© Universiti Tun Hussein Onn Malaysia Publisher's Office

IJIE

Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/ijie
ISSN : 2229-838X e-ISSN : 2600-7916

The International Journal of Integrated Engineering

An Efficient Model for Data Classification Based on SVM Grid Parameter Optimization and PSO Feature Weight Selection

Ahmed Hussein Ali[1*], Mahmood Zaki Abdullah[2]

[1]ICCI, Informatics Institute for Postgraduate Studies, Baghdad, IRAQ

[2]Computer Engineering, College of Engineering, Mustansiriyah University, Baghdad, IRAQ

*Corresponding Author

DOI: https://doi.org/10.30880/ijie.2020.12.01.001
Received 1 July 2018; Accepted 8 October 2018; Available online 31 January 2020

Abstract: The support vector machine (SVM) is a classifier with different applications due to its perfect experimental performance compared to other machine learning algorithms. It has been used mostly in pattern recognition, fault diagnosis, and text categorization. The performance of SVM is extremely dependent on the sufficient setting of its parameters such as SVM max-iteration, SVM kernel-type, gamma and C value. Therefore, the choice of suitable initial parameters of SVM will result in a good performance and classification result. This paper introduces a new schema for optimizing SVM parameters using grid search and particle swarm optimization PSO feature weighting. The experimental results demonstrate that the new method obtained a high accuracy compared to the traditional SVM and other SVM-optimization methods.

Keywords: Machine learning, Support vector machine, Optimization; PSO, Grid search

1. Introduction

Classification refers to the learning of a target function f (also known as a classification model) that can map each attribute set x to any of the predefined class labels y [1]. The classification of different categories of large data sets has become one of the most important computing and machine learning problems of this decade. It has wide applications, including identifying gene sequences in DNA, processing credit-card applications, face recognition, character recognition, and drug recovery. The main objective of classification is to identify patterns in a large data set which can help to analyze the data to make decisions and identify new concepts hidden in large and inconsistent data [2, 3]. Learning models can be classified as batch learning and online learning. Batch learning is when the learning algorithm is provided with all the data prior to the learning process, while in online learning, the algorithm is fed with the examples one at a time, then, it estimates the output before being fed with the correct value. Generalization refers to the ability of a hypothesis provide a precise classification of a data that is not contained in the training set, and the major aim is to optimize this feature. The SVM has shown better generalization properties than other learning methods [4]. The general approach towards solving classification problems is depicted in Fig. 1. First, a set of training data is required; this set of data is made up of known record class label. The training dataset is used to establish the classification model. The established classification model is applied to the test dataset which is made up of an unknown class label record. The data

*Corresponding author: ahmed.ali@aliraqia.edu.iq

structural model is randomly divided into testing and training sets. In the training phase, the training dataset is used to analyze the tuples from the property description database for the model construction. Such models consist of the Decision tree, classification rules, and mathematical formula.

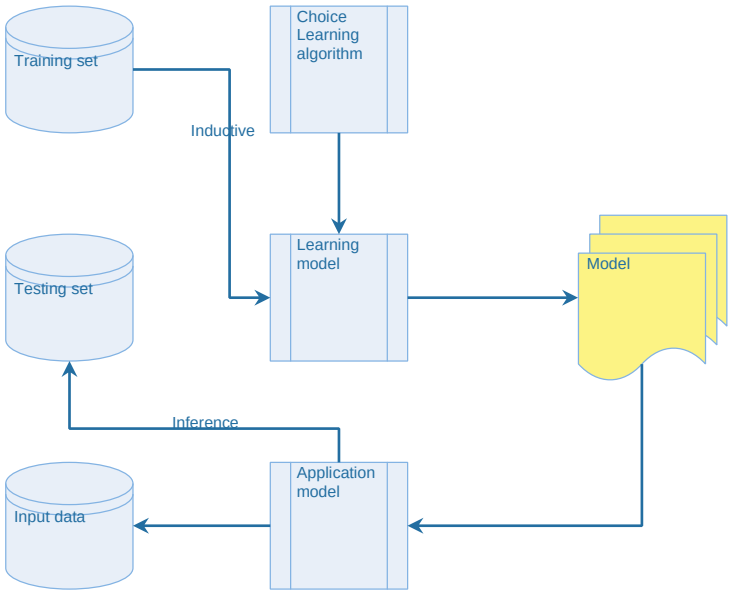


Fig. 1 - Classification method

The testing phase involves the use of the test dataset to check the models' classification accuracy. Hyper-parameter setting has a significant impact on the performance of a given machine learning algorithm. This paper introduces PGSVM a new schema for optimizing SVM parameters using grid search and particle swarm optimization PSO feature weighting. This model can be used to classify the tuples of other datasets. The proposed method in this study was applied on eight standard datasets from UCI machine learning repository.

This paper is organized in sections; section 2 discussed the used classification and optimization algorithms (SVM, PSO, and Grid search). Section 3 present the related work while the proposed methodology and results analysis discussed in section 4. Section 5 presented the conclusion.

2. Basic Principle of SVM and Optimization Algorithms

2.1. SVM Data Classification

The SVM was developed in the mid-1990s from the statistical learning theory of Boser, Guyon, and Vapnik[5]. This theory stipulates that the bounds on the generalization error rate of a learning machine can be obtained when applied on an unseen data using a linear classification method. The SVM belongs to the supervised learning method. Vapnik showed the possibility of minimizing the bounds on the generalization error rate of a learning by maximizing the class separation margins. Hence, a good classification accuracy on an unseen dataset can be achieved, thus, resulting in a high generalization. But, since the class separation margin does not depend on the data dimensionality, the classification performance of a highly dimensional dataset can be improved. Therefore, during the design of the SVM, the major aim is to maximize the separation margin between the classes of interest and introduce a linear separating hyperplane between them. In such way, the classification accuracy of SVM can be enhanced, and with an improved generalization, capacity compared to the other classifiers that only aim to reduce the training error rate, like the neural network classifiers. The SVM can also adapt itself and assume a nonlinear classifier configuration by simply spreading out the data through mapping into a higher dimensional feature space before applying SVM. This is similar to the application of a nonlinear classifier to the original input space. The effect of Hughes phenomenon can be minimized with the SVM since it can use a limited number of training samples to adequately classify a dataset in a higher dimensional feature space. The SVM has been recently applied to several classification tasks, such as text categorization and detection, face detection, digit recognition, and hand-written character. The SVM was initially proposed by Vapnik and Lemer [6] as a large margin classifier but has now been extended through improvements to the non-linear case [7]. Meanwhile, the SVM does not only aim at class separation (just as the Perceptron algorithm), it also maximizes the separating margin between two classes, as shown in Fig. 2.

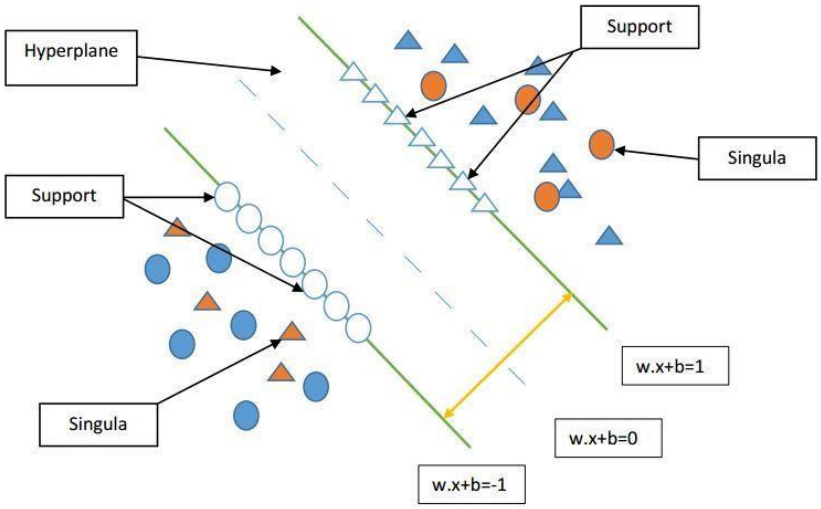


Fig. 2 - SVM hyperplane

2.2. Grid Search Optimization

Several studies have been conducted on the selection of kernel parameters [8, 9]. These studies have proposed methods such as the grid search method and its variant, as well as the bilinear method for kernel parameters selection. The PSO and grid search are the two methods that often give the best values of the kernel parameters (C,γ) as earlier reported [10]. The grid search method takes the combined values of M and N for C and γ (M*N for C,γ) and trains them respectively before estimating the promote recognition rate. The optimal parameter will be selected based on the value with the maximum recognition rate in M*N combinations of C,γ [11, 12]. The grid search method usually presents a high learning accuracy, but its larger computation takes much time. The grid search is more preferable because it can parallelly participate in the training of every SVM since they do not depend on each other. Hence, it takes much time to execute a complete grid search. Hsu et al [13] suggested an improved grid search method with the basic idea of first getting the optimal values of C and γ using a large step combination of C and γ before conducting a detailed grid search within a range close to the C and γ values. The bilinear grid search method was suggested by Li et al [14] with the basic idea of getting the optimal combination of parameters that can give an improved bilinear grid search before searching for the optimum values in a certain range close to the optimal parameter combination.

2.3. Optimize Weights (PSO)

The recent studies on swarm intelligence have focused on computational intelligence approaches [15] that are mainly inspired by the psychosocial behavior of the human or animal groups [16, 17]. Swarm intelligence refers to the study of computational systems comprised of relatively simple agents that moves and locally interact within a simulated space to achieve system self-organization. Their design is typically inspired by social behavior of social animals. The PSO [18] imitates the flocking behavior of birds and their means of information exchange. The Ant colony is inspired by the interaction of individual ants with others through their pheromone trails. In the basic PSO [11], each particle is initialized to a random position in a solution space. A solution space for an abstract problem is the set of all possible values or scenarios that can be generated from the inputs. The positions and velocities are updated using the relation:

$$V = wv + c_p r_1 (pbest - x) + c_g r_2 (gbest - x) \quad (1)$$

$$X = x + v \quad (2)$$

where v and x represent the particles' velocity and position, pbest is the particles' ever-achieved best position, while gbest is the best-achieved position among the whole particles. r1 and r2 are randomly generated numbers in the range of (0,1), and w represents the coefficients (personal, social, and momentum) with constant predefined values.

3. Related Work

There are several optimization approaches with SVM proposed by the research community. Among the existing efforts on optimization hyper-parameters and feature, Yingsheng Ji et al[4] develop EnKF method to optimize hyperparameters and features for support vector machine algorithm. Miranda et al[19] proposed a new hybrid meta-learning architecture for multi-objective optimization of SVM parameters. Kumar Jain et al[20] develop a novel method for optimizing SVM hyper-parameter using self-organizing map.

4. Proposed Methodology and Results Analysis

Cross-validation is a statistical approach for the verification of a classifiers' performance. During the process, the original data is divided into two sets; one set is used as the training set while the other part is used as a validation set. The classifier is first trained with the training set before using the validation set to test the model for classification accuracy.

4.1.Model

The training and testing sets are first extracted from the original dataset before a pre-processing step. The appropriate features were selected with the PSO based on the weight of the features, see Table 1.

Table 1 - Features weight (PSO)

Attribute	Weight
Attribute_41	0.984
Attribute_59	0.974
Attribute_8	0.968
Attribute_10	0.937
Attribute_26	0.901
Attribute_60	0.883
Attribute_48	0.873
Attribute_57	0.865
Attribute_2	0.863
Attribute_28	0.808
Attribute_33	0.795
Attribute_37	0.795
:	:

The standard grid search method was used to optimize the parameters of SVM see Fig.3 and Fig.4. Based on Matlab environment, the effectiveness of the improved method was evaluated and compared to the accuracy of the latest methods. The SVM was trained with the training dataset before using the models to predict the category labels of the test set. The proposed PGSVM model is shown in Fig. 5. The two main steps that proposed to improve SVM classification algorithm are:

First, the feature weighting is important to eliminate the irrelevant features or features with little weight. The PSO feature weighting allow us to select the suitable features or feature with high impact. PSO feature weighting starts with randomize initialization of particles (population). The features (attributes) of the selected dataset act as a particle of PSO algorithm. The position of particles depends on min-max value of each attribute of the dataset. When each value of the algorithm initialized, then the evaluation function starts to evaluate each particle in PSO to find pbest value.

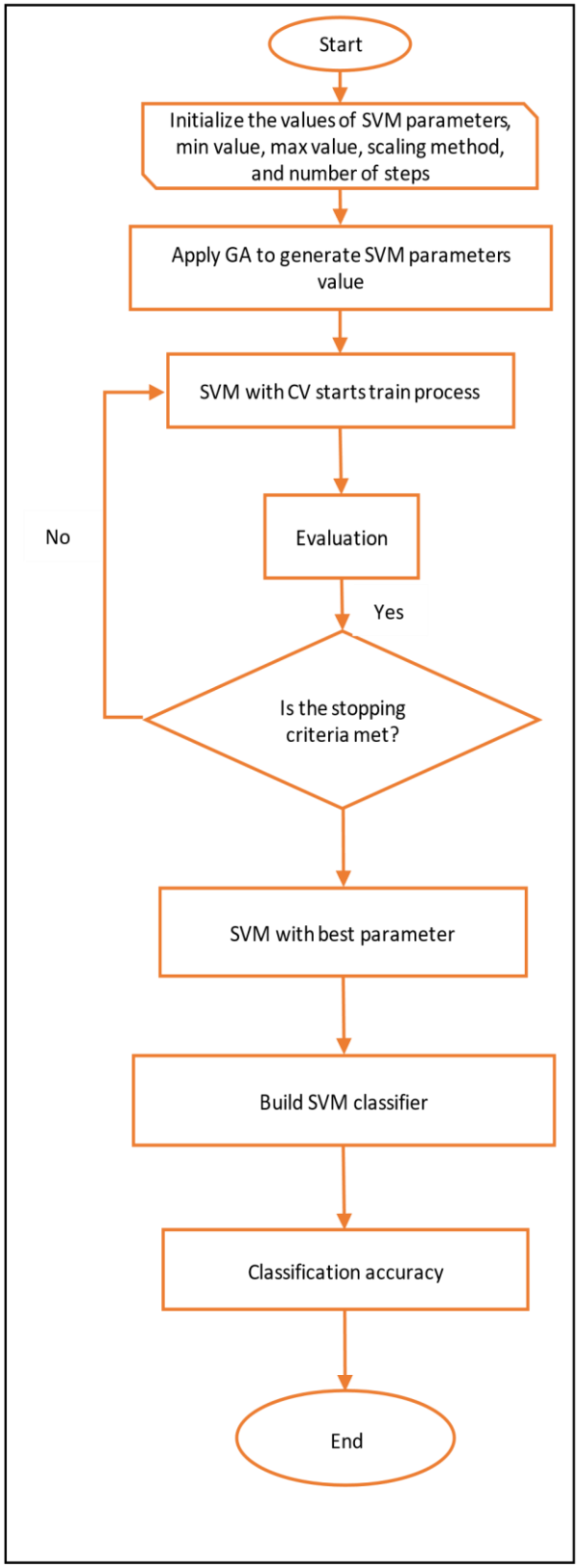


Fig. 3 - optimize SVM parameter by GA

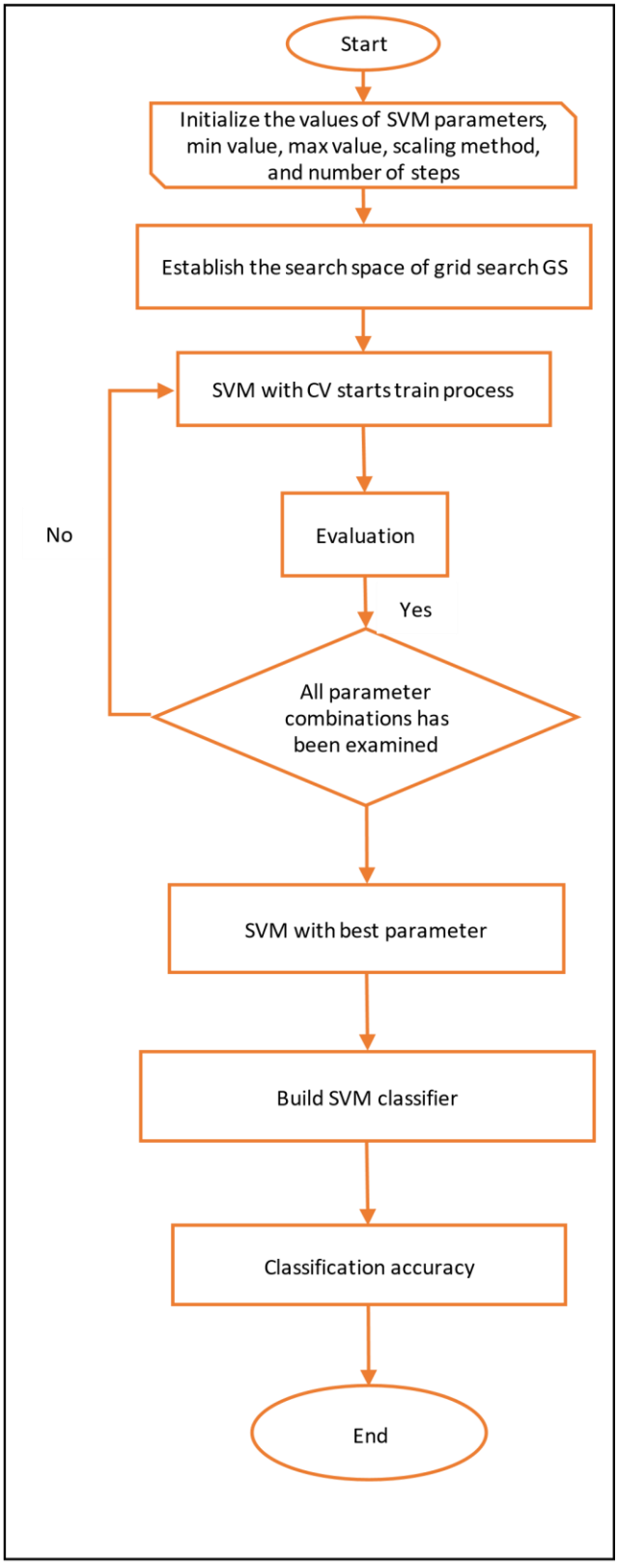


Fig. 4 - optimize SVM parameter by GS

After obtained the pbest then the gbest value can be calculated accurately. So the particle position still updated until reach the termination case. Finally, the PSO phase yielded a set of dataset features labeled by a value called feature weight. The output feature weight ranged from [0 to 1]. The feature with value (0) means that the feature irrelevant for predication and its impact is zero. In contrast the value (1) means the feature is very important for prediction and its impact is very high, see Table 1.

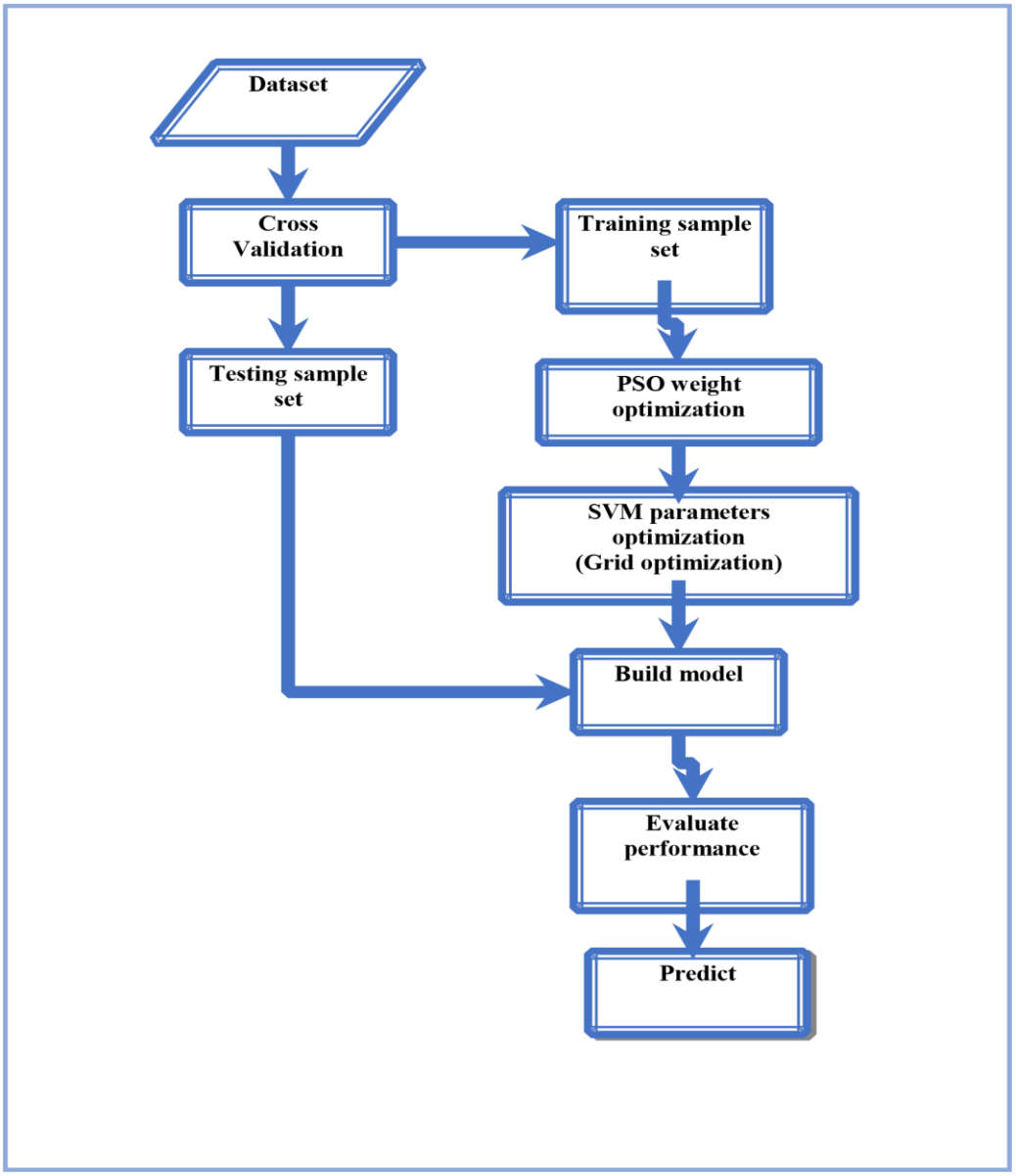


Fig. 5 - The proposed PGSVM model

Second, to prevent the support vector machine can be categorized into two main steps:

1. Diversion of input space to higher dimension output space using nonlinear TF ø.
2. Build the best separating plane of the output space. The best separating plane can be generated by proper kernel function. The SVM common kernel functions include:

Linear kernel function:

$$K(x_i, x_j) = x_i^T . x_j \quad (3)$$

Polynomial kernel function:

$$K(x_i, x_j) = (x_i^T . x_j + 1)^d \quad (4)$$

RBF kernel function:

$$K(x_i, x_j) = exp(-\| x_i . x_j \|^2 / \sigma^2) \quad (5)$$

Sigmoid kernel function:

$$K(x_i, x_j) = \tanh[b. (x_i^T . x_j) + \theta] \quad (6)$$

The impact of kernel function selection and hyperparameter optimization directly affected the SVM performance and overfitting phenomena. So utilizing grid search in this phase will prevent overfitting and generate the optimal separating plane by get the proper kernel. The proposed algorithm after cross validation will initialize the values of SVM parameter, min-max value, scaling method, and the number of steps. Then the search space established in grid form. The support vector machine in cross validation CV starts train on the data. In each training step, we made accuracy evaluation, so if all parameter combinations have been examined then the best kernel parameter will build the SVM. Finally, classification accuracy of the obtained SVM will be evaluated.

4.2. Dataset

The proposed method in this paper was applied on SONAR dataset obtained from Github repository. The dataset falls under a multivariate category as consists of 209 samples and two class, with 60 features measured from each sample. To ensure the validity of our proposed model, we made the experiment with additional seven datasets. Table. 2 describe the collection of datasets.

Table 2 - Dataset description

Dataset	No. of Attributes	No. of Objects
Cryotherapy	6	90
Wine	13	179
Immunotherapy	7	91
Sonar	60	208
Iris	5	150
Transfusion	4	749
Haberman	3	306
Wifi_localization	7	2000

4.3. Results Comparison

In this section, the standard SVM, Grid-SVM, and the proposed PGSVM algorithm were evaluated to verify the need for parameters tuning, as well as the effect of feature reduction on the PGSVM algorithm. The evaluation results are presented in Table 4. Table 4 present a comparison between the proposed approach with native SVM and GridSVM approaches in terms of main machine learning statistical metrics. Without parameter tuning, the performance of SVM degenerated severely. The main contribution of this work is to adopt the robust and efficient PGSVM algorithm for parameter optimization and feature selection in order to obtain the maximum SVM performance in classification tasks.

The proposed algorithm evaluated using 8 datasets from UCI to appraise the performance. The performance evaluation metrics used in this experiment are kappa statistics, mean squared error, accuracy, f-measure, recall (sensitivity), selectivity and precision. Which be seen in table 3.

Table 3 - Classification performance measurement

Measure	Formula
Precision	$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$
Recall / Sensitivity	$Recall/Sensitivity = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$
Selectivity	$Selectivity = \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}}$
Accuracy	$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$
F-Measure	$F - Measure = \frac{2 * \mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$

Experiments were performed on an Intel® CORE8 CPU HQ processor,16G RAM and 64-bit operating system. Generally, C and γ are the most important hyperparameter affected on the fitting of SVM. We use CV-K fold set k=10, to split the given dataset into k subset. To ensure the accuracy, we use CV accuracy as the following formula:

$$CV_{accuracy} = \frac{\sum_i test_{accuracy}}{\mathrm{K}} \quad (7)$$

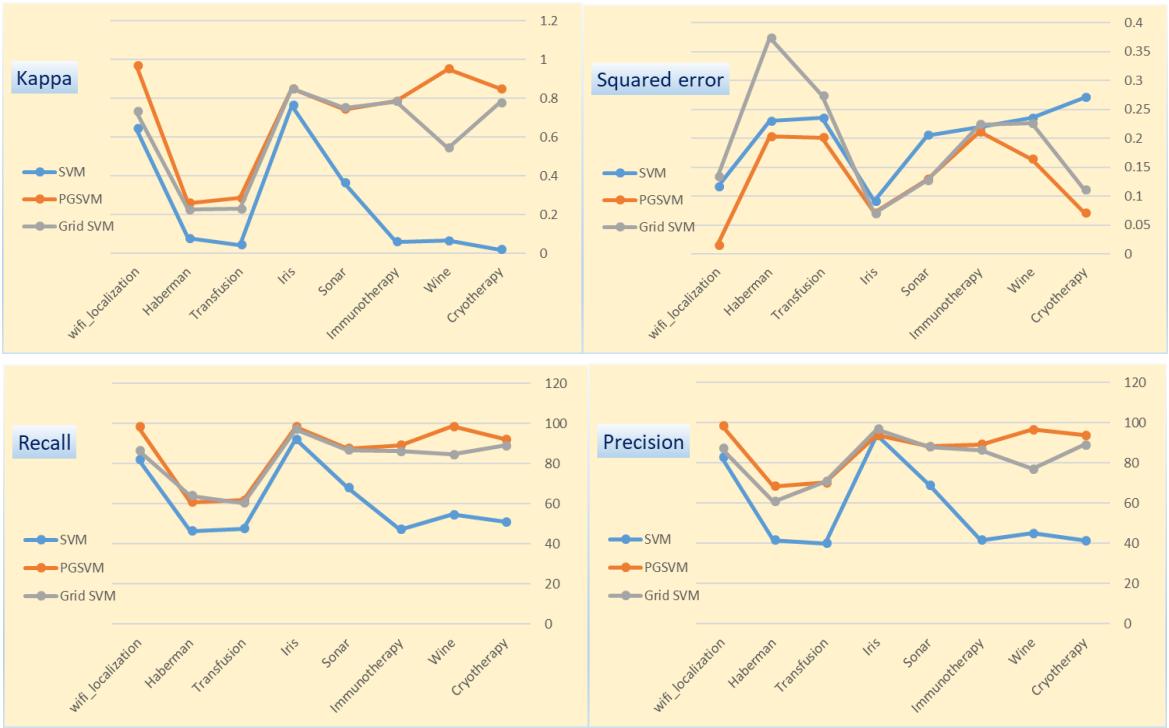
$$test_{accuracy} = \frac{\text{no. of predicted data}}{\text{total testing data}} * 100\% \quad (8)$$

The three set of kernel optimized are linear kernel that have one parameter to optimized in our algorithm, this one parameter is C. the second set is sigmoid kernel and RBF kernel with two hyperparameter C and γ. The third set is polynomial kernel can be optimized with three hyperparameter C, γ, and degree.

Table 4 - Comparison between the proposed approach with native SVM and Grid-SVM approaches in terms of machine learning statistical metrics

Data set	The Statistical Metrics	SVM	Proposed PGSVM	Grid-SVM
Cryotherapy	Accuracy	50.00	92.59	88.89
	kappa	0.020	0.851	0.778
	weighted_mean_recall	51.04	92.31	89.01
	weighted_mean_precision	41.38	93.75	89.01
	squared_error	0.271	0.071	0.111
Wine	Accuracy	67.38	98.11	77.36
	kappa	0.066	0.953	0.543
	weighted_mean_recall	54.63	98.72	84.62
	weighted_mean_precision	44.88	96.67	76.92
	squared_error	0.235	0.164	0.226
Immunotherapy	Accuracy	67.78	92.59	90.56
	kappa	0.059	0.786	0.785
	weighted_mean_recall	47.16	89.29	86.12
	weighted_mean_precision	41.57	89.29	86.34
	squared_error	0.219	0.211	0.224
Sonar	Accuracy	68.80	87.10	86.32
	kappa	0.364	0.744	0.753
	weighted_mean_recall	68.01	87.67	86.86

	weighted_mean_precision	69.09	88.15	88.01
	squared_error	0.205	0.129	0.127
Iris	Accuracy	95.76	99.87	99.87
	kappa	0.764	0.851	0.851
	weighted_mean_recall	92.31	98.43	97.31
	weighted_mean_precision	93.75	93.79	96.75
	squared_error	0.091	0.071	0.070
Transfusion	Accuracy	69.14	78.57	78.00
	kappa	0.043	0.286	0.230
	weighted_mean_recall	47.63	61.88	60.33
	weighted_mean_precision	40.07	70.24	70.87
	squared_error	0.235	0.201	0.274
Haberman	Accuracy	63.64	75.82	62.64
	kappa	0.078	0.259	0.225
	weighted_mean_recall	46.31	60.85	63.93
	weighted_mean_precision	41.48	68.41	60.88
	squared_error	0.230	0.203	0.374
wifi_localization	Accuracy	82.14	98.50	86.67
	kappa	0.643	0.970	0.733
	weighted_mean_recall	82.15	98.50	86.67
	weighted_mean_precision	82.76	98.53	87.26
	squared_error	0.116	0.015	0.133



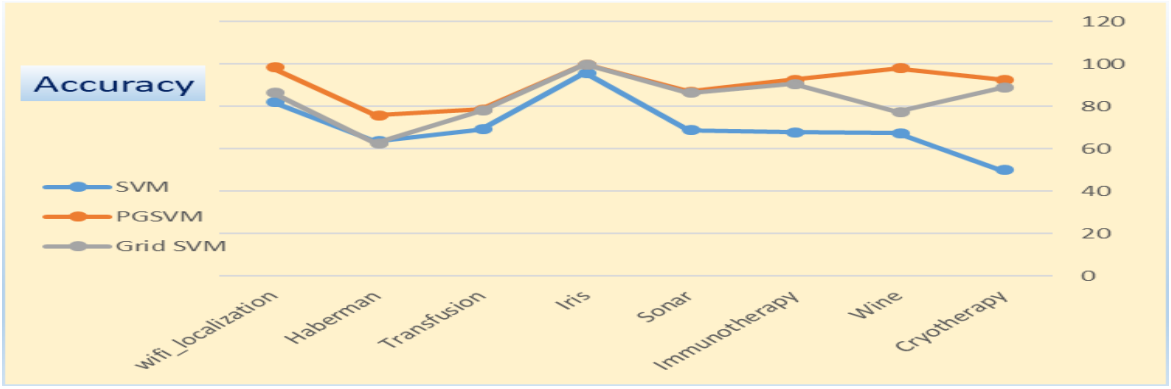


Fig. 6 - Obtained results comparison

Table 5 - Comparison between the proposed approach and GA in term of accuracy

Dataset	Proposed PGSVM	Genetic Algorithm GA
Cryotherapy	92.59	85.19
Wine	98.11	69.81
Immunotherapy	92.59	59.26
Sonar	87.10	83.87
Iris	99.87	99.87
Transfusion	78.57	41.96
Haberman	75.82	59.34
Wifi_localization	98.50	87.33

Table 6 - Comparison between the proposed approach and GA in term of squared error

Dataset	Proposed PGSVM	Genetic Algorithm GA
Cryotherapy	0.071	0.148
Wine	0.164	0.302
Immunotherapy	0.211	0.407
Sonar	0.129	0.147
Iris	0.071	0.070
Transfusion	0.201	0.580
Haberman	0.203	0.407
Wifi_localization	0.115	0.127

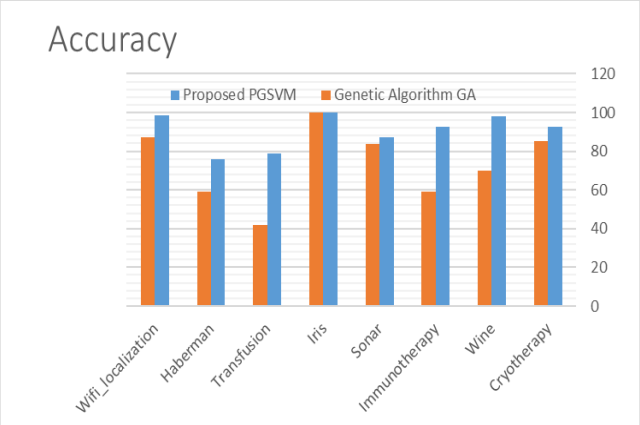


Fig. 7 - Table 2. Comparison between the proposed approach and GA in term of accuracy

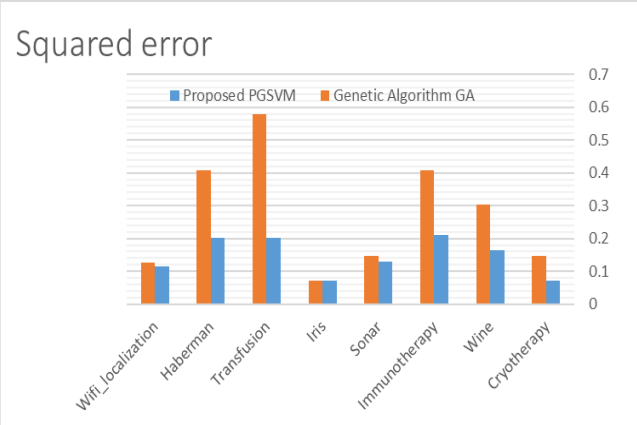


Fig. 8 - Table 2. Comparison between the proposed approach and GA in term of squared error

The experimental results prove that the proposed method always find the optimal parameter for SVM, but it is time consuming. Therefore, we have two ways to eliminate that with large scale or big data. First we can depend on the best combination obtained in the small datasets with different distribution. Second, by implementing parallel grid optimization to accelerate the process of selecting best parameter.

5. Conclusions and Future Work

The parameters of SVM are of great importance to the success of the SVM. To optimize the parameters properly and gain a higher classification accuracy, this study proposed PGSVM utilize a PSO feature weight and standard grid search (GS). In the proposed PSO-GS, GS was used to explore the search space and detect the potential parameters with the optimum solutions, while the PSO algorithm was used to conduct an effective refinement of the features that will be suitable for the process of classification. The experiments showed that the PGSVM method selected the parameters with higher data classification efficiency compared to random selection. The expected results were obtained as it achieved the global optimal solution although an expansion of the parameters optimization scope may take more time in finding the optimal parameters. One approach may be to avoid searching all the parameter data points. The global optimal solution may also be found in order to minimize the required time for parameters optimization. The classification concept can be applicable to big data using map-reduce technique. This will be the focus of future studies in the regard.

Acknowledgement

The authors are grateful to ICCI, Informatics Institute for Postgraduate Studies (IIPS_IRAQ) for their moral support. Many thanks for Mustansitiyah university (www.uomustansiriyah.edu.iq) for its support to finish this paper. Special thanks to the anonymous reviewers for their valuable suggestions and constructive comments.

References

[1] G. Kesavaraj and S. Sukumaran, "A study on classification techniques in data mining," in *Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on*, 2013, pp. 1-7.

[2] A. Ahlawat and B. Suri, "Improving classification in data mining using hybrid algorithm," in *Information Processing (IICIP), 2016 1st India International Conference on*, 2016, pp. 1-4.

[3] H. Li and Q. Lu, "K-CV parameter optimization method in the application of SVM classification data," in *Big Data Analysis (ICBDA), 2017 IEEE 2nd International Conference on*, 2017, pp. 25-29.

[4] Y. Ji, Y. Chen, H. Fu, and G. Yang, "An EnKF-based scheme to optimize hyper-parameters and features for SVM classifier," *Pattern Recognition*, vol. 62, pp. 202-213, 2017.

[5] E. Alickovic and A. Subasi, "Ensemble SVM Method for Automatic Sleep Stage Classification," *Submitted to IEEE TIM*.

[6] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273-297, 1995.

[7] Y. Liu and J. Du, "Parameter Optimization of the SVM for Big Data," in *Computational Intelligence and Design (ISCID), 2015 8th International Symposium on*, 2015, pp. 341-344.

[8] M. Claesen and B. De Moor, "Hyperparameter search in machine learning," *arXiv preprint arXiv:1502.02127*, 2015.

[9] Q. Huang, J. Mao, and Y. Liu, "An improved grid search algorithm of SVR parameters optimization," in *Communication Technology (ICCT), 2012 IEEE 14th International Conference on*, 2012, pp. 1022-1026.

[10] G. JayaBrindha and E. G. Subbu, "Ant Colony Technique for Optimizing the Order of Cascaded SVM Classifier for Sunflower Seed Classification."

[11] L. Grama, L. Tuns, and C. Rusu, "On the optimization of SVM kernel parameters for improving audio classification accuracy," in *Engineering of Modern Electric Systems (EMES), 2017 14th International Conference on*, 2017, pp. 224-227.

[12] Saba Abdul-baqi Salman, Ahmed Hussein Ali , Mohammad Khamees Khaleel, Mostafa Abdulghfoor Mohammed, "A New Model for Iris Classification Based on Naïve Bayes Grid Parameters".

[13] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," 2003.

[14] L. Lin, Z. Xiaolong, Z. Kai, and L. Jun, "Bilinear grid search strategy based support vector machines learning method," *Informatica*, vol. 38, 2014.

[15] J.-Y. Wu and C.-J. Lu, "Computational intelligence approaches for stock price forecasting," in *Computer, Consumer and Control (IS3C), 2012 International Symposium on*, 2012, pp. 52-55.

[16] M. A. Mohammed and N. ŢĂPUŞ, "A Novel Approach of Reducing Energy Consumption by Utilizing Enthalpy in Mobile Cloud Computing," *Studies in Informatics and Control*, vol. 26, pp. 425-434, 2017.

[17] R. A. Hasan and M. N. Mohammed, "A Krill Herd Behaviour Inspired Load Balancing of Tasks in Cloud Computing," *Studies in Informatics and Control*, vol. 26, pp. 413-424, 2017.

[18] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on*, 1995, pp. 39-43.

[19] P. B. Miranda, R. B. Prudêncio, A. P. De Carvalho, and C. Soares, "A hybrid meta-learning architecture for multi-objective optimization of SVM parameters," *Neurocomputing*, vol. 143, pp. 27-43, 2014.

[20] D. K. Jain, S. B. Dubey, R. K. Choubey, A. Sinhal, S. K. Arjaria, A. Jain, *et al.*, "An approach for hyperspectral image classification by optimizing SVM using self organizing map," *Journal of Computational Science*, vol. 25, pp. 252-259, 2018.