Investigating the Effect of Different Kernel Functions on the Performance of SVM for Recognizing Arabic Characters

Sayed Fadel[1,2], Said Ghoniemy[1,2], Mohamed Abdallah[1,3], Hussein Abu Sorra[1], Amira Ashour[1,4], Asif Ansary[1]

[1]College of Computers and Information Technology, Taif University, KSA
[2]Faculty of Computers and Information Sciences, Ain Shams University, Egypt
[3]Faculty of Electronic Engineering, Menoufia University, Egypt
[4]Faculty of Engineering, Tanta University, Egypt

Abstract—A considerable progress in the recognition techniques of Latin and Chinese characters has been achieved. By contrast, Arabic Optical character Recognition is still lagging in spite that the interest and research in this area is becoming more intensive than before. This is because the Arabic is a cursive language, written from right to left, each character has two to four different forms according to its position in the word, and several characters are associated with complementary parts above, below, or inside the character. Support Vector Machines (SVMs) are used successfully for recognizing Latin, and Chinese characters. This paper studies the effect of different kernel functions on the performance of SVMs for recognizing Arabic characters. Eleven different kernel functions are used throughout this study. The objective is to specify which type of kernel functions gives the best recognition rate. The resulting kernel functions can be considered as base for future studies aiming at enhancing their performance. The obtained results show that Exponential and Laplacian Kernels give excellent performance, while others, like multi-quadric kernel, fail to recognize the characters, speciallywith increased level of noise.

Keywords—SVM; Kernel Functions; Arabic Character Recognition

I. INTRODUCTION

Optical character recognition systems improve human-machine interaction and are urgently required for many governmental and commercial organizations. A considerable progress in the recognition techniques of Latin and Chinese characters has been achieved. By contrast, Arabic Optical character Recognition (AOCR) is still lagging because of the special characteristics of the Arabic Language. It is a cursive language, written from right to left, each character has two to four different forms according to its position in the word, and most characters are associated with complementary parts above, below, or inside the character. However, the interest and research in this area is becoming more intensive than before.

SVMs is a powerful tool compared to other supervised classification techniques. It is based on statistical learning theory developed by Vladimir Naumovich Vapnik [1] back in 1963 and since then, his original ideas have been perfected by a series of new techniques and algorithms including that of Olusayo D. Fenwa et al. [2], who evaluated the Performance Of PSO-Based Kernel Support Vector Machine in Offline Digit Recognition. One of the most important design choices for SVMs is the kernel-parameter, which implicitly defines the structure of the high dimensional feature space where a maximal margin hyperplane will be found. However, before this stage is reached in the use of SVMs, the actual kernel must be chosen, as different kernels may exhibit different performance.

This paper investigates the effect of eleven different kernels on the performance of SVMs in recognizing Arabic characters. The objective is to extract the kernels giving the best performance. Future work can then be elaborated for enhancing the performance of these kernels. The rest of the paper is organized as follows. Section II reviews the literature. Section III describes the used algorithm. The analysis and discussion of the obtained results are summarized in Section IV. Conclusions of this work with some future lines of research are presented in the last section.

II. LITERATURE SURVEY

Support Vector Machine (SVM) is a state-of-the-art classification method that belongs to the general category of kernel methods. A kernel method is an algorithm that depends on the data only through dot-products. When this is the case, the dot product can be replaced by a kernel function which computes a dot product in some possibly high-dimensional feature space. This approach has two advantages: First, the ability to generate nonlinear decision boundaries using methods designed for linear classifiers. Second, the use of kernel functions allows the user to apply a classifier to data that have no obvious fixed-dimensional vector space representation.

In recent years, Kernel methods have received major attention, particularly due to the increased popularity of the Support Vector Machines. Kernel functions can be used in many applications as they provide a simple bridge from linearity to non-linearity for algorithms which can be expressed in terms of dot products [3].

A linear support vector machine is composed of a set of given support vectors $\mathbf{y}$ and a set of weights $\mathbf{w}$. The computation for the output of a given SVM with N support vectors $y_1, y_2, \dots y_N$ and weights $w_1, w_2, \dots, w_N$ is then given by:

$$F(x) = \sum_{i=1}^{N} w_i(y_i, x) + b \qquad (1)$$

Using kernels, the original formulation for the SVM given SVM with support vectors $y_1, y_2, \ldots, y_N$ and weights $w_1, w_2, \ldots, w_N$ is now given by:

$$F(x) = \sum_{i=1}^{N} w_i k(y_i, x) + b \qquad (2)$$

This work investigates the effect of eleven different kernels on the performance of SVM for recognizing Arabic characters. The most popular kernels for real-valued vector inputs are shown below [4]:

1) Linear (trivial) Kernel:

Linear kernel, the simplest kernel function, is given by the inner product <x,y> plus an optional constant **c**. Algorithms using a linear kernel are often equivalent to their non-kernel counterparts.

$$k(x, y) = x^T y + c \qquad (3)$$

2) Quadratic Kernel:
$$k(x, y) = (S(x, y) + c)^2 \qquad (4)$$

where, c and S are kernel-specific parameters

3) Rational Quadratic Kernel
$$k(x, y) = 1 - \frac{||x-y||^2}{||x-y||^2 + c} \qquad (5)$$

4) Multiquadric Kernel

The Multiquadric kernel can be used in the same situations as the Rational Quadratic kernel.

$$k(x, y) = \sqrt{||x - y||^2 + c^2} \qquad (6)$$

5) Inverse Multiquadric Kernel
$$k(x, y) = \frac{1}{\sqrt{||x-y||^2 + c^2}} \qquad (7)$$

6) Polynomial Kernel

The Polynomial kernel is a non-stationary kernel that is well suited for problems where all the training data is normalized.

$$k(x, y) = (\alpha x^T y + c)^d \qquad (8)$$

Adjustable parameters are the slope **α**, the constant term **c** and the polynomial degree **d**.

7) Gaussian Kernel

The Gaussian kernel is an example of radial basis function kernel.

$$k(x, y) = \exp\left(-\frac{||x-y||^2}{2\sigma^2}\right) \qquad (9)$$

The adjustable parameter ϭ plays a major role in the performance of the kernel, and should be carefully tuned to the problem at hand.

8) Exponential Kernel

The exponential kernel is closely related to the Gaussian kernel, with only the square of the norm left out. It is also a radial basis function kernel.

$$k(x, y) = \exp\left(-\frac{||x-y||}{2\sigma^2}\right) \qquad (10)$$

9) Laplacian Kernel

The Laplace Kernel is completely equivalent to the exponential kernel, except for being less sensitive for changes in the sigma parameter. Being equivalent, it is also a radial basis function kernel.

$$k(x, y) = \exp\left(-\frac{||x-y||}{\sigma}\right) \qquad (11)$$

It is important to note that the observations made about the sigma parameter for the Gaussian kernel also apply to the Exponential and Laplacian kernels.

10) Hyperbolic Tangent (Sigmoid) Kernel

The Hyperbolic Tangent Kernel is also known as the Sigmoid Kernel and as the Multilayer Perceptron (MLP) kernel.

$$k(x, y) = \tanh(\alpha x^T y + c) \qquad (12)$$

It is interesting to note that a SVM model using a sigmoid kernel function is equivalent to a two-layer perceptron neural network. This kernel was quite popular for support vector machines due to its origin from neural network theory. There are two adjustable parameters in the sigmoid kernel, the slope **alpha** and the intercept constant **c**. A common value for alpha is 1/N, where N is the data dimension.

11) Multi-Layer Perceptron:

The long established MLP, with a single hidden layer, also has a valid kernel representation.

$$k(x, y) = \tanh((x, y) + c) \qquad (13)$$

Many authors tried the investigation of using SVMs and similar tools for recognizing Arabic characters and categorizing Arabic text. Mahmoud Zennaki et al., in [5], presented a comparative study of SVM models for learning handwritten Arabic Characters. Eugen-Dumitru Tautu and Flrin Leon, [6] examined the effect of 4 kernels on the performance of SVM for recognizing English handwritten characters. The examined kernels are: linear, RBF, polynomial, and sigmoid functions. They found that the type of the kernel function affects the recognition accuracy. Behjat Siddiquie et al, [7] tried combining multiple Kernels for efficient image classification. S.F. Bahgat et al, [8] proposed a Hybrid Technique for Recognizing Arabic Characters.

III. PROPOSED APPROACH

The process starts with creating the database of the Arabic alphabetic character images used throughout the study. The database contains the character images, the feature vectors of noise-free images, as well as the feature vectors of character images corrupted by additional salt-and-pepper noise with levels ranging from 0.1 to 0.9. The SVM is first trained using the noise-free feature vectors for each used kernel. The SVM is then used for the classification of noisy character set for each kernel. A comparative analysis is then carried out to conclude which kernels are suitable and which are not. A detailed description of the used algorithm is as shown below.

Algorithm for determining the effect of different kernels on the performance of SVMs for recognizing Arabic characters:

INPUT: *Noise free database for the Arabic 28 character images.*

OUTPUT: *Recognition rate of Arabic characters using 11different kernels.*

1. *Initialize: clear all data from the workspace.*
2. *Calculate the statistical features of the noise free characters and train the SVM.*

    for l=1:11 (No. of kernels)

       for i=1:28 (No. of characters)

          for j = 1:100 (repeat 100 times)

             Generate 100 statistical feature vector matrices

          end for

          Train the SVM for all characters using this kernel

       end for

       Repeat training for each kernel

    end for

3. *Calculate noisy statistical feature vectors for all characters.*

    for k=1:9 (Salt & pepper level noise from 0.1 – 0.9)

       for i=28 (No. of characters)

          Generate noisy database.

          Calculate noisy statistical features for all characters.

       end for

       Repeat for all noise levels.

    end for

4. *Calculate the performance of SVMs for recognizing Arabic characters for each kernel.*

    for l=1:11   (no. of kernels).

       for m=1:10   (Take the average of 10 repetitions).

          for k =1:9

             SVM classification for different noise levels

             Store results.

          end for

          Repeat 10 times.

       end for

       Calculate the average recognition rate for each kernel.

    end for

5. *Exit.*

IV.   RESULTS ANALYSIS AND DISCUSSION

This section investigates the performance of SVMs in recognizing the Arabic characters, corrupted by salt and pepper noise with levels starting from 0.1 to 0.9, using eleven different kernels. Samples of the free and noisy Arabic character images used in this study are shown in Fig. 1.

As shown in TABLE I and Fig. 2., it is clear that Laplacian, Exponential, Gaussian, Inverse Multiquadric, and Rational Quadratic Kernels give excellent results for all noise levels. On the other hand, the multi-quadric kernel gives very bad results for all noise levels. Linear, Quadratic, Polynomial, Hyperbolic Tangent, and Multilayer Perceptron Kernels, have monotonically decreasing performance with increasing the noise level.

Excluding the Multiquadric kernel, TABLE II and Fig. 3. show that there are two groups of kernels according to their performance. Group 1, which has the best performance for all noise levels, includes Laplacian, Exponential, Gaussian, Inverse multiquadric, and Rational Quadratic Kernels. Group 2, which has a monotonically decreasing performance with increasing noise levels, includes Linear, Quadratic, Polynomial, Hyperbolic Tangent, and Multilayer Perceptron Kernels.
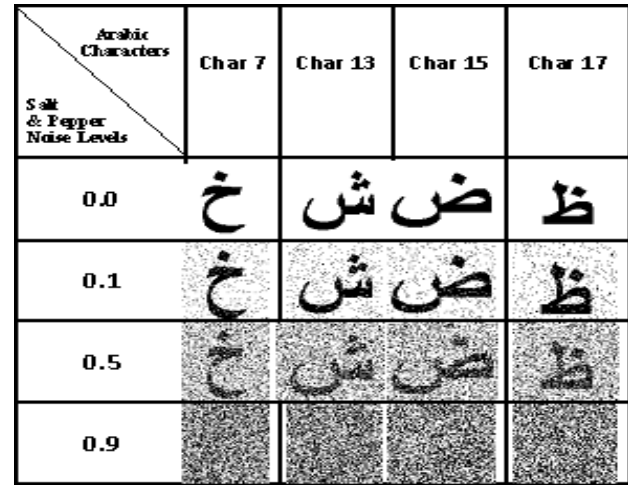

Fig. 1.   Samples of the used database

TABLE I.   RECOGNITION RATE OF DIFFERENT KERNELS (%) IN THE PRESENCE OF SALT AND PEPPER NOISE

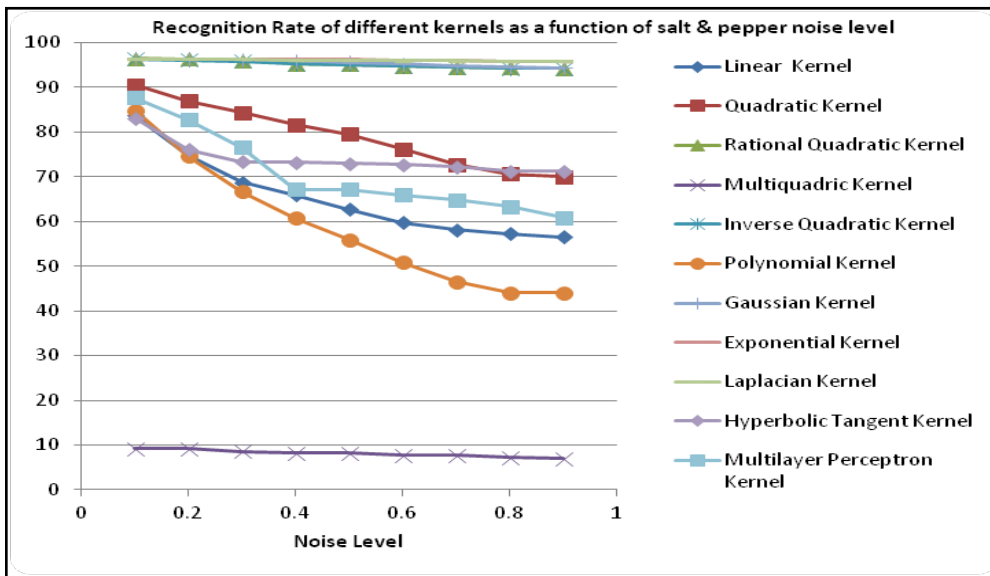| Salt & Pepper Noise Level	Linear Kernel	Quadratic Kernel	Rational Quadratic Kernel	Multiquadric Kernel	Inverse Quadratic Kernel	Polynomial Kernel	Gaussian Kernel	Exponential Kernel	Laplacian Kernel	Hyperbolic Tangent Kernel	Multilayer Perceptron Kernel
0.1	83.7	90.6	96.5	9.17	96.4	84.8	96.4	96.4	96.4	83.1	87.7
0.2	74.7	86.9	96.3	9.20	96.1	74.6	96.4	96.4	96.4	76.1	82.7
0.3	68.8	84.3	96.0	8.55	95.9	66.7	96.2	96.4	96.4	73.4	76.6
0.4	65.9	81.7	95.3	8.23	95.3	60.8	96.0	96.4	96.1	73.2	67.2
0.5	62.7	79.6	95.1	8.17	95.1	55.9	95.6	96.4	96.1	73.0	67.2
0.6	59.8	76.2	94.8	7.76	94.8	50.8	95.2	96.1	96.0	72.7	65.9
0.7	58.1	72.8	94.5	7.69	94.5	46.5	94.9	96.0	96.0	72.3	64.8
0.8	57.3	70.6	94.4	7.16	94.3	44.1	94.6	95.9	95.9	71.3	63.4
0.9	56.5	70.0	94.2	6.98	94.2	44.1	94.4	95.8	95.8	71.2	60.9

Fig. 2.   Recognition rate of Arabic characters using SVM with different kernels as a function of salt & pepper noise
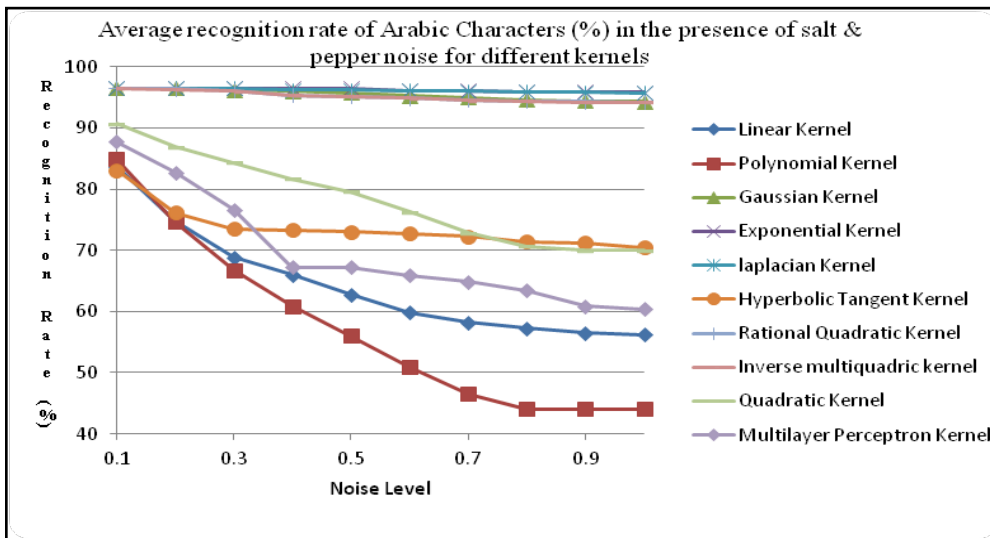

Fig. 3.   Recognition rate of Arabic characters after excluding the multi-quadric kernel
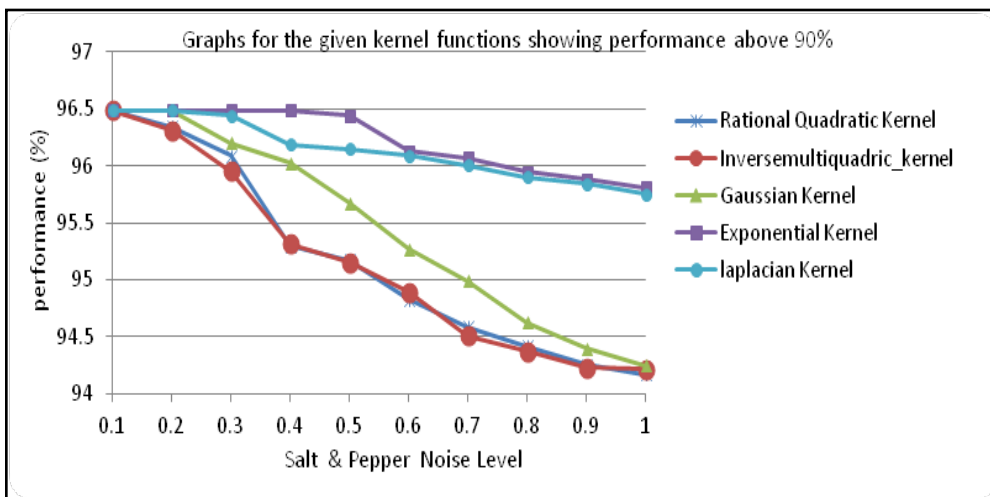

Fig. 4.   Performance comparison for the five best kernels

TABLE II.   RECOGNITION RATE OF ARABIC CHARACTERS AFTER EXCLUDING THE MULTI-QUADRIC KERNEL

Salt & Pepper Noise Level	Linear Kernel	Polynomial Kernel	Gaussian Kernel	Exponential Kernel	laplacian Kernel	Hyperbolic Tangent Kernel	Rational Quadratic Kernel	Inverse multiquadrickernel	Quadratic Kernel	Multilayer Perceptron Kernel
0.1	83.7	84.9	96.5	96.5	96.5	83.1	96.5	96.5	90.7	87.7
0.2	74.7	74.7	96.5	96.5	96.5	76.2	96.3	96.3	86.9	82.7
0.3	68.8	66.8	96.2	96.5	96.4	73.5	96.1	96.0	84.3	76.6
0.4	66.0	60.9	96.0	96.5	96.2	73.3	95.3	95.3	81.7	67.3
0.5	62.8	56.0	95.7	96.4	96.2	73.0	95.2	95.2	79.6	67.2
0.6	59.9	50.9	95.3	96.1	96.1	72.8	94.8	94.9	76.2	65.9
0.7	58.2	46.6	95.0	96.1	96.0	72.3	94.6	94.5	72.9	64.9
0.8	57.3	44.1	94.6	96.0	95.9	71.4	94.4	94.4	70.6	63.5
0.9	56.5	44.1	94.4	95.9	95.8	71.3	94.3	94.2	70.0	60.9

Focusing on the group having the best performance, TABLE III and Fig. 4., show that the exponential and Laplacian kernels give the best performance in the set of the examined kernels. However, there are another 14 kernels that will be examined to extract the final result.

V.   CONCLUSION

SVMs are used as a classification tool for the recognition of Arabic characters. However, before this stage is reached in the use of SVMs, the actual kernel must be chosen, as different kernels may exhibit different performance. This paper studies the effect of eleven different kernel functions on the performance of SVMs for recognizing Arabic characters.

The obtained results show that Laplacian, Exponential, Gaussian, Inverse Multiquadric, and Rational Quadratic Kernels give excellent results for all noise levels. On the other hand, the multi-quadric kernel gives very bad results for all noise levels. Linear, Quadratic, Polynomial, Hyperbolic Tangent, and Multilayer Perceptron Kernels, have monotonically decreasing performance with increasing the noise level. Further investigation showed that Exponential and Laplacian kernels give the best performance. Future work will stress on another set of kernels to extract the most suitable kernel functions for recognizing Arabic characters.

TABLE III.   KERNEL FUNCTIONS SHOWING PERFORMANCE ABOVE 90%

Salt & Pepper noise value	Gaussian Kernel	Exponential Kernel	laplacian Kernel	Rational Quadratic Kernel	Inverse multiquadric kernel
0.1	96.5	96.5	96.5	96.5	96.5
0.2	96.5	96.5	96.5	96.3	96.3
0.3	96.2	96.5	96.4	96.1	95.9
0.4	96.0	96.5	96.2	95.3	95.3
0.5	95.6	96.4	96.2	95.2	95.1
0.6	95.3	96.1	96.1	94.8	94.9
0.7	94.9	96.1	96.0	94.6	94.5
0.8	94.6	95.9	95.9	94.4	94.4
0.9	94.4	95.8	95.8	94.3	94.2

REFERENCES

[1] Vapnik, V. and A. Lerner, "Pattern recognition using generalized portrait method", Automation and Remote Control, 24, 774–780, 1963.

[2] Olusayo D. Fenwa, Modupe O. Alade, Oluyinka T. Adedeji, " Evaluation of Performance Of PSO-Based Kernel Support Vector Machine in Offline Digit Recognition ", International Journal for Advanced Research in Engineering and Technology, Volume 2, Issue VII , pp 74 – 80, July 2014, ISSN 2320-6802.

[3] Parveen Kumar, Nitin Sharma and Arun Rana. Article: Handwritten Character Recognition using Different Kernel based SVM Classifier and MLP Neural Network (A COMPARISON). International Journal of Computer Applications 53(11):25-31, September 2012.

[4] Souza, César R. "Kernel Functions for Machine Learning Applications." 17 Mar. 2010. http://crsouza.blogspot.com/2010/03/kernel-functions-for-machine-learning.html.

[5] Mahmoud Zennaki, Mamoun Mamouni, Kaddour Sadouni, "A Comparative Study of SVM Models for Learning Handwritten Arabic Characters", WSEAS Transactions On Advances In Engineering Education, Issue 2, Volume 10, July 2013.

[6] Eugen-Dumitru Tăutu And Florin Leon, "Optical Character Recognition System Using Support Vector Machines", Buletinul Institutului Politehnic Din Iasi, Publicat de Universitatea Tehnică „Gheorghe Asachi" din Iasi, Tomul LVIII (LXII), Fasc. 2, 2012, SecNia, Automatică si Calculatoare

[7] Behjat Siddiquie , Shiv N. Vitaladevuni and Larry S. Davis, "Combining Multiple Kernels for Efficient Image Classification", www.cs.umd.edu/~behjat/papers/WACV09.pdf

[8] S.F. Bahgat, S.Ghomiemy, S. Aljahdali, and M. Alotaibi," A Proposed Hybrid Technique for Recognizing Arabic Characters," (IJARAI) International Journal of Advanced Research in Artificial Intelligence,Vol. 1, No. 4, 2012.