, J., Alberola, A., & Valls, G. (2012). Feature selection using support vector machines and bootstrap methods for ventricular fibrillation detection. *Expert System with Applications*, *39*(2), pp. 1956–1967.

[105] Unler, A., Murat, A., & Chinnam, R. (2011). mr^{2}PSO: A maximum relevance minimum redundancy feature selection method based on Swarm Intelligence for Support Vector Machine classification. *Information Sciences*, *181*(20), 4625–4641.

[106] Yunxiang L. & Jiongjun D. (2015). Parameter optimization of the SVM for big data. *Proceedings of the 8th International Symposium on Computational Intelligence and Design*, 12–13 Dec. 2015, pp. 341–344, Hangzhou, Zhejiang, China.

[107] Akay, M. (2009). Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Systems with Applications*, *36*(2, Part 2), 3240–3247.

[108] Polat, K. & Güneş, S. (2009). A new feature selection method on classification of medical datasets: kernel F-score feature selection. *Expert Systems with Applications*, *36*(7), 10367–10373.

[109] Maldonado, S. & Weber, R. (2009). A wrapper method for feature selection using Support Vector Machines. *Information Sciences*, *179*(13), 2208–2217.

[110] Zhang, J., Chen, W., Zhong, J., Tan, X., & Li, Y. (2009). Continuous function optimization using hybrid ant colony approach with orthogonal design scheme. In Tzai-Der W. (Ed.), *Simulated Evolution and Learning, The Sixth International Symposium on Neural Networks (ISNN 2009)* (pp. 126–133). Berlin Heidelberg: Springer.

[111] Nguyen, M. & Torre, F. (2010). Optimal feature selection for support vector machines. *Pattern Recognition*, *43*(3), 584–591.

[112] Peng, Y., Wu, Z. & Jiang, Z. (2010). A novel feature selection approach for biomedical data classification. *Journal of Biomedical Informatics*, *43*(1), pp. 15–23.
[113] Schaffernicht, E., Stephan, V., & Groß, H. (2007). An efficient search strategy for feature selection using chow-liu trees. In J. de Sá, L. Alexandre, W. Duch, & D. Mandic (Eds.). *Artificial Neural Networks* (pp. 190–199). Berlin Heidelberg: Springer.
[114] Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., & Vapnik, V. (2001). *Feature selection for SVMs*. Retrieved December 20, 2010, from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.102.7476.
[115] Basiri, M., Aghaee, N., & Aghdam, M. (2008). Using ant colony optimization-based selected features for predicting post-synaptic activity in proteins. In E. Marchiori & J. Moore (Eds.), *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*, (pp. 12–23). Berlin Heidelberg: Springer.
[116] Kanan, H., Faez, K., & Taheri, S. (2007). Feature selection using ant colony optimization (ACO): A new method and comparative study in the application of face recognition system. In P. Perner (Ed), *Advances in Data Mining. Theoretical Aspects and Applications*, (pp. 63–76). Berlin Heidelberg: Springer.
[117] Foithong, S., Pinngern, O., & Attachoo, B. (2012). Feature subset selection wrapper based on mutual information and rough sets. *Expert Systems with Applications*, *39*(1), 574–584.
[118] Lian, H. (2012). On feature selection with principle component analysis for one-class SVM. *Pattern Recognition Letters*, *33*(9), 1027–1031.
[119] Su, C. & Yang, C. (2008). Feature selection for the SVM: An application to hypertension diagnosis. *Expert Systems with Applications*, *34*(1), 754–763.
[120] Liu, Y. & Zheng, F. (2006). FS-SFS: A novel feature selection method for support vector machines. *Pattern Recognition*, *39*(7), 1333–1345.
[121] Hall, M. (2000). *Feature selection for discrete and numeric class machine learning*. Retrieved December 20, 2010, from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.148.6025.
[122] Oliveira, L., Sabourin, R., Bortolozzi, F., & Suen, C. (2003). A methodology for feature selection using multi-objective genetic algorithms for handwritten digit string recognition. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, *17*(6), 1–35.
[123] Hidalgo-Muñoz, A.R., López, M.M., Santos, I.M., Pereira, A.T., Vázquez-Marrufo, M., Galvao-Carmona, A., & Tomé, A.M. (2013). Application of SVM-RFE on EEG signals for detecting the most relevant scalp regions linked to affective valence processing. *Expert Systems with Applications*, *40*(6), 2101–2108.
[124] Garbarine, E., DePasquale, J., Gadia, V., Polikar, R., & Rosen, G. (2011). Information-theoretic approaches to SVM feature selection for metagenome read classification. *Computational Biology and Chemistry*, *35*(3), 199–209.
[125] Ye F. (2016). Simultaneous support vector selection and parameter optimization using Support Vector Machines for sentiment classification. *Proceedings of the 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pp. 59–62, 26–28 Aug. 2016.
[126] Abdi, M. & Giveki, D. (2013). Automatic detection of erythemato-squamous diseases using PSO–SVM based on association rules. *Engineering Applications of Artificial Intelligence*, *26*(1), 603–608.
[127] Azadeh, A., Saberi, M., Kazem, A., Ebrahimipour, V., Nourmohammadzadeh, A., & Saberi, Z. (2013). A flexible algorithm for fault diagnosis in a centrifugal pump with corrupted data and noise based on ANN and support vector machine with hyper-parameters optimization. *Applied Soft Computing*, *13*(3), 1478–1485.
[128] Lei, L. & Qiao, G. (2012). Text categorization using SVM with exponent weighted ACO. *Proceedings of the 31st Chinese Control Conference*, 25–27 July 2012, Hefei, 3763–3768.
[129] Wang, S. & Meng, B. (2011). Parameter selection algorithm for support vector machine. *Procedia Environmental Sciences*, *11*(Part B), pp. 538–544, 2011.
[130] Sartakhti, J. S., Zangooei, M. H., & Mozafari, K. (2011). Hepatitis disease diagnosis using a novel hybrid method based on Support Vector Machine and Simulated Annealing (SVM-SA). *Computer Methods and Programs in Biomedicine*, *108*(2), 570–579.
[131] Samadzadegan, F., Soleymani, A., & Abbaspour, A. (2010). Evaluation of genetic algorithms for tuning SVM parameters in multi-class problems. *Proceedings of the 11th International Symposium on Computational Intelligence and Informatics*, 18–20 November 2010, Budapest, 323–328.
[132] Fang, X. & Bai, T. (2009). Share price prediction using wavelet transform and ant colony algorithm for parameters optimization in SVM. *Proceedings of The Global Congress on Intelligent Systems*, 19–21 May 2009, Xiamen, 288–292.
[133] Zhou, J., Zhang, A., & Bai, T. (2008). Client classification on credit risk using rough set theory and ACO-based support vector machine. *Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing*, 12–14 October 2008, Dalian, 1–4.
[134] Wang, R. (2012). AdaBoost for feature selection, classification and its relation with SVM: A review. *Physics Procedia*, *25*, 800–807.
[135] Verleysen, M., Rossi, F., & Francois, D. (2009). Advances in feature selection with mutual information. In M. Biehl, B. Hammer, M. Verleysen & T. Villmann (Eds.), *Similarity-based clustering recent developments and biomedical applications* (pp. 52–69). Berlin Heidelberg: Springer.

[136] Jensen, R. & Shen, Q. (2004). Fuzzy-rough data reduction with ant colony optimization. *Fuzzy Sets and* Systems, *149*(1), 5–20.

[137] Partridge, D., Wang, W., & Jones, P. (2001). Efficient and effective feature selection in the presence of feature interaction and noise. In S. Singh, N. Murshed & W. Kropatsch (Eds.), *Advances in pattern recognition* (pp. 196–204), Berlin Heidelberg: Springer.

[138] Shahjahan, M. & Murase, K. (2009). An efficient feature selection using ant colony optimization algorithm. In C. Leung & J. Chan (Eds.), *Neural information processing* (pp. 242–252). Berlin Heidelberg: Springer.

[139] Xiong, S., Wang, J.-Y., & Lin, H. (2010). Hybrid feature selection algorithm based on dynamic weighted ant colony algorithm. *Proceedings of the 9th International Conference on Machine Learning and Cybernetics*, 11–14 July 2010, Qingdao, 518–522.

[140] Mesleh, A. & Kanaan, G. (2008). Support Vector Machines text classification system: Using Ant Colony Optimization based feature subset selection. *Proceedings of The International Conference on Computer Engineering & Systems*, 25–27 November 2008, Cairo, Egypt, 143–148.

[141] H.B. Alwan & K.R. Ku-Mahamud (2017). Mixed-variable ant colony optimisation algorithm for feature subset selection and tuning support vector machine parameter, *Int. J. Bio-Inspired Computation, 9*(1), 53-63.

[142] Mohammed A., Salih T., & Mohammad T. K. (2017). A new hybrid approach for feature selection and Support Vector Machine model selection based on self-adaptive cohort intelligence", *Expert Systems with Applications*, *88*(2017), pp. 118–131.

[143] Sarafrazi, S. & Nezamabadi-pour, H. (2013). Facing the classification of binary problems with a GSA-SVM hybrid system. *Mathematical and Computer Modelling*, *57*(1–2), 270–278.

[144] Han, H., Hou-Jun, W. & Xiucheng, D. (2011). Transformer fault dignosis based on feature selection and parameter optimization. *Energy Procedia*, *12*, 662–668.

[145] Lin, K. & Chien, H. (2009). CSO-based feature selection and parameter optimization for support vector machine. *Proceedings of The Joint Conference on Pervasive Computing*, 3–5 December 2009, Tamsui, Taipei, 783–788.

[146] Lin, S., Ying, K., Chen, S., & Lee, Z. (2008). Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications*, *35*(4), 1817–1824.

[147] Vieira, S., Mendonça, L., Farinha, G., & Sousa, J. (2013). Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients. *Applied Soft Computing*, *13*(8), 3494–3504.

[148] Abd-Alsabour, N. (2010). Feature selection for classification using an ant system approach. In M. Hinchey, B. Kleinjohann, L. Kleinjohann, P. Lindsay, F. Ramming, J. Timmis, & M. Wolf (Eds.), *Distributed, Parallel and Biologically Inspired Systems*, (pp. 233–241). Berlin Heidelberg: Springer.

[149] Maldonado, S., Weber, R., & Basak, J. (2011). Simultaneous feature selection and classification using kernel-penalized Support Vector Machines. *Information Sciences*, *181*(1), 115–128.