# An Observation and Experimental Evaluation of Image Spam Detection

**Mallikka Rajalingam, M. Balamurugan**

***Abstract:*** ***In belonging to other supports duel beside researchers of image spam detections, unsolicited mail have newly developed the image based spam dodge to construct the investigation of e-mails' content of text unsuccessful. To avoid signature based recognition, it involves in implanting the unsolicited text or message into an appendage image, which is frequently arbitrarily customized. Identifying image based spam emails tries out to be an motivating illustration of the problem text embedded in images were subjected to noise such as background pattern, color, font variations and imperfections in a font size so as to eliminate the chances of being identified as unsolicited e-mail by classification techniques. In this research paper we spring a exhaustive review and categorization of machine learning and classification systems suggested so far in contradiction of image based spam email, and make an empirical investigation and correlation of few of them on real, widely accessible data sets.***

***Index Terms: Image spam detection, recognition, segmentation, and support vector machine.***

## I. INTRODUCTION

Currently, text embedded in image (image spam) detection of email is a problematic of countless importance, inclined the previously enormous in addition however aggregate extent of multimedia (interactive) contents (e-mail) existing taking place in World Wide Web, and the opportunity of dissimilar types of mobile gadget skilled to create and get into specific contents. Amongst remains, this increases security concerns, alike the chance of committing scams (i.e., phishing of e-mail), and the right to use illegitimate or inappropriate contents of email by adolescents. Machine learning and classification techniques are in advance an appropriate part in view of this claim area. Certain illustration of this issue, which we have motivated on this work, is image-based unsolicited email (i-spam/image spam) occurrence. Few years back, e-mail content of unsolicited message was merely of text message. Consequently, spam (unsolicited) filters considered only the content of email (text) including information of header to distinguish amongst legitimate emails and unsolicited emails. With this issues in contradiction of the creators of spam (unsolicited) filters, in 2006 spammers familiarized image-based (unsolicited) spam bogus, that involves in eliminating irrelevant (spam/unwanted) information in e-mail's content and entrenching into illustration of image that is showed as attachment. This permits to deceive the investigation of electronic mails' content of text. To identify image-based spam e-mail, machine learning and detection techniques are also required, and certainly numerous methods and techniques have been proposed in recent times. Conversely, the solutions suggested so far shows some flaws, and their accuracy or efficiency of detecting image based unsolicited email has not been methodically investigated so far. This paper contains the entire assessment of image based unwanted e-mail and the variety of triumphing electronic message detection techniques used for electronic mail detection with effective analysis in their benefits and disadvantages. Moreover, the paper illuminates the belief of textual content primarily based e-mail category with numerous gadget getting to know methods like decision Tree, SVM (Support Vector Machine) and Naïve Bayes. This paper also discussed an in depth rationalization of segmentation of text from images, recognition of a character and detection of image based electronic mail that's considered as the premise of the modern research studies work.On this effort, we make a review or survey on the machine learning techniques and detection methods proposed so far against image-based unsolicited e-mail. We recommend an organization of several methods or techniques and debate their ability benefits and drawbacks which including weakness of data handling. We additionally perform an experimental assessment and assessment of a few of those strategies, which is missing within the literature. Particularly, proposed researches are done on huge and widely accessible data sets of image emails which have been connected to actual genuine and unsolicited e-mails. This paper is organized as follows. An outline summary of image based spam and image spam recognition techniques are discussed in section II. In the literature, several techniques are proposed for image spam detection are reviewed with the susceptibilities of image based spam detection techniques in section III. The investigational evaluation is stated in section IV and finally, this paper concludes with future enhancement of the work are conversed in section V.

## II. ANALYSIS OF IMAGE-BASED SPAM AND ITS TECHNIQUES

The phrase "Spam" can be clarified as [1] **"**unsolicited or spam mail is abundant the internet with uncountable facsimiles of the same message,

**Revised Manuscript Received on September 15, 2019**

**Mallikka Rajalingam**, Department of Computer Science and Engineering, Bharathidasan University, Trichy, India. E-Mail: mallikka2002@gmail.com

**M. Balamurugan**, Department of Computer Science and Engineering, Bharathidasan University, Trichy, India. E-Mail: mmbalmurugan@gmail.com

in an exertion to force the message on community who would not otherwise designate to receive it". Unsolicited e-mails typically encompass commercial campaigns of indeterminate products, get-rich-quick schemes, dating services, or commercial services.An electronic mail message is the utmost protuberant way of interacting with others. Worldwide electronic mail account increased from 3.3 billion in 2012 to 4.3 billion in 2016 [2] with yearly growing rate of 6%. As a consequence, in 2005, spammers have hosted yet another concept of image spam which embeds spam text content into graphical images. In Image spam e-mail, contains text message (unsolicited) that is entrenched into images that send as an e-mail attachment. Meanwhile maximum of email consumers will show the image content file openly to the handler, the image spam communication is carried as quickly as the electronic mail is open, furthermore there is unnecessary to open image attachment document.Unsolicited or spam emails consume more network bandwidth for the duration of broadcast, this occupies consumer time of searching. As of December 2014, statistical intelligences illustrate that worldwide unsolicited user account holder for 66.41% of e-mail traffic and Asia establishes 54% [3]. Current research [4] discloses the detail that maximum of the e-mail consumers have more unsolicited than legitimate e-mails. Web security threats [17] for phishing and loot users' information and discoursed about 2FA protocol scheme to secure reliable process to inhibit users' accounts. They used RSA cryptography to ensure the verification of web users' and server. An essential continues to develop an innovative technique which could identify image spam emails which inspired the investigator to recognize the several methods or techniques used up to date and the growth of an innovative algorithm based method or technique to identify legitimate and unsolicited image mails. A methodology [5] have discussed for extracting or removing text information from images such as scene images, document images, etc. and DWT used for removing text content from difficult/complex images. A robust approach [6] have presented for text extraction and identification of images.

There are two critical capabilities in e-mail type that are generally separated into various sub-tasks. Primarily, compilation of facts and demonstration are often complicated in particular electronic mail communications. Secondly, e-mail characteristic feature selection process and character feature deduction encounter to decrease the capabilities amount for durable venture steps. Reliable mapping in the training and testing set has been recognized by e-mail classification part. Machine learning techniques used to serve the above-mentioned tasks are expanded in the following section. Character segmentation is characterized into two subgroups: classifier based and non-classifier based techniques. Researchers tested [16] machine learning approaches can identify various anonymous worms and compared with Naïve Bayes and KNN classifiers to show better performance. The dataset used in testing and training the algorithm by 18 authentic worm variants. The proposed method compared with other methods to prove classification accuracy. Based on the experiment they concluded that KNN is slightly better than NB algorithm. Fig. 1 shows overview of image spam email detection techniques.

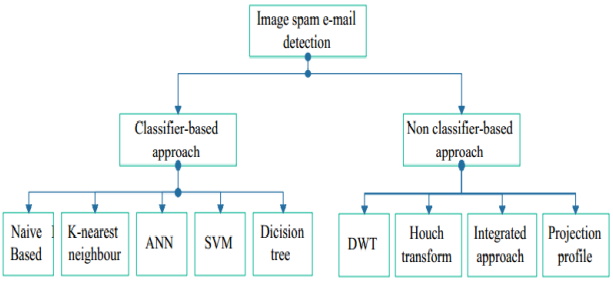


**Fig. 1. Overview of detection techniques**

## III. LITERATURE ANALYSIS

This study incorporates the complete study of spam or unsolicited e-mails and various existent email detection methods used for image email detection by robust examination of their benefits and disadvantages. Additionally, the review clarifies the idea of text e-mail identification with several machine learning procedures like SVM (Support Vector Machine), Naïve Bayes, and Decision Tree. The section also deliberated a thorough narrative of image segmentation, character recognition and I-spam/image spam detection, which considered as the foundation of the current research.

### A. Classifier Based Approach

***Naïve-Bayes/probabilistic Classifier*:** In 1998, first probabilistic classifier for unsolicited or identification is proposed. Bayesian classifier operates on proceedings which are reliant and possibility of happening that may arise in the upcoming or it can be identified the former happening of the alike occasion [7].There are two clusters of emails. It could be either unsolicited or legitimate. In effect, all the spam filters grounded on stats employ Bayesian likelihood computation to include definite token's information to a universal score. On the basis of the universal score, conclusions are drawn.

***K-nearest neighbor classifier*:** Finding out the nearest neighbours can quicken the employment of conventional indexing procedures. The categorization of spam or ham messages is determined with the category of mail messages are so near. The assessment among the vectors is an actual method [8]. This is the notion of the k-nearest neighbour process:

**Stage-1: Training**

Keep the training email or communication.

**Stage-2: Filtering**

Provide a message x or communication, command its k-nearest neighbours betwixt the emails in the preparation group. If any further unsolicited email or spam preparation group or if there are further spam's betwixt these neighbours, categorize provided email or message as spam or else, categorize it as ham.

***ANN (Artificial Neural Networks classifier)*:** An ANN alias "Neural Network" (NN), is a measurable grounded simulation of biological neural network methods that functions on the rule of studying by instance [9]. Below are the artificial neural networks algorithms:

**Step-1: Training**

Modify w and b to random value or to 0.

Determine a training sample (x,c) for which sign ($w^T$ x + b)
If there is no such sample, formerly training is ended
Save the last w and halt
Or else move to following step
Renew (w,b): w:=w+cx, b:=b+c and move to earlier phase.

**Step-2: Filtering**

Provided a message x, find its category as sign ($w^T$x+b)

***SVM (Support Vector Machine)*:** A k-fold cross verification unsystematically divides the sample data set into k roughly same size subsets, drops one subset, constructs a classifier on the balance models in order to assess the categorization execution on the new subset [10]. Here procedure is iterated k- times on behalf of every subset to attain the cross verification execution above the entire exercise data set. If the exercise/training data set is huge, a diminutive subset can be employed for cross verification to reduce calculating charges. The algorithm mentioned below could be employed in current categorization procedure.

**Input:** Trial x to categorize
Training/sample set T={(x1,y1), (x2,y2),…. (xn,yn)};
Number of nearest neighbors k.

**Outcome:** Result yp∈{-1,1}
Find k sample (xi,yi) with minimal values of K(xi,xi) – 2 *K(xi,x)
Train an SVM model on the k selected samples
Categorize x using current model, to obtain the outcome yp
Return yp

***Decision Tree* :** The quality with the most standardized data is selected to mark the conclusion. The J48 algorithm formerly reoccurs on the less significant sub lists [11].

This process has some improper instances:

- Entire models in the account pertain to identical category. Once this takes place, plainly produces a leaf node for decision.
- Not any element gives a little information. According to example, J48 produces a result node advanced the tree employing the anticipated assessment of the group of classes.
- Case of imaginary category confronted. Moreover, J48 produces a resultant node advanced the ranking tree employing the anticipated assessment.

### B. Non-Classifier Based Approach

***DWT and Hough Transform*:** Analysts [12] acquiesced a fusion of text or character fragmentation method incorporated with Discrete Wavelet Transform and Hough Transform to take out text character from images. For training and testing, Ling-Spam Corpus database was employed. Primarily, pictures in colour are transformed into grayscale. Employing Otsu's procedure, the grayscale is changed into binary picture. After that, all linked elements that are below 15 pixels are eliminated from the binary picture. Following binarization, the lines and characters are partitioned by employing the advanced mixture method. The advanced sample was tried for exactness, False Negative, True Negative, True Positive, False Positive, recollect, accuracy, F-measure, and was established to be 100%, 0.99, 0.18, 0.81, and 0.008, respectively. Nevertheless, the dimension of training and testing data set was diminutive.

***Integrated Approach*:** A combined method of License plate detection is suggested by [13] employing Harris Corner and character partition from a picture. As the result of open structure, an Automatic License Plate Recognition (ALPR) has turned out being a crucial investigation focal point. Many arrangements were presented for license plate recognition, and each procedure had its own specific aims of concern and restrictions. The important measure in ALPR arrangement is the elaborate constraint of number plate, partition, identification. Harris corner algorithm finishes being energetic in altering movement and brightened lightning circumstances. The accuracy of license plate limitation is nurtured forward to the partition stage. The partition is carried out by a procedure of linked element study united with pixel count, aspect proportion, and height of characters.

***Projection Profile-based Technique:*** Projection Profile-grounded Method is a procedure for text partition employed right away in run-length contracted, printed English text documents [14]. Line partition is carried out employing the projection profile method. Furthermore, partition into words and characters is achieved by tracking the white runs by the foundation area of the text line. A character segmentation procedure employing projection profile-grounded method was originated initially by [15]. Primary view decision tree algorithm for cursive script identification grounded on the usage of histogram as a projection profile method was originated. The problems were related with quality and image handlings such as noise, distortion, variation in style, the shift of the character, size of the character, rotation, variation in thickness, and variation in texture.

## IV. EXPERIMENTAL EVALUATION

The proposed algorithm is tried with 150 images. For this, email images with text in it along with images taken from image spam data set have employed. A set of various characters of different font type, font style, font size, noise in background image, with low resolution, occluded images, special characters, and special symbol are taken for experimentation. As ther are 150 images, 410 lines and 5,280 characters involved, not all of them can be listed. Only a few images are shown as output of character segmentation. The number of characters in each image is not the same after segmentation.

The information that is accessible in the spam & ham-base data set is in both numeric and string arrangement. The sixty qualities in the data set depict proportionate frequencies of different prominent words and characters in emails. We wish to change these to Boolean values for the experiment. The quality will take a value 1 if the word or character is there in the email and 0 if it is not there in the email. To do this, we use a Numeric to Binary filter that will change all the numeric values to the binary. The changed data set is employed to train the classifier to discern spam from normal email by verifying the amount of occasions of every term or word for all the unsolicited and non-spam emails. The Precision, F-measure, Accuracy and Recall are estimated for every respective classifier.

From the outcomes measured, the classifiers grounded on the percentage of rightly categorized cases.

**Table 1.Output results for SPAM images**

| Correct rate(CR) | Error rate | Sensitivity | Specificity | Precision | Recall | F-measure | Accuracy |
|---|---|---|---|---|---|---|---|
| 82.3 | 17.7 | 100 | 78 | 0.909 | 1 | 0.95 | 96.7 |
| 85.4 | 14.5 | 100 | 82 | 0.909 | 1 | 0.95 | 96.7 |
| 82.26 | 17.74 | 83.33 | 82 | 0.909 | 0.833 | 0.87 | 95.1 |
| 77.42 | 22.6 | 100 | 78 | 0.91 | 0.75 | 0.82 | 96.8 |
| 83.9 | 16.13 | 100 | 86 | 0.91 | 0.75 | 0.82 | 93.5 |

**Table 2.Output results for HAM images**

| Correct rate(CR) | Error rate | Sensitivity | Specificity | Precision | Recall | F-measure | Accuracy |
|---|---|---|---|---|---|---|---|
| 82.26 | 17.72 | 100 | 79 | 0.89 | 1 | 0.95 | 95.16 |
| 82.32 | 17.62 | 100 | 78 | 0.90 | 1 | 0.95 | 95.6 |
| 74.2 | 25.81 | 100 | 68 | 0.91 | 1 | 0.95 | 95.2 |
| 82.25 | 17.75 | 66.6 | 86 | 0.88 | 0.67 | 0.75 | 95.13 |
| 80.64 | 19.35 | 91.67 | 78 | 0.86 | 0.91 | 0.91 | 93.54 |

**Table 3. Comparison with existing approach**

| Author | Method | Accuracy in % |
|---|---|---|
| Zhang et al. (2014) | ANN<br>SVM<br>Decision tree | 94.38<br>94.42<br>94.27 |
| Wu (2009) | ADTree | 91.60 |
| Lekha and Prakasam (2016) | SMO | 92.63 |
| Sharma and Arora (2013) | RANDOM TREE | 91.54 |
| Rusland et al. (2017) | Naïve Bayes Classifier | 82.88 |
| Author (2019) | Proposed method | 95.79 |

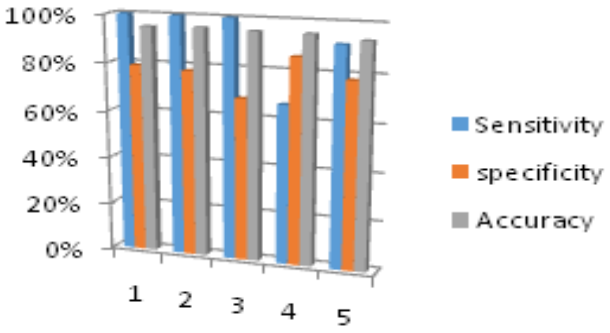


**Fig. 2. Performance measure of sensitivity, specificity and accuracy**

The output of the five different SPAM images after the segmentation process is tabulated in the table 1. The Average value of performance metrics obtained for this proposed algorithm is about 82.2 CR, 17.73 ER, 86.6% sensitivity, 81% specificity, 0.909 precision values, 0.866 recall, F-measure of about 0.883 and Accuracy of about 95.79%.The model improves output quality in provisions of both specificity and sensitivity. The output of five different HAM images after the segmentation process is tabulated in the table 2 which has the Average value obtained with these proposed algorithms are of 82.2 CR, 17.73 ER, 86.6% sensitivity, 81% specificity, 0.909 precision, 0.866 recall, F-measure of about 0.883 and Accuracy of about 95.79%. An algorithm which has a highest accuracy will be considered as the enhanced approach with better classification capability. Therefore, table 3 compared the accuracy range of few existing classifier studies.

The performance measure of specificity, sensitivity, and accuracy of image spam are shown in fig. 2. Based on graph analysis, it is transparent that the first input dataset achieved the sensitivity of 100%, specificity of about 78%and accuracy range of about 96.70%. Similarly the sensitivity, specificity, accuracy for the second dataset is 100%, 75% and 96.77% respectively and for the third dataset, sensitivity, specificity, accuracy is in the range of 100%, 69% and 95.16%.

Moreover for the fourth and fifth dataset, sensitivity, specificity, accuracy range is 67%, 86% 96.80% and 92%, 78% and 93.56% respectively.

## V. CONCLUSION

The spam emails not only waste computing resources and network bandwidth of the internet users', but it also, affects the larger scale, interrupt enterprises' of standard system process. In order to detect image spam email, a novel framework is proposed which is combination of character segmentation, recognition and classification technique (CSRC). The proposed framework exploits to take an advantage of processing low level features and extraction of embedded text data. As findings, proposed method able to reduce the impact of spam email and effectively filter out image spam messages. Secondly, considerably faster due to the input of spam dataset, this includes seed pixels which considerably decrease the space of possible classification process. The recommendation of future direction can combine the text and document reconstruction approach with character classifier towards a weighted decision about the class. Also extract the non-text content extracted features like hyperlinks, header and embedded images.

## REFERENCES


1. Kamboj, R. (2010). A Rule Based Approach for Spam Detection. Ph. D. Thesis, Computer Science and Engineering Department, Thapar University (TU), Patiala.
2. Radicati, S. & Hoang, Q. (2012). *Email Statistics Report*. [Online]. PALO ALTO. Available from: http://www.radicati.com/wp/wp_content/uploads/2012/04/Email-statistics-Report-2012-2016-Executive-Summary.pdf.
3. Statista (2017). *Global spam volume as percentage of total e-mail traffic from January 2014 to September 2016, by month*. [Online]. 2017. The Statistics Portal. Available from: http://www.statista.com/statistics/420391/spam-email-traffic-share/. [Accessed: 3 January 2017].
4. Biggio, B., Fumera, G., Pillai, I. & Roli, F. (2011). A survey and experimental evaluation of image spam filtering techniques. *Pattern Recognition Letters*. [Online]. 32 (10). pp. 1436–1446. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0167865511000936.
5. Syal, N. & Garg, N.K. (2014). Text Extraction in Images Using DWT, Gradient Method and SVM Classifier. International Journal of Emerging Technology and Advanced Engineering, vol. 4, no. 6.
6. Chandrasekaran, R. & Chandrasekaran, R.M. (2011). Morphology based text extraction in images. *International Journal of Computer Science and Technology*. 2 (4). pp. 103–107.
7. Almeida, T.A., Almeida, J. & Yamakami, A. (2011). Spam filtering: how the dimensionality reduction affects the accuracy of Naive Bayes classifiers. *Journal of Internet Services and Applications*. [Online]. 1 (3). pp. 183–200. Available from: http://www.springerlink.com/index/10.1007/s13174-010-0014-7.
8. Gerardnico (2017). *Machine Learning - K-Nearest Neighbors (KNN) algorithm - Instance based learning*. [Online]. 2017. Available from: https://gerardnico.com/wiki/data_mining/knn.
9. Mehdy, M.M., Ng, P.Y., Shair, E.F., Saleh, N.I.M. & Gomes, C. (2017). Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer. *Computational and Mathematical Methods in Medicine*. [Online]. 2017 (1). pp. 1–15. Available from:https://www.hindawi.com/journals/cmmm/2017/2610628/.
10. Anderson, C., Figa-Saldana, J., Wilson, J.J.W. & Ticconi, F. (2017). Validation and Cross-Validation Methods for ASCAT. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. [Online]. 10 (5). pp. 2232–2239. Available from: http://ieeexplore.ieee.org/document/7809062/.
11. Kumar, R., K.Suresh, S.Ramakrishna & M.Padmavathamma (2017). Development Of Data Mining System To Compute The Performance Of Improved Random Tree And J48 Classification Tree Learning Algorithms. In: *International Conference on Innovative Applications in Engineering and Information Technology(ICIAEIT-2017)*. 2017, pp. 128–132.
12. Rajalingam, M. & Sumari, P. (2016). An enhanced character segmentation and extraction method in image-based email detection. *International Journal of Control Theory and Applications*. 9 (26). pp. 171–179.
13. Panchal, T., Patel, H. & Panchal, A. (2016). License Plate Detection using Harris Corner and Character Segmentation by Integrated Approach from an Image. *International Conference on Communication, Computing and Virtualization*. [Online]. 79 (1). pp. 419–425. Available from: http://www.academia.edu/24314198/License_Plate_Detection_using_Harris_Corner_and_Character_Segmentation_by_Integrated_Approach_from_an_Image.
14. Javed, M., Nagabhushan, P. & Chaudhuri, B.B. (2013). Extraction of Projection Profile, Run-Histogram and Entropy Features Straight from Run-Length Compressed Text-Documents. In: *2013 2nd IAPR Asian Conference on Pattern Recognition*. [Online]. November 2013, IEEE, pp. 813–817. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6778437.
15. Rodrigues, R.J., Vianna, G.K. & Thomé, A.C.G. (2001). *Character Feature Extraction Using Polygonal Projection Sweep (Contour Detection)*. In: [Online]. pp. 687–695. Available from: http://link.springer.com/10.1007/3-540-45723-2_83.
16. Abdulla, S., Ramadass, S., Altaher, A., & Al-Nassiri, A. (2014). Employing Machine Learning Algorithms to Detect Unknown Scanning and Email Worms. *International Arab Journal of Information Technology*. 11 (2). pp. 140-148.
17. Yasin, A.& Abuhasan, A. (2018). Enhancing Anti-phishing by a Robust Multi-Level Authentication Technique (EARMAT). *The International Arab Journal of Information Technology*. 15 (6). pp. 990-999.


## AUTHORS PROFILE

**Mallikka Rajalingam** received her M.Sc Information Technology from Bharathidasan University, Tiruchirappalli, India in 2005, M.Phil Computer Science from Madurai Kamaraj University, Madurai, India in 2008, M.Tech Computer Science & Engineering from SASTRA University, Thanjavur, India in 2009. She worked as a Research Officer (RO) at School of Computer Science, Universiti Sains Malaysia (USM), Malaysia. She is currently pursuing the Ph.D. degree at the Department of Computer Science & Engineering, Bharathidasan University, Trichy, India. Her research interests include image processing, computer vision, pattern recognition, character recognition, document image analysis, text analysis and multimedia networking.

**Dr. M. Balamurugan** is currently working as Professor and Head in the Department of Computer Science and Engineering of Bharathidasan University, Trichy, India. He has credits of 20+ international and national conferences publications. He has published 30+ research papers in national and international journals. His research interests are mainly focused on the area of Data Science. He has supervised several research scholars in these areas.