A Review on Machine Learning Techniques for Image Based Spam Emails Detection

Muhammad Abdullahi
Department of Computer Science, Federal University of Technology, Minna, Nigeria
mibnmagaji22@gmail.com

Sulaimon A. Bashir
Department of Computer Science, Federal University of Technology, Minna, Nigeria
bashirsulaimon@futminna.edu.ng

Abdulmalik D. Mohammed
Department of Computer Science, Federal University of Technology, Minna, Nigeria
drmalik@futminna.edu.ng

Opeyemi O. Abisoye
Department of Computer Science, Federal University of Technology, Minna, Nigeria
o.abisoye@futminna.edu.ng

***Abstract*—Sending and receiving e-mails have continued to take the lead being the easiest and fastest way of e-communication despite the presence of other forms of e-communication such as social networking. The rise in online transactions through email has globally contributed to the increasing rate of spam emails relatively which has been a major problem in the field of computing. In this note, there are many machine learning techniques available for detecting these unwanted spams. In spite of the significant progress made in the figures of literature reviewed, there is no machine learning method that has achieve 100% accuracy. Each algorithm only utilizes limited features and properties for classification. Therefore, identifying the best algorithm is an important task as their strengths need to be weighed against their limitations. In this paper we explored different machine learning techniques relevant to the spam detection and discussed the contributions provided by researchers for controlling the spamming problem using machine learning classifiers by conducting a comparative studies of the selected machine learning algorithms such as: Naive Bayes, Clustering techniques, Random Forest, Decision Tree and Support Vector Machine (SVM)**

Keywords—Spam Image, Email Classification, Filtering Techniques.

I. INTRODUCTION

Email almost serves as a requirement for e-transactions. Sending and receiving e-mails have continued to take the lead being the easiest and fastest way of e-communication despite the presence of different types of e-communications. The rise in the applications of email and online transaction through emails has globally contributed to high rate of email spamming which has been a major problem in the field of computing. There are many machine learning methods available for detecting these unwanted spams. In spite of the significant progress made in the figures of literature reviewed, there is no machine learning method that has achieve 100% accuracy [1]. Each algorithm only utilizes limited features and properties for classification.

The successful and increasing use of the internet has encouraged a quick and easy types of online transactions and different ways of e-communication, the common example of this is emailing. However, it has become very common to send and receive emails as a major means of communication [1]. The increasing rate of spam mails is continuous and alarming, i.e. the bulk distribution of unwanted emails mostly of a commercial purpose with unpleasant content has subjected the service providers to a major problem [1] which endangers the confidentiality of the users and causes loss of resources. Since they are causing enormous misfortunes for the organizations, starting with the waist of bandwidth, mail server load to the profitability of clients due to the time spent identifying and handling spam mail senders. Spam messages do not only increase device correspondence and loss of storage facility but also used for numerous attacks and to bridge security measures. This violence can be used to abuse the client data and take their valuable sensitive data such as passwords and financial details [1, 3].

The latest survey study on email server revealed that 60% of all email traffic is spam, therefore making it mandatory to create an anti-spam filters. The current spam filters are developed for detecting different spam mails based on the features. In particular, the technique of text categorization is used to filter email spam. But spammers has employed a new way of succeeding the available filters by attaching a textual based content on image in the mail, experiencing image spam a another trick which is so far the most modern kind spam mail with obfuscation. Notwithstanding, emails have continue to maintain success in the area of online business transaction and are now are now a necessity for other means of online communication. Practically, almost all human uses emails. The author in [12] estimated that by the end of 2020, next to half of the global population expected to use emails.

The emails popularity and increase in its application for electronic communication has resulted to an increase in the amount of spam emails globally. Spam emails which are also known as junk emails are unsolicited message content sent by email to several recipients and not requested. Researchers in [13] opined that the spammers had no previous relationship with the recipient but send the spam mails on destructive purpose after collecting addresses from various sources such as tagged filled forms, phone book and spam messages. Spamming is a rapidly growing means of attack such as phishing, worms and virus as the most dangerous threat to the users of email [14, 15].

Supervised methods such as classification or prediction task aimed at discovering the hidden classes between the independent variable and target class are popularly used method for data processing according to [16]. Classifiers allow the observations to be assigned to tags for supervised learning, so that unobserved data can be classified in accordant with the trained data. Spam detection systems are focused on using estimate of arrangements to quantify messages as either spam or not.

In recent times undesirable business messages such as spam have become a major issue associated with the network. It alludes to the person sending spam messages as the spammers who collect email addresses from various web pages and chat rooms [22]. Spam ensures that the users are

not able to make proper utilization of time and storage space to the maximum rate. The considerable amount of spam mail that flows through computer networks has detrimental impact on email server memory space, bandwidth, user time and processing power [23]. The threat of spam email is growing on annual base and account for more than 77% of the entire traffic of email globally [23].

The rest part of this paper is structured as follows: Section 2 provides a brief review of related literature in the field of classification algorithms for the detection and filtering of spam email. Section 3 demonstrates the emerging spam filtering approaches and the essential description about the selected machine learning techniques for spam email classification. Section 4 conduct a comparison on the areas of their strength and limitations using performance metrics and present the result and discussion and lastly, Section 5 present the conclusion.

II. RELATED STUDIES

In the interest of the global research community, the rapid rise in email spam filtering is attributed to the increase in spam emails which has led many comparative studies by the researchers on the efficacy of spam image based email classification techniques using hybridized metrics. Hence it is important to identify the technique that can work better on a particular metric to support correct separation of emails to either be a spam or not. Here, we take an over view of the related and current scientific research works presented in the literature under the scope of approaches to filtering image spam-based emails that are low-level methods, Optical Character Recognition (OCR) based methods and those that involve both methods.

Chopra et al. [1] applied two stage approach for classifying the textual part of a given image to identify words in the mail as either spam or non-spam. In the first stage, OCR tool was used and Bayesian algorithm was used in the researchers stated in their paper titled “The Image and text spam filtering” that spammers has introduced new technique to embed spam mails into the image attached to the package. In an attempt to deal with this problem, the researchers are led to propose the method. The method was suggested, based on the hybridization of KNN and SVM. The fundamental concept is to classify the nearest neighbors to a verification problem and to prepare a close by SVM for the task of separation on neighbor array. Their work experiment was conducted using Dredze dataset and public dataset which shows that the results are approximately improved to 98% but limited to only accuracy as a performance metric.

Sadat M. and Rahmati in [4] suggested a method in their paper “A process for image spam detection using texture feature” where they used the image texture function to identify the spam image. In this study, the co-occurrence gray level matrix (GLCM) was applied as one of the texture characteristics to each image. Then to identify images with feature that each image acquired. The neighbors classifier k-nearest and the Bayesian naïve are used. The properties obtained are 22 attributes, and then the classifiers evaluate the images obtained from Dredze and Image Spam Hunter datasets. The dataset is divided by cross validation methods in to training set and test sets [4]. The result obtained from the classification covering four performance metrics: accuracy, precision, recall and F-measures in their experiment indicate an improvement in this research domain and compare with previous work, there is a substantial reduction in runtime but the study is limited to using only two classifiers.

Kumaresan T. et al. [5] proposed a solution that removes particularly low-level features such as image metadata and histogram features. Due to the extracted features, a SVM classifier is applied with the aid of a function of kernel to detect image spam, the accuracy obtained with the method is 90% but their work is limited because of the time complexity. In this paper, they used multiple image features to build classifiers for image spam. The classifiers used are the combination of SVM and PSO. PSO improves the output by iteratively scanning candidate solutions and also ensure that the particle in the search space are moved. Again, due to its computational complexity, PSO is conveniently applicable only to the dataset that are relatively small as compared to SVM [5].

Authors in [6] suggested an approach combining the properties of spam images with the density of corner points in the images to filter the spam image. The algorithm’s simple idea depends on the images proportion in the corner to determine whether it is a spam or not. The researcher presented that most of the technical approaches available for spam filtering are not effective for test messages imbedded into images and have identify this as a major problem hindering the performance of online transactions. The development of the proposed approach was done involving color edge detection, image binarization, and corner point detection. And after the experimental evaluation of the proposed approach, the result show that the detection rate of spam images is 90.5%. The 8-bit RGB mode is used for the analysis. The major point in this experiment is to identify the corner and conduct a statistical analysis and the limitation of this approach is that, it cannot handle crafty spams.

Meghali D. et al. [7] suggested a method for classifying the embedded image as spam or as a legitimate mail. The technique is based on an interpretation of the image containing only one region of text and the dataset used is Dredze dataset, Classification methods are applied in a hybridized manner. Particle Swarm Optimization is combined with Artificial Neural network for selection of features while the classifier for employed for spam classification and separation is Support Vector Machine. The learning ability of filters is the major strength of this method because every filter is different in terms of the data stored and model learned if every user receives different email but limited by complexity. The proposed framework is designed to handle both low level features and further processing of embedded text extraction. Their approach has been contrasted against other approaches and the result shows that AUC used in the proposed system for performance assessment is better than others methods [7].

Many conventional methods for detecting spam emails including the Bayesian method, the rule based system, Heuristic based filter IP blacklist, DNS black and white list holes have been made known[19]. They applied a neural system strategy where neurons were trained and proposed an efficient techniques based on neural network for spam classification component to enhance the exactness, accuracy and F-review. The proposed system is contrasted with SVM and the result indicate that system is doing relatively better. The performance metrics used for the comparison are precision and accuracy. The approach of the plan is introduced to improve the accuracy quotient of the current methods [19]. Approximately 1000 spam terms is included in the report. Due to the average performance of the proposed algorithm, it can be used with other algorithms to improve the spam detection.

Rathi and Pareek in [20] analyzed many methods of data mining for dataset containg 57 attributes with a single target feature in a discreet mode for the purpose of identifying the best approach for email identification and separation. The researchers analyzed the performance of different techniques for the classification operation in this paper. It was confirmed that the result showed a success in terms of accuracy when

the process of selecting the features was incorporated during the experiment. It was also noticed that the best classifier for spam mail detection with the accuracy of 99.72% was Random Tree and Random Forest to be the second in performance with an accuracy of 99.52% [20]. Researchers in [24] also focused mainly on spam email classification using machine learning techniques. The research is centered on concepts, actions, efficacy and patterns in spam filtering as well as the common machine learning approaches employed to combat the threat of spam.

III. METHODOLOGY

A. Research Questions

The purpose of this paper is to conduct a survey on spam detection approaches proposed by various researchers. In this study, two research questions were formulated which are:

- What are the major advantages and limitations affecting the performance of the current machine learning techniques?
- Which of the techniques performs better in terms of accuracy, precision, recall and the F-measure each?
- Does difference in dataset affect the performance of a classifier?

B. Research Objectives

And based on the research questions above, three research objectives were formulated.

- The first objective is to review and identify some common limitations of machine learning techniques.
- The second objective is to discover the technique that have the higher performance in terms accuracy precision, recall and F-measures on spam detection in research domain.
- The third objective is to investigate whether differences in dataset affect the classifiers' performance or not.

Publications on spam image-based detection techniques were searched, reviewed and eleven papers were selected in ascending order based on the number of citations in order to achieve the first, second and third objectives. Here, current machine learning techniques for spam detection are reviewed and their advantages and limitations were identified. The second and third objective was achieved by tabulating the performance result of the selected spam image-based detection techniques from the reviewed literature in tabular form and from which the technique with higher performance was identified with concentration on four major performance metrics such as: accuracy, precision, recall and F-measure was identified.

C. Methods

Here we discussed the methods that are used to recognize spam images, these methods are grouped into three distinct classifications including; header base, contest based and OCR based techniques [4]:

Header Based Techniques

Presently, it is common observed that email users always hide the client's header, however, this is the reason why most people cannot see their email header. Therefore, the header is produced along with the content of email. It is usual for e-mail messages to be used as an alternative to either activate display of e-mail or not. The major logic of this technique is to determine the piece of the email course wasted. The email header involves a number of fields that provide an important information margin [2, 4].

Content Based Techniques

These techniques are based on the extraction of features and the analysis of image content. These types of filters are used to examine and analyze the substance and techniques of the image [4]. The technique is geared towards the analysis of the different properties of the image and these characteristics are undesirably represented by the features of the image. It handles attribute and content such as image shape. The email body check for those properties used by the spammers. Email may be in form of image or text or even an image and a text combined. Text-based filtering approaches are often reliant on all forms of information and are reflective of the primary process and common ways to eliminate spam but spammers always seem to engage in a new tactic to trick the detection measures.

OCR based techniques

These techniques are usually applied to extract text embedding in the image using the OCR tool [4]. OCR is an electronic or mechanical representation of validated images that are manually typed, typewritten or content printed of machine encoded text. It is usually apply to turn books and records into electronic files to a modern record keeping model in an institute or to share the file of the site. OCR is able to find and alter the text, check for word and even phrase, store tightly, display or produce a copy free of scanning artifacts and then, apply techniques such as machine interpretation, text mining and speech text [7].

D. Machine Learning Techniques

In this section we provide description of some selected machine learning methods which have been applied to spam email classification and conduct a comparison on the areas of their strength and limitations.

Naïve Bayes

The classification process of this technique is an example of a learning techniques and also a predictive classification technique. It works as a basic probabilistic method which enable us to capture the clarity of the concept in an ethical manner by analyzing the likelihood of the result. It is applied to provide answers to analytical and quantitative problems [25]. Bayesian technique is named after a researcher who suggested the algorithm in person of Thomas Bayes (1702-1761). Classification provides functional learning methods and advanced information and analytical evidence may be combined. Bayesian classification provides a valuable framework for interpreting ad analyzing a variety of learning approaches. It determines the exact possibility for postulation and is resilient to noise in input data. It is a simple probabilistic method that is developed on the Bayes analysis, with valid assumptions which are independent in nature.

Clustering Technique

Clustering works by aggregating pattern classes in to a related group of classes. Clustering belongs to a category of approaches that divides case studies in to clusters

comparatively. This techniques has call the attention of scientific researchers and academics and have been used in various fields of practice. These techniques are unsupervised learning techniques and are used on the dataset of email spam with a true labels. Given that suitable representations are available, a good number of clustering techniques have the ability to classify email spam datasets in either spam or ham clusters. Whissel and Clarke in [26] have shown this in their research paper which was specifically written on email spam clustering.

Support Vector Machine

Support Vector Machines (CSVM) are controlled learning algorithms which have been established to perform better compared to the other learning algorithms aid. SVM is a category of algorithms that are introduced for handling classification and regression problems. SVM has used application while offering solutions of quadratic programming problems which have inequality weaknesses and sequential equality by differentiating different classes through hyper plane. It utilizes full advantage of the boundary [27]. Although the SVM may not be as swift as other classification algorithms, the algorithm draws it advantage from its high accuracy due to its ability to use multidimensional border of the model which is not linear or sequential.

Decision Tree

A Decision Tree (DT) is a classifier that uses a similar pattern with a tree structure. According to authors in [24, 26], decision tree induction is a distinctive method which contributes to information on classification. Decision tree nodes is either a leaf node that specifies the meaning of the intended function (class) or be a decision node that suggests that a certain verification is to be carried out with one branch and a sub tree as subset of the larger tree representing any likely test output. Decision tree learning is a technique that has been effectively used for filtering spam email. The aim of this approach is to produce a model of DT and train the model so as to predict the value of a target variable based on the total number of input variables.

Random Forest (RF)

This is a popular instance of an ensemble learning technique that is suitable for classification of data in to classes [26]. For the first time, random forest was proposed by researcher in [27]. The technique makes a specialized predictions using a tree structure. At the stage of training, some decision trees are created by the writer of the program. These decision trees are then applied for the task of predicting the group; this is done by considering the chosen groups of each tree and the category. These decision trees are then used for the purpose of predicting the group, this is done by taking into consideration the selected groups in each tree and the group with the highest number of votes is taken as an output. Random forest approach is gaining more prominence these days and has been applied in a number of field and literature to solve the analogous problem according to [26].

IV. RESULTS AND DISCUSSION

A summary of the reviewed machine learning techniques is presented in this section from the literature. Table 1. present a tabular form of the summary after achieving the first objective in section D. The details summaries consist of research year, reference number, classification techniques, advantages and the limitation of each technique.

TABLE 1: SUMMARY OF THE CLASSIFICATION TECHNIQUES.

Pub. year	Ref. No	Techniques	Advantage(s)	Limitation(s)
2017	[25]	Naïve Bayes classifier	-Handling of ambiguity by ethically influencing the probability of the results.	-Dependent on Bayesian filtering assumption (that events occurred independently in nature)
2016	[24]	Decision Tree	-Very short training period.	Not flexible for adjustment.
2016	[26]	Random Forests	-Higher performance with lesser classification error -Efficient mechanism during the data lost.	Longer training period
2015	[27]	Support vector machine	Capacity to modeI muItidimensio naI borderIines that are not sequentiaI or straightforward .	Slow classification,
2016	[26]	Clustering technique	- Ability to process encrypted messages, while preserving confidentiality.	-Inability to locate sensitive comparators. (its success depends on its ability to locate sensitive comparators)

While table 2 present the performance of the techniques relative to the dataset used. In order to achieve the second and third objective of this review, the detail summaries consist of publication year, reference no, dataset employed, the techniques, accuracy, precision, recall and F-measures.

TABLE 2: SUMMARY OF THE TECHNIQUES AS COMPARED FROM THE RELATED STUDIES.

Year	Ref. No	Dataset	Techniques	Accu-racy	Precis-ion	Re-call	F-Measure
2015	[1]	DredzeDataset	SVM and PSO	90%	-	-	-
2015	[3]	Spambase	Naïve Bayes	84%	89%	78%	-
2015	[4]	Dredze	KNN	91/41	87/03	99/53	92/86
			Naïve Bayes	75/49	78/98	82/12	80/52
		ISH Dataset	KNN	93/74	97/96	91/01	94/35
			Naïve Bayes	99/19	98/50	98/52	99/25
2013	[20]	Amazon.com	Random Forest	99.52%	-	-	-
			Random Tree	99.72%	-	-	-
2018	[21]	Spam_base	Random Forest	94.2 %	94.2%	94.2 %	94%
			Naïve Bayes	88.2 %	88.5%	88.5 %	88.5%
			Multilayer perceptron	93.2 %	93.3%	93.2 %	93%
			J48	92.3 %	92.3%	92.3 %	92.3%
2017	[28]	Dredze	Naïve Bayes classifier	98%	-	-	-
2013	[29]	Spambase	Random Forests	93.89%	95.87 %	94.10%	-
2016	[30]	Enron	Decision Tree	96%	98%	94%	-
2015	[31]	Spambase	SVM	79.50%	79.02 %	68.69%	-
			Naïve Bayes	76.24%	70.59 %	72.05%	-
2018	[32]	Spambase	ANN	92.41%	92.40 %	92.40	-

As provided in table 1, this study has investigated the strength and limitations of spam email detection techniques and identified handling ambiguity, short training period, high performance, capacity to modeI muItidimensionaI borderIines and capacity to model encrypted messages as the advantages while, limitations are complexity, slow classification, classification error and longer training period and inability to locate sensitive comparators. Also found that Random Tree has the highest percentage of accuracy of 99.72%, therefore it is the best classifier in terms of accuracy and we also discovered that accuracy is the most used performance metric in the literature. Decision tree has the highest precision of 98%, KNN has the best recall with 99/52 while Naïve Bayes is the best in terms of F-measures with 99/25. The investigation also shows that differences in dataset affect the performance of classifiers.

V. CONCLUSION AND FUTURE WORK

There are many machine learning techniques available for detecting these unwanted spams. In spite of the significant progress made in the volume and figure of literature reviewed, there is no machine learning method that has achieve 100% accuracy. Each algorithm only utilizes limited features and properties for classification. Therefore, identifying the best algorithm is an important task as their strengths need to be weighed against their limitations. It was noted that significant progress have been made based on the volume and figure of literature reviewed, hence, more research is required to improve the performance of hybrid system on Artificial immune system and to focus on the availability of well labeled dataset to ensure effective spam filtering. It has been also noted that there is an increasing use of internet and that, the increase in the use and application of internet is relative to the increasing rise of spam image.

REFERENCES

[1] Chopra, Nisha D., and K. P. Gaikwad (2015). "Image and text spam mail filtering." Int. J. Comput. Technol. Electron. Eng (IJCTEE) 5, no. 3.

[2] Ravikumar K, Gandhimathi P. A (2014) Review on Different Spam Detection Approaches.

[3] Renuka, D.K.; Visalakshi, P.; Sankar, T.J.I.J.C.A. Improving E-mail spam classification using ant colony optimization algorithm. Int. J. Comput. Appl. 2015, 2, 22–26.

[4] Sadat Hosseini M, Rahmati M. (2015)A Method for Image Spam Detection Using Texture Features.

[5] Kumaresan, T., Sanjushree, S., Suhasini, K. and Palanisamy, C., (2015). Image spam filtering using support vector machine and particle swarm optimization. Int. J. Comput. Appl, 1, pp.17-21.

[6] Wang, Jianyi, and Kazuki Katagishi (2014) "Image Content-Based" Email Spam Image" Filtering." Journal of Advances in Computer Networks 2, no. 2: 110-114.

[7] Das, Meghali, and Vijay Prasad (2014). "Analysis of an Image Spam in Email Based on Content Analysis." International Journal on Natural Language Computing (IJNLC) 3, no. 3, pp. 129-140.

[8] Liu, Tzong-Jye, Cheng-Nan Wu, Chia-Lin Lee, and Ching-Wen Chen (2014). "A self-adaptable image spam filtering system." Journal of the Chinese Institute of Engineers 37, no. 4, pp. 517-528.

[9] Foqaha, Mohammed Awad1and Monir.(2016) "EMAIL SPAM CLASSIFICATION USING HYBRID APPROACH OF RBF NEURAL NETWORK AND PARTICLE SWARM OPTIMIZATION.".

[10]. D.Sasikala, R.Roshiniya, Sarishnaratnakaran, Tapati Deb," Texture Analysis Of Plaque In Carotid Artery" International Journal Of Innovations In Scientificand Engineering Research(Ijiser), Vol 4 Issue 2 Feb 2017, Pp.66-70.

[11] J. M. Carmona-cejudo, G. Castillo, M. Baena-garcía, and R. Morales-bueno, "Knowledge-Based Systems A comparative study on feature selection and adaptive strategies for email foldering using the ABC-DynF framework," vol. 46, pp. 81–94, 2013.

[12] R. Group, "Email Statistics Report , 2016-2020," vol. 44, no. 0, pp. 0–3, 2016.

[13] A. Sharaff, N. . Nagwani, and A. Dhadse, "Comparative Study of Classification Algorithms for Spam Email Detection," Springer, no. January, 2016.

[14] A. F. Yasin, "Spam Reduction by using E-mail History and Authentication (SREHA),"Int. J. Inf. Technol. Comput. Sci., vol. Vol.8, no. No.7, p. pp.17-22, 2016.

[15] M. Iqbal, M. A. Malik, A. Mushtaq, and K. Faisal, "Study on the Effectiveness of Spam Detection Technologies,"mInt. J. Inf. Technol. Comput. Sci., vol. Vol.8, no. 1, pp. 11–21, 2016.

[16] S. M. Abdulhamid et al., "A Review on Mobile SMS Spam Filtering Techniques," IEEE Access, 2017.

[17] M. Zavvar, M. Rezaei, and S. Garavand, "Email Spam Detection Using Combination of Particle Swarm Optimization and Artificial Neural Network and Support Vector Machine," Int. J. Mod. Educ. Comput. Sci., vol. 7, no. July, pp. 68–74, 2016.

[18] P. Parveen and P. G. Halse, "Spam Mail Detection using Classification," vol. 5, no. 6, pp. 347–349, 2016.

[19] R. Sharma and G. Kaur, "E-Mail Spam Detection Using SVM and RBF," no. April, pp. 57–63, 2016.

[20] M. Rathi and V. Pareek, "Spam Mail Detection through Data Mining – A Comparative Performance Analysis," Int. J. Mod. Educ. Comput. Sci., vol. 5, no. December, pp. 31–39, 2013.

[21] S. M. Abdulhamid, M. Shuaib, O. Osho, I Ismaila, J. K. Alhassan,"Comparative Analysis of Classification Algorithms for Email Spam Detection", International Journal of Computer Network and Information Security(IJCNIS), Vol.10, No.1, pp.60-67, 2018.DOI: 10.5815/ijcnis.2018.01.07

[22] M. Awad, M. Foqaha, Email spam classification using hybrid approach of RBF neural network and particle swarm optimization, Int. J. Netw. Secur. Appl. 8 (4) (2016).

[23] D.M. Fonseca, O.H. Fazzion, E. Cunha, I. Las-Casas, P.D. Guedes, W. Meira,

M. Chaves, Measuring characterizing, and avoiding spam traffic costs, IEEE Int. Comp. 99 (2016).

[24] A. Bhowmick, S.M. Hazarika, Machine Learning for E-Mail Spam Filtering: Review, Techniques and Trends, arXiv:1606.01042v1 [cs.LG] 3 Jun 2016, 2016, pp. 1–27.

[25] Available at, Mail Server Solution, 2017, http://telco-soft.in/mailserver.php.

[26] S. Dipika, D. Kanchan, Spam e-mails filtering techniques, Int. J. Tech. Res. Appl. 4 (6) (2016) 7–11.

[27] Z.S. Torabi, M.H. Nadimi-Shahraki, A. Nabiollahi, Efficient support vector machines for spam detection: a survey. (IJCSIS), Int. J. Comput. Sci. Inf. Secur. 13 (1) (2015) 11–28.

[28] Al-Duwairi, Basheer, Ismail Khater, and Omar Al-Jarrah (2012). "Detecting image spam using image texture features." International Journal for Information Security Research (IJISR) 2, no. 3/4, pp. 344-35

[29] Sharma, S.; Arora, A. Adaptive approach for spam detection. Int. J. Comput. Sci. Issues 2013, 10, 23.

[30] Khan, Z.; Qamar, U. Text Mining Approach to Detect Spam in Emails. In Proceedings of the International Conference on Innovations in Intelligent Systems and Computing Technologies (ICIISCT2016), Las Piñas, Philippines, 24–26 February 2016; p. 45.

[31] Karthika, R.; Visalakshi, P.J.W.T.C. A hybrid ACO based feature selection method for email spam classification. WSEAS Trans. Comput. 2015, 14, 171–177.

[32] Bassiouni, M.; Ali, M.; El-Dahshan, E.A. Ham and Spam E-Mails Classification Using Machine Learning Techniques. J. Appl. Secur. Res. 2018, 13, 315–331.