



An Adaptive Gaussian Kernel for Support Vector Machine

Abdullah Elen¹ · Selçuk Baş² · Cemil Közkurt³

Received: 2 October 2021 / Accepted: 30 January 2022 / Published online: 5 March 2022
© King Fahd University of Petroleum & Minerals 2022

Abstract

The most commonly used kernel function of support vector machine (SVM) in nonlinear separable dataset in machine learning is Gaussian kernel, also known as radial basis function. The Gaussian kernel decays exponentially in the input feature space and uniformly in all directions around the support vector, causing hyper-spherical contours of kernel function. In this study, an adaptive kernel function is designed based on the Gaussian kernel, which is used in SVM. While the sigma parameter is determined as an arbitrary value in the traditional Gaussian kernel, a modified Gaussian kernel method is used that calculates an adaptive value depending on the input vectors in the proposed kernel function. The proposed kernel function is compared with the linear, polynomial and Gaussian kernels commonly used in support vector machines. The results show that the proposed kernel function performs well on separable linear and nonlinear datasets compared to other kernel functions. It is also compared to state-of-the-art support vector machine kernels.

Keywords Support vector machine · Gaussian Kernel · Radial basis function · Data classification

1 Introduction

Support vector machine (SVM) [1, 2] may be called the flagship of classification in machine learning. Since SVM offers one of the most robust and accurate methods among all well-known algorithms such as k-nearest neighborhood (k-NN), artificial neural networks (ANN), decision trees (DT), it is considered to be tried in classification. It has a solid theoretical basis, requires few samples for training and works stable with a large number of dimensions. SVM [3] has been one of the most developed and used methods of recent years, based

on solid mathematical principles such as Lagrangian duality optimization [4] and kernel function [5] trick approaches, in classification tasks of supervised machine learning methods. Classification with supervised learning [6–8] is a method of machine learning [1, 9] task that categorically predicts class to which a particular instance belongs. Kernel function methods attract the attention of machine learning researchers due to their robust mathematical infrastructure, higher accuracy rates and relatively fast training times compared to other machine learning methods such as k-NN, ANN and DT in solving nonlinear problems. Although there are quite a lot of kernel function varieties, basic ones are linear, polynomial, Gaussian or in other words radial basis function (RBF) and sigmoid kernels [9] in classification. However, the kernel selection [2] issue, aimed at which kernel function is required to classify data most accurately, is an important research topic in recent years. Also, novel kernel function types [10–17] are produced continuously from large mathematical functions space since a wide variety of nonlinearly separable dataset are available in the real world. Kernels were applied to many areas including different dataset in literature [18–20]. In order to select the most suitable kernel for the topological structure of the dataset, kernel variety in literature should be increased.

Authors in [21] proposed a framework which can be applied on kernels and transform them more accurately

✉ Cemil Közkurt
ckozkurt@bandirma.edu.tr

Abdullah Elen
aelen@bandirma.edu.tr

Selçuk Baş
sbas@bandirma.edu.tr

¹ Department of Software Engineering, Faculty of Engineering and Natural Sciences, Bandirma Onyedü Eylül University, 10200 Bandirma, Balıkesir, Turkey

² Department of Accounting and Tax, Bandirma Vocational School, Bandirma Onyedü Eylül University, 10200 Bandirma, Balıkesir, Turkey

³ Department of Mechatronic Engineering, Faculty of Engineering and Natural Sciences, Bandirma Onyedü Eylül University, 10200 Bandirma, Balıkesir, Turkey



in terms of clustering, classification and dimensionality reduction of dataset. They prepared a synthetic dataset in 3D spiral form and used a common repository dataset. In experiments, their geometry aware kernels improved performance. In study [22], a modified fuzzy *c*-means clustering algorithm based on Mahalanobis distance that takes into samples' correlation account has been proposed and success achieved in terms of reducing effect of outliers and obtaining high classification accuracy. In [12], construction methods of some orthogonal polynomial kernels and proposed triangular-Chebyshev and triangular-Legendre kernels were compared with universal kernels via classification and regression scenarios. Proposed ones could achieve classification and regression tasks more accurately. Jiang et. al [14] proposed stationary Mahalanobis distance-induced kernels in SVM with applications in credit risk evaluation. They focused on stationary kernel construction where they assumed distance between two simultaneously translated vectors would be the same as the one without translation. Proposed kernels were appropriate for credit risk evaluation and over classical machine learning methods by meaning of performance. Baek and Kim [23] studied SVM's computational expense and proposed a new look up table based on a less costly optimal additive kernel. However, results showed that their proposed kernel is not higher than Gaussian kernel in accuracy. Ding et al. have proposed a kernel modifying the traditional RBF kernel of SVM, which they call random radial basis function (RRBF), in which kernel parameters can be randomly assigned and expand a single parameter to multivalued parameters of finite length. In experimental studies, datasets with 13 binary and 5 multi-class labels were used. As a result of experimental comparisons, they stated that RRBF outperforms RBF, polynomial, extreme learning machine (ELM), Chebyshev and Hermite kernels for both binary and multi-class classification datasets [24]. Another study by Ding et al. is the random compact Gaussian (RCG) kernel, which they proposed by combining random feature mapping used in the extreme learning machine (ELM) kernel with a traditional Gaussian kernel. Unlike their proposed method ELM kernel, all parameters in RCG kernel are randomly assigned. Random feature mapping saves the parameter selection time of RCG core and implicit mapping saves core computation time of RCG core. Experiments on both classification and regression datasets have reported that RCG tends to achieve better generalization performance than other kernels [25].

In the studies reviewed above, it is seen that improved new kernels via modification where kernels modified accordingly Mercer condition, distance metric where used metric provides positive semi-definition criteria and numerical trick like lookup tables. The Gaussian kernel in classification problems is referenced in the literature very often since it accurately classifies most of common datasets. Moreover, recent studies

on kernels for the SVM are compared with Gaussian kernels especially on nonlinearly separable datasets. In this study, we have proposed a new kernel named adaptive Gaussian (AG). AG uses a mean-based method for preprocess of input vectors of training dataset and ρ parameter selection method based on standard deviation. The contributions of this study are listed as follows:

- (1) The proposed method offers an adaptive kernel method by calculating the standard deviation value depending on the norms of the input vectors instead of the variance parameter in the Gaussian kernel.
- (2) It performs automatic calculation by disabling the manually determined parameter specific to the problem in the traditional Gaussian kernel.
- (3) Since it does not need problem-specific parameters, it can be applied directly and practically to all classification problems.

This paper is organized as follows; in the first section, the literature about SVM and Gaussian kernels has been introduced. In the second section, SVM has been briefly explained. In the third section, the theoretical approach of the proposed kernel function has been discussed. In the fourth section, experimental results have been presented and compared to well-known kernel functions and in the last section, this paper is summarized.

2 Support Vector Machine

SVM is a supervised learning model in machine learning used to classify binary or multiple datasets of linear or nonlinear separable type. Since the Lagrangian dual problem is used as an optimization approach in the SVM classifier, the number of training processes is saved, and a significant speed advantage is obtained compared to other algorithms. Thus, SVM is successful in high-volume datasets, as well as in high-dimensional problems with few data. Support vectors with hyperplanes are used for a linearly separable dataset as given in Fig. 1.

Given m training pairs $(x_1, y_1), \dots, (x_m, y_m)$, where $x_i \in R^n$, with $i = 1, 2, \dots, m$, is an input vector labeled by $x_i y_i \in \{-1, +1\}$, the linear SVM classifier searches for an optimal separating hyperplane given in Fig. 1.

$$\langle \omega, x \rangle + b = 0 \quad (1)$$

$\omega \in R^n$ is the normal vector to the hyperplane and b is a scalar. Equation (1) is obtained by solving the following

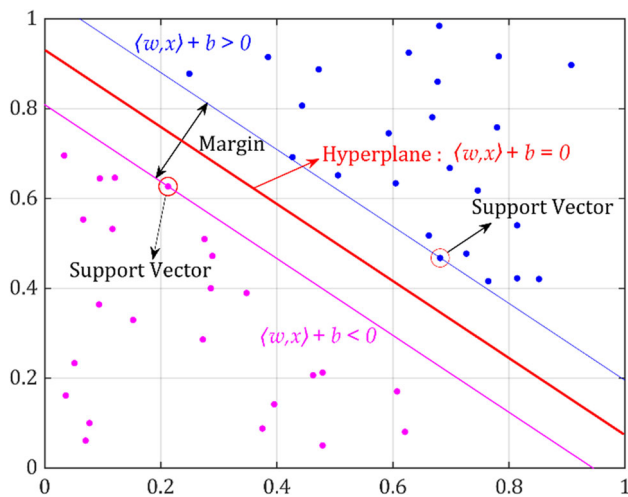


Fig. 1 Linear SVM classifier with hyperplane and margin convex quadratic programming problem below:

$$\min_{\omega, \tau_i} \frac{\|\omega\|^2}{2} + C \left(\sum_{i=1}^n \tau_i \right) \tag{2}$$

$$\text{subject to: } \begin{cases} y(\langle \omega, x \rangle + b) \geq 1 - \tau_i & \text{with, } i = 1, 2, \dots, m, \\ \tau_i \geq 0 \end{cases} \tag{3}$$

where C is a regularity parameter greater than zero, which balances the significance between maximization of margin width and minimization of training error. The slack variable $\tau_i \in R$ is the soft margin error of the i th training sample. The solution of Eqs. 2 and 3 is obtained by finding the saddle point of the Lagrange function and a decision function of the form given in Eq. 4.

$$f(y) = \text{sgn}(\langle \omega, x \rangle + b) = \text{sgn} \left(\sum_{i=1}^m a_i y_i x_i^T x + b \right) \tag{4}$$

The problems, caused by the nonlinear topology of datasets, can be solved by mapping datasets to multi-dimensional space by the kernel functions, which can be divided into a linearly separable problem. SVMs with kernel functions are created for nonlinearly separable data. These kernel functions are basically polynomial, Gaussian and sigmoid. The Gaussian kernel function allows the separation of nonlinearly separable data by mapping the input vector to Hilbert space. The Gaussian kernel is an exponential function including norm and real constant given in Eq. 5.

$$K_G(u, v) = \exp \left(-\frac{\|u - v\|_2^2}{2\rho^2} \right) \tag{5}$$

where u and v are input vectors, the Euclidean norm [26] in the numerator part of the exponential expression is obtained

with input vectors, and ρ in the denominator part is a real constant, which is an arbitrary value. Gaussian kernel function decays exponentially in the input feature space and uniformly in all directions around the support vector, inducing to hyperspherical contours of the kernel function. When using the Gaussian kernel in SVM, input vectors of the kernel are generally raw form; however in [27, 28] applied a correction method as the mean of input vectors. Width of each bell-shaped surface is directly proportional to ρ . The parameter ρ that gives optimal width can be found by trying. Experimental process of finding ρ is an iterative approach that consumes time and power. Therefore, the development of an efficient method for tuning ρ to an optimal width for the data would be an important solution for SVM classification problems. A stationary state of a Gaussian kernel can be represented as bell shaped with three-dimensional surface and data distributions given in Fig. 2.

Other basic kernel functions used in SVM are linear kernel given in Eq. 6 and polynomial kernel given in Eq. 7.

$$K_{\text{Lin}}(u, v) = \langle u, v \rangle \tag{6}$$

$$K_{\text{Pol}}(u, v) = (\langle u, v \rangle + 1)^n \tag{7}$$

where n is degree of polynomial.

3 The Kernel Approximation

Let $\{x_n\}$ be a sequence in the metric space (X, d) . If there is an N number in the form of

$$d(x_n, x_m) < \epsilon \tag{8}$$

for $\forall \epsilon > 0$ and $\forall k, l \geq N$, the $\{x_n\}$, a sequence is called the Cauchy sequence. The metric space to which every Cauchy sequence is convergent is the complete metric space. On the other hand, a normed space and $d(x_n, x_m)$ is defined by

$$d(x_n, x_m) = \|x_n - x_m\| \tag{9}$$

in that case d is a metric. Further, the complete inner product space as metric space is named Hilbert space H .

The kernel function $K : X \times X \rightarrow R$ is defined by

$$K(u, v) = \langle \varphi(u), \varphi(v) \rangle_H \tag{10}$$

where the space H is a feature space of K and $\varphi : X \times X \rightarrow H$ is a feature map. Also, from the property of the inner product $h, g \in H$ be given by.

$$g(x) = \sum_{i=1}^m \eta_i K(v_i, u) \text{ and } h(x) = \sum_{i=1}^n \lambda_i K(v_i, u) \tag{11}$$

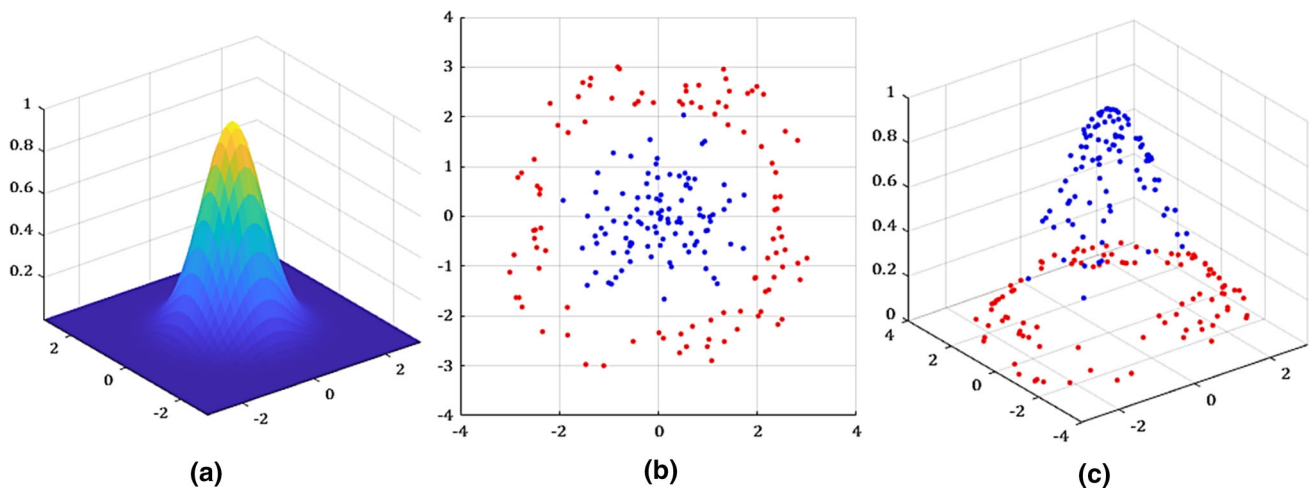


Fig. 2 Behavior of SVM with Gaussian kernel on nonlinear separable dataset; **a** 3D Gaussian surface **b** nonlinearly separable dataset **c** dataset mapped to Hilbert space by Gaussian kernel

Then

$$\begin{aligned} \langle g, h \rangle &= \sum_{i=1}^m \sum_{j=1}^n \eta_i \lambda_j K(u_i, v_j) = \sum_{i=1}^m \eta_i h(u_i) \\ &= \sum_{j=1}^n \lambda_j g(v_j) \end{aligned} \tag{12}$$

Theorem 1 (Mercer) [3] To be a valid SVM kernel, for $\varphi(u)$, the following integration should always be nonnegative for the given kernel function $K(u, v)$

$$\int \int K(u, v) \varphi(u) \varphi(v) du dv \geq 0 \tag{13}$$

If $K(u, v)$ is a kernel and κ is the kernel matrix with $\kappa_{i, j} = K(u_i, v_j)$,

$$\begin{aligned} \zeta^T \kappa \zeta &= \sum_{i, j} \zeta_i \zeta_j K(u_i, v_j) = \langle \varphi(u), \varphi(v) \rangle \\ &= \left\| \sum_i \zeta_i \varphi(u_i) \right\|^2 \geq 0 \end{aligned} \tag{14}$$

4 Proposed Method

In SVM and statistics, we take advantage of kernels, which are measures of the similarity between two examples. Most features of SVM are determined by the choice of kernel function. The use of different kernel functions represents the inner product in different feature space. The Gaussian kernel in classification problems is referenced in literature very

often since it accurately classifies most of common datasets. Gaussian kernel function decays exponentially in the input feature space and uniformly in all directions around the support vector, inducing to hyper-spherical contours of the kernel function. The Gaussian kernel is defined by

$$K(u, v) = \exp\left(-\gamma \|u - v\|_2^2\right), \tag{15}$$

where $\gamma = \frac{1}{2\rho^2}$ and ρ is a positive real constant and $\|\cdot\|$ is the Euclidean norm for vectors. In addition to possessing the advantages of the Gaussian kernel, we present a new kernel named adaptive Gaussian (AG) in Eq. 16.

$$K_{AG}(u, v) = \exp\left(-\frac{\|u - v\|^2 - \delta(\|u - v\|^2) + \Omega}{\delta(\|u - v\|^2)}\right) \tag{16}$$

where $\delta(\bullet)$ is the standard deviation function and Ω is as shown in Eq. 17.

$$\Omega = \begin{cases} |\min(\|u - v\|^2 - \delta(\|u - v\|^2))|, & \min(\|u - v\|^2 - \delta(\|u - v\|^2)) < 0 \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

To avoid division by zero, a term is added to the denominator of the equation as in Eq.18.

$$K_{AG}(u, v) = \exp\left(-\frac{\|u - v\|^2 - \delta(\|u - v\|^2) + \Omega}{\delta(\|u - v\|^2) + \varepsilon}\right) \tag{18}$$

ε is a negligible number to avoid a division by zero. Equation 19 is defined to simplify reading complexity.

$$\xi = \|u - v\|^2 - \delta(\|u - v\|^2) \tag{19}$$

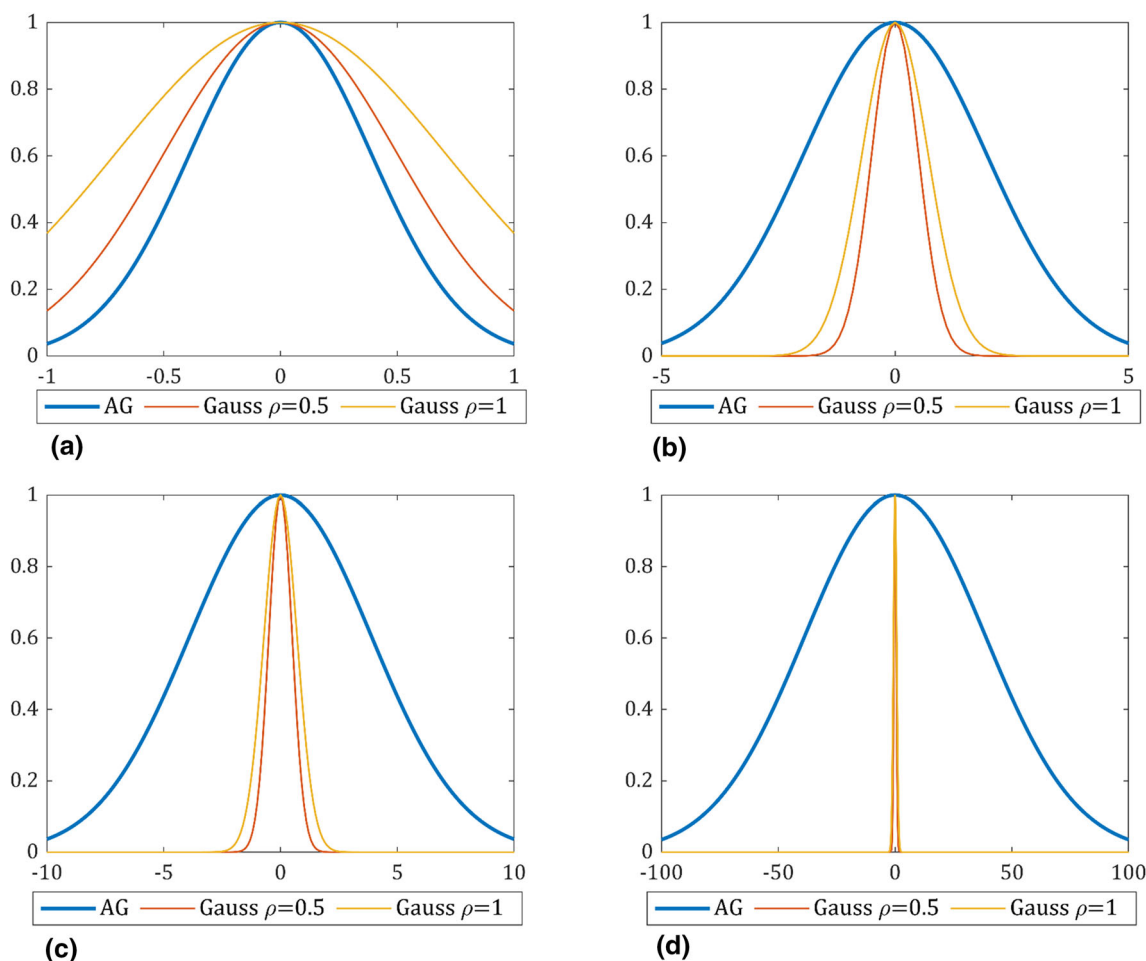


Fig. 3 AG vs. Gaussian kernel function curves **a** data between -1 and 1 **b** data between -5 and 5 **c** data between -10 and 10 **d** data between -100 and 100

As a result, the equation of the proposed kernel function, Eq. 20, is obtained.

$$K_{AG}(u, v) = \exp\left(-\frac{\xi + \Omega}{\delta(\|u - v\|^2) + \varepsilon}\right) \tag{20}$$

The denominator of the proposed kernel function adapts to the input vectors. At the same time, adaptation is provided by calculating squared norm of the differences of input vectors and their standard deviations in the numerator part. Therefore, the proposed kernel function is considered adaptive. With the piecewise function Ω used in the numerator part, if its minimum value is below zero, its absolute value is calculated and an offset operation is applied to the entire dataset.

To simplify the understanding of the adaptive property, visualization can be made on a dataset whose property space is two dimensional. As a result of all the vectors entering the kernel according to the vector in the center, in other words, in the central static state, distributed vectors on a three-dimensional bell-shaped surface are obtained. These vectors

in Hilbert space can now be easily classified with SVM. The bell curve of the AG corresponding to input vectors in different value ranges and the bell curves of the Gaussian kernel are compared in Fig. 3.

The AG kernel function adapts to different data intervals and creates an optimal Gaussian bell curve. With this adaptive feature of the AG, the problem of finding optimum value of the parameter ρ which is an arbitrary real constant value in the Gauss kernel function is eliminated. The time complexity of a conventional Gaussian kernel is $O(dN_{SV})$, where N_{SV} is the number of support vectors and d is the number of features [25]. In the time complexity of the proposed method, the standard deviation of the norm of the input vectors is calculated in addition to the Gaussian kernel. In the time complexity, the d -dimensional standard deviation calculation cost is added for each support vector, and thus, the time complexity of the proposed method is found as $O((d + d)N_{SV})$.

Table 1 Datasets used in experimental studies

No	Dataset	Repository	Features	Instances	Classes
1	Balance	KEEL	4	625	3
2	Bands	KEEL	19	365	2
3	Cleveland	KEEL	13	297	5
4	Dermatology	KEEL	34	358	6
5	German	KEEL	20	1000	2
6	Haberman	KEEL	3	306	2
7	Heart	KEEL	13	270	2
8	Hepatitis	KEEL	19	80	2
9	Housevotes	KEEL	16	232	2
10	Mammographic	KEEL	5	830	2
11	Optdigits	KEEL	64	5620	10
12	Sonar	KEEL	60	208	2
13	Spectfheart	KEEL	44	267	2
14	Titanic	KEEL	3	2201	2
15	Vehicle	KEEL	18	846	4
16	Wine	KEEL	13	178	3
17	Wisconsin	KEEL	9	683	2
18	Australian	UCI	14	690	2
19	Ionosphere	UCI	34	351	2
20	Monks 1	UCI	6	124	2
21	Monks 2	UCI	6	169	2
22	Monks 3	UCI	6	122	2
23	Transfusion	UCI	4	748	2
24	WDBC	UCI	30	569	2
25	WPBC	UCI	33	194	2

5 Experimental Results

In order to test the accuracy of the proposed method in experimental studies, a total of 25 datasets, 17 of which are KEEL-dataset repository and 8 of them UCI machine learning repository, were used as shown in Table 1. The MATLAB platform was used to implement the AG into SVM. The real constant of Gauss kernel functions γ values is taken as default value 1. In addition, the degree of polynomial kernel function is taken as the default value of 2.

Figure 4 shows the k -fold cross-validation technique for the datasets given to the classifier model. Here, the k -value is determined as 10. We use a common procedure of tenfold cross-validation to run the experiments. Cross-validation is the process of creating grouped datasets by selecting different amounts of samples from the same dataset.

The cross-validation technique is used to measure the predictive ability of a machine learning model on previously unseen data with less bias. Each dataset used in our experiments is divided into ten subgroups with an approximately equivalent number of samples. Thus, ten runs are performed

for each dataset during the training and testing process of machine learning models. Then, by calculating their mean accuracy values, the overall success rate of machine learning models was obtained.

Accuracy (ACC), F-measure (FM) and Matthews correlation coefficient (MCC) metrics were used to measure classification success of the proposed method. ACC is a commonly used metric to measure the classification success of a model. The ACC value is equal to the ratio of correct predictions of a classifier to the total number of samples in a dataset. However, only model accuracy is not sufficient in imbalanced datasets.

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (21)$$

F1-score is obtained by calculating the harmonic mean of the precision and recall values. It is more appropriate to use this scale instead of the ACC to avoid choosing a wrong classifier model in imbalanced datasets.

$$\text{FM} = \frac{2}{\text{TPR}^{-1} + \text{PPV}^{-1}} \quad (22)$$



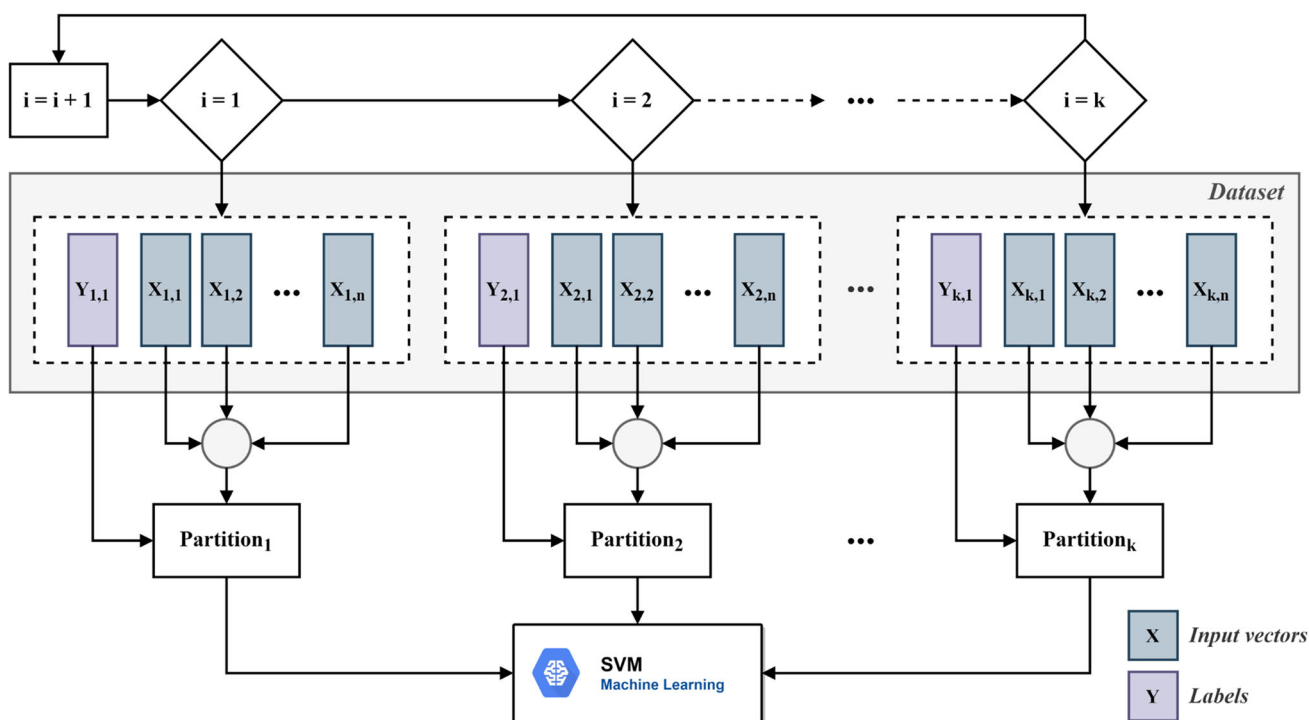


Fig. 4 k-fold cross-validation of the dataset

Matthews correlation coefficient (MCC) is a correlation coefficient between observed and predicted binary classifications. Like most correlation coefficients, the MCC ranges from -1 to 1 . A higher correlation between observed and predicted values is better estimation. Coefficient -1 represents the total disagreement between prediction and observation, coefficient 0 is not better than a random prediction, and $+1$ represents a perfect estimate.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (23)$$

Table 2 shows the proposed method and the results of the classification processes applied on 25 datasets of other SVM kernels. As seen from the table, it is obvious that the proposed kernel ACC and FM are better than other methods. The proposed kernel and polynomial kernel have very close scores in terms of MCC.

Figure 5 shows the general classification accuracies of SVM kernel according to dataset repositories. It is seen that the proposed method has higher classification accuracy on both KEEL and UCI dataset repositories compared to other kernels.

In Table 3, Wilcoxon signed-rank test is given to test whether there is a difference between the two measurement results obtained from the datasets in experimental studies. In the hypothesis test, significance level was determined as

0.05 . According to the test results obtained from the hypothesis test, it is seen that the proposed method is significantly different from other SVM kernels.

In recent years, many kernel functions have been proposed for the SVM. Comparison results of some of them with the proposed method are given in Table 4.

6 Conclusion

The use of kernel functions in SVM is one of the popular studies of recent years, and efforts to develop new kernel functions are given in detail in the introduction section. Especially the success of the Gaussian kernel function, which is one of the basic kernel functions, on nonlinear separable datasets is obvious. This kernel function is introduced and its usage in SVM is explained. The fact that the Gauss function depends on an arbitrary real constant parameter causes the problem of finding the optimal value of the parameter in the training process. In this study, an adaptive kernel function is proposed by making an improvement on the Gaussian kernel, which is an effective kernel function. The proposed kernel function was used in SVM classifiers and applied in 25 well-known datasets and comparisons were made with the basic kernel functions: Gaussian, linear and polynomial. As a result of the comparisons of the experimental studies, the proposed model achieved a higher success than other kernel functions with 84.7% ACC and 73.6% FM. Although the Gaussian

Table 2 Comparison of classification results

Datasets	SVM-AG			SVM-G			SVM-LIN			SVM-POL		
	ACC	FM	MCC	ACC	FM	MCC	ACC	FM	MCC	ACC	FM	MCC
Balance	0,937	0,058	0,863	0,933	0,062	0,826	0,939	0,042	0,807	0,984	0,009	0,952
Bands	0,719	0,675	0,378	0,709	0,635	0,340	0,692	0,616	0,299	0,713	0,691	0,387
Cleveland	0,824	0,257	0,174	0,833	0,260	0,244	0,829	0,307	0,235	0,798	0,262	0,138
Dermatology	0,984	0,951	0,944	0,824	0,343	0,477	0,989	0,962	0,957	0,992	0,972	0,970
German	0,762	0,684	0,392	0,716	0,504	0,176	0,760	0,685	0,389	0,678	0,624	0,253
Haberman	0,748	0,594	0,262	0,737	0,506	0,164	0,728	0,426	0,073	0,744	0,534	0,197
Heart	0,802	0,797	0,607	0,763	0,755	0,529	0,833	0,829	0,667	0,764	0,762	0,531
Hepatitis	0,846	0,693	0,505	0,838	0,455	0,000	0,842	0,688	0,519	0,850	0,615	0,481
Housevotes	0,965	0,965	0,933	0,776	0,745	0,601	0,957	0,956	0,916	0,947	0,946	0,896
Mammographic	0,803	0,803	0,610	0,802	0,802	0,611	0,798	0,797	0,603	0,809	0,809	0,623
Optdigits	0,992	0,962	0,960	0,945	0,777	0,779	0,996	0,982	0,981	0,998	0,990	0,989
Sonar	0,865	0,864	0,751	0,719	0,676	0,506	0,790	0,786	0,585	0,886	0,884	0,778
Spectfheart	0,793	0,619	0,296	0,791	0,572	0,239	0,811	0,658	0,360	0,789	0,672	0,358
Titanic	0,781	0,720	0,469	0,786	0,697	0,493	0,776	0,714	0,456	0,780	0,699	0,472
Vehicle	0,861	0,706	0,625	0,876	0,748	0,669	0,885	0,766	0,692	0,927	0,855	0,808
Wine	0,985	0,976	0,967	0,981	0,970	0,959	0,984	0,974	0,965	0,979	0,967	0,954
Wisconsin	0,955	0,952	0,910	0,965	0,962	0,926	0,967	0,964	0,930	0,938	0,931	0,866
Australian	0,854	0,852	0,710	0,843	0,840	0,684	0,852	0,852	0,716	0,805	0,803	0,609
Ionosphere	0,927	0,919	0,843	0,921	0,917	0,845	0,883	0,863	0,750	0,881	0,864	0,741
Monks 1	0,743	0,735	0,506	0,668	0,659	0,351	0,651	0,642	0,308	0,770	0,763	0,554
Monks 2	0,677	0,650	0,320	0,723	0,679	0,404	0,621	0,383	0,000	0,787	0,772	0,564
Monks 3	0,891	0,888	0,799	0,808	0,803	0,639	0,781	0,775	0,576	0,819	0,813	0,665
Transfusion	0,774	0,539	0,208	0,766	0,477	0,146	0,762	0,436	0,081	0,768	0,496	0,164
WDBC	0,921	0,919	0,851	0,971	0,969	0,938	0,978	0,977	0,954	0,961	0,958	0,919
WPBC	0,775	0,616	0,298	0,763	0,464	0,158	0,793	0,645	0,380	0,734	0,652	0,327
Averages	0,847	0,736	0,607	0,818	0,651	0,508	0,836	0,709	0,568	0,844	0,734	0,608

Fig. 5 Overall classification accuracy of SVM kernels according to dataset repositories

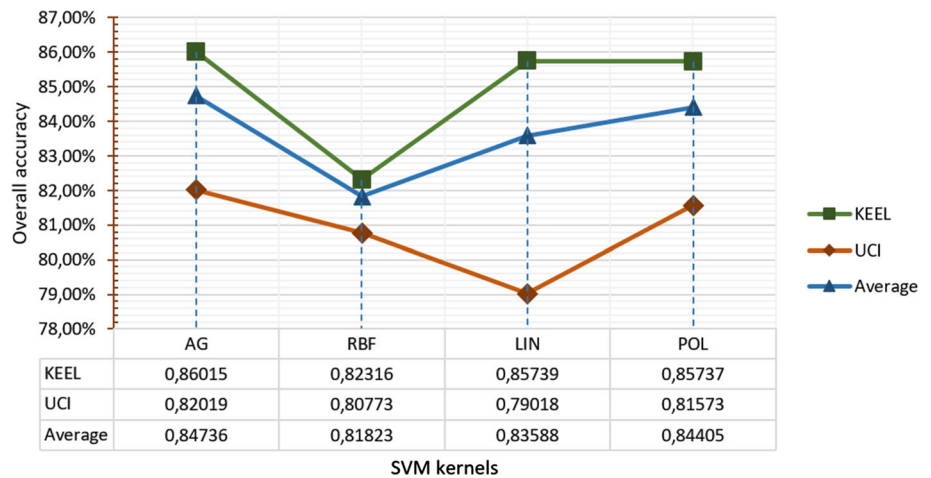


Table 3 Wilcoxon signed-rank test results ($\alpha = 0.05$)

Datasets	SVM-AG vs. SVM-G	SVM-AG vs. SVM-LIN	SVM-AG vs. SVM-POL
Balance	9,5494E-01	8,5837E-01	9,0330E-01
Bands	1,7053E-03	1,2668E-02	6,0176E-03
Cleveland	1,1106E-08	1,6733E-01	1,6051E-03
Dermatology	8,6890E-39	4,2248E-01	4,0898E-01
German	3,3996E-28	9,5616E-01	8,0450E-09
Haberman	2,5674E-03	2,7668E-08	7,3006E-03
Heart	8,6235E-01	7,2824E-01	3,4231E-01
Hepatitis	1,7952E-04	8,3028E-01	1,5364E-01
Housevotes	6,5492E-08	7,8089E-01	7,1057E-01
Mammographic	2,3847E-01	1,1584E-01	4,0389E-01
Optdigits	4,4412E-70	1,7798E-04	1,9019E-04
Sonar	1,7086E-15	3,9571E-02	6,1070E-03
Spectfheart	7,7331E-02	9,0022E-01	3,0345E-02
Titanic	2,3082E-11	8,8289E-01	1,9252E-04
Vehicle	5,1134E-01	4,3810E-01	1,0213E-01
Wine	8,8594E-01	9,4834E-01	8,8605E-01
Wisconsin	3,4388E-01	9,3197E-02	3,2885E-02
Australian	5,1734E-01	5,9620E-02	4,4993E-01
Ionosphere	5,2137E-02	2,1802E-02	1,4533E-01
Monks 1	9,9355E-02	3,7363E-01	4,2648E-02
Monks 2	3,3249E-02	1,4945E-17	3,7074E-01
Monks 3	9,6810E-02	3,7151E-01	2,0171E-01
Transfusion	2,3025E-03	4,3223E-10	1,5965E-02
WDBC	9,5001E-03	3,7300E-03	1,8818E-02
WPBC	5,1695E-05	8,7931E-01	2,8521E-05

Table 4 Comparison of classification results of the proposed method and reported results of state-of-the-art SVM kernels

Datasets	Padierna et al. [29]	Moghaddam et al. [13]			Tian et al. [12]		Jafarzadeh et al. [30]		Zhou et al. [31]	Proposed method
		H-C	H-G	H-W	Tri-Che	Tri-Leg	Che-G	Che-W		
German	0,769	–	–	–	0,734	0,743	–	–	0,766	0,762
Haberman	0,751	0,728	0,759	0,743	–	–	–	–	–	0,748
Heart	–	0,805	0,829	0,828	0,766	0,760	0,809	0,809	0,848	0,802
Sonar	0,923	0,838	0,868	0,848	0,927	0,929	0,895	0,885	–	0,865
Titanic	–	–	–	–	–	–	–	–	0,777	0,781
Vehicle	0,855	–	–	–	–	–	–	–	–	0,861
Australian	0,865	–	–	–	0,838	0,828	–	–	–	0,854
Ionosphere	0,934	0,959	0,960	0,927	0,880	0,882	0,969	0,966	–	0,927
Monks 1	0,975	–	–	–	0,606	0,647	–	–	–	0,743
Monks 2	0,856	–	–	–	0,693	0,691	–	–	–	0,677
Monks 3	0,935	–	–	–	0,940	0,942	–	–	–	0,891
Transfusion	–	0,689	0,695	0,696	–	–	–	–	–	0,774
WDBC	0,982	0,959	0,978	0,958	0,960	0,962	0,981	0,983	–	0,921

kernel function provides 81.9% overall classification accuracy, the proposed method offers approximately 3% better than Gaussian kernel. It has been shown that the proposed method in the study can be used effectively in SVM classification problems. According to the test results obtained from the hypothesis test, it is seen that the proposed method is significantly different from other SVM kernels. The proposed kernel function can be applied in future studies in different metric spaces.

References

- Vapnik, V.: The Nature of Statistical Learning Theory. Springer, New York (1995) <https://doi.org/10.1007/978-1-4757-3264-1>
- Ding, L., Liao, S., Liu, Y., Liu, L., Zhu, F., Yao, Y., & Gao, X. (2020). Approximate kernel selection via matrix approximation. *IEEE Trans. Neural Netw. Learn. Syst.* 31(11), 4881–4891. <https://doi.org/10.1109/TNNLS.2019.2958922>.
- Cortes, C.; Vapnik, V.: Support-vector networks. *Mach. Learn.* 20(3), 273–297 (1995). <https://doi.org/10.1007/BF00994018>
- Gasimov, R.N.: Augmented Lagrangian duality and nondifferentiable optimization methods in nonconvex programming. *J. Global Optim.* 24(2), 187–203 (2002). <https://doi.org/10.1023/A:1020261001771>
- Bao, Y., Wang, T., & Qiu, G. (2014). Research on applicability of svm kernel functions used in binary classification. In: *Proceedings of International Conference on Computer Science and Information Technology* (pp. 833–844). Springer, New Delhi. https://doi.org/10.1007/978-81-322-1759-6_95
- Bzdok, D.; Krzywinski, M.; Altman, N.: Machine learning: supervised methods. *Nat Methods* 15, 5–6 (2018). <https://doi.org/10.1038/nmeth.4551>
- Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. *Int. J. Comp. Trends Technol. (IJCTT)*, 48(3), 128–138. <https://doi.org/10.14445/22312803/IJCTT-V48P126>.
- Elen, A.; Avuçlu, E.: standardized variable distances: a distance-based machine learning method. *Appl. Soft Comput.* 98, 106855 (2021). <https://doi.org/10.1016/j.asoc.2020.106855>
- Bi, Q.; Goodman, K.E.; Kaminsky, J.; Lessler, J.: What is machine learning? a primer for the epidemiologist. *Am. J. Epidemiol.* 188(12), 2222–2239 (2019). <https://doi.org/10.1093/aje/kwz189>
- Amari, S.I.; Wu, S.: Improving support vector machine classifiers by modifying kernel functions. *Neural Netw.* 12(6), 783–789 (1999). [https://doi.org/10.1016/S0893-6080\(99\)00032-5](https://doi.org/10.1016/S0893-6080(99)00032-5)
- Ozer, S.; Chen, C.H.; Cirpan, H.A.: A set of new Chebyshev kernel functions for support vector machine pattern classification. *Pattern Recogn.* 44(7), 1435–1447 (2011). <https://doi.org/10.1016/j.patcog.2010.12.017>
- Tian, M.; Wang, W.: Some sets of orthogonal polynomial kernel functions. *Appl. Soft Comput.* 61, 742–756 (2017). <https://doi.org/10.1016/j.asoc.2017.08.010>
- Moghaddam, V.H.; Hamidzadeh, J.: New Hermite orthogonal polynomial kernel and combined kernels in support vector machine classifier. *Pattern Recogn.* 60, 921–935 (2016). <https://doi.org/10.1016/j.patcog.2016.07.004>
- Jiang, H.; Ching, W.K.; Yiu, K.F.C.; Qiu, Y.: Stationary Mahalanobis kernel SVM for credit risk evaluation. *Appl. Soft Comput.* 71, 407–417 (2018). <https://doi.org/10.1016/j.asoc.2018.07.005>
- Shankar, K.; Lakshmanaprabu, S.K.; Gupta, D.; Maselena, A.; De Albuquerque, V.H.C.: Optimal feature-based multi-kernel SVM approach for thyroid disease classification. *J. Supercomput.* 76(2), 1128–1143 (2020). <https://doi.org/10.1007/s11227-018-2469-4>
- Ye, N., Sun, R., Liu, Y., & Cao, L. (2006) Support vector machine with orthogonal Chebyshev kernel. In: *18th International Conference on Pattern Recognition (ICPR'06)* (Vol. 2, pp. 752–755). IEEE. <https://doi.org/10.1109/ICPR.2006.1096>
- Zanaty, E.A.; Afifi, A.: Support vector machines (SVMs) with universal kernels. *Appl. Artif. Intell.* 25(7), 575–589 (2011). <https://doi.org/10.1080/08839514.2011.595280>
- Ozguven, M. M., Yilmaz, G., Adem, K., & Kozkurt, C. Use of Support Vector Machines and Artificial Neural Network Methods in Variety Improvement Studies: Potato Example. *Curr Inves Agri Curr Res* 6 (1)-2019. CIACR. MS. ID, 229, <https://doi.org/10.32474/CIACR.2019.06.000229>.
- Elen, A. & Turan, M. K. (2019). Classifying white blood cells using machine learning algorithms. *Int. J. Eng. Res Develop.* 11(1): 141–152. <https://doi.org/10.29137/umagd.498372>.
- Yöntem, M. K., & Adem, K. (2019). Prediction of the level of alexithymia through machine learning methods applied to automatic thoughts. *Current Approaches Psychiat.* <https://doi.org/10.18863/pgy.554788>.
- Pan, B.; Chen, W.S.; Xu, C.; Chen, B.: A novel framework for learning geometry-aware kernels. *IEEE Trans. Neural Networks Learn. Syst.* 27(5), 939–951 (2015). <https://doi.org/10.1109/TNNLS.2015.2429682>
- Zhang, Y.; Xie, F.; Huang, D.; Ji, M.: Support vector classifier based on fuzzy c-means and Mahalanobis distance. *J. Intell. Inf. Syst.* 35(2), 333–345 (2010). <https://doi.org/10.1007/s10844-009-0102-y>
- Baek, J.; Kim, E.: A new support vector machine with an optimal additive kernel. *Neurocomputing* 329, 279–299 (2019). <https://doi.org/10.1016/j.neucom.2018.10.032>
- Ding, X., Liu, J., Yang, F., & Cao, J. (2021) Random radial basis function kernel-based support vector machine. *J. Franklin Instit.*, (In press). <https://doi.org/10.1016/j.jfranklin.2021.10.005>
- Ding, X.; Liu, J.; Yang, F.; Cao, J.: Random compact Gaussian kernel: application to ELM classification and regression. *Knowl.-Based Syst.* 217, 106848 (2021). <https://doi.org/10.1016/j.knsys.2021.106848>
- Baş, S.; Körpınar, T.: Modified roller coaster surface in space. *Mathematics* 7(2), 195 (2019). <https://doi.org/10.3390/math7020195>
- Mustaqeem, M.; Saqib, M.: Principal component based support vector machine (PC-SVM): a hybrid technique for software defect detection. *Clust. Comput.* (2021). <https://doi.org/10.1007/s10586-021-03282-8>
- Xue, S.; Yan, X.: A new kernel function of support vector regression combined with probability distribution and its application in chemometrics and the QSAR modeling. *Chemom. Intell. Lab. Syst.* 167, 96–101 (2017). <https://doi.org/10.1016/j.chemolab.2017.05.005>
- Padierna, L.C.; Carpio, M.; Rojas-Domínguez, A.; Puga, H.; Fraire, H.: A novel formulation of orthogonal polynomial kernel functions for SVM classifiers: the Gegenbauer family. *Pattern Recogn.* 84, 211–225 (2018). <https://doi.org/10.1016/j.patcog.2018.07.010>
- Jafarzadeh, S. Z., Aminian, M., & Efati, S. A set of new kernel function for support vector machines: An approach based on Chebyshev polynomials. In: *ICCKE 2013* (pp. 412–416). IEEE. <https://doi.org/10.1109/iccke.2013.6682848>
- Zhou, S.-S.; Liu, H.-W.; Ye, F.: Variant of Gaussian kernel and parameter setting method for nonlinear SVM. *Neurocomputing* 72(13–15), 2931–2937 (2009). <https://doi.org/10.1016/j.neucom.2008.07.016>

