Sajad Fathi Hafshejani<sup>1,2\*</sup> and Zahra Moaberfard<sup>1</sup>

<sup>1\*</sup>Department of Applied Mathematics, Shiraz University of Technology, Modarres Boulevard, Shiraz, 71557-13876, Fars, Iran.

<sup>2\*</sup>Department of Math and Computer Science, University of Lethbridge, 4401 University Dr W, Lethbridge, T1K 3M4, AB, Canada.

\*Corresponding author(s). E-mail(s): s.fathi@sutech.ac.ir; Contributing authors: zahra.moaberfard@gmail.com;

#### Abstract

In the last few years, various types of machine learning algorithms, such as Support Vector Machine (SVM), Support Vector Regression (SVR), and Non-negative Matrix Factorization (NMF) have been introduced. The kernel approach is an effective method for increasing the classification accuracy of machine learning algorithms. This paper introduces a family of one-parameter kernel functions for improving the accuracy of SVM classification. The proposed kernel function consists of a trigonometric term and differs from all existing kernel functions. We show this function is a positive definite kernel function. Finally, we evaluate the SVM method based on the new trigonometric kernel, the Gaussian kernel, the polynomial kernel, and a convex combination of the new kernel function and the Gaussian kernel function on various types of datasets. Empirical results show that the SVM based on the new trigonometric kernel function and the mixed kernel function achieve the best classification accuracy. Moreover, some numerical results of performing the SVR based on the new trigonometric kernel function and the mixed kernel function are presented.

**Keywords:** Support vector machine, Kernel-method, Trigonometric kernel function

2 A new trigonometric kernel function for support vector machine

### 1 Introduction

Support Vector Machine (SVM) is a supervised learning algorithm mostly used for classification, but it can also be applied for regression. Vapnik [1] proposed the SVM method for the first time, and it has been utilized in a wide range of real-world problems such as bioinformatics [2], biometrics [3], power systems [4], and chemoinformatics [5]. In SVM, the training data are used for training and building the classification model. This model is then used to classify unknown samples.

SVM achieves competitive results when the data are linearly separable. However, kernel functions can be considered for non-separable data. They map the data into a vector space and use linear algebra and geometry to find out the structure of the data. There are some reasons to map the data into a feature space. By mapping the original data into higher-dimensional space, it is possible to transform nonlinear relations within the data into linear ones [6].

It has been proven that the kernel's theory is based on structural risk minimization by using the maximum margin idea [7]. The kernel function, which is a key factor in determining algorithm performance, is at the heart of many machine learning algorithms. Kernels allow data to be mapped into high-dimensional feature space to increase the computational power of linear machines. Thus, it is a way of extending linear hypotheses to non-linear ones, and this step can be performed implicitly. SVM can be classified into linear and nonlinear approaches [8]. Trying to learn a non-linear separating boundary in the input space increases the computational requirements during the optimization phase because the separating surface will be of at least the second order. Instead, SVM maps the data, using predefined kernel functions, into a new but higherdimensional space, where a linear separator would be able to discriminate between the different classes. The SVM's optimization phase will thus entail learning only a linear discriminant for surfacing the mapped space. Of course, the kernel function's selection and settings are critical for SVM performance. Kernel methods, in fact, transfer the original data into another space, the so-called "feature space," and reveal the data's hidden features [9].

Various kernel functions for machine learning algorithms have been introduced, and some of their significant properties have been explored [10–12]. The Gaussian kernel function is a popular kernel function used in most machine learning algorithms [13]. Table 1 demonstrates four popular kernel functions used in most classification papers.

i	name	$K(x_i, x_j)$
1	Polynomial kernel	$K_1(x_i, x_j) = (1 + x_i x_j)^p$
$2 \\ 3 \\ 4$	Gaussian kernel RBF Kernel Sigmoid kernel	$K_{2}(x_{i}, x_{j}) = \exp(-\frac{\ x_{i} - x_{j}\ ^{2}}{2\sigma^{2}})$ $K_{3}(x_{i}, x_{j}) = \exp(-\gamma \ x_{i} - x_{j}\ ^{2})$ $K_{4}(x_{i}, x_{j}) = \tanh(\alpha + \beta x_{i}^{T} x_{j})$

Table 1: Four common kernel functions

The number of kernel parameters is one of the prime issues in many kernelbased methods. Finding the best values for kernel parameters is a challenging problem because their values have a significant impact on their performance. In fact, when the kernel has at least two parameters, the kernel-based method consumes more time in the training phase to find the best values for these parameters. During these years, researchers have been focusing on introducing new kernel functions with one parameter. For example, due to the existence of two parameters in the Sigmoid kernel function in Table 1, machine learning algorithms based on this kernel function spend more time finding the best values for the parameters  $\alpha$  and  $\beta$ . The Gaussian kernel function is an example with one parameter and has been widely utilized in most machine learning algorithms. However, the accuracy of the SVM classification based on this kernel function on some datasets is still not good. In this regard, many efforts have been made to introduce new kernel functions to increase the classification accuracy of the SVM. Recently, Tharwat [14] suggested a new approach for estimating the best values for kernel parameters. Their method works based on the maximum and minimum distances between samples in each class of dataset.

In the two past decades, significant efforts have been devoted to the development of kernel functions for kernel-based methods. Combining kernel functions is a successful approach for kernel-based methods. It has been proven that the combination of two or more kernel functions is also a kernel function, and the new combined kernel function can significantly improve the performance of the SVM [15].

#### 1.1 Contribution

- This paper gives a new family of one-parameter trigonometric kernel function to improve accuracy of the SVM and SVR approaches. We demonstrate that the new kernel function satisfies the kernel conditions. Moreover, we prove that the new proposed kernel function is a positive definite, which is a critical condition for kernel functions.
- We propose a new mixed kernel function, which is obtained by combining the new trigonometric function and the Gaussian function.
- We adapt the proposed method in [14] to the new trigonometric kernel-SVM and show how to predict the best value for parameter  $\sigma$ .
- We perform an extensive implementation to demonstrate the efficiency of the new trigonometric kernel function. We evaluate the accuracy of the SVM based on the new trigonometric kernel function, the Gaussian kernel function, the polynomial kernel function, as well as a convex combination of the trigonometric kernel function and the Gaussian kernel function on 24 datasets with different ranges of dimensions and numbers of features. Empirical results show that the new proposed kernel function gives good classification accuracy in nearly all the data sets, especially those of high dimensions. It is shown that the new mixed kernel function is very efficient for data classification. Moreover, we examine the efficiency of the SVR method

based on the new trigonometric kernel function and the new mixed kernel function.

The paper is organized as: In Section 2, we first recall that some properties of kernel function and then introduce the new trigonometric kernel function. We also investigate some properties of the new kernel function. In section 3, we present a new mixed kernel function by using a convex combination of the new kernel function and the Gaussian kernel function. Some numerical results of performing kernel-SVM kernel-SVR on 24 datasets are shown in Section 3. We finally end up the paper by providing some concluding remarks.

We use the following notations conventions throughout the paper:  $\|.\|$  denotes the Euclidean norm of a vector, the non-negative and positive orthants are denoted by  $\mathbb{R}^n_+$  and  $\mathbb{R}^n_{++}$  respectively.

## 2 A new kernel function

This section is devoted to presenting a new trigonometric kernel function for SVM and investigating various important properties of the new proposed kernel function.

We start this section by examining an example. Let us investigate exactly how a kernel function makes data classification easier. Assume X is a randomly generated dataset with two class labels of 1, 0. This dataset has 400 samples and consists of two circles. We plot the original data in Fig. 1 (left). It is obvious that a linear SVM is not able to classify data with a good accuracy. To address this problem, we utilize a kernel function and map data into a new space, Fig. 1 (right). After mapping, a linear SVM can classify data with high accuracy.



Fig. 1: Behavior of kernel function

Our next goal is to define a new kernel function. To this end, we first remind that the definition of the kernel function demonstrated in [16].

**Definition 1** Let  $X \in \mathbb{R}^d$  be a non-empty set. A function  $K : X \times X \to \mathbb{R}$  on X is a kernel function if there exists a  $\mathbb{K}$ -Hilbert space H and a map  $\Phi : X \to H$  such that:

$$K(x, x') = \langle \Phi(x), \Phi(x') \rangle, \qquad \forall \ x, x' \in X.$$
(1)

where  $\Phi$  is a nonlinear (or sometimes linear) map from the input space X to the feature space  $\mathcal{F}$ , and  $\langle ., . \rangle$  is an inner product. Besides, the function  $\Phi$  is so-called a *feature map* and H called a *feature space* of K.

**Definition 2** A kernel function is shift-invariant if it has the following form:

$$K(x, x') = K(x - x').$$
 (2)

Now we can introduce a new kernel function for SVM. In this regard, consider the following trigonometric function:

$$\psi(x) = \sin(h(x)), \qquad h(x) = \frac{\pi}{2 + \sigma x^2},$$
(3)

where  $\sigma$  is a positive real number. The behaviour of the functions  $\psi(x)$ ,  $\psi'(x)$ , and h(x) with different values of parameter  $\sigma$  are shown in Fig. 2.



**Fig. 2**: The behaviour of the function  $\psi(x)$ ,  $\psi'(x)$ , and h(x) with different values of parameter  $\sigma$ .

From Fig. 2, we can conclude that:

- The new kernel function has a different value for each value of parameter  $\sigma$ . It implies that, we can figure out the best value for the  $\sigma$  for each dataset.
- The kernel function's first derivative is negative, implying that  $\psi(x)$  is a decreasing function for all  $x \ge 0$ .
- We have  $h(x) \in (0, \frac{\pi}{2}]$ , for all  $x \ge 0$ .

Next lemma presents some properties of the function given by Eq. 3.

Lemma 1 Let  $\psi$  be a function defined by Eq. 3. Then, we have:

- $\psi(0) = 1.$
- $\psi(x) > 0$  for all  $x \in \mathbb{R}$ .
- $\psi'(x) \leq 0$ , for all  $x \in \mathbb{R}_+$ .
- $\psi$  is bounded, i.e., for all  $x \in \mathbb{R}$ , we have  $\|\psi(x)\| \le \psi(0)$ .

Proof Using this fact  $\sin(\frac{\pi}{2}) = 1$ , we can conclude that the first item is true. To prove the second item, we know that  $\|.\|$  is non-negative, so  $h_1(x) = \frac{\pi}{2+\sigma\|x\|^2} \in (0, \frac{\pi}{2}]$  for all x, which implies  $\sin(h_1(x)) \in (0, 1]$ . To prove the third item, we have:

$$\psi'(x) = h'(x)\cos(h_1(t))$$

It is clear that h(x) is a decreasing function, so we can conclude that h'(x) < 0. On the other hand,  $\cos(h(x))$  is a positive function for all  $x \in (0, \frac{\pi}{2}]$ , i.e., the first derivative of the function  $\psi(x)$  given by Eq. 3 is negative.

The fourth item of the lemma is obtained by using the fact that  $\psi(0) = 1$ ,  $\psi(x) \leq 0$ and  $\psi(x) \geq 0$  for all  $x \geq 0$ .

Now, we are in the position to define a new trigonometric kernel function for the SVM approach for the first time.

**Definition 3** Suppose  $x, x' \in X$ , and the feature map function  $\psi(x)$  given by Eq. 3, then a new trigonometric kernel function is defined by:

$$K(x, x') = \sin(\frac{\pi}{2 + \sigma ||x - x'||^2})$$
(4)

where  $\sigma$  is a positive real constant.

Using the symmetric property for the inner product, the first property of the new proposed kernel function can be expressed as follows:

$$K(x, x') = \sin(\frac{\pi}{2 + \sigma \|x - x'\|^2}) = \sin(\frac{\pi}{2 + \sigma \|x' - x\|^2}) = K(x', x), \quad (5)$$

It implies that the new trigonometric kernel function satisfies the symmetric property.

The following definition recalls another property of the kernel function called positive definite.

**Definition 4** (Positive definite kernel [16]) Let X be a nonempty set.  $K : X \times X \to \mathbb{R}$  is called a positive kernel function if and only if for any  $c \in \mathbb{R}^m$ , the following inequality is true:

$$\sum_{i,j=1}^{m} c_i c_j K_{i,j} \ge 0,$$
(6)

in which  $K_{i,j} = K(x_i, x_j)$ .

Remark 1 Suppose that  $K: X \times X \to \mathbb{R}$  is a kernel function. Then K is-so called strictly positive definite if and only if for any  $c \in \mathbb{R}^m$ , we have:

$$\sum_{i,j=1}^{m} c_i c_j K_{i,j} > 0.$$
(7)

Chasing the mentioned conditions is very challenging for some kernel functions. A common way to address these challenges is to utilize the matrix of the kernel function. Van Den Berg et al. [16] proved that a kernel function is positive definite if and only if the symmetric matrix K is positive definite. Note that, components of the matrix  $K_{i,j}$  are  $K(x_i, x_j)$ , as well as  $K_{i,i} = K(x_i, x_i) = 1$ for i = 1, 2, ..., n. It shows all elements on the main diagonal matrix K are one. Now, we are in the position to present the positive definite property for the new proposed kernel function, which delineates some properties of the new proposed kernel function.

Lemma 2 ([16]) Suppose  $K: X \times X \to \mathbb{R}$  is a kernel function. Then K is positive definite if and only if

$$\det(K(x_i, x_j)_{i,j \le n}) \ge 0, \qquad \{x_1, x_2, ..., x_n\} \subseteq X.$$

**Theorem 1** The new proposed kernel functions defined by Eq. 3 is positive definite.

*Proof* We begin by induction on n. Considering the structure of matrix K as:

$$K = \begin{pmatrix} 1 & l_{12} & \cdots & l_{1n} \\ l_{21} & 1 & \cdots & l_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ l_{n1} & l_{2n} & \cdots & 1 \end{pmatrix},$$
(8)

where  $0 \leq l_{ij} < 1$ ,  $i \neq j, 1 \leq i, j \leq n$ , and  $l_{ij} = K(x_i, x_j)$ . The property  $K(x_i, x_j) = K(x_j, x_i)$  implies K is a symmetric matrix. For the matrix K, we have:

$$n = 1, \qquad det(K_{1,1}) = 1 > 0;$$
  

$$n = 2, \qquad det(K_{2,2}) = 1 - l_{12}^2 > 0$$

in which  $K_{i,i}$  denotes a square matrix with dimension *i*. Let us suppose the theorem be true for n-1. We prove that  $det(K_{n,n}) > 0$ . Using the fact that  $K_{11} = 1 > 0$ and by subtracting  $K_{1,1}$  times the first column from the K-th column,  $K = 2, \ldots, n$ , the new matrix where the first column remained unchanged and other columns are changed by using the relation  $K'_{jk} = K_{jk} - K_{1k}K_{j1}$  for  $k \ge 2$  can be obtained. In addition, the new matrix has the same principal minors as the matrix K, so we have:

$$det(K'_{j,k \le p}) = det(K_{j,k \le p}), \quad p = 1, 2, \dots, n.$$
(9)

So, the matrix K' can be rewritten as:

$$K' = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & k'_{22} & \cdots & k'_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & k'_{2n} & \cdots & k'_{nn} \end{pmatrix},$$
(10)

Determinant of the matrix K can be calculated as follows:

$$det(K) = 1 * det \begin{pmatrix} k_{2,2} \cdots k_{2,n} \\ \vdots & \dots & \vdots \\ k'_{2,n} \cdots & k'_{n,n} \end{pmatrix} > 0,$$

where the inequality is obtained from the fact that determinate of matrix  $K_{n-1,n-1}$  is positive.

Theorem 1 implies that the proposed kernel functions is positive definite.

## 3 A new mixed kernel function

Mixing or combining several kernel functions into one kernel function was used to improve the performance of a single kernel function [17]. Here, we utilize this idea and suggest a new mixed kernel function, that is a convex combination of the Gaussian kernel function and the new trigonometric kernel function given by Eq. 1. For this purpose, we first recall a lemma allowing us to combine two kernel functions.

Lemma 3 Let f and g be two kernel functions. Then, a convex combination of f and g, i.e.,  $\beta f + (1 - \beta)g$  is also a kernel function for all  $\beta \in [0, 1]$ .

Using Lemma 3, we introduce a new mixed kernel function.

**Proposition 2** If  $\beta \in [0,1]$ , then the new mixed kernel function is given by:  $K_c(x,x') := \beta \sin(h_1(x,x')) + (1-\beta)e^{\frac{\|x-x'\|^2}{2\sigma^2}}, \quad \text{where} \quad h_1(x,x') = \frac{\pi}{2+\sigma\|x-x'\|^2}.$ (11)

Using Lemma 3, we have the following Remark.

 $Remark \ 2$  The kernel function defined by Eq. 11 is a positive definite kernel function.

Note that Eq. 11 is a kernel function and satisfies the properties of the kernel function. To compare behaviour of the new mixed kernel with the Gaussian kernel function and new trigonometric kernel function, we plot behaviour of these functions in Fig 3. This figure clearly shows that the value of the mixed fraction is always between the new proposed kernel function and the Gaussian function. In fact, the mixed kernel function can control the effects of changing the parameter  $\sigma$  well.

## 4 Numerical results

This section gives some numerical results of performing the kernel-SVM and kernel-SVR on various data sets.



**Fig. 3**: Behavior of the new mixed kernel function, the Gaussian kernel function and new trigonometric kernel function.

#### 4.1 SVM problem

The kernel SVM optimization problem can be express as:

$$\min_{\alpha} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{n} \alpha_j$$
  
S.t. 
$$\sum_{i=1}^{n} y_i \alpha_i = 0, \quad 0 \le \alpha_i \le C,$$

where n denotes the number of samples, C is the penalty or regularization parameter, and  $K(x_i, x_j)$  is a kernel function.

#### 4.2 Kernel functions

We implement the SVM based on the following kernel functions:

$$K_1(x, x_j) = (1 + xx_j)^p, \quad p = 2$$
 (12)

$$K_2(x, x_j) = \exp(-\frac{\|x - x_j\|^2}{2\sigma^2})$$
(13)

$$K_3(x, x') = \sin(\frac{\pi}{2 + \sigma ||x - x'||^2})$$
(14)

$$K_4(x, x') = \beta \sin(\frac{\pi}{2 + \sigma \|x - x'\|^2}) + (1 - \beta) e^{\frac{\|x - x'\|^2}{2\sigma^2}}, \quad \beta = \frac{1}{2}.$$
 (15)

#### 4.3 Parameter selection

We use the idea presented in [14] to find the best values for the parameters C and sigma. In fact, we select the best values for sigma and C from the following set:

$$c, \ \sigma \in \{2^{-5}, 2^{-4}, ..., 2^{10}\}$$

To apply [14]'s idea, we first calculate the minimum and maximum distances between samples in each class. After that, we choose large values for  $\sigma$  when

9

the dataset is compact and small values when the dataset is sparse. In this regard, we consider the following three datasets.

• Gaussian dataset: Gaussian dataset has 400 samples and two classes. Each sample has two features. This is a compact dataset because the maximum distances between samples of the first and second classes are 4.736 and 5.8477, respectively, and the minimum distances between samples of the first and second classes are 0.0412 and 0.1427, respectively. We run the kernel-SVM with different values of the parameters  $\sigma$  and C. The obtained results for C = 1 and C = 10 are shown in Tables 2 and 3. Large values for  $\sigma$  produce better results because the distance between samples is small.

**Table 2**: The number of misclassified training samples (# TrE.), number of misclassified testing samples (# TsE.), and number of support vectors (# SVs) of the new trigonometric kernel SVM using different values of parameter  $\sigma$  with C = 1 with Gaussian dataset.

Results	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$	$\sigma = 10$	$\sigma = 50$	$\sigma = 100$	$\sigma = 1000$
# SV	2	156	315	128	173	215	315
# TrE.	63	48	43	27	17	10	3
# TsE.	17	35	35	27	22	18	14

**Table 3**: The number of misclassified training samples (# TrE.), number of misclassified testing samples (# TsE.), and number of support vectors (# SVs) of the new trigonometric kernel SVM using different values of parameter  $\sigma$  with C = 10 with Gaussian dataset.

Results	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$	$\sigma = 10$	$\sigma = 50$	$\sigma = 100$	$\sigma = 1000$
# SV	267	180	199	93	317	179	312
# TrE.	108	66	62	31	14	3	0
# TsE.	31	24	22	33	35	19	16

• College dataset: The college dataset has 777 samples and two classes. Each sample has 17 features. The first class has 565 samples, and the second has 212 samples. We discovered that the maximum distances between samples of the first and second classes are 20, 111 and 47, 861, respectively, and the minimum distances between samples of the first and second classes are 0. Small values for  $\sigma$  can produce better results in this case. We run the kernel-SVM with various values of the sigma and C parameters. The obtained results for C = 1 and C = 10 are shown in Tables 4 and 5.

**Table 4**: The number of misclassified training samples (# TrE.), number of misclassified testing samples (# TsE.), and number of support vectors (# SVs) of the new trigonometric kernel SVM using different values of parameter  $\sigma$  with C = 1 with College dataset.

Results	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$	$\sigma = 10$	$\sigma = 50$	$\sigma = 100$	$\sigma = 1000$
# SV	621	621	621	621	621	621	621
# TrE.	0	0	0	0	0	0	0
# TsE.	33	29	30	32	31	33	39

**Table 5**: The number of misclassified training samples (# TrE.), number of misclassified testing samples (# TsE.), and number of support vectors (# SVs) of the new trigonometric kernel SVM using different values of parameter  $\sigma$  with C = 10 with College dataset

Results	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$	$\sigma = 10$	$\sigma = 50$	$\sigma = 100$	$\sigma = 1000$
# SV	621	621	621	621	621	621	621
# TrE.	0	0	0	0	0	0	0
# TsE.	31	31	28	228	34	34	38

• **Tic-Tac-Toc dataset:** This dataset has two classes: the first class has 323 samples, and the second has 626 samples and nine features. The maximum distances between samples of the first and second classes were 4.90 and 4.47, respectively. In this case, we need large values for parameter  $\sigma$ . We present a summary of the results in Table 6.

**Table 6**: The number of misclassified training samples (# TrE.), number of misclassified testing samples (# TsE.), and number of support vectors (# SVs) of the new trigonometric kernel SVM using different values of parameter  $\sigma$  with C = 1 with the Tic-Tac-Toc dataset

Results	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$	$\sigma = 10$	$\sigma = 50$	$\sigma = 100$	$\sigma = 1000$
# SV	674	348	476	751	766	766	766
# TrE.	37	8	0	0	0	0	0
# TsE.	47	35	28	24	19	16	22

#### 4.4 Datasets

We perform the SVM on the 24 datasets with different number of samples and features<sup>1</sup>. There are two data samples are used in all experiments; 80% of the data were used to train SVM and 20% of the data were used to test model. Table 7 gives details about 24 datasets used in this section. Note that "Name"

 $<sup>^{1}</sup> https://www.openml.org/search?q=gisette\&type=data\&sort=runs\&status=active$ 

denotes the name of the dataset, "F" presents the number of the features of the dataset, and "N" is the number of samples.

i	Name	$\#\mathbf{F}$	₩N	i	Name	₩F	₩N
1	Banana	2	5300	13	Sonar	60	208
2	Bill-authentication	4	1372	14	SPECT	22	80
3	Carseats	10	400	15	Weekly	8	1089
4	College	17	777	16	Wisc-bc-data	31	569
5	Column-2C-weka1	6	310	17	PhpVDlhKL	230	64
6	Dataset-spine	12	309	18	Fri-c1-100-10	11	100
7	Ex2data1	2	100	19	Svmguide1	4	3089
8	Ex2data2	2	118	20	tic-tac-toe.data	10	958
9	GaussianData	2	400	21	analcatdata_lawsuit	5	264
10	GSE58606-data1	1927	133	22	phpAmSP4g	31	569
11	Numeric-sequence	28	2400	23	kc2	22	5228
12	Pimalndians1	8	390	24	haberman	4	306

Table 7: Data Sets

#### 4.5 Results of performin SVM

Classification accuracy for SVM based on four kernel functions given by Eqs. 12–15 are demonstrated in Fig 4. Note that x-axis denotes the number of dataset (i.e., *i*) in Table 7. In addition, " $K_1$ " shows the accuracy results for polynomial kernel-SVM, " $K_2$ " denotes the results for the Gaussian kernel function, and " $K_3$ " and " $K_4$ " are results for the new trigonometric kernel function and mixed kernel function given by Eq. 15, respectively.



Fig. 4: Accuracy results of performing SVM based on the four kernel functions on 24 data-set denoted in Table 7.

#### 4.6 Kernel SVR

Here we investigate efficiency of the proposed kernel functions for SVR approach. In this regard, we implement the SVR based on the kernel functions given by Eqs. 14 and 15 on a set of 200 data points generated by using the following function:

$$f(x) = \sin(x) * \exp(-0.2 * x) + \epsilon$$

where  $\epsilon$  is a random number.

Fig. 5 demonstrates the behaviour of the SVR based on two mentioned kernel functions. This figure shows SVR based on the new proposed kernel function and mixed kernel function can fit the best curve within an  $\epsilon$  tube to predict the model.



**Fig. 5**: Behaviour of SVR based on the kernel functions given by Eq. (14) and Eq. (15).

#### 4.7 Discussion

Based on Tables 2-7, Figs. 4 and 5, we can conclude that:

- A variety of datasets with different numbers of samples and features have been used.
- The average classification accuracies of the SVM method based on the polynomial, Gaussian, new trigonometric, and mixed kernel functions are 77.93%, 82.79%, 84.88%, and 87.64% respectively.
- The new trigonometric kernel-SVM improved the accuracy of the Gaussian kernel-SVM (from 82.79% to 84.88%) based on average accuracy.
- The new mixed kernel function improved the classification accuracy of the SVM. According to the results, it improves the 4.85% accuracy of the Gaussian kernel-SVM and the 2.66% accuracy of the new proposed kernel-SVM.
- The new kernel-SVM achieves better classification accuracy for 10 datasets than the Gaussian kernel-SVM.
- The new proposed kernel function achieves the best results with small values of  $\sigma$  when the dataset is sparse and with large values of  $\sigma$  when the dataset is compact.

• The SVR based on the new proposed kernel functions is able to predict an appropriate graph with the minimum error.

## 5 Concluding remarks

In this paper, we introduced a new trigonometric kernel function for machine learning. Various properties of the proposed kernel function were investigated. Then we combined the proposed kernel function with the Gaussian kernel function and introduced a new mixed kernel function. The numerical results confirmed that the new proposed trigonometric kernel function improved the classification accuracy of the SVM for the most considered datasets. Moreover, the mixed kernel function achieved the best results in terms of classification accuracy.

## Declarations

### Acknowledgements

The authors would like to thank the editors and anonymous reviewers for their constructive comments.

## Funding

No funding was received for conducting this study.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and material

The dataset used in this paper is available from the following link  $\rm https://paperswithcode.com/dataset/orl$ 

### **Conflict of Interests**

The authors have no conflicts of interest to declare that are relevant to the content of this article.

### **Competing interests**

The authors have declared that no competing interests exist.

## References

- Cortes, C., Vapnik, V.: Support-vector networks. Machine learning 20(3), 273–297 (1995)
- [2] Byvatov, E., Schneider, G.: Support vector machine applications in bioinformatics. Applied bioinformatics 2(2), 67–77 (2003)
- [3] Vatsa, M., Singh, R., Noore, A.: Improving biometric recognition accuracy and robustness using dwt and svm watermarking. IEICE Electronics Express 2(12), 362–367 (2005)
- [4] Moulin, L., Da Silva, A.A., El-Sharkawi, M., Marks, R.J.: Support vector machines for transient stability analysis of large-scale power systems. IEEE Transactions on Power Systems 19(2), 818–825 (2004)
- [5] Doucet, J.-P., Barbault, F., Xia, H., Panaye, A., Fan, B.: Nonlinear svm approaches to qspr/qsar studies and drug design. Current Computer-Aided Drug Design 3(4), 263–289 (2007)
- [6] Tharwat, A.: Parameter investigation of support vector machine classifier with kernel functions. Knowledge and Information Systems 61(3), 1269– 1302 (2019)
- [7] Zhou, J., Tian, Y., Luo, J., Zhai, Q.: Novel non-kernel quadratic surface support vector machines based on optimal margin distribution. Soft Computing 26(18), 9215–9227 (2022)
- [8] Liang, Z., Shao, S., Lv, Z., Li, D., Sleigh, J.W., Li, X., Zhang, C., He, J.: Constructing a consciousness meter based on the combination of non-linear measurements and genetic algorithm-based support vector machine. IEEE Transactions on Neural Systems and Rehabilitation Engineering 28(2), 399–408 (2020)
- [9] Apsemidis, A., Psarakis, S., Moguerza, J.M.: A review of machine learning kernel methods in statistical process monitoring. Computers & Industrial Engineering 142, 106376 (2020)
- [10] Smola, A.: Advances in kernel methods: Support vector learning (1999)
- [11] Wang, Z., Zhu, C., Niu, Z., Gao, D., Feng, X.: Multi-kernel classification machine with reduced complexity. Knowledge-Based Systems 65, 83–95 (2014)
- [12] 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 Recognition 84, 211–225 (2018)

- 16 A new trigonometric kernel function for support vector machine
- [13] Hoang, D.T., Kang, H.J.: A motor current signal-based bearing fault diagnosis using deep learning and information fusion. IEEE Transactions on Instrumentation and Measurement 69(6), 3325–3333 (2019)
- [14] Tharwat, A.: Parameter investigation of support vector machine classifier with kernel functions. Knowledge and Information Systems 61(3), 1269– 1302 (2019)
- [15] Feng, X., Ling, X., Zheng, H., Chen, Z., Xu, Y.: Adaptive multi-kernel svm with spatial-temporal correlation for short-term traffic flow prediction. IEEE Transactions on Intelligent Transportation Systems 20(6), 2001–2013 (2018)
- [16] Van Den Berg, C., Christensen, J.P.R., Ressel, P.: Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions vol. 100. Springer, ??? (2012)
- [17] Kung, S.Y.: Kernel Methods and Machine Learning. Cambridge University Press, ??? (2014)