GENERALIZED LEGENDRE POLYNOMIALS FOR SUPPORT VECTOR MACHINES (SVMS) CLASSIFICATION

Ashraf Afifi[1] and E.A.Zanaty[2]

[1]Department of Computer Engineering, Computers and Information Technology College, Taif University, Al-Hawiya 21974, Kingdom of Saudi Arabia

[2]Computer Science Dept., Faculty of Science, Sohag University, Sohag, Egypt.

ABSTRACT

In this paper, we introduce a set of new kernel functions derived from the generalized Legendre polynomials to obtain more robust and higher support vector machine (SVM) classification accuracy. The generalized Legendre kernel functions are suggested to provide a value of how two given vectors are like each other by changing the inner product of these two vectors into a greater dimensional space. The proposed kernel functions satisfy the Mercer's condition and orthogonality properties for reaching the optimal result with low number support vector (SV). For that, the new set of Legendre kernel functions could be utilized in classification applications as effective substitutes to those generally used like Gaussian, Polynomial and Wavelet kernel functions. The suggested kernel functions are calculated in compared to the current kernels such as Gaussian, Polynomial, Wavelets and Chebyshev kernels by application to various non-separable data sets with some attributes. It is seen that the suggested kernel functions could give competitive classification outcomes in comparison with other kernel functions. Thus, on the basis test outcomes, we show that the suggested kernel functions are more robust about the kernel parameter change and reach the minimal SV number for classification generally.

KEYWORDS

Legendre Polynomials, Kernel Functions, Functional Analysis, SVMS, Classification Problem.

1. INTRODUCTION

Support Vector Machines (SVMs) has become famous machines for data classification as a result of use for the vast data set and practical for application [1-3]. The operation of SVMs is based upon selecting kernel functions [4-6]. Picking various kernel functions will give out various SVMs [7- 9] and may turn out to be in various performances [10-11]. Some effort has been carried out on curbing kernels by handling prior knowledge; however, the optimal selection of a kernel for a provided problem is yet a free research crisis [12]. Chapelle and Schölkopf [13] suggested a kernel to use constant transformations. The disadvantage here is that they are most probably just suitable for linear SVM classifiers. Hastie et al. [14] had given comparisons among multi-class SVMs algorithms when implied to defy data set. Zanaty et al. [15-17] mixed GF and RBF functions to attain new kernel functions that can make use of their corresponding power. In [18-19], the Hermite kernel functions were defined for advancing the operation of SVMs in a variety of applications. Meng and Wenjian [20] proposed orthogonal polynomials to advance generalization performance in both classification and regression duties. The particular estimation of crossing kernel SVMs which is logarithmic in time was shown in Maji et al. [21]. They proved that the procedure is approximately in complex and the classification efficacy is passable, but the

runtimes are symbolically boosted in comparison with the implanted radial bases function (RBF) and polynomial kernel (POLY) because of the great number of SVs for every classifier [14, 21]. Ozer et al. [22] presented kernel functions coming from the Chebyshev polynomials. They built various kernel functions so that they can catch the highly non-linear boundaries in the Euclidian space. In Jiang and Ching [23], the managed kernel learning with an SVM classifier was outstandingly applied in biomedical diagnosis like segregating various types of tumour tissues for noisy Raman Spectra, see [24-25] for further details.

The problems of data classification remain in picking the most convenient kernel of SVMs for a specific application, specifically since various functions and parameters can have vastly different operations [19-22]. A vital research field in SVMs is to establish an effective kernel function for constructing SVMs in a particular application, specifically because of variable current application which will demand various methods [22].

In this paper, Legendre kernel functions are constructed to advance the classification certainty of SVMs for both linear and non-linear data groups. We sustain a group of Legendre kernel functions based on advancing SVMs classification certainty. The class of Legendre kernel functions fulfils mercer conditions and gives competitive operation in comparison with all other typical kernel functions with the same standard of the simulation datasets. The suggested kernels can be used for categorizing compound data which have numerous properties.

The remainder of the paper is arranged in this way: In section 2, SVM classifications are elaborated. The kernel theory is deliberated in section 3. The generalized Legendre kernels are introduced in section 4. Section 5 shows functional examination on the presented Legendre kernels. Experimental and comparative outcomes are given in section 6. Lastly, section 7 presents the conclusion.

2. SUPPORT VECTOR MACHINE (SVM)

SVMs became popular for data classification and regression. They used to segregate the data by constructing two hyperplanes. If you have a set of N points $x_k \in R^n, k = 1, ..., N$ to be conjoined with a label $y_k \in \{+1, -1\}$ which can be categorized the data into one of two groups. According to SVMs formulation, the classifier $y(x)$ will be the design of a hyper-plane $w^T x_k + b$ which shows optimum separation $\dfrac{2}{\|w\|^2}$ between points x_k belonging to the two classes. This introduces an optimization issue of the form:

$$Q : \min_{w,b} \frac{1}{2} w^T w \text{ s.t. } y_k [w^T x_k + b] \geq 1,$$

(1)

where the $\dfrac{1}{2} w^T w$ term stands for a cost function to be minimized to maximize segregation.

The restraints are formulated so that the nearest points x_k with labels [either +1 or -1] are (with suitable input space scaling) at least $\dfrac{1}{\|w\|^2}$ away from the separating hyper-plane. Nevertheless,

for the Least-Squares SVM classification, changes are done so that at the target value, an error variable ek is enabled so that misclassifying can be gone along with in case of coinciding distributions and the subsequent optimization problem is framed in the primal weight space for a given training set $\{x_k, y_k\}_{k=1}^{N}$

$$Q : \min_{w,b,c} J_p(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2$$

(2)

Along with the N restrains as given in Eq.(3). This formula included the tradeoff between a cost function term and a sum of squared errors governed by the trade-off parameter γ.

$$y_k [w^T \phi(x_k) + b] = 1 - e_k, \ k = 1,..,N$$

(3)

To solve this 'primal minimization' issue, we design the dual maximization of Eq.(2) using the Lagrangian form:

$$D : \max_{\alpha} \ L(w,b,c;\alpha),$$

(4)

Where

$$L = J_p(w,e) - \sum_{k=1}^{N} \alpha_k \{ y_k [w^T \varphi(x_k) + b] - 1 + e_k \}$$

(5)

and αk are Lagrange multipliers. The circumstances for optimality are given by

$$\begin{cases} \dfrac{\partial L}{\partial w} = 0 \ \rightarrow w = \sum_{k-1}^{N} \alpha_k y_k \varphi(x_k) \quad \ldots \\[2mm] \dfrac{\partial L}{\partial b} = 0 \ \rightarrow \sum_{k=1}^{N} \alpha_k y_k = 0 \\[2mm] \dfrac{\partial L}{\partial e_k} = 0 \ \rightarrow \alpha_k = \gamma e_k. \quad k = 1,..,N \\[2mm] \dfrac{\partial L}{\partial \alpha_k} = 0 \ \rightarrow y_k w^T \varphi(x_k) + b] - 1 + e_k = 0, \\[2mm] \qquad \text{where}, \quad k = 1,......,N \end{cases}$$

(6)

After removal of the variables w and e we get this solution:

$$\begin{bmatrix} 0 & y^T \\ y & \Omega + I/\gamma \end{bmatrix} \begin{bmatrix} b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 1_v \end{bmatrix}$$

(7)

where $y = y_1,..; y_N]$, $I_v = [1;...;1]$ and $\alpha = [\alpha_1;...;\alpha_N]$;

The kernel trick is applied here as follows:

$$\Omega_{kl} = y_k y_L \varphi(x_k)^T \varphi(x_L) = y_k y_L K(x_k, x_L) \quad k, L = 1,...,N$$

(8)

The outcoming LS-SVM model for classifier turns out:

$$y(x) = sign\left[\sum_{k=1}^{N} \alpha_k y_k K(x_k, x) + b\right] \tag{9}$$

where α_k, b are the answer to the linear system presented by Eq.(7) and N stand for the number of non-zero Lagrange multipliers α_k, called SVs.

According to Eq.(9), the kernel functions have been applied on the pairs of elements separately, for a given pair of two input vectors x and z, the outcoming kernel can be formulated as:

$$K_j(x,z) = \langle \varphi(x_j), \varphi(z_j) \rangle \tag{10}$$

Where $K_j(.)$ is the kernel function that is evaluated on the jth elements of the vector pair $x \, and \, z$. More various kernel functions were found in literature as in Table (1).

Table 1: Expressions of kernels list

kernel	kernel expression
Polynomials kernel [14]:	$K(x.z) = \left(\dfrac{\langle x\,z \rangle + 1}{\beta}\right)^n$
Gaussian kernel [25]	$K(x.z) = e^{\dfrac{-\|x_i - x_j\|^2}{\sigma^2}}$
Wavelet kernel [22]:	$K(x.z) = \prod_{j=1}^{m}\left(Cos\left(1.75\dfrac{x_j - z_j}{a}\right)\exp\left(-\dfrac{\|x_j - z_j\|^2}{2a^2}\right)\right)$, m is order of polynomials.
Chebyshev kernel [7]:	$K(x.z) = \prod_{j=1}^{m}\dfrac{\sum_{i=0}^{n} T_i(x_j)T_i(z_j)}{\sqrt{1 - x_j z_j}}$, m is polynomials order and $T(x)$ is the Chebyshev polynomials.

3. PROPOSED KERNEL FUNCTIONS

In the suggested modified Henon map will be defined in terms of two basic processes namely ciphering and deciphering. To advance the classification certainty of SVMs, various kernel functions are required for various applications. We figured that Legendre function will ensure to be effective kernels for numerous applications. From the solution of Legendre's differential equation, the formula of Legendre polynomials may be written down using Rodrigues' formula:

$$Q_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}[(x^2 - 1)^n] \tag{11}$$

By differentiating $(n + 1)$ times both sides of the identity:

$$(x^2 - 1)\frac{d}{dx}(x^2 - 1)^n = 2nx(x^2 - 1)^n$$

and applying the general Leibniz rule for repeated differentiation.

3.1. Legendre Recurrence

From Eq.(16), expanding for the first two terms gives:

$$Q_0(x) = 1, \qquad Q_1(x) = x,$$

Eq.(11) is discriminated concerning t on both sides to acquire more terms with no use of direct broadening of the Taylor series, and reorganized to attain:

1.
$$\frac{x - t}{\sqrt{1 - 2xt + t^2}} = (1 - 2xt + t^2)\sum_{n=0}^{\infty} nQ_n(x)t^{n-1}$$

Replacing the quotient of the square root with its description in (11), and equating the coefficients of powers of t in the outcoming expansion gives Bonnet's recursion formula:

$$\theta(t, x) = \frac{1}{\sqrt{1 - 2xt + t^2}} \equiv \sum_{n=0}^{\infty} Q_n(x)t^n \tag{12}$$

Theorem 1: Taylor series expansion of Legendre's differential equation:

$$\theta(t, x) = \frac{1}{\sqrt{1 - 2xt + t^2}} \equiv \sum_{n=0}^{\infty} Q_n(x)t^n$$

can be represented as the following recurrence relation:

2.
$$(n + 1)Q_{n+1}(x) = (2n + 1)xQ_n(x) - nQ_{n-1}(x).$$

Proof: differentiate Eq. (12) with respect to t .

$$\frac{\partial g}{\partial t} = \frac{2(t - x)}{-2(1 - 2xt + t^2)^{\frac{3}{2}}} = \sum_{j=1}^{\infty} jQ_j(x)t^{j-1}$$

$$\Rightarrow (x - t)\sum_{j=0}^{\infty}(j + 1)Q_{j+1}(x)t^j = (1 - 2xt + t^2)\sum_{j=1}^{\infty} jQ_j(x)t^{j-1}$$

$$\Rightarrow x\sum_{j=0}^{\infty} Q_j(x)t^j - \sum_{j=1}^{\infty} Q_{j-1}(x)t^j = \sum_{j=0}^{\infty}(j + 1)Q_{j+1}(x)t^j - 2x\sum_{j=1}^{\infty} jQ_j(x)t^j + \sum_{j=2}^{\infty}(j - 1)Q_{j-1}(x)t^j$$

Compare the coefficients of t^j:

(j=0)

$$Q_0(x) = 1, \qquad Q_1(x) = x,$$

(j=1)

$$xQ_1 - Q_0 = 2Q_2 - 2xQ_1$$

$$\Rightarrow x^2 - 1 = 2Q_2 - 2x^2 \Rightarrow Q_2 = (3x^2 - 1)/2$$

$(j \geq 2)$

$$xQ_j - Q_{j-1} = (j+1)Q_{j+1} - 2xjQ_j + (j-1)Q_{j-1}$$

$$\Rightarrow (j+1)Q_{j+1} = (2j+1)xQ_j - jQ_{j-1}$$

3.2. Orthogonally of Legendre Function

Legendre polynomials $Q_n(x)$, $n = 0, 1, 2, 3, ..., N$ form a whole orthogonal group on the interval $-1 \leq x \leq 1$.

Theorem 2: Let $P_n(x)$ denote the Legendre polynomial of degree n. Then:

$$\int_{-1}^{1} Q_m(x)Q_n(x)dx = 0 \quad if \ m \neq n$$

Proof: We know that the polynomials $Q_n(x)$ and $Q_m(x)$ satisfy:

$$((1-x^2)Q'_n(x))' + n(n+1)Q_n(x) = 0 \quad and \tag{13}$$

$$((1-x^2)Q'_m(x))' + m(m+1)Q_m(x) = 0 \tag{14}$$

Multiplying equation (1) by $Q_m(x)$ and equation (2) by $Q_n(x)$, we can obtain:

$$(n(n+1) - m(m+1)Q_n(x)Q_m(x) = ((1-x^2)Q'_m(x))'Q_n(x) - ((1-x^2)Q'_n(x))'Q_m(x),$$

Therefore,

$$(n(n+1) - m(m+1)\int_{-1}^{1} Q_n(x)Q_m(x)dx = \int_{-1}^{1}\left(((1-x^2)Q'_m(x))'Q_n(x) - ((1-x^2)Q'_n(x))'Q_m(x)\right)dx$$

$$= -\int_{-1}^{1}(1-x^2)Q'_m(x))'Q_n(x)dx + (1-x^2)Q'_n(x))'Q_m(x)\Big|_{x=-1}^{x=1}$$

$$+ \int_{-1}^{1}(1-x^2)Q'_m(x))'Q_n(x)dx + (1-x^2)Q'_n(x))'Q_m(x)\Big|_{x=-1}^{x=1} = 0.$$

Since

$$n \neq m, n(n+1) \neq m(m+1)$$

Then we have:

$$\int_{-1}^{1} Q_m(x)Q_n(x)dx = 0 \quad if \ m \neq n$$

This completes the proof.

3.3. Generalized Legendre Kernels

Here, we are suggesting a general method of conveying the kernel function to resolve the vagueness on how to apply Legendre kernels. As of what we know, there was a preceding work illustrating the Legendre polynomials for vector inputs recursively. thus for vector inputs, we illustrate the generalized Legendre polynomials as:

$$Q_0(x) = 1$$
$$Q_1(x) = x$$
$$(j+1)Q_{j+1} = (2j+1)xQ_j - jQ_{j-1}$$
$$Q_{j+1}(x) = ((2j+1)/(j+1))xQ_j(x) - (j/(j+1))Q_{j-1}(x), \forall j \geq 0$$

Therefore, the generalized Legendre, $Q_j(x)$, yields a row vector; if not, it gives a scalar value. Therefore, with the use of generalized Legendre polynomials, we describe generalized n^{th} order Legendre kernel as

$$K(x,z) = \sum_{j=0}^{n} Q_j(x)Q_j^T(z) \tag{15}$$

Where x and z are m-dimensional vectors. In Eq. (13), the denominator should be greater than zero. To fulfill (14), as every element in x and z vectors has a value in between [-1,1], the maximum value for the inner product $< x.z >$ is equal to $\sum_{I=1}^{m} I = m$ thus minimum a value will be equal to m which is the dimension of input vector x.

Consequently, the 5th order generalized Legendre kernel could be presented as:

$$k(x,z) = 1 + c + \frac{(5a-2)(5b-2)}{3} + \frac{c(35a-23)(35b-23)}{144} + \frac{(1181.25a^2 - 38.325a + 2.66)(1181.25b^2 - 38.325b + 2.66)}{50625}$$

$$+ c\frac{(93555a^2 - 205686a - 17568)(93555b^2 - 205686b - 17568)}{9720} +$$

$$\frac{(243243a^3 - 598571.1a^2 + 124132a - 55370.88)(243243b^3 - 598571.1b^2 + 12413b - 55370.88)}{13500}$$

Where $a = < x.x >, b = < z.z >$, and $c = < x.z >$. In addition, the first 4th order kernel functions are listed in Table 1.

Table 2: List of the generated Legendre kernel functions up to 4th order.

Parameter : n	Kernel function: $k(x, z)$, Parameter : n
0	1
1	1+c
2	$1+c+\dfrac{(5a-2)(5b-2)}{3}$
3	$1+c+\dfrac{(5a-2)(5b-2)}{3}+\dfrac{c(35a-23)(35b-23)}{144}$
4	$1+c+\dfrac{(5a-2)(5b-2)}{3}+\dfrac{c(35a-23)(35b-23)}{144}+$ $\dfrac{(1181.25a^2-38.325a+2.66)(1181.25b^2-38.325b+2.66)}{50625}$

4. FUNCTIONAL ANALYSIS

In order to reproduce proposed kernels, we use functional analysis as described in [26] to prove Mercer's theorem conditions. The mapping ϕ could be designed from the eigen function decomposition of k. With respect to Mercer's work [27-28], it is known that if k is the symmetrical and continuous kernel of an integral operator $O_k : L^2 \to L^2$, in such a way that:

$$O_{kg}(x) = \int K(x, z)\theta(z)dz$$

is positive, i.e.,

$$\int K(x, z)\theta(x)\theta(z)dxdz \geq 0 \qquad \forall \theta \in L^2 ,$$

then k can be expanded into a uniformly convergent series $K(x, z) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(x)\varphi_i(z)$, with $\lambda_i \geq 0$. In this case, the mapping from input space to feature space produced by the kernel is expressed as $\varphi : x \mapsto (\sqrt{\lambda_1}\varphi_1(x), \sqrt{\lambda_2}\varphi_2(x),...)$ such that k acts as the given dot product, i.e., $(\varphi(x), \varphi(z)) = \varphi^T(x)\varphi(z) = K(x, z).$

Theorem 3: A kernel is a valid SVM kernel; if it satisfies the Mercer Conditions [29].

Proof: If the kernel does not fulfill the Mercer Conditions, SVM might not derive the best parameters, but instead it might bring up suboptimal parameters. Additionally, in case of the Mercer conditions not being fulfilled, the Hessian matrix for the optimization portion might not be positive straightforward. Thus we inspect if the generalized Legendre kernel fulfill the Mercer conditions.

To be a credible SVM kernel, for any finite function $\theta(x)$, the subsequent integration must always be non-negative for the given kernel function $k(x, z)$ [1]:

$$\iint K(x, z)\theta(x)\theta(z)dxdz \geq 0$$

(16)

Where

$$k(x, z) = \sum_{j=0}^{n} Q_j(x)Q_j^T(z) = Q_0(x)Q_0^T(z) + Q_1(x)Q_1^T(z) + \dots\dots + Q_n(x)Q_n^T(z) \quad (17)$$

Regard that $\theta(x)$ is a function where $\theta: R^m \rightarrow R$, then we can calculate and validate the Mercer condition for $k(x, z)$ as follows by speculating each element is independent of others:

$$\iint k(x, z)\theta(x)\theta(z)dxdz = \iint \sum_{j=0}^{n} Q_j(x)Q_j^T(z)\theta(x)\theta(z)dxdz$$

$$= \sum_{j=0}^{n} Q_j(x)Q_j^T(z)\theta(x)\theta(z)dxdz = \sum_{j=0}^{n} \left[\int Q_j(x)g(x)dx \int Q_j^T(z)\theta(z)a \right. \quad (18)$$

$$= \sum_{j=0}^{n} \left[\left(\int Q_j(x)\theta(x)dx \right)\left(\int Q_j^T(x)\theta(x)dx \right) \right] \geq 0$$

Proof: Let $K_i (i = 1,..., M)$ be Mercer kernels and let

$$K(x, z) = \sum_{i=1}^{M} a_i K_i(x, z),$$

where $a_i \geq 0$ is a nonnegative constant. According to Mercer's theorem, we have

$$\int K_i(x, z)\theta(x)\theta(z)dxdz \geq 0, \forall \theta \in L^2, i = 1,..., M.$$

(19)

By taking the sum of the positive combination of (19) with coefficients a_i over i, one obtains

$$\sum_{i=1}^{M} a_i \int K_i(x, z)\theta(x)\theta(z)dxdz \geq 0, \quad \forall g \in L^2.$$

(20)

Therefore, one reaches

$$\int K(x, z)\theta(x)\theta(z)dxdz \geq 0, \quad \forall g \in L^2.$$

In particular, if $\sum_{i=1}^{M} a_i = 1, (a_i \geq 0)$, then we consider $K(x, z)$ in (20) as the convex mixing of the positive definite kernels $K_i(x, z)$. This kind of kernel can find many applications in practice.

Theorem 5: The product of Mercer kernels is also a Mercer kernel.

Proof: is like that of the previous theorem.

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

The multi-class problem is described as the categorizing problem that has numerous classes. To prolong these classifiers to take care of many classes, the target of this method is to map the generalization capabilities of the binary classifiers to the multi-class domain. Multi-class SVMs are ordinarily applied by mixing a couple of two class SVMs. In multi-class experimentations, we have trained SVM for every class individually so that one is against all. In every experiment, we utilized the SVM toolbox accessible at [32]. The multi-class problem is described as a classification problem which has numerous classes or characters. Current SVMs [24] are binary classifiers, i.e., they could categorize two classes. To be capable of dealing with various classes (greater than 2), the current classifiers should be prolonged. The target is to depict the generalization abilities of the binary classifiers to the multi-class domain. Multi-class SVMs are usually applied by merging several two - class SVMs. The classifier is constructed to read two input data files, the training data and the test data (for more details see [11, 18]). Every file is arranged as records, each of which is made up of a vector of attributes x: $x_{1i} = (x_1, x_2, ..., x_m)$ followed by the target y: $x_{2i} \in (y_1, y_2, ..., y_c)$ where c is the number of classes m and is the number of attributes. The SVM designs binary classifiers, and utilizes the training data to find the maximum separating hyperplane.

The classification experimentations are carried out on number image segmentation data sets like Brickface, Foliage, Sky, Cement, Window, Path and Grass data set [30-31]. The data has 7 diverse image classes. It has 210 data for training and a different 2100 data for evaluating. Every vector has 18 elements having diverse maximum and minimum values. For the training, we got 30 data for the class (+1) and 180 data for the class (1) and likely for testing, we have 300 against 1800 data, correspondingly for every class. With the test step, the kernel functions presented various performance values on various classes and there was winning kernel presenting the optimal performance one very class as shown in Table (3). The suggested Legendre kernel operated better than the standard. The optimal performance values having the least SV numbers are presented in bold. Table (3) illustrates the test results for every class with various kernel functions. We carry out some evaluations to compare the suggested kernel with its preceding opposite as well as the Gaussian (GF) [25], polynomial (POLY) [14], Wavelet [7] and Chebyshev [22] kernel functions. The operation of the suggested kernel with SVMs according to classification accuracy (ACC) and kernel parameter against SV, is calculated by implementation to data sets in Table (3). As shown in Table (3), the generalized Legendre kernel results present better generalization capability than the current GF, POLY, Wavelet and Chebyshev kernels. For instance, the optimal ACC is retrieved for Brickface, Foliage, Window, Path, Grass with the least SV to be 17, 6, 4, 6 and 15 correspondingly. Even though the ACC of the generalized Legendre kernel and GF kernel for Foliage data is similar, we emphasize that SV of the generalized Legendre kernel has the least. In Figs.(1-7), we give the ACC against SV of the POLY, GF, Wavelet, Chebyshev, and generalized Legendre kernels for Brickface, Foliage, Sky, Cement, Window, Path and Grass data set. The generalized Legendre kernel functions present the minimal SV 4 while maintaining the generalization ability right for the dataset. The relation between ACC and SV showed that, as the SV increases, the ACC increases and asymptotically arrive at a high performance value.

Figs. (8-12) describe the relation between kernel parameter vs. SV for GF, POLY, Wavelet and Chebyshev kernel functions when these methods are implemented on Brickface, Foliage, Sky, Cement, Window, Path and Grass data set correspondingly. These figures prove that as the kernel parameter increases, the Chebyshev, Legendre and wavelet kernel increase their operation and asymptotically reach allow performance value. While as the kernel parameter increases, the Chebyshev and wavelet kernels need more SV than Legendre kernel. During the evaluations, we witnessed that the generalized Legendre kernel function reaches the minimal SV number in general.

Table 3: Data classification results with different kernel functions.

| Data sets	GF			POLY			Wavelet			Chebyshev			Legendre		
	SV No.	ACC	σ	SV no.	ACC	n	SV no.	ACC	a	SV no.	ACC	n	SV no.	ACC	n
Brickface	61	0.960	0.4	112	0.969	2	44	0.997	1.4	23	0.997	**0**	**17**	**0.999**	15
Sky	7	1.0	4.3	14	1.0	2	**5**	**1.0**	1.5	5	0.99	**0**	6	1.0	3
Foliage	106	0.98	1.4	122	0.999	8	99	0.989	2.4	122	0.971	**0**	**6**	**0.999**	2
Cement	88	0.989	0.5	78	0.971	12	77	0.997	**1.4**	**65**	**0.999**	4	16	0.998	1
Window	49	0.957	0.4	80	0.969	7	34	0.99	1.4	34	0.985	**4**	**4**	**0.99**	**2**
Path	44	0.969	0.43	68	0.959	6	66	0.98	1.5	22	0.976	**3**	**6**	**0.98**	**1**
Grass	22	0.986	11	25	0.964	2	45	0.979	1.5	54	0.975	**0**	**15**	**0.990**	2
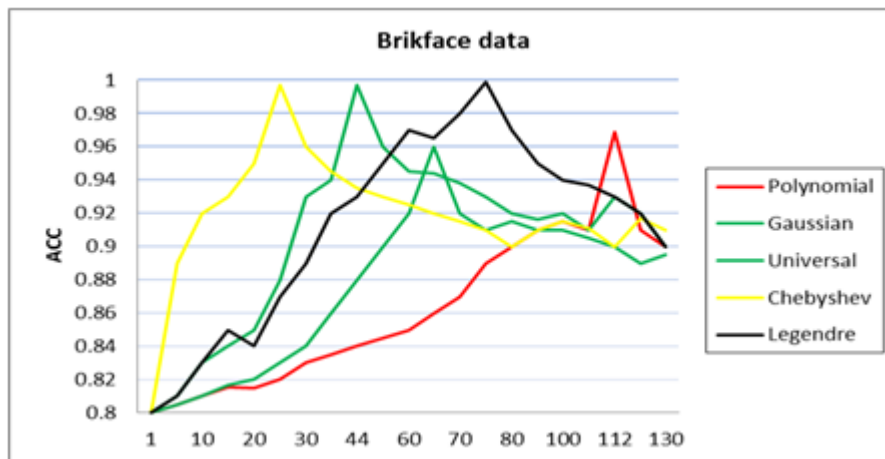
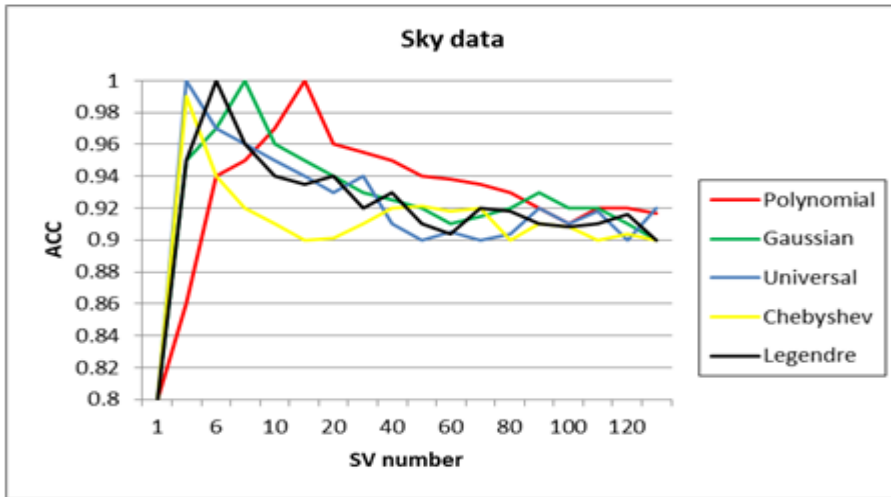Figure 1: SV vs. ACC for Brickface data.
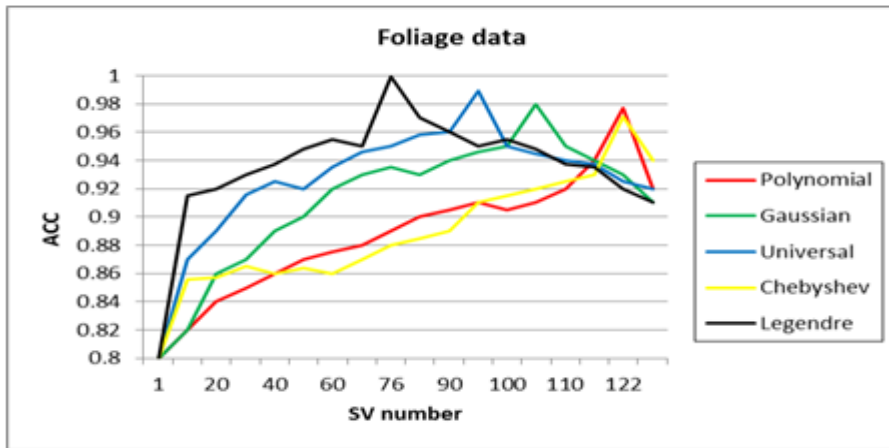
Figure 2: SV vs. ACC for Sky data.
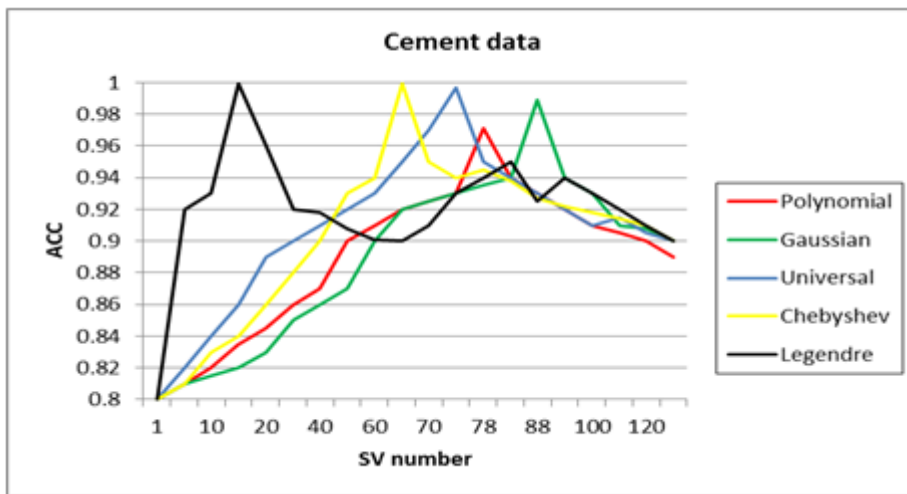

Figure 3: SV vs. ACC for Foliage data.

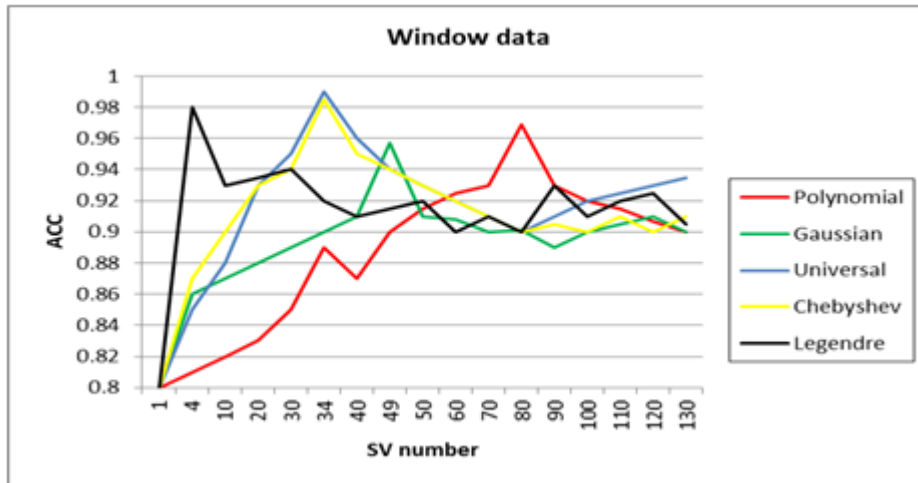Figure 4: SV vs. ACC for Cement data
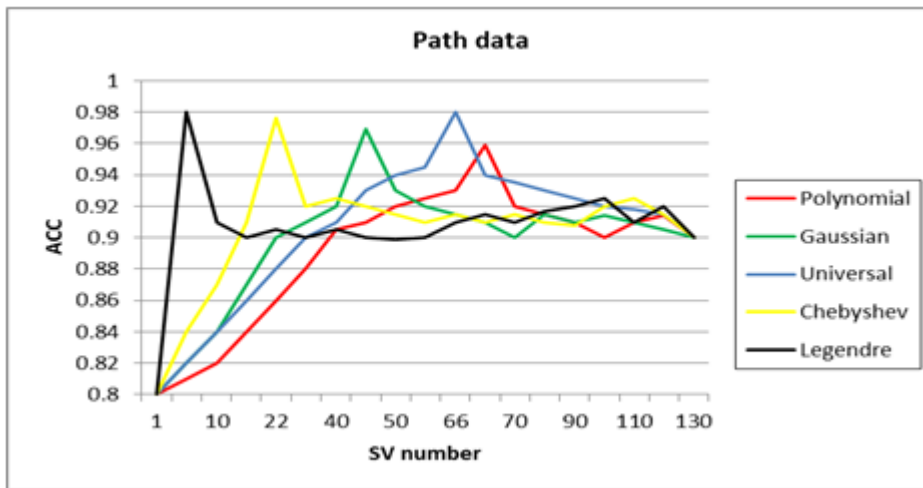
Figure 5: SV vs. ACC for Window data.
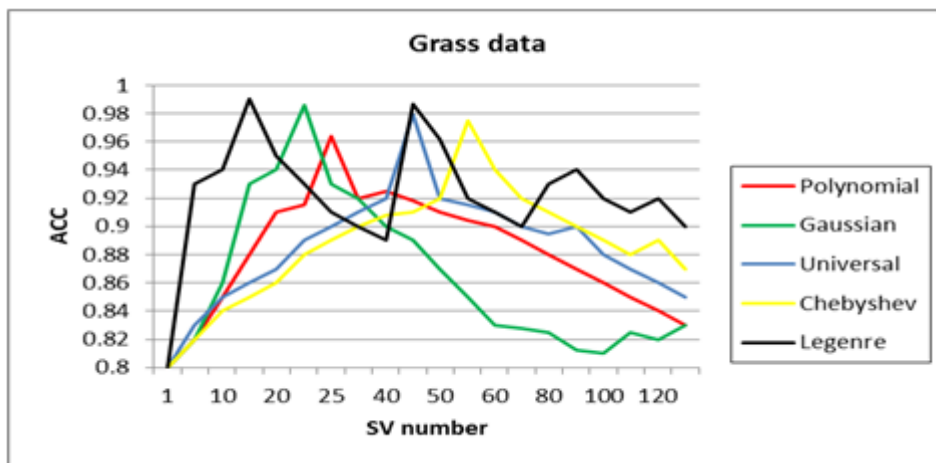

Figure 6: SV vs. ACC for Path data.

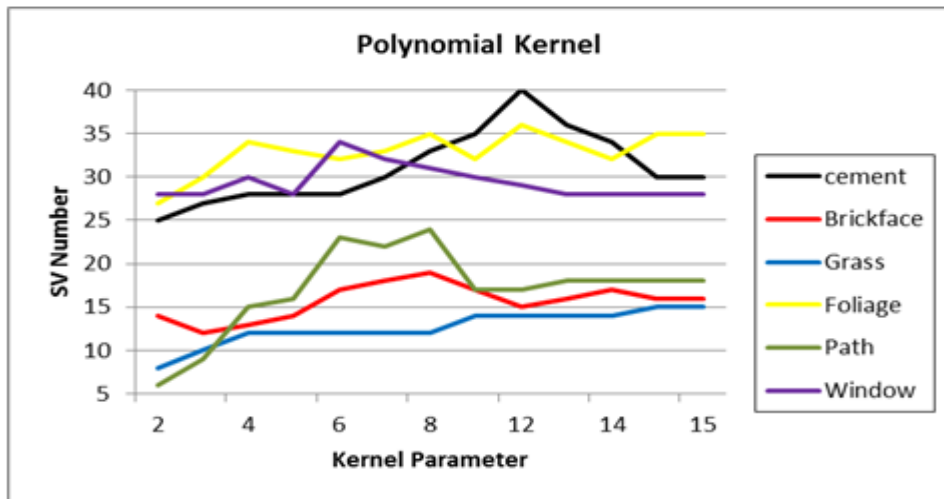Figure 7: SV vs. ACC for Grass data

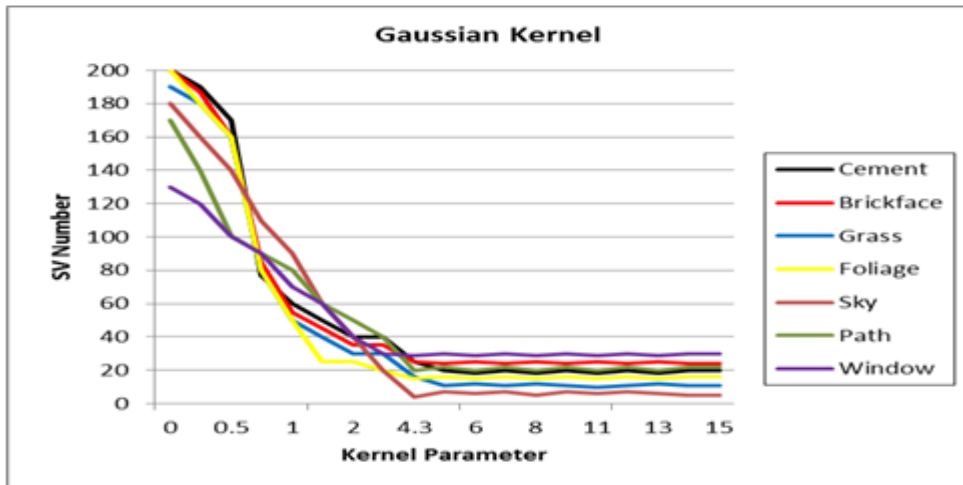Figure 8: Polynomial kernel parameter vs. SV number.

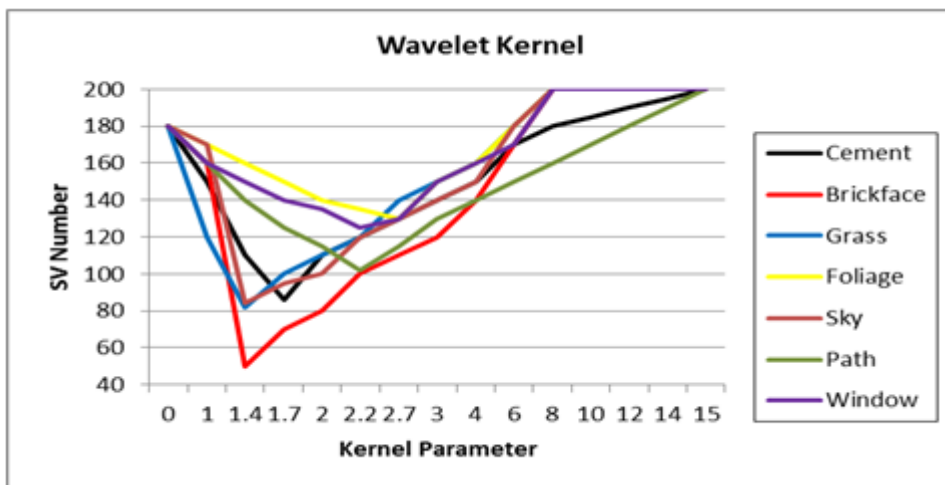Figure 9: Gaussian kernel parameter vs. SV number

Figure 10 : Wavelet kernel parameter vs. SV.
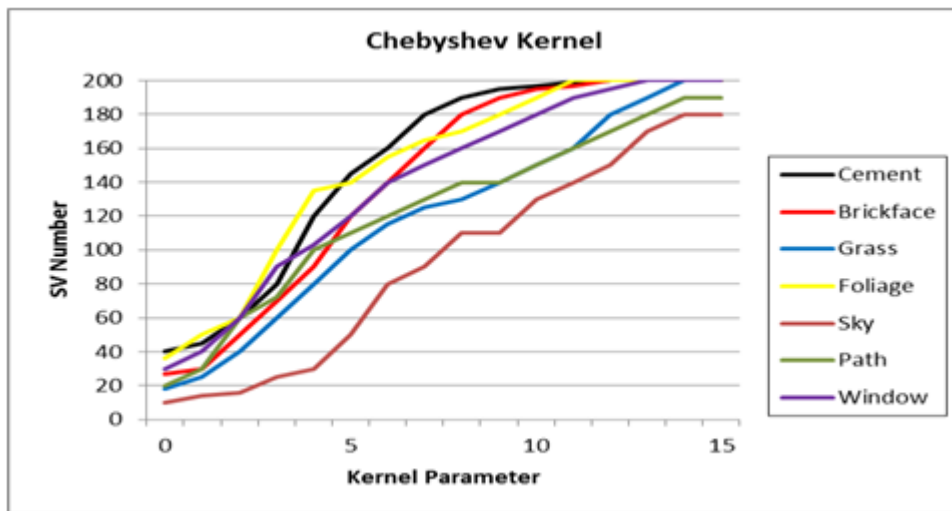
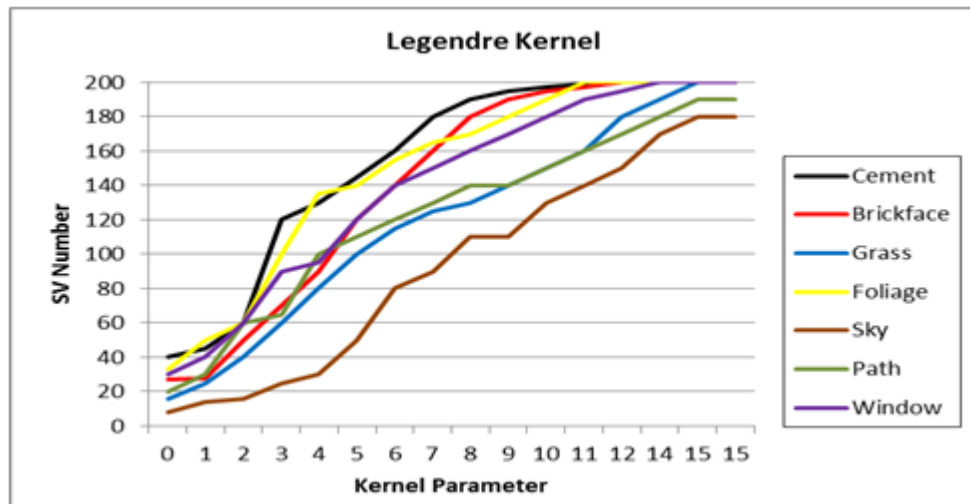Figure 11: Chebyshev kernel parameter vs. SV.

Figure 12: Legendre kernel parameter vs. SV.

6. CONCLUSION

Presenting the current paper, the classification certainty of SVMs has become advanced by mapping the training data into a feature space by the help groups of Legendre functions. A class of Legendre kernel functions based upon the properties of the common kernels is suggested, being able to recognize many applications in training. Normalization takes a vital job for generalized Legendre kernel, and thus the whole data should be normalized between [-1,1] before utilizing the kernel function. Upon the simulation results, it can be said that picking order of Legendre polynomials from an integer group is usually sufficient to acquire a good classification consequence from the generalized Legendre kernel function.

We have made a comparison between the classification efficacy of the Legendre kernel function and the current kernels like the current GF, POLY and Wavelet kernels. In accordance with the test outcomes, the generalized Legendre kernel shows the lowest number of support vectors on almost every evaluation. In strictly, we suggest this is derived from the orthogonally characteristic of the Legendre polynomials. This character of the kernel function can be vital and helpful in

several applications where the support vector number is greatly vital as in feature selection. Therefore generalized Legendre kernel functions can be regarded as a valid substitute to the GF, POLY, and Wavelet kernel functions for a couple of particular datasets. The test outcomes imply that their outcomes have been analogs to the kernel functions developed from the generalized Legendre polynomials of the primer kind. Thus, we have not comprised the family of kernel functions and their outcomes in this study. Also, since handling the properties of the generalized Legendre polynomials is beyond this study, we have not studied these properties specifically even though this study could be helpful to design new kernel functions developed from generalized Legendre polynomials and that could be the goal of upcoming work.

REFERENCES

[1] Vapnik V. N., (1995) "The nature of statistical learning theory", Springer-Verlag, New York, NY, USA.

[2] Kim H., Pang S., Je H., Kim D., Bang S.Y., (2003) "Constructing support vector machine ensemble", Pattern Recognition", vol.36, no.12, pp.2757–2767.

[3] Du P., Peng J., Terlaky T.. (2009) "Self-adaptive support vector machines", modeling and experiments Computational Management Science, vol. 6, no.1, pp. 41–51.

[4] Boser B. E., Guyon I. M., Vapnik V. N., (1995) "A training algorithm for optimal margin classifiers", Proc. Fifth Ann. Workshop Computing Learning Theory, pp. 144-152.

[5] Vapnik V.N. , (1999 "An overview of statistical learning theory)", IEEE Trans. Neural networks, vol. 10, no. 5, pp. 988-999.

[6] Cortes C., Vapnik V.N., (1995) "Support-vector networks", Machine Learning, vol. 20, pp. 273- 297.

[7] Scholkopf B., (1997) "Support vector learning", PhD dissertation, Technische Universitat Berlin, Germany.

[8] Vapnik V.N., Golowich S., and Smola A., (1997) "Support vector method for function approximation, regression estimation and signal processing", Advances in Neural Information processing Systems, vol. 9, Cambridge, Mass.: MIT Press.

[9] Mangasarian O.L., Musicant D.R., (1999) "Successive over relaxation for support vector machines", IEEE Trans. Neural Networks, vol. 10, no. 5, pp. 1032-1037.

[10] Aronszajn N. , (1950) "Theory of Reproducing Kernels", Trans. Am. Math. Soc., vol. 68, pp. 337-404.

[11] Shawe-Taylor J., Bartlett P.L., Willianmson R.C., Anthony M.,   (1998) "Structural risk minimization over data-dependent hierarchies", IEEE Trans. Information Theory, vol. 44, no. 5, pp. 1926-1940.

[12] Williamson R. C., Smola A., Schölkopf B., (1999) "Entropy numbers, operators and support vector kernels", MA: MIT Press, Cambridge, pp. 44-127.

[13] Chapelle O. and Schölkopf B., (2002) "Incorporating invariances in non-linear support vector machines", In T. G. Dietterich, S. Becker, and Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems, vol. 14, pp. 594-609, Cambridge, MA: MIT Press.

[14] Hastie T., Hsu C-W., Lin C-J, (2005) "A comparison of methods for multi-class support vector Machines", IEEE Transactions on Neural Networks, no. 13, pp.415–425.

[15] Zanaty , E.A. ,and Sultan Aljahdali, (2008) "Improving the accuracy of support vector machines", Proceedings of 23rd International conference on computers and their application April, pp. 75-83, Cancun, Mexico.

[16] Zanaty E.A, Sultan Aljahdali, R.J. Cripps, (2009) "Accurate support vector machines for data classification", Int. J. Rapid Manufacturing, vol. 1, no. 2, pp. 114-127.

[17] Zanaty E.A, Ashraf Afifi, (2011) "Support vector machines (SVMs) with universal kernels ", in International journal of Artificial Intelligence, vol. 25 , pp.575-589.

[18] E. A. Zanaty &Ashraf Afifi, Jun (2018) "Generalized Hermite kernel function for support vector machine classifications", International Journal of Computer Applications, In press, Accepted 07 Jun 2018.

[19] Vahid H. M., Javad H.,., (2016) "New Hermite Orthogonal Polynomial Kernel and Combined Kernels in Support Vector Machine Classifier", Elsevier, Pattern Recognition, vol. 60, pp. 921–935.

[20] Meng T., Wenjian W., (2017) "Some Sets of Orthognal Polynomial Kernal Functions", Elsevier, Applied Soft Computing, vol. 61, pp. 741–756.

[21] Maji S., Berg A.C., Malik J. ,(2008) "Classification using intersection kernel support vector machines is efficient", IEEE Conference on Computer Vision and Pattern Recognition, June pp.1–8.

[22] Ozer Sedat, Chen ChiH., Cirpan HakanA., (2011) "A set of new Chebyshev kernel functions for support vector machine pattern classification", Pattern Recognition, no. 44, pp.1435–1447.

[23] Hao Jiang, Wai-Ki Ching, (2012) "Correlation kernels for support vector machines classification with applications in cancer data", Computational and Mathematical Methods in Medicine, doi:10.1155/2012/205025, pp.1-7.

[24] A. Kyriakides, E. Kastanos, K. Hadjigeorgiou, and C. Pitris, (2011) "Classification of Raman spectra using the correlation kernel" , Journal of Raman Spectroscopy, vol. 42, no. 5, pp. 904–909.

[25] A. Kyriakides, E. Kastanos, K. Hadjigeorgiou, and C. Pitris, (January 2011) "Support vector machines with the correlation kernel for the classification of Raman spectra", Advanced Biomedical and Clinical Diagnostic Systems IX, vol. 7890 of Proceedings of SPIE, pp. 78901B-1–78901B-7, San Francisco, Calif, USA.

[26] Zhi-B., Hong C., Xin H., Luis C. P., Martin C., Alfonso, Rojas D., Hector P., Hector F., (2018) "A novel Formulation of Orthognal Polynomial Kernal Functions for SVM Classifiers : The Gegenbauer family", Elsevier, Pattern Recognition, vol. 84, pp. 211–225.

[27] Mercer T. , (1909)"Functions of positive and negative type and their connection with the theory of integral equations", Philosophical Trans. of the Royal Soc. of London, Series A, pp. 415-446.

[28] Williamson R.C., Smola A.J., and Scholkopf B., (1998) "Generalization performance of regularization networks and support vector machines via entropy numbers of compact operators", Technical Report 19, NeuroCOLT.

[29] Scholkopf B., A.J. Smola, (2001) "Learning with kernels: support vector machines, regularization, optimization", and Beyond, MIT Press.

[30] Web: http://www.liacc.up.pt/ML/old/statlog/datasets.html

[31] Web: http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/

[32] Web:http://www. http://svm.sourceforge.net/docs/3.00/api/

AUTHOR

Ashraf Afifi received the B.Sc.(Hons.), M.Sc., and Ph.D. degrees in Electronics and Communication Engineering from Zagazig University, Egypt, in 1987, 1995, and 2002, respectively. He is currently Associate Professor with the Department of Computer Engineering, Faculty of Computers and Information Technology, Taif University, Saudi Arabia. He is a co-author of about 30 papers in international journals and conference proceedings. His research interests cover communication security, image processing, and image encryption. Email: a.afifi@tu.edu.sa, Mobile: 00966507275265.

Prof. E.A.Zanaty is a professor and of Computer Science at the Faculty of Science, Sohag University, Sohag, Egypt. He received his Bachelor of Mathematics degree from Sohag Faculty of Science, Assuit University, Egypt, 1992. He received his MSC Degree in Computer Science in 1997 from South Valley University, Sohag, Egypt and started his career as Assistant Lecturer in November 1997 at the same Faculty. Prof. Zanaty completed his Ph.D Studies with Prof. Guido Bruneet supervision at TU-Chemnitz, Germany, during the period 2000-2003. During this period, he worked as Assistant researcher in Computer Graphics and Visualization Department, College of Computer Science, Chemnitz University, Germany. After that, he backed to Egypt to work as senior of Computer Science laboratory at Sohag University, Egypt. During 2007 till 2016, Prof. Zanaty contracted with College of Computers and Information Technology, Taif University, Taif, Saudi Arabia. He supervised them on several committees and he became the Head of Computer Science Department in 2009 while he earned a professor job in February 2015 from Sohag University, Egypt and also in May 2015 from Taif University, Taif, Saudi Arabia.

Prof. Zanaty is an Editorial board of several journals and member of KACST and IAENG. He also is the Editor in Chief of International Journal of Informatics and Medical Data Processing (IJIMDP). His research activities are focused on artificial intelligence, data classification, reverse engineering, data reduction, medical image segmentation, and reconstruction. In these areas, he has published several technical papers in refereed international journals or conference proceedings.