Spam image email filtering using K-NN and SVM

Yasmine Khalid Zamil, Suhad A. Ali, Mohammed Abdullah Naser
Department of Computer Science, College of Science for Women, University of Babylon, Iraq

Article Info	ABSTRACT
	The developing utilization of web has advanced a simple and quick method for e-correspondence. The outstanding case for this is e-mail. Presently days sending and accepting email as a method for correspondence is prominently utilized. Be that as it may, at that point there stand up an issue in particular, Spam mails. Spam sends are the messages send by some obscure sender just to hamper the improvement of Internet e.g. Advertisement and many more. Spammers introduced the new technique of embedding the spam mails in the attached image in the mail. In this paper, we proposed a method based on combination of SVM and KNN. SVM tend to set aside a long opportunity to prepare with an expansive information set. On the off chance that "excess" examples are recognized and erased in pre-handling, the preparation time could be diminished fundamentally. We propose a k-nearest neighbor (k-NN) based example determination strategy. The strategy tries to select the examples that are close to the choice limit and that are effectively named. The fundamental thought is to discover close neighbors to a question test and prepare a nearby SVM that jelly the separation work on the gathering of neighbors. Our experimental studies based on a public available dataset (Dredze) show that results are improved to approximately 98%.

Corresponding Author:

Yasmine Khalid Zamil,
Department of Computer Science,
University of Babylon, Babylon, Iraq.
Email: yasakhalid88@yahoo.com

1. INTRODUCTION

Email is a widespread technology nowadays because of its speed time added to its cheap. Email Spam defined as unsolicited bulk email, it is a major problem for internet networks [1], [2], [3]. With the proliferation of malicious software, spammers have been able to launch large and widespread campaigns that cause economic losses and increase traffic. Late investigations uncovered that spam movement constitutes over 89% of internet activity, As of late spammers have embraced a new style of spam, that is the spam image trick to make the examination of messages' body content inefficient. Spam image is an endeavor by ~~messages in a joined~~ image that is intelligible by human and hidden from a text-based filter and becomes more difficult to detect. Spammer uses images in an e-mail message, which includes the goal of the spammer. The cost of managing spam is greater compared to the cost of transmission. This cost is due to waste of network resources, increased traffic and significant economic losses, and a decrease in employee productivity [1]. After the adoption of the splash on the unwanted images in the inclusion of their goal became filters based on the text is ineffective in the detection of unwanted images led to the need for filters based on images.

The main issue in the spam image filtering is to create an efficient algorithm of the spam image filtering to separate the spam email image from other popular images in the email. Many techniques have been proposed in filtering this type of image in email, all spam image filtering techniques belong to three main groups [4], [5] these are the header based strategies of e-mail consists of many fields that provide a useful information margin [4], OCR based techniques using OCR tool to extract the text embedding into

image [5], [6], [4] and content-based strategies. In content-based strategies, the analysis and studying about the image substance and features, for example, shading, edge, surface, and so on, are used for filter spam image from other normal images [6], [5], [4]. In this paper, we proposed filtering method based on gray level co-occurrence matrix (GLCM) to extract image texture features. The classification between image spam or ham using Support Vector Machine (SVM), k-nearest neighbor (KNN) and also combination of the two techniques (SVM and KNN). Figure 1 shows samples of spam images.

Figure 1. Examples of the image spam

The rest of the paper is organized as follows. In section 2, a brief review of present related works. Section 3 provides a proposed system. Section 4 presents performance evaluation. In section 5 presents the result. Finally, Section 6 concludes conclusions.

2.   LITERATURE SURVEY

Many discussions have been carried out previously on image spam detection. This section of the paper provides an overview of relevant research work in image spam classification. In 2017 Rui Chan proposed system includes three-layer spam filtering. Spam is filtered by analyzing both the header and the image. The structure of the model explicates carefully the idea of the design and many technologies related to the model. Experimental results show that this system has a satisfactory filtering effect [7].

In 2015 Monireh sadat Hosseinia et. al Suggested a method for spam image filtering, and image texture feature was used to classify the spam image. The gray level co-occurrence matrix has been applied to each image. The properties obtained are 22 features and then the k-nearest neighbor classifier and naive Bayesian are used to evaluate the images obtained from the both of works database Dredze and Image Spam Hunter [4]. In 2015 T. Kumaresan et. al suggested a scheme which extracts the features especially low-level features (like metadata and histogram features of images). An SVM classifier with kernel function is used to identify a spam image based on extracted features, the accuracy of this method 90%, but the time complexity still is a problem in this work [8].

In 2014 Jianyi Wang et. al proposed an approach that was based on combines the characteristics of spam images with the corner point density to detect. The general idea of the algorithm is based on the corner proportion of the images to judge if it is a spam or not spam [9]. In 2015 Nisha D. Chopra et. al used two methods to classify spam images. The first method using OCR tool for separating text from the image, and the second method is used a Bayesian algorithm to detect the words in the mail are spam or not spam [10]. In 2014 Meghali Das et. al proposed a method that based on analyzing the image that contains only a text region. Then classify the embedded image as spam or legitimate accordingly, they tested their method on Dredze dataset [11].

3.   RESEARCH METHOD

In this section, we discuss the main steps of our proposed system. The goal of our works is to create a system that is able to distinguish between ham images and spam images based on texture and content characteristics. The procedure of extracting features from the image attached to an email is delineated in Figure 2. This procedure consist of the following stages:
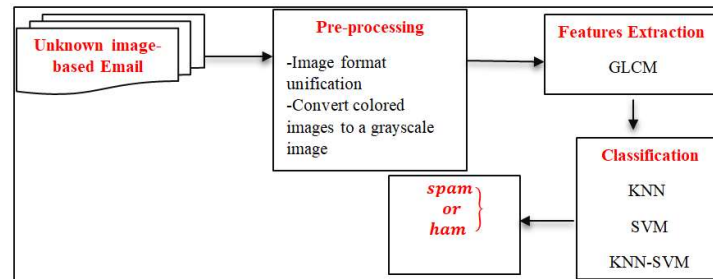

Figure 2. Proposed system general architecture

3.1. Dataset

Dataset is used in our work is Dredze. [12] This dataset contains e-mail images with different sizes which are (3299) spam images of e-mail and (2021) images of legitimate (ham) e-mail. A set of images has been deleted during the processing phase because these images do not provide enough information and its size is very small close to tens of bytes, or some of these images are already empty does not contain information texture. This led to 3264 for spam image and 1783 for ham image.

3.2. Pre-processing stage

Preprocessing stage has the main advantage which is organizing the data in order to simplify classification. All operations that apply to a scanned image is called preprocessing process, in order to reduce or eliminate noise data and keep only the desired information to make the next operation (feature extraction process) easy to implement. The pre-processing stage consists of many operations such as:

3.2.1. Image format unification in JPEG format

JPEG is one of the most recognizable and popular raster image formats. This format appeared as a result of the "Joint Photographic Experts" work. The selection of JPEG format because it is proven to be an effective format in classification process [13].

3.2.2. Convert colored images to a grayscale image

The process that converts the color images to grayscale is aimed to save as much information about the original color image as possible. The conversion process from a color image to a grayscale image requires more knowledge about the color image. A pixel color in an image is a combination of three colors Red, Green, and Blue (RGB). The conversion of a color image into a grayscale image is converting the RGB values (24 bit) into grayscale value (8 bit) [14]. When the image is denoted in the RGB model, it has Red, Green, and Blue components: let R, G and B are the value of these components, respectively then the gray value can be obtained by using Equation 1.

$$RGB = .2989* R+.5870*G+.1440*B \qquad (1)$$

3.2.3. Resizing images

In this step, all images in the dataset are unified to the same size to prepare it for another process which is features extraction. Through our experience, we found that resizing of images to [65×65] gave the best results.

3.3. Features extraction

After the pre-processing stage has been achieved, feature extraction has applied on the image to extract some feature and represented it as feature vector there are many feature extraction methods that are used in differing applications. Some of them may succeed in one application and fail in another. The selected

feature extraction method is an important step in order to achieve a high classification rate; in our experiment, we used the Gray-Level Co-occurrence matrix (GLCM) method.

3.3.1. Gray-level co-occurrence matrix method

The texture could be a characteristic sight of the surface and is a crucial characteristic to explain the various elements of the image. The aim of the study of texture to seek out how to explain the essential options of the image and displays them in an exceedingly single and straightforward kind which might be wont to accurately classify. The GLCM, is a two dimensional matrix g (I, j) that reveals properties the spatial distribution of the gray-levels within the texture image, Where the element (i, j) of the matrix is the number of times the pair of pixels with the value of *i* and the other pixel in values *j* and the distance between them is *d*. The number of rows and columns in the array is equal to the number of gray levels in the original image. In our work, we used the three corners of the matrix (0, 90 and 135) between the pixel and the neighbor pixel. The probability for each pair (i, j) is computed according to the following equation.

$$\mathbf{g(i,j)} = \mathbf{g(i,j)} / \sum_i \sum_j \mathbf{g}(i,j) \tag{2}$$

From the co-occurrence matrix ($g_{d,}\theta$) twelve features can be derived are Energy, Entropy, Contrast, Homogeneity, correlation, and others as shown in Table 1.

Table 1. Gray-Level Co-occurrence Matrix (GLCM) Features

Feature number	Measure
F1	$\text{Energy} = \sum_i \sum_j \mathbf{g}(i,j)^2$
F2	$\text{Entropy} = -\sum_i \sum_j \mathbf{g}(i,j) \log_2 \mathbf{g}(i,j)$
F3	$\text{Contrast} = \sum_i \sum_j (\mathbf{i-j})^2 \mathbf{g}(i,j)$
F4	$\text{Homogeneity} = \sum_i \sum_j \frac{\mathbf{1}}{\mathbf{1+(i-j)^2}} \mathbf{g}(i,j)$
F5	$\text{Dissimlarity} = \sum_i \sum_j \mathbf{g}(i,j) * \mathbf{|i-j|}$
F6	$\text{Mean } \mu_I = \sum_i \sum_j \mathbf{i} * \mathbf{g}(i,j)$
F7	$\text{Mean } \mu_j = \sum_i \sum_j j * g(i,j)$
F8	$\text{Variance I } \sigma_I = \sum_i \sum_j g(i,j) * |i - \mu_I|^2$
F9	$\text{Variance J } \sigma_j = \sum_i \sum_j g(i,j) * |j - \mu_j|^2$
F10	$\text{standard deviation I} = \sqrt{\sigma_I}$
F11	$\text{Standard devition j} = \sqrt{\sigma_j}$
F12	$\text{Maximum probability} = \max \mathbf{g}(i,j)$

3.3.2. Normalization

Normalization is considered as an imperative information preprocessing to stay away from properties in more prominent numeric reaches overwhelming those in littler numeric reaches, Highlight normalization, or feature scaling, is an essential system for information pre-processing. With a reasonable inspiration to roughly even out the range and weight of information traits [15], there are several ways to normalization but one of the least difficult and most broadly utilized detailing is the in the range (Min, Max). Assume that:

$$\mathbf{I}: \{ \mathbb{X} \subseteq \mathbb{R}^n \} \to \{\mathbf{Min}, .., \mathbf{Max}\}$$

Normalization transforms an n-dimensional grayscale image (I) with intensity values in the range (Min, Max), into a new image.

$$\text{IN}: \{\, \mathbb{X} \subseteq \mathbb{R}^n \} \rightarrow \{\textbf{newMin}, .., \textbf{newMax}\}$$

With intensity values in the range (newMin, newMax). The linear normalization of a grayscale digital image is performed according to the formula [16]:

$$\text{IN} = (\text{I} - \text{Min})\frac{newMax - newMin}{Max - Min} + newMin \tag{3}$$

3.4. Features extraction

Classifiers are used for different purposes [17], in this paper are used for classifying the image into two classes as ham or spam by comparing its features with one of a given set of classes. A classifier is used to identify an object by using its features, and then these features are compared and saved as models for the classes trained. In the testing phase, it will identify the unknown object by extracting its features and then compared with the features, In our experiments, we used the class SVM as well as the KNN as well as our work combination between the SVM and the KNN for several reasons, such as to improve the puncture and reduce the time and storage and will be presented in detail in the section SVM-KNN.

3.4.1. SVM

Support vector machine is powerful classification systems in data classification, it includes solving quadratic problems and this requires a great time for training and big memory for huge scale issues [18], a support vector machine (S:VM) can be utilized when our information has completely two classes. An SVM characterizes information by finding the ideal hyperplane that isolates all information purposes of one class from those of alternate class. The hyperplane for an SVM implies the one with the biggest edge between the two classes [19]. Margin implies the maximal width of the bit parallel to the hyperplane that has no inside information focuses [8], SVM has a place with a group of generalized linear classifiers and it can be translated as an expansion of the perception [20]. A unique property is that they at the same time limit the empirical classification error and amplify the geometric margin thus they are otherwise are named maximum margin, Figure 3 shows SVM Shown classifier.
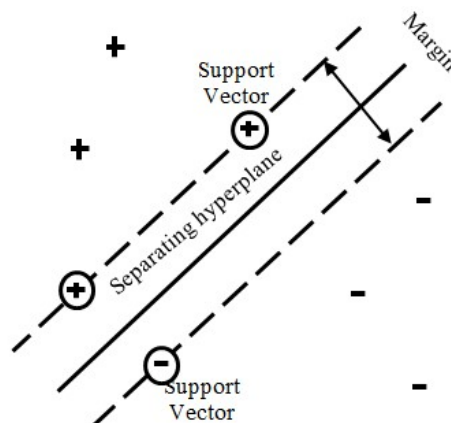

Figure 3. Support vector machine [8]

3.4.2. KNN

K-Nearest Neighbor algorithm (KNN) is a type of supervised learning which is used in several applications in the field of image classification, data mining, and many others. KNN can be calculated by several distance metrics the best metrics are Euclidean distance can be calculated as follow [14]. Xi, xj are two vector $x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}\ldots\ldots x_{in})$ and $x_j = (x_{j1}, x_{j2}, x_{j3}, x_{j4}, x_{j5}\ldots x_{jn})$ distance calculated as follow:

$$\text{D}(\text{X}_I, \text{XJ}) = \sqrt{\sum_{K=1}^{n}\left(\left(x_{ik} - x_{jk}\right)\right)^2} \tag{4}$$

The K-NN calculation is powerful and clear to actualize. In any case, one of the primary disadvantage of K-NN is its inefficiency for large-scale high dimensional data sets [21], The principle purpose behind its the downside is its "lazy" learning algorithm natures calculation and it is since it doesn't have a genuine learning stage and that comes about a high computational cost at the characterization time.

3.4.3 KNN-SVM

The SVM has a good performance but contains some problems which take a great time and the use of the CPU and the use of the actual memory, considering the training and classification, especially when the dimensions between the data is high, adding that when training requires a few data, this mean the number of data for training less from data for test , while the way KNN classification performs the simple and low-cost [21] so we found through our work to classify spam images in email to simplify the process of training and optimization of the SVM algorithm and to obtain very efficient results using KNN with SVM. Figure 4 shows the proposed combination of KNN-SVM flowcharts to classify email images.

The steps of this technique are:
1.   Compute distances of the query to all training examples.
2.   If the k neighbors have all the same labels, the query is labeled and exit; else, compute the pair-wise distances between the k neighbors;
3.   Convert the distance matrix to a kernel matrix and apply multiclass SVM;
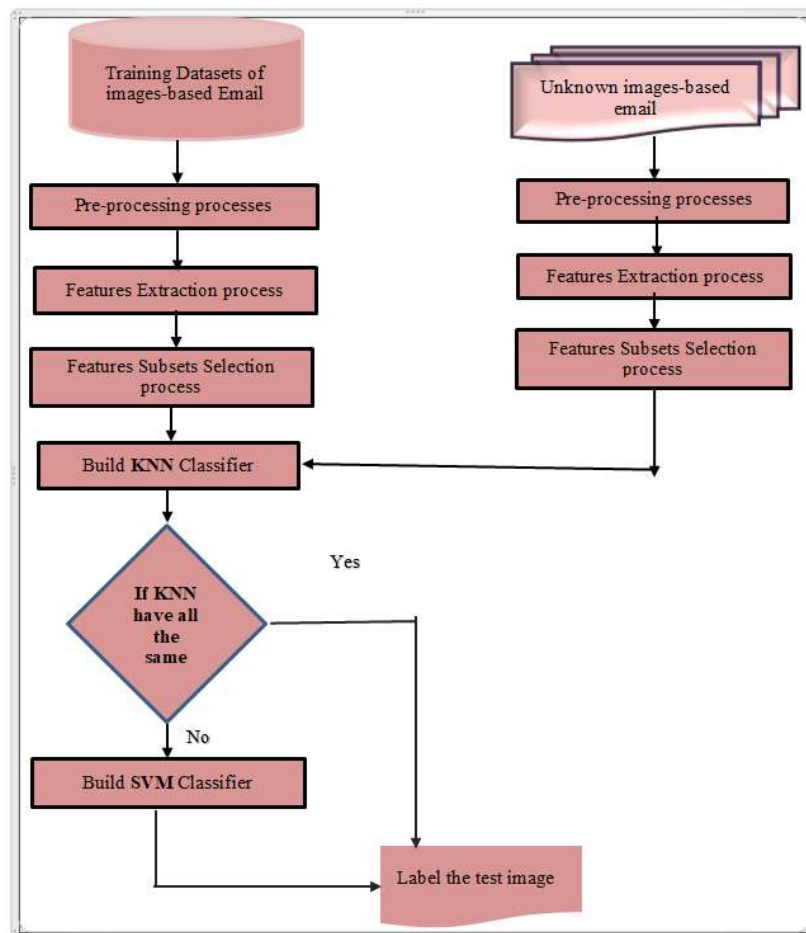4.   Use the resulting classifier to label the query.

Figure 4. Proposed system architecture of SVM and KNN classifiers

4.   PERFORMANCE EVALUATION METRICS

The following standard performance metrics to evaluate the proposed method: accuracy, precision, recall, F-measure, which are defined as follows in Table 2 [1], [4].

Table 2. Performance Evaluation Metrics

Measure	Defined as:	What it means
Accuracy	$\dfrac{TP + TN}{FN + FP + TN + TP}$	Percentage of predictions that are correct [22]
Precision	$\dfrac{TP}{TP + FP}$	Precision is the level of the right forecast (for spam email) [1].
Recall	$\dfrac{TP}{TP + FN}$	Spam Recall looks at the likelihood of true positive examples being recovered (completeness of the retrieval process) [1].
F-measure	$\dfrac{2 * Precision \text{ x } Recall}{Precision + Recall}$	F-measure consolidates these two measurements in a single condition which can be deciphered as a weighted average of precision and recall [1].

Where FP, FN, TP, TN are characterized as follows [1], [22], [4].
1. False Positive (FP): The number of messages for ham e-mail that are classified incorrectly.
2. False Negative (FN): The number of spam e-mail messages that are classified incorrectly.
3. True Positive (TP): The correct classification of spam mail.
4. True Negative (TN): The correct classification of ham mail.

5. RESULTS AND ANALYSIS

A GLCM based feature point extraction method for image spam classification system is built. In the next, we conduct three sets of experiments to verify the effectiveness and efficiency of our approach. In the first set of experiments, we verify the classification performance under the measures of accuracy using SVM as a classifier. In the second set of experiments, the classification performance under the measures of accuracy using KNN as a classifier, and in the third experiment the classification performance under the measures of accuracy using a combination of KNN-SVM as classifier. Finally, we compare the performance of three approaches.

5.1. Results with applying SVM

By using SVM classifier, we obtained the average accuracy 0.497 when the train data are (1100, 1770) for ham and spam image respectively. Table 3 shows the results with different numbers of the training samples.

Table 3. Result of SVM with Different Training Samples

Spam image (3264 images)			Ham image (1783 image)			Average Accuracy
Train	Test	Accuracy of spam image	Train	Test	Accuracy for ham image	
50	1494	90.36	50	683	91.51	90.93
100	1494	89.76	100	683	91.95	90.85
150	1494	92.10	150	683	91.80	91.95
200	1494	93.31	200	683	91.95	92.63
1770	1494	0	1100	683	99.56	49.78

It can be noted from Table 3, that SVM classifier give appropriated result when the number of training samples is small, and the accuracy decrease for spam images equal to (0) when a number of training (1770) samples. Figure 5 shows the average accuracy of SVM with a different number of training samples.
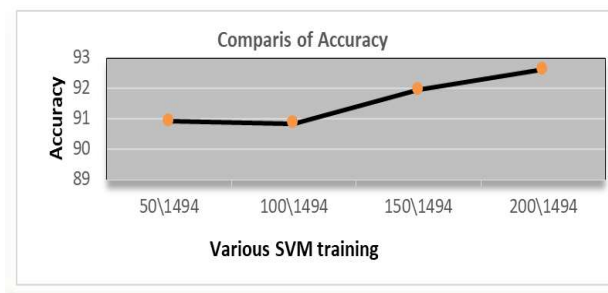

Figure 5. Accuracy for SVM with different number of training samples

5.2. Results with applying KNN

Using KNN for different K value, 3264 for Spam image (train data 1770 and testing 1494 image) and 1783 for Ham image (train data 1100 and testing 683 images), the results are showing in the Table 4 for the values of k between 15 to 40. From Table 4, it can be noted that best average accuracy obtained for K in the range (15-20).

Table 4. Result for KNN with Different Values for K

K values	Accuracy for spam image	Accuracy for ham image	Average Accuracy
16	92.97	95.92	94.45
17	95.92	92.09	94.01
20	96.05	91.95	94
25	95.25	90.63	92.94
30	95.45	91.36	93.41
35	95.18	89.60	92.39
40	94.78	90.92	92.85

5.3. Combination of KNN- SVM

The proposed method tries to select the patterns that are located near the boundary and are correctly labeled. In order to do that, A pattern near the decision boundary tends to have neighbors with mixed class labels. Thus, the of K-nearest neighbors' class labels can estimate the K patterns which will be input to SVM. Table 5 shows the results of average accuracy for spam and ham images. It can be noted from results that combination of KNN-SVM gives best results. Figure 6 shows the performance evaluation metrics for our proposed method and Figure 7. Show comparison for performance metrics accuracy, precision, recall, and f-measure. The accuracy of our proposed based texture features and some other methods are reported in Table 6 to prove the efficiency of our proposed system.

Table 5. Result for SVM-KNN with Different Values of k

K values	Accuracy for spam image	Accuracy for ham image	Accuracy
15	98.80	95.31	97.06
16	98.80	95.61	97.20
17	98.80	95.61	97.20
20	98.93	95.61	97.27
25	98.80	95.61	97.20
30	98.93	95.17	97.05
35	99	95.17	97.08
40	98.19	94.88	96.54

Table 6. The Accuracy that Achieved by Our Proposed Method and Other Methods for Email Image Classification

| Related work	Public year	Techniques used for image spam filtering/classification	Classification Accuracy
[7]	2017	Multi-layer algorithm	96.2%
[4]	2015	k-nearest neighbor classifier (KNN) and naive Bayesian (NB)	91forKNN and 75 for NB classifier
[8]	2015	using Support Vector Machine and Particle Swarm Optimization	90%
[9]	2014	Thresholding	91.3%
[10]	2015	OCR and Bayesian Algorithm	Not defined
[11]	2014	Content analysis	Not defined
Proposed method		Texture-based features using a combination of SVM-KNN	97.20
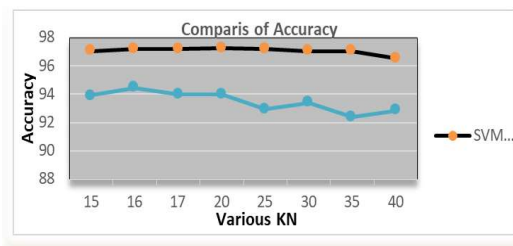
Figure 6. Shows the performance evaluation metrics for our proposed method
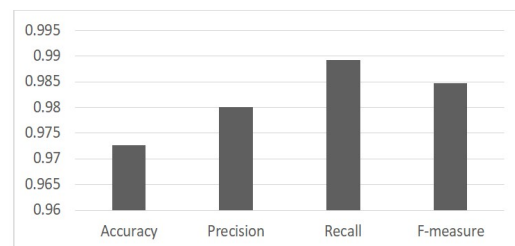

Figure 7. Show Comparison for performance metrics accuracy, precision, recall, and f-measure

6. CONCLUSION

In this paper, our proposed method for distinguishing the ham and spam images was presented using GLCM, which is one of the image texture features. For each image, the 12 features are extracted in three directions. These features are the entropy, energy, mean, etc. At first we apply SVM to classify the images as ham or spam, But because of the problems of SVM represented by a great time for training and big memory for huge scale issues [18], we resorted to KNN to get the best results but also have problems is the pruning of the data with high spacing. To improve the SVM performance a combination of SVM and KNN applied to get the best accuracy. As shown from Table 5 the average accuracy is 97.27 when the value of K is 20.

REFERENCES

[1] Al-Duwairi, et al., "Detecting Image Spam Using Image Texture Features," *International Journal for Information Security Research (IJISR)*, vol. 2, pp. 344-35, 2012.

[2] M.Soranamageswari and Dr.C. Meena, "A Novel Approach towards Image Spam Classification," *International Journal of Computer Theory and Engineering*, vol. 3, pp. 84-88, 2011.

[3] Ketari, et al., "A Study of Image Spam Filtering Techniques." *In Computational Intelligence and Communication Networks (CICN)*, 2012 Fourth International Conference, pp. 245-250, 2012.

[4] Sadat Hosseini, *et al.*, "A Method for Image Spam Detection Using Texture Features," *International Academic Journal of Science and Engineering,* vol. 2, pp. 51-58, 2015.

[5] Attar A, et al., "A Survey of Image Spamming and Filtering Techniques," *Artificial Intelligence Review*, pp. 71-105, 2013.

[6] Caruana, et al., "A Survey of Emerging Approaches to Spam Filtering," *ACM Computing Surveys (CSUR)*, vol. 44, pp. 1-27, 2012.

[7] Chang R., "Application of a Multi-Layer Algorithm on Image Spam Filtering," *Modern Physics Letters B*, pp.19-21, 2017.

[8] M.sadat Hosseini, and M. Rahmati. "A Method for Image Spam Detection Using Texture Features," *International Academic Journal of Science and Engineering,* vol. 2, pp. 51-58, 2015.

[9] Jianyi Wang and Kazuki Katagishi. "Image Content-Based Email Spam Image Filtering," *Journal of Advances in Computer Networks 2*, pp. 110-114, 2014,

[10] Nisha D. Chopra, Prof. K. P. Gaikwad, "Image and Text Spam Mail Filtering," *International Journal of Computer Technology and Electronics Engineering (IJCTEE),* vol 5, pp.15-18, June 2015.

[11] Meghali Das and Vijay Prasad, "Analysis of an Image Spam in Email Based on Content Analysis," *International Journal on Natural Language Computing (IJNLC),* vol. 3, pp.129-140, 2014.

[12] R. G. Mark Dredze and A. Elias-Bachrach, "Learning Fast Classifiers for Image Spam," *presented at the in Proc. CEAS 2007*, Mountain View, California, August2-3, 2007.

[13] Paola, Justin D., and Robert A. Schowengerdt., "The Effect of Lossy Image Compression on Image Classification," *In Geoscience and Remote Sensing Symposium*, 1995. IGARSS'95.'Quantitative Remote Sensing for Science and Applications', International, vol. 1, pp. 118-120. IEEE, 1995.

[14] Aman Kataria and M. D. Singh, "A Review of Data Classification Using K-Nearest Neighbour Algorithm*,*" *International Journal of Emerging Technology and Advanced Engineering* 3, Vol. 3, pp. 354-360, June 2013.

[15] Lingxi Xie1, Qi Tian2, and Bo Zhang1, "Feature Normalization for Part-Based Image Classification", *In Image Processing (ICIP)*, 2013 *20th IEEE International Conference* on, pp. 2607-2611, 2013.

[16] Kumar, et al, "Features Fusion for Classification of Logos," *Procedia Computer Science* 85, pp.370-379, 2016.

[17] Dmitriy Fradkin and Ilya Muchnik, "Support Vector Machines for Classification", *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* 70, pp.1-9, 2006.

[18] Wang, et al., "Training Data Selection for Support Vector Machines," *In International Conference on Natural Computation*, pp. 554-564. Springer, 2005.

[19] M. Hamiane and F. Saeed, "SVM Classification of MRI Brain Images for Computer- Assisted Diagnosis*,*" *International Journal of Electrical and Computer Engineering (IJECE),* vol. 7, pp. 2555–2564, 2017.

[20] K. V. V. Kumar and P. V. V. Kishore, "Indian Classical Dance Mudra Classification Using HOG Features and SVM Classifier," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, pp. 2537, 2017.

[21] Lam Hong et al.*,* "A Review of Nearest Neighbor-Support Vector Machines Hybrid Classification Models," *Journal of Applied Sciences* 10, pp. 1841-1858, 2010.

[22] N.S. Kumar, et al, "Detecting E-mail Spam Using Spam Word Associations," *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, pp. 222-226, April 2012.

BIOGRAPHIES OF AUTHORS

Yasmine Khalid Zamil is Master student in Science College for Women, University of Babylon, Iraq. received a B.Sc. degree from the Science College for Women, University of Babylon, Iraq in 2010. Her areas of interest are digital image processing, data mining and information hiding.

Suhad Ahmed Ali is working as Assistant Professor in Computer Science Department, Science College for Women, University of Babylon, Iraq. She received M.S. and Ph.D. degrees from Department of Computer Science, Babylon University in 2002 and 2014, respectively. Her areas of interest are digital image processing, pattern recognition and information hiding.

Mohammed Abdullah Naser was born in Hilla, Babylon City, Iraq, on February 1, 1976. He received the B.Sc. degree in computer science on 1998 from the University of Technology, Iraq. He received the M.Sc. degree in Computer Security on 2001 from the same university. He received the Ph.D. degree in Computer Security and Data mining on 2006 from the Iraqi Commission for Computers and Informatics, Iraq. Currently, he is associated Professor in the University of Babylon. His research interests include computer and data security, Data mining applying the data mining techniques in security fields and applications.