Phishing Image Spam Classification Research Trends: Survey and Open Issues

Ovye John Abari[1], Nor Fazlida Mohd Sani[2], Fatimah Khalid[3],
Mohd Yunus Bin Sharum[4], Noor Afiza Mohd Ariffin[5]
Faculty of Computer Science & Information Technology,
Universiti Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia

***Abstract*—A phishing email is an attack that focused completely on people to circumvent existing traditional security algorithms. The email appears to be a dependable, appropriate, and solid communication medium for internet users. At present, the email is submerged with spam content, both in text-based form or undesired text planted inside the images. This study reviews articles on phishing image spam classification published from 2006 to 2020 based on spam classification application domains, datasets, features sets, spam classification methods, and the measurement metrics adopted in the existing studies. More than 50 articles, both from Web of Science and Scopus databases were picked. Achieving the study's target, we carried out a broad survey and analysis to identify the domains where spam classification was applied. Furthermore, several public data sets, features set, classification methods, and measuring metrics are found and the popular once were pinpointed. The study revealed that Personal Collection, Dredze, and Spam Archives datasets are the most commonly used datasets in image spam classification research. Low-level and image metadata are the most widely used features set. The methods of image spam classification as identified in this study are supervised machine learning, unsupervised machine learning, semi-supervised machine learning, content-based and statistical learning. Among these methods, the most commonly utilized is the Support Vector Machine (SVM) which falls under supervised machine learning. This is followed by Naïve Bayes and K-Nearest Neighbor. The commonly adopted metrics for the performance evaluation of the existing image spam classifiers are also identified and briefly discussed. We compared the performance of the state-of-the-art image spam models. Lastly, we pointed out promising directions for future research.**

***Keywords*—Phishing; spam; image spam classification; machine learning; deep learning**

I. Introduction

Phishing is a social engineering attack against people in a helpless society by controlling human beings into giving their confidential information to the cheats, called phishers. It is a criminal way of stealing internet users' private information using deceptive emails and counterfeit websites [1]. Phishing is also defined by [2] as a criminal instrument that utilizes both social engineering and specialized deception to take consumers' individual personality information and monetary account credentials. The coming of the Internet and the increasing number of its users have made email to be an important medium of communication. As of late, there has been an expanding utilization of emails and this has driven to the appearance of issues caused by phishing emails and spam. A typical email user gets around 40-50 emails per day [3].

According to [4], the entire number of phish identified in 1Q 2018 was 263,538. This was more than 45% from the 180,577 taken note in 4Q 2017. It was moreover higher than the 190,942 recorded in 3Q 2017. Likewise, the whole number of phishing identified in 2Q 2018 was 233,040, related to 263,538 in 1Q 2018. These sums are more than the 180,577 recorded in 4Q 2017 and the 190,942 watched in 3Q 2017. The phishing identified in 2Q and 3Q of 2019 were 112,163 and 122,359 respectively. Although there is a significant decrease in the phishing activities when compared with the figures of the previous years (2018 and 2017); however the request for phishing identification in our contemporary society is still a necessity to protect end-users from malicious emails. Phishing attacks are growing speedily in size and it's attacks expanding dynamically. This results in a serious economic loss around the world [1]. Fig. 1 depicts the statistics of phishing attacks in the 1Q of 2019 while Fig. 2 illustrates the most-targeted industry sectors in 2Q of 2019 [4].

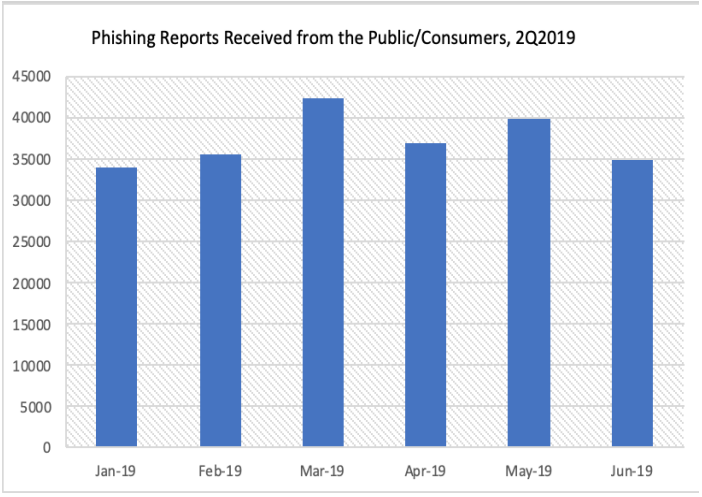

Fig. 1. Phishing Report of 2Q 2019 [4].

The past decade has seen the internet and emails to be flooded with spam content [5]. Regardless of constant awareness and the number of anti-spam algorithms emerging, spam contents are in increase [6]. Sending a large volume of spam contents at the server-side causes delays in service response, reducing the authenticity of the mail and consume a large portion of the storage space. At the user side, grouping the spam into valid and not valid, considering the large number

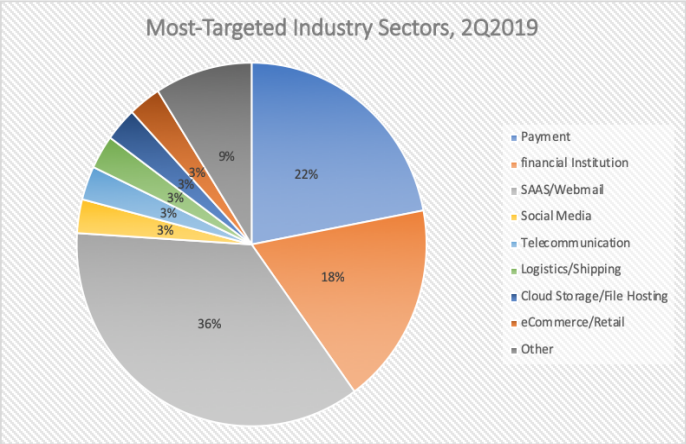

Fig. 2. Most Targeted Industry Sectors in 2Q 2019 [4].

of electronic mails that a user gets per day need devoting a substantial amount of time [7]. Spam messages are not restricted to email. Many people are exposed to spam content when they visit social networks like Telegram, Facebook, Instagram, Twitter, and so on. A study revealed that more than 70% of the total internet users use these social networks and are exposed to spam content [8].

Various algorithms have been designed to solve the problem of text-based spam. At present, spammers are sending these messages in the form of an image to confuse and possibly overpower these algorithms. Image spam is a concept that began in early 2005. More than 50% of the spam was made up of images by the end of 2006 [9], [6]. Image spam is another modern challenge in a phishing email. Image spam is email spam where a text content inserted into images to confuse conventional text-based spam channels [10]. It is a complex type of spam that is tempting and strenuous for the user to notice [5], [11]. Fig. 3 shows examples of spam images.

Fig. 3. Examples of Image Spams [10].

The objective of image spam is clearly to bypass the investigation of the content of text-based email performed by the existing spam algorithms. For this reason, spammers usually include some bogus text to the email together with the attached image such as a length of words that are persuasive or cogent to surface in genuine emails and not in spam [10].

Machine Learning (ML) is a branch of artificial intelligence that involved in creating algorithms that can modify itself using structured data without human intervention to yield expected results [12]. Examples are Linear Regression, Logistic Regression, Decision Tree, Support Vector Machine (SVM), Naïve Bayes, K-Means, and Random Forest. Deep Learning (DL) is a branch of machine learning in which algorithms are developed and function similar to those in machine learning, but there are multiple layers of these algorithms, and each providing a different meaning to the data it feeds on [12]. These algorithms include the Artificial Neural Networks (ANN), Deep Neural Network (DNN) and Convolutional Neural Network (CNN) [13]. In summary, machine learning algorithms need structural data, that is they are built to learn to do things by understanding labeled data, then use it to produce further outputs with more sets of data. However, they need to be retrained through human intervention when the actual output is not the desired one. While deep learning algorithms depend on layers of the artificial neural network. They do not require human intervention as the nested layers in the neural networks put data through hierarchies of different concepts, which eventually learn through their own errors [12].

There are different types of techniques used in classifying image spam as shown in Fig. 4 [3]. These are grouped into Supervised Machine Learning, Unsupervised Machine Learning, Semi-supervised Machine Learning, Content-based Learning, and Statistical Learning. Numerous researchers utilized these approaches for phishing email classification and detection. Depending on the nature of the data to be classified, choosing suitable and appropriate techniques is exceptionally crucial. The supervised machine learning algorithms often used from the surveyed literature are Decision Tree, Fuzzy Logic, Support Vector Machine, Neural Networks, Bayesian Network, and Genetic Algorithm. Some researchers compared two or more of these techniques to see which one produces better results [14], [15]. Deep learning approaches have not been well exploited in image spam classification since their advent [16]. They have the capability to handle large datasets and can extract image features more accurately than the existing image processing techniques [5].

Unlike other survey articles, we achieve comparisons of the performance of the existing state-of-the-art image spam models. Also, this review can help researchers working in the field of image spam classification by answering the following research questions:

(a) What are the various areas of application where image spam classification has been utilized?
(b) Which publicly available datasets can be accessed for the various areas of application of image spam classification?
(c) What are the commonly used features set in the existing image spam classification models?
(d) What performance evaluation parameters are applied to determine the effectiveness of the image spam classification algorithm?
(e) What are the challenges and research directions for future researchers working in the field of image spam classification?

The organization of the paper is as follows. Section 2 review the existing literatures or related works. Section 3

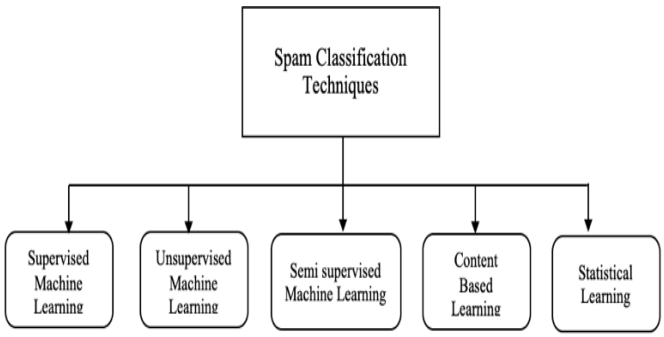

Fig. 4. Types of Techniques in Spam Classification [3].

discuss the future research directions. Section 4 gives the summary of the paper.

II. Related Works

The review of the related works is discussed under the following headings: Identification of spam classification application areas, spam classification dataset analysis and review, feature set analysis and review, and the analysis and review of spam classification techniques.

A. Identification of Spam Classification Application Areas

Basically, spams are categorized into Text-based and Image-based [5]. Spams are further divided into content-based spam and non-content-based spam. Content-based spam is the first-generation image spam [17]. This includes the spam in emails in text-based form. In this category, the extracted content from the body, headers, and keywords of emails are used by the classification algorithms to classify the images[5]. A wide range of machine learning techniques can handle this type of spam classification [6]. Non-content-based spam include complex kind of email spam and this falls into the second and third generation of image spam [17]. In this category, the undesired text is embedded in images. To classify the image spam, we can rely on the attributes of the image but recently, the advent of deep learning techniques make it possible to classify these images based on their raw byte form [5].

Images that fall in the first generation contain simple spam images hence they can be easily recognized by the optical character recognition (OCR) tools. In the second and third generation, the images contain noise and superimposing background to confuse and make them unrecognizable by the OCR. The OCR tools have the ability to partition the portions of the image that contain particular objects for the purpose of text extraction and detection [17], [5]. The background noise included with the text inside an image is a challenging task for OCR [17]. In this study, we are going to look at the application areas of spam classification under two (2) domains. Text-based and Image-based spam as shown in Table I.

B. Spam Classification Dataset Analysis and Review

This section shows the datasets that were used in spam classification and the detailed analysis. The researchers used public datasets in their works. They used one or more personal collections, Dredze, spam archive, Princeton spam corpus, image spam hunter, and so on as their datasets. For example, [33] used only Dredze dataset. The detailed analysis of data sets used in both text-based and image-based spam classification is shown in Table II and their locations in Table III.

Table II depicts the name of the datasets and sample size, the number of studies, and their references (where a specific dataset is used). This study reviewed that the Dredze dataset is the most commonly used datasets in image spam classification. This dataset consists of a total of 5789 spams (with 3239 spams and 2550 ham). Ten (10) studies adopted Dredze dataset, followed by Spam Archive dataset (with seven studies), image spam hunter (four studies), Trec07, ICDAR2003 and Char74k (two studies each), while the others datasets (Enron corpus, SMS spam, Princeton spam corpus, LingSpam, SpamAssassin and Indian corpus) have one studies each. Seventeen (17) studies used personal collection datasets from twitter. The location where the datasets can be downloaded and utilize are also presented and showed in Table III. Fig. 5 shows the name of datasets with the corresponding number of articles that adopted the datasets.

C. Feature Set Analysis and Review

This section discusses the feature sets used in all the studies under review. A feature describes the specific or distinctive attributes of image spam during processing. One of the essential steps to design efficient and accurate algorithms in spam classification is the feature extraction and selection [3]. A brief overview of these features is explained below. Table V shows the features used in image spam classification and Fig. 6 presents the graph of the number of articles versus the image features.

- Text area: This is the boundary the text occupied in an image. It is also called a text boundary. This is a way of identifying the presence of text in an image.
- Low-level (Color): These attributes are entropy values of the image RGB color, brightness, hue, and saturation. Other values include variance, skew, and the mean. The mean value represents the average pixel value of the image and it is applied to define the background of an image. In these features, there are distinct histogram attributes for a spam and ham image. Skewness is used in identifying the surfaces of an image. Spam images normally have high kurtosis values than ham images.
- Image similarity (Texture): The local binary pattern (LBP) is useful in measuring the similarity and information of adjacent pixels in an image. LBP is a powerful tool for identifying image spam which is simply text placed on a white background.
- Image region Similarity (Shape, Edge): Histogram of Oriented Gradients (HOG) determines how intensity gradient varies in an image. Edges are features used to detect spam images. It helps to identify boundaries in an image. A canny algorithm is an edge filter that is mostly used to determine the edges in an image.

TABLE I. DISTRIBUTION OF ARTICLES BASED ON THE APPLICATION DOMAINS.

Domain	No. of Studies	Reference
Text-based	9	[18], [19], [20], [21], [22], [23], [24], [25], [26]
Image-based	16	[27], [28], [29], [9], [30], [31], [32], [33], [34], [35], [17], [5], [36], [37], [38], [39]

TABLE II. DATASETS USED IN BOTH TEXT-BASED AND IMAGE-BASED SPAM CLASSIFICATION.

Dataset	No. of Studies	Sample Size	References
Spam archive	7	12053 spam (spam,=9503 & ham = 2550)	[20], [30], [34], [5], [37], [38], [39]
Dredze	10	5789 spam (spam =3239 & ham = 2550)	[30], [31], [33], [34], [40], [17], [5], [37], [28], [21]
Enron corpus	1	Not specified	[22]
SMS spam	1	Not specified	[18]
Princeton spam corpus	1	Total spam = 1004	[37]
Image spam hunter	4	1730 spam (spam =920 & ham = 810)	[21], [35], [40], [17], [5]
Trec07	2	Not specified	[21], [25]
ICDAR2003	2	11615 spam (train data =6185 & test data = 5430)	[29], [9]
Char74k	2	Total 62992 spam (train data = 44094 & test data = 18897)	[29], [9]
LingSpam	1	Not specified	[25]
SpamAssassin	1	Not specified	[25]
Personal collection & twitter	17	5326 spam (spam =3299 & ham = 2027)	[20], [24], [30], [31], [41], [32], [34], [35], [17], [40], [5], [36], [37], [38], [39], [27], [18]
Indian corpus	1	Not specified	[19]

- Image metadata: These attributes contain the depth, width, height, and compression ration of the image files. Mathematically, in an image, the compression ratio (CR) is given as:

$$CR = \frac{height * width * channels}{size\ of\ file} \quad (1)$$

- Text Obfuscation (Noise): Signal to noise ratio (SNR) and entropy of noise are the two attributes of noise. Spam images usually contain less noise than ham images. The percentage of mean to standard deviation of an image is the SNR.

Several researchers as showed in Table IV used image features to identify an image spams [30]. For instance, [30] proposed an image spam classifier using Maximum Entropy, Decision Tree, and Naïve Bayes methods. They focus only on the low level and image metadata features of the image for the classification and achieved an average accuracy of 95% with a computation time of 2.5-4.4ms. They considered a few features set for the training of the algorithm. Features reduction and elimination techniques such as principal component analysis (PCA), recursive features elimination (RFE), and univariate features selection (UFS) are very vital in optimizing or reducing the number of features in an image in order to achieve better feature classification and accuracy. Author in [35] used PCA and SVM to developed a classifier for image spam. They used a few image spam hunter and personally collected datasets to trained their classifier and claimed 70-97% accuracy. They did not take the processing time into consideration. Author in [17] used the same feature reduction and elimination approach in their work. The authors looked at 38 features of the image and used RFE and UFS to reduce the undesirable features. They employed the SVM method to train their classifier using 920 spam and 810 ham of image spam hunter dataset and 1089 spam and 1029 ham of Dredze and personal collected dataset. Accuracy of 54-98% and false-positive of 0.01-0.79 were obtained. The time taken for the classification was not considered.

D. Spam Classification Techniques Analysis and Review

Spam Email classification techniques as depicted in Fig. 4 are categorized into five (5) groups. These are supervised machine learning, unsupervised machine learning, semi-supervised machine learning, content-based learning, and statistical learning [3], [42], [43]. In supervised machine learning, input instances are given for the learning procedure and the output labels do not conveniently recognize a function that approximates this behavior. Supervised machine learning techniques include Decision Tree, Naïve Bayes, Support Vector Machine, K-Nearest Neighbor, Bayesian Network,

TABLE III. LIST OF PUBLICLY AVAILABLE DATASETS AND THEIR CORRESPONDING LINKS.

#	Dataset	Location
1	Spam archive	https://archive.ics.uci.edu/ml/datasets/sms+spam+collection
2	Dredze	http://www.cs.jhu.edu/∼mdredze/data/
3	Enron corpus	http://www.aueb.gr/users/ion/data/enron-spam/
4	SMS spam	https://www.kaggle.com/uciml/sms-spam-collection-dataset
5	Princeton spam corpus	https://www.cs.princeton.edu/cass/spam/
6	Image spam hunter	https://users.cs.northwestern.edu/∼yga751/ML/ISH.htm#dataset
7	Trec07	http://plg.uwaterloo.ca/∼gvcormac/treccorpus07/
8	ICDAR2003	http://algoval.essex.ac.uk/icdar/Datasets.html
9	Char74k	http://www.ee.surrey.ac.uk/CVSSP/demos/chars74k/
10	LingSpam	http://www.csmining.org/index.php/ling-spam-datasets.html
11	SpamAssassin	http://spamassassin.apache.org/publiccorpus
12	Personal collection & twitter	Not available
13	Indian corpus	Not available

Random Forest, Fuzzy logic, Multilayer Perceptron, Neural Networks, and deep learning methods such as Convolution Neural Network. In unsupervised machine learning, the learning procedure is equipped with input instances but with output labels. Here, the leaning procedure tries to recognize related patterns through input instances to determine output. An example of unsupervised machine learning is k-means clustering [3]. Semi-supervised machine learning is a combination of supervised and unsupervised machine learning. In semi-supervised machine learning, some of the input datasets are labels and the learning procedure requires large labelled data. Active learning is one of the examples of semi-supervised machine learning. Content-based techniques use keywords in classifying the spam email [3]. Examples are optical character recognition (OCR) and Sobel filters. In statistical learning, each keyword is assigning a probability and the overall probability is used to classify the image spam. Supervised machine learning is the most frequently used techniques in spam classification even though researchers used all the other types of techniques. Table IX presents the distribution of spam classification techniques [3]. Thirty (30) studies adopted supervised machine learning techniques, four (4) used unsupervised techniques, eight (8) and five (5) studies adopted content-based learning and statistical learning respectively.

Fumera et al. [20] developed an algorithm for detecting and classifying text-based spam using optical character recognition (OCR) tool where they used 445 spam and 4852 ham of spam archive dataset and 5608 spam and 9526 ham of personally collected dataset to train their model using support vector machine (SVM). The authors focus only on the true positive and false positive rate and the result obtained are 0.81 and 0.01 respectively. They did not consider the time taken for the classifier to detect and classify a spam email and the method used is inefficient since it cannot handle large datasets conveniently. The proposed classifier cannot detect image spam email. The same OCR tool was used in the work of some researchers [21], [22], [24], [23]. They examined and applied OCR software to filter image spam email. While [22] used KNN, Naïve Bayes and Reverse DBSCAN in his work, [24] used Sobel operators (filters) to process the image as displayed in Table VI.

Image spam classifiers have been proposed using a near-duplicate detection approach but with different distance measurements [39], [38], [37], [36]. They both considered low level and image similarity features of the image spam in training their models. While [39] used Visual and Object Semantics as a distance measure to classified the image spam and achieved an accuracy of 96 %, [38] used Histogram and Euclidean distance measures to obtain a better result of 98% accuracy. The reason for the difference observed in the two results was because of the former used a larger dataset than the later. The computation time was not considered except in the study of [36]. The time taken to detect image spam and classify it as either spam or ham in this research is 50ms. This is displayed in table VII. Table VIII presents the keys of the abbreviations as used in Tables IV, V, VI and VII

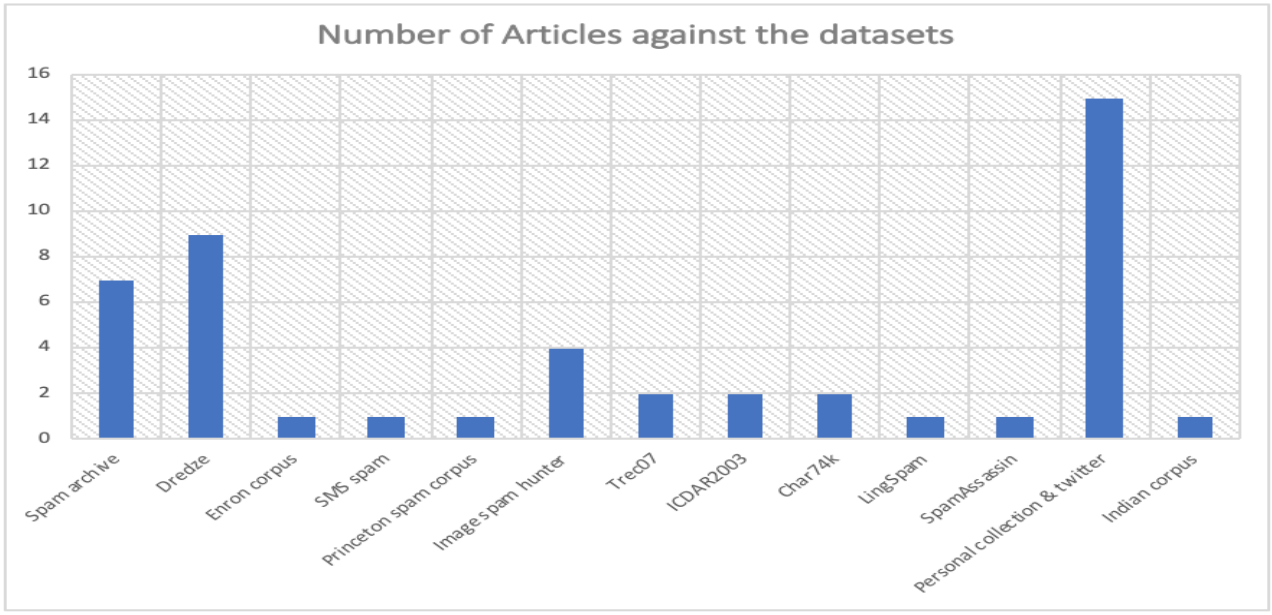

Fig. 5. Number of Articles versus Datasets

TABLE IV. RESEARCHES ON IMAGE-BASED CLASSIFICATION USING IMAGE FEATURES.

Ref	Features						Method	Dataset & Size				Results	Time(ms)	Remark
	TA	LL	IS	IRS	IM	TO		DS	Spam	Ham	Both			
[30]		y			y		Decision Tree, Naïve Bayes	d, sa pc	3239 9503	12742 2550		Acc=90 – 99	2.5 – 4.4	Few features
[31]	y	y				y	SVM	d pc	3239 8549	2550 2006		TP=0.94 - 0.98 FP=0.02 - 0.05	1200	Real-time not achieved
[32]		y			y		SVM & Active Learning	pc	1190	—	—	Acc= 99.0- 99.3	—	Few features
[33]		y				y	SVM	d	—	—		Unspecified	—	Few LL features
[34]		y			y		SVM	d sa pc	3203 9280	1786 1371		Acc=95	—	Only LL features
[35]		y	y		y		PCA, SVM	ish pc	920 1000	810		Acc=0.70 - 0.97 FP=0.04-0.25	—	Method inefficient
[17]		y			y	y	SVM	ish d pc	920 1089	810 1029		Acc=0.54 - 0.98 FP=0.01 - 0.79	—	Method inefficient
[5]		y			y	y	NN, DNN, CNN	d ish pc sa		2681	19920 1000	Acc=95.63 - 98.95	—	RFE,UFS not used

Support Vector Machine (SVM) method is one of the most commonly used classification algorithms in image spam classification [3] and has been adopted by many researchers in their works [44], [35], [17], [33], [32], [31]. SVM is suitable for binary classification problems but difficult to handle large datasets [27]. In the work of [31], in order to identify the image as spam or ham, they considered 3 features domain namely, text area, low-level features (image color), and text obfuscation (noise) of the image. They claimed to have obtained 94-98% accuracy with 1200ms computation time.

Singh [5] proposed an image spam algorithm using deep learning algorithms. They did not consider the time it took to identify and classify image spam and used only a few datasets concentrating on low level, image metadata and image obfuscation (noise) features of the image. They obtained 95.63 to 98.95% accuracy. An approach to object segmentation was not used to detect the segmented spam area. After their advent, deep learning has not been well exploited in classifying image spam. Deep learning has the ability to handle large dataset and can more accurately extract image features than existing image processing techniques [5].

Web content-based approaches can be combined with machine learning techniques to build a system for phishing website and email detection [45]. The author in [45] used this approach to designed a 92% accuracy detection system known as CANTINA+. Web structured-based method using Google PageRank has been

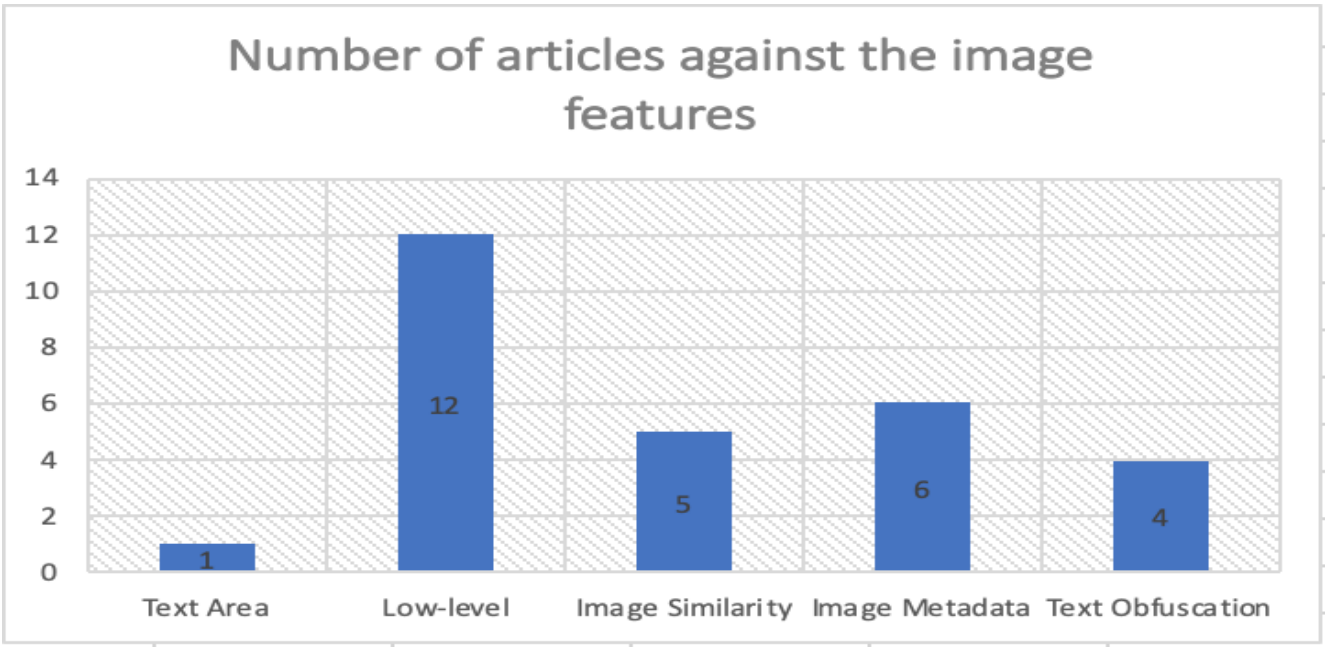

Fig. 6. Number of Articles versus Image Features.

TABLE V. FEATURES USED IN IMAGE SPAM CLASSIFICATION. SEE TABLE VIII FOR THE ABBREVIATIONS OF THE FEATURES WHERE "+" MEANS 'USED' AND "-" MEANS 'NOT USED'.

Image features						Ref
TA	LL	IS	IRS	IM	TO	
-	+	-	-	+	-	[30]
-	+	+	-	-	-	[36]
+	+	-	-	-	+	[31]
-	+	+	-	-	-	[37]
-	+	+	-	-	-	[39]
-	+	+	-	-	-	[38]
-	+	-	-	+	-	[34]
-	+	-	-	+	-	[32]
-	+	+	-	+	-	[33]
-	+	+	-	+	-	[35]
-	+	-	-	+	+	[17]
-	+	-	-	+	+	[5]
-	-	-	+	-	-	[29]
-	-	-	+	-	-	[9]
1	12	5	2	6	4	Total

TABLE VI. RESEARCHES ON TEXT-BASED CLASSIFICATION USING OCR METHOD.

Ref	Method	Dataset and size			Results	Time(ms)	Remark
		DS	Spam	Ham			
[20]	OCR, SVM	sa	445	4852	TP= 0.77-0.81	__	Method not efficient
		pc	5608	9526	FP=0.01		
[21]	OCR	Trec07, d, ish	__	__	Acc=99.83%	__	The OCR not suitable
[22]	OCR,KNN, Naïve Bayes, Reverse DBSCAN	Enron corpus	__	__	Acc= 87%	__	The OCR not reliable, restricted to certain fonts
[23]	OCR	__	__	__	__	__	OCR not suitable
[24]	Sobel filters, OCR	pc	3299	2027	Acc=45.30 - 90.12	2.6	Method not efficient

TABLE VII. RESEARCHES ON IMAGE CLASSIFICATION USING NEAR-DUPLICATE APPROACH.

Ref	Features						Distance measure	Dataset & Size			Results	Time(ms)
	TA	LL	IS	IRS	IM	TO		DS	Spam	Ham	Both	
[36]		y		y			Manhattan	pc	1071	107	TP=0.63 - 0.96 FT=0 - 0.173	50
[37]		y		y			Jensen-Shannon	sa psc, d, pc	1004	__	Acc=95 - 98 TP=0.76 - 0.84	-
[38]		y		y			Histogram, Euclidean	pc, sa	1977	8000	Acc=81-98	-
[39]		y		y			Visual and Object Semantics	sa pc	6459	1473	Acc=96.66 FP=3.34	-

used to achieve 98% accuracy in classification [46]. A Bayesian algorithm and the incremental forgetting weight algorithm were used to create a model that effectively tackled idea drift and data bias in the classification of spam emails [25]. It is possible to Combine statistical analysis of website URLs with machine learning techniques to develop a classification algorithm with a better precision rate [47].

Many researchers work on detecting and classifying email phishing but did not focus on spam emails. [14], for example, used the dataset gathered from twitter and implemented an algorithm using SVM, KNN, Random Forest, and classification features to improve the accuracy of phishing tweets detection. Their findings yield 94.75% classification accuracy with only 11 selected features, which is higher than 94.56% obtained by other researchers who used more than 11 features for the same dataset. To build a phishing detection model and solve the complexities of phishing attacks

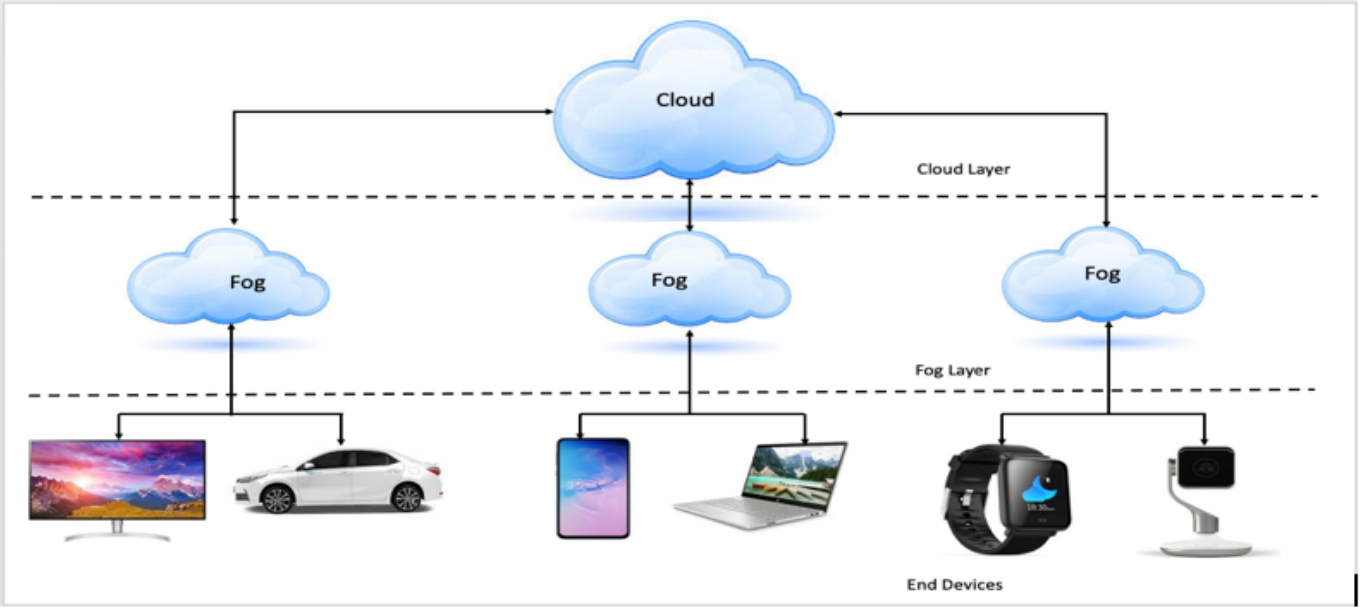

Fig. 7. Fog Computing Architecture [48].

TABLE VIII. KEYS USED IN TABLES IV, V, VI, & VII.

Features	
TA	Text area
LL	Low-level
IS	Image similarity
IRS	Image regions similarity
IM	Image metadata
TO	Text obfuscation

Datasets	
pc	Personal collection(s)
d	Dredze
sa	Spam Archive
psc	Princeton Spam Corpus
ish	Image Spam Hunter

Results	
Acc	Accuracy
TP	True positive rate
FP	False positive rate

in the real world, deep packet inspection, and software-defined networking (SDN) techniques with artificial neural networks (ANN) were applied. They reported a 98.39% accuracy and their model can provide an effective and efficient solution for detecting and minimizing phishing emails [49].

One of the hybridized approaches used in email phishing detection is neuro-fuzzy, which is the combination of fuzzy logic and neural network. [1] used this approach to developed an anti-phishing model and obtained an improved detection accuracy of 98.36%. A better result of 99.29% accuracy was obtained using the same method [50]. While [50] research did not focus on missed detection and false alarm rates, a high rate of missed detection and a false alarm was reported by [1].

In the literature, decision tree data mining techniques such as associative rule mining and classification were well used. A classification algorithm has been proposed using these methods to derive new rules from the phishing data sets [51], [52]. The main challenge with this approach is that the set of rules is not objective and largely depends on the programmer [1]. A classifier that can categorize emails written in Chinese into spam or ham based on a specific feature was created using the same method [26]. Data mining knowledge discovery procedures were used to develop an intelligent classification model that was tested using Random Forest, J48, SVM, MLP, and Bayes Net. Using the Random Forest and J48 algorithm, an accuracy of 99.1% and 98.4% was achieved respectively [53].

Convolutional Neural Network has recently been used to create a text-based spam classifier with the introduction of long short time memory neural network (LSTM NN) and an accuracy of more than 92-98% has been achieved [18]. [28], [44] used KNN and Naïve Bayes to implemented his work with the Dredze image dataset. The authors used a distributed associative memory tree to extract features of the image. This feature extraction method performs best in comparison with other distributed approaches with a relatively small amount of resources for spam detection. A 98% accuracy has been reached [28]. A Random Forest has the best accuracy, precision, recall, and F-measure than SVM and multilayer perceptron when PCA was used to construct a twitter-dataset image spam model. An accuracy of 96.3% has been achieved in this study [27].

Naiemi et al. [9] proposed a new algorithm to recognize characters in image spam by improving the existing feature extraction of HOG using SVM as the classifier. The study improved scale and translation robust HOG (STRHOG) developed with the Chars74K dataset with an accuracy of 72.2% [29]. In STRHOG, the matrices of the oriented gradient for input images of different sizes have a high computation value and a large part of this matrix does not have any effect in recognizing the image. [9] were able to overcome these problems in their work and obtained a detection accuracy of 84.91%. Some of this study's weaknesses are briefly debated. Support Vector Machine (SVM) adopted in the work is good and suitable for problems in binary classification [27]. SVM works perfectly when dealing with 2,3,4 classes but the Char74K dataset used in the work has 62 classes and is therefore a multiclass problem. Additionally, we are trying as much as possible not to lose data in machine and deep learning. In fact, generating data for any missing attribute within a dataset is advisable. In HOG, the image passes through cropping, and in the process, data is loose. Finally, the study did not consider the

time it took to detect and classify the image spam. Because of its complex computation, the canny algorithm used for the edge detection consumes a lot of time and it will be hard to implement to hit the real-time response.

In most of the reviewed articles, the computational time was not considered. Table X shows the reference of the articles that considered time in text-based and image-based spam classification.

TABLE X. LIST OF ARTICLES THAT CONSIDERED COMPUTATIONAL TIME IN SPAM CLASSIFICATION.

#	Ref	Application Domain	Computation Time (ms)
1	[24]	Text	2.6
2	[30]	Image	2.5 - 4.4
3	[31]	Image	1200
4	[36]	Image	50

E. Performance Metrics Review and Analysis

Confusion matrix (CM) as shown in Fig. 8 measure the performance of a classification algorithm in terms of accuracy, recall, precision, and F-measure. These definitions are enumerated below. CM is a matrix between True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). TP is when the image is a spam image and the classifier label it as spam. TN is when the image is a ham image and the classifier label it as a ham. FP is when the image is a ham image and the classifier label it as spam. FN is when the image is a spam image and the classifier label it as a ham [5].

Fig. 8. Confusion Matrix [5].

The often-utilized performance metrics and their formulas as highlighted in the works of [3], [55] are discussed below.

(a) Accuracy: This is the percentage of predictions that are correct. It is used to determine how well a classifier works. It is defined mathematically as:

$$Accuracy = \frac{TP + TN}{P + N} \tag{2}$$

where P = TP + FN and N = TN + FP

(b) Precision: This is the percentage of image spam classified correctly as ham. It is calculated as:

$$Precision = \frac{FPR}{100} = \frac{\frac{FP}{N}}{100} = \frac{\frac{FP}{FP+TN}}{100} \tag{3}$$

(c) Recall: This is the percentage of image spam classified correctly as spam. It is defined as:

$$Recall = \frac{TPR}{100} = \frac{\frac{TP}{N}}{100} = \frac{\frac{TP}{TP+FN}}{100} \tag{4}$$

(d) F-Measure: This is how effectively a classifier identifies positive labels. It is the weighted average of precision and recall. F-Measure is calculated as:

$$F - Measure = \frac{2 * Recall * Precision}{Recall + Precision} \tag{5}$$

(e) Simplicity: This is how effectively a classifier identifies negative labels. It is defined as:

$$Simplicity = \frac{TN}{FP + TN} \tag{6}$$

(f) Area Under Curve (AUC): This is the ability of a classifier to prevent incorrect classification. It is given as:

$$AUC = \frac{1}{2}(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}) \tag{7}$$

The performance of the existing state-of-the-art image spam models using the above metrics is shown in Table XI. The existing works considered one or more of the performance metrics. For instance, [36], [37], [31], [40], [5], [56] considered only the accuracy.

III. FUTURE RESEARCH DIRECTIONS

We discuss some of the challenges and open issues in the existing studies on image spam classification research in this section.

- Dataset: Image spam classification is a binary classification problem (ham or spam). Some of the datasets used in the reviewed articles have four or more classes and these types of datasets are suitable and work perfectly for a multiclass problem and not for binary problems. A more challenging dataset is required for future image spam classification research.
- Optical Character Recognition (OCR) Approach: Most of the existing works used the OCR technique. In the OCR method, data is lost by cropping the image during the pre-processing stage. The images don't have the same dimension and are forced to be of the same size, thereby losing some of the important data. Machine learning algorithms tries as much as possible not to lose data. In fact, it generates data for any missing attribute in a dataset. More suitable techniques are needed for the extraction of the features of an image in future research on image spam classification.
- Deep Learning Technique: The state-of-the-art image spam classifiers developed using machine learning techniques, which work with few datasets have difficulty in extracting the relevant features of the images and this has negative effects on the overall output of the classification. Deep learning models have the capability to handle large datasets and can extract image features more accurately than machine learning techniques [5]. This approach has not been well exploited in image spam classification since its advent [16]. With this in mind, the future image spam classifier can be implemented using deep learning techniques like deep neural networks, and convolutional neural networks to make the classifier more powerful and improve the performance in terms of the accuracy and precision of the classification algorithms.
- Fog Architecture: Fog Computing, also known as fog networking or fogging is a newly introduced concept. It is an internet of thing (IoT) architecture that expands the cloud so that it is closer to end devices. It supplies information, computing asset like storage and application services to the end devices. More also, at the edge of networks, fog bolsters high versatility because it pulls services given at places close to the end-users [61]. Fig. 7 shows the architecture of fog computing where it clearly depicts the three (3) layers namely, end devices (IoT) layer, fog layers, and cloud layers [48]. This concept which was recently used by [1] to detect phishing websites produced high detection accuracy. Also, the authors revealed that fog-based services are faster than cloud-based services and that it is manageable and easy

TABLE IX. SUMMARY OF SPAM CLASSIFICATION TECHNIQUES WHERE "Y" MEANS 'YES' AND "X" MEANS 'NO' IN RESPECT TO TEXT-BASED AND IMAGE-BASED CLASSIFICATION.

Category	Method	No. of Articles	Text-Based	Image-Based	Reference
Supervised Machine Learning	1. Decision Tree	2	y	y	[30], [26]
	2. Naïve Bayes	3	y	y	[22], [30], [28]
	3. Support Vector Machine (SVM)	9	y	y	[20], [31], [32], [33], [34], [35], [17], [27], [9], [44]
	4. K-Nearest Neighbor	3	y	y	[22], [28], [29], [44]
	5. Bayesian Network	1	y	x	[25]
	6. Random Forest	1	x	y	[27]
	7. Fuzzy Logic	0	x	x	-
	8. Multilayer Perceptron	1	x	y	[27]
	9. Neural Networks	2	x	y	[28], [5]
	10. Deep Neural Netwoks	1	x	y	[5]
	11. CNN	3	y	y	[5], [18], [19]
	12. CNN+LSTM	2	y	x	[19], [18]
Unsupervised Machine Learning	13. K-Means Clustering	1	x	y	[27]
	14. Reverse DBSCAN	1	y	x	[22]
	15. Manhattan Distance	1	x	y	[36]
	16. Visual and Object Semantic Distance	1	x	y	[54]
Semi-Supervised Learning	17. SVM + Active Learning	1	x	y	[32]
Content-Based Learning	18. OCR Filter	5	y	x	[20], [21], [22], [23], [24]
	19. HOG	2	x	y	[29], [9]
	20. Sobel Filter	1	y	x	[24]
Statistical Learning	21. PCA	2	x	y	[27], [35]
	22. Jensen-Shannon	1	x	y	[37]
	23. Histogram/ Euclidean Distance	1	x	y	[38]
	24. Distributed Associative Memory Tree	1	x	y	[28]

to implement a machine learning algorithm on fog nodes than on the cloud. In view of this, an algorithm can be implemented on a fog node to increase the detection speed of the image spam classification.

- Computation Time: Image spam detection and classification should be a real-time process in order to minimize response delay. In the reviewed articles, the time taken to classified the image is neglected. The canny algorithm mostly used for edge detection in the histogram of oriented gradients (HOG) method consumes a lot of time due to its complex computation. It is difficult to implement to reach the real-time response. Future research should consider reducing the processing and classification time using recent hardware technology.

IV. CONCLUSION

This study provides a thorough overview of image spam classification studies to help researchers in this field in gaining excellent knowledge and understanding of current image spam classification solutions in the major areas. Journal articles published between 2006 to 2020 on image spam detection and classification were thoroughly studied and grouped into two application domains; text-based and image-based. The selected papers were analyzed from five dimensions of rationality: spam classification application domains, datasets adopted and features sets utilized in the two application domains, the methods used, and the matrices considered for the performance evaluation. More than 50 articles on spam classification were energetically picked and examined. A comprehensive analysis of several techniques, features set, datasets, and performance evaluation metrics used in spam detection and classification were summarized. The survey revealed that Personal Collection, Dredze, and Spam Archives datasets are the most commonly adopted datasets. Similarly, low-level and image metadata features are the most widely used features sets in spam classification research. The various methods of image spam classification as pinpointed in this study are supervised machine learning, unsupervised machine learning, semi-supervised machine learning, content-based and statistical learning. Among these methods, the most commonly used is the supervised machine learning method. Support Vector Machine (SVM) provides the best performance and it is often used in supervised learning. This is followed by Naïve Bayes and K-Nearest Neighbor techniques. The commonly investigated matrices for the performance evaluation are accuracy, recall, precision, f-measure, simplicity, and confusion matrix that depicts the relationship between TP, TN, FP, and FN. Finally, we present promising directions for future research.

FUNDING

This work was kindly supported by Putra Grant Scheme under project no: 9621600

TABLE XI. PERFORMANCE OF EXISTING STATE-OF-THE-ART IMAGE SPAM CLASSIFIERS USING THE ABOVE METRICS

Ref	Performance Metrics				Method	Dataset
	Accuracy	Precision	Recall	F-Measure		
[36]	0.92	–	–	–	Nearest Neighbour (Manhattan distance)	Personal Collection
[30]	0.91 0.80 0.87	– – –	– – –	0.93 0.83 0.89	MaxEntropy Naive Bayes Decision Tree	Dredze, Spam Archive
[37]	0.97	-	-	-	Nearest Neighbour (Jansen Shannon), and SVM	Spam Archive, Dredze, Princeton, and Personal collection
[31]	0.96	–	–	–	SVM	Dredze and Personal collection
[21]	0.99 0.88 0.99 0.88 0.68	1 0.99 1 0.99 1	0.99 0.83 0.99 0.83 0.53	– – – – –	RF KNN DT Naive Bayes SVM	Dredze and Image Spam Hunter
[57]	0.94	0.95	0.99	0.97	CNN	SMS Spam and Twitter
[40]	0.98	–	–	–	SVM	Dredze
[28]	0.98	0.99	–	–	NN	Dredze
[58]	0.95	0.96	0.98	0.97	LSTM	SMS Spam and Twitter
[5]	0.99	–	–	–	CNN	Dredze
[18]	0.95	0.96	0.99	0.97	CNN and LSTM	SMS Spam and Twitter
[56]	0.92	–	–	–	SVM and Particle Swarm Optimization	Spam Archive
[59]	0.97 0.97 0.99	0.98 0.98 0.99	0.96 0.95 1	0.97 0.96 0.99	CNN	Image Spam Hunter, Dredze, Personal collection
[60]	0.79 0.96 0.99	– – –	– – –	– – –	SVM Multilayer Perceptrons CNN	Image Spam Hunter, Personal collection

ACKNOWLEDGMENT

The authors would like to acknowledge the Universiti Putra Malaysia for supporting this research.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] C. Pham, L. A. Nguyen, N. H. Tran, E.-N. Huh, and C. S. Hong, "Phishing-aware: A neuro-fuzzy approach for anti-phishing on fog networks," *IEEE Transactions on Network and Service Management*, vol. 15, no. 3, pp. 1076–1089, 2018.

[2] A.-P. W. Group, "Phish activity trends report," APWG, Tech. Rep., 3rd Quarter 2018.

[3] G. Mujtaba, L. Shuib, R. G. Raj, N. Majeed, and M. A. Al-Garadi, "Email classification research trends: Review and open issues," *IEEE Access*, vol. 5, pp. 9044–9064, 2017.

[4] A.-P. W. G. (APWG), "Phishing activity trends report 2nd quarter 2019," Tech. Rep., 2019.

[5] A. P. Singh, "Image spam classification using deep learning," Master's thesis, San Jose State University, 2018.

[6] S. Dhanaraj and V. Karthikeyani, "A study on e-mail image spam filtering techniques," in *2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*. IEEE, 2013, pp. 49–55.

[7] A. Bhowmick and S. M. Hazarika, "Machine learning for e-mail spam filtering: review, techniques and trends," *arXiv preprint arXiv:1606.01042 (2016)*, 2016.

[8] J. Constine, "Facebook climbs to 1.59 billion users and crushes q4 estimates with $5.8 b revenue," *TechCrunch, (2016)*, vol. 27, 2016.

[9] F. Naiemi, V. Ghods, and H. Khalesi, "An efficient character recognition method using enhanced hog for spam image detection," *Soft Computing*, pp. 1–16, 2019.

[10] S. Gao, C. Zhang, and W.-B. Chen, "Identifying image spam authorship with variable bin-width histogram-based projective clustering," in *2011 IEEE International Conference on Multimedia and Expo*. IEEE, 2011, pp. 1–6.

[11] A. Attar, R. M. Rad, and R. E. Atani, "A survey of image spamming and filtering techniques," *Artificial Intelligence Review*, vol. 40, no. 1, pp. 71–105, 2013.

[12] A. kapoor. (2019, May) deep-learning-vs-machine-learning-a-simple-explanation. [Online]. Available: https://mc.ai/deep-learning-vs-machine-learning-a-simple-explanation/

[13] T. Guo, J. Dong, H. Li, and Y. Gao, "Simple convolutional neural network on image classification," in *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)(*. IEEE, 2017, pp. 721–724.

[14] S. W. Liew, N. F. M. Sani, M. T. Abdullah, R. Yaakob, and M. Y. Sharum, "An effective security alert mechanism for real-time phishing tweet detection on twitter," *Computers & Security*, vol. 83, pp. 201–207, 2019.

[15] ——, "Improvement of classification features to increase phishing tweets detection accuracy." *Journal of Theoretical & Applied Information Technology*, vol. 96, no. 10, 2018.

[16] B. Biggio and F. Roli, "Wild patterns: Ten years after the rise of adversarial machine learning," *Pattern Recognition*, vol. 84, pp. 317–331, 2018.

[17] F. D. T. Aneri Chavda, Katerina Potika and M. Stamp, "Support vector machines for image spam analysis," in *Support Vector Machines for Image Spam Analysis*, vol. 1, no. 978-989-758-319-3, In Proceedings of the 15th International Joint Conference on e-Business and Telecommunications (ICETE 2018). Copyright © 2018 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved, 2018, pp. 431–441.

[18] G. Jain, M. Sharma, and B. Agarwal, "Spam detection in social media using convolutional and long short term memory neural network," *Annals of Mathematics and Artificial Intelligence*, vol. 85, no. 1, pp. 21–44, 2019.

[19] M. S. C. S. Dongre, "E-mail spam classification using long short-

term memory method," *International Journal of Scientific Research & Engineering Trends, (2019)*, vol. 5, no. ISSN (online) 2395-566X, 2019.

[20] G. Fumera, I. Pillai, and F. Roli, "Spam filtering based on the analysis of text information embedded into images," *Journal of Machine Learning Research*, vol. 7, no. Dec, pp. 2699–2720, 2006.

[21] A. S. Manek, D. Shamini, V. H. Bhat, P. D. Shenoy, M. C. Mohan, K. Venugopal, and L. Patnaik, "Rep-etd: A repetitive preprocessing technique for embedded text detection from images in spam emails," in *2014 IEEE International Advance Computing Conference (IACC)*. IEEE, 2014, pp. 568–573.

[22] A. Harisinghaney, A. Dixit, S. Gupta, and A. Arora, "Text and image based spam email classification using knn, naïve bayes and reverse dbscan algorithm," in *2014 International Conference on Reliability Optimization and Information Technology (ICROIT)*. IEEE, 2014, pp. 153–155.

[23] D. Yamakawa and N. Yoshiura, "Applying tesseract-ocr to detection of image spam mails," in *2012 14th Asia-Pacific Network Operations and Management Symposium (APNOMS)*. IEEE, 2012, pp. 1–4.

[24] P. Wan and M. Uehara, "Spam detection using sobel operators and ocr," in *2012 26th International Conference on Advanced Information Networking and Applications Workshops*. IEEE, 2012, pp. 1017–1022.

[25] C. Jou, "Spam e-mail classification based on the ifwb algorithm," in *Asian Conference on Intelligent Information and Database Systems*. Springer, 2013, pp. 314–324.

[26] H. Chen, Y. Zhan, and Y. Li, "The application of decision tree in chinese email classification," in *2010 International Conference on Machine Learning and Cybernetics*, vol. 1. IEEE, 2010, pp. 305–308.

[27] K. S. Adewole, T. Han, W. Wu, H. Song, and A. K. Sangaiah, "Twitter spam account detection based on clustering and classification methods," *The Journal of Supercomputing*, pp. 1–36, 2018.

[28] A. Amir, B. Srinivasan, and A. I. Khan, "Distributed classification for image spam detection," *Multimedia Tools and Applications*, vol. 77, no. 11, pp. 13 249–13 278, 2018.

[29] J. Chen, H. Zhao, J. Yang, J. Zhang, T. Li, and K. Wang, "An intelligent character recognition method to filter spam images on cloud," *Soft Computing*, vol. 21, no. 3, pp. 753–763, 2017.

[30] M. Dredze, R. Gevaryahu, and A. Elias-Bachrach, "Learning fast classifiers for image spam." in *CEAS*, 2007, pp. 2007–487.

[31] B. Biggio, G. Fumera, I. Pillai, and F. Roli, "Improving image spam filtering using image text features," in *Proc of the fifth conf on email and anti-spam*, 2008.

[32] Y. Gao, A. Choudhary, and G. Hua, "A comprehensive approach to image spam detection: from server to client solution," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 826–836, 2010.

[33] M. Das and V. Prasad, "Analysis of an image spam in email based on content analysis," in *Proc. Int. Conf. On Natural Language Processing And Cognitive Computing*, vol. 201, no. 4, 2014.

[34] C. Wang, F. Zhang, F. Li, and Q. Liu, "Image spam classification based on low-level image features," in *2010 International Conference on Communications, Circuits and Systems (ICCCAS)*. IEEE, 2010, pp. 290–293.

[35] A. Annadatha and M. Stamp, "Image spam analysis and detection," *Journal of Computer Virology and Hacking Techniques*, vol. 14, no. 1, pp. 39–52, 2018.

[36] Z. Wang, W. K. Josephson, Q. Lv, M. Charikar, and K. Li, "Filtering image spam with near-duplicate detection." in *CEAS*, 2007.

[37] B. Mehta, S. Nangia, M. Gupta, and W. Nejdl, "Detecting image spam using visual features and near duplicate detection," in *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008, pp. 497–506.

[38] P. He, X. Wen, and W. Zheng, "A simple method for filtering image spam," in *2009 Eighth IEEE/ACIS International Conference on Computer and Information Science*. IEEE, 2009, pp. 910–913.

[39] Z. Qu and Y. Zhang, "Filtering image spam using image semantics and near-duplicate detection," in *2009 Second International Conference on Intelligent Computation Technology and Automation*, vol. 1. IEEE, 2009, pp. 600–603.

[40] A. Chavda, "Image spam detection," Master's thesis, San Jose State University, 2017.

[41] Y. Gao, A. Choudhary, and G. Hua, "A nonnegative sparsity induced similarity measure with application to cluster analysis of spam images," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 5594–5597.

[42] R. Mitchell, J. Michalski, and T. Carbonell, *An artificial intelligence approach*. Springer, 2013.

[43] M. Balakumar and V. Vaidehi, "Ontology based classification and categorization of email," in *2008 International Conference on Signal Processing, Communications and Networking*. IEEE, 2008, pp. 199–202.

[44] Y. K. Zamil, S. A. Ali, and M. A. Naser, "Spam image email filtering using k-nn and svm." *International Journal of Electrical & Computer Engineering (2019), 2088-8708,*, vol. 9, no. 1, 2019.

[45] G. Xiang, J. Hong, C. P. Rose, and L. Cranor, "Cantina+: A feature-rich machine learning framework for detecting phishing web sites," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 2, p. 21, 2011.

[46] A. N. V. Sunil and A. Sardana, "A pagerank based detection technique for phishing web sites," in *2012 IEEE Symposium on Computers & Informatics (ISCI)*. IEEE, 2012, pp. 58–63.

[47] R. Verma and K. Dyer, "On the character of phishing urls: Accurate and robust statistical learning classifiers," in *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*. ACM, 2015, pp. 111–122.

[48] H. Atlam, R. Walters, and G. Wills, "Fog computing and the internet of things: a review," *Big Data and Cognitive Computing*, vol. 2, no. 2, p. 10, 2018.

[49] T. Chin, K. Xiong, and C. Hu, "Phishlimiter: A phishing detection and mitigation approach using software-defined networking," *IEEE Access*, vol. 6, pp. 42 516–42 531, 2018.

[50] L. A. T. Nguyen and H. K. Nguyen, "Phishing identification using a novel non-rule neuro-fuzzy model," *International Journal of Computer Science and Information Security*, vol. 14, no. 4, p. 8, 2016.

[51] W. Hadi, F. Aburub, and S. Alhawari, "A new fast associative classification algorithm for detecting phishing websites," *Applied Soft Computing*, vol. 48, pp. 729–734, 2016.

[52] N. Abdelhamid, "Multi-label rules for phishing classification," *Applied Computing and Informatics*, vol. 11, no. 1, pp. 29–46, 2015.

[53] A. Yasin and A. Abuhasan, "An intelligent classification model for phishing email detection," *arXiv preprint arXiv:1608.02196 (2016)*, 2016.

[54] Z. Qu and Y. Zhang, "A new near-duplicate detection system using object semantics for filtering image spam," in *2009 International Conference on Information Management, Innovation Management and Industrial Engineering*, vol. 3. IEEE, 2009, pp. 607–610.

[55] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management*, vol. 45, no. 4, pp. 427–437, 2009.

[56] T. Kumaresan, P. Subramanian, D. S. Alex, M. T. Hussan, and B. Stalin, "Email image spam detection using fast support vector machine and fast convergence particle swarm optimization," *International Journal of Recent Technology and Engineering (IJRTE) (2019)*, vol. 8, May 2019.

[57] G. Jain, M. Sharma, and B. Agarwal, "Spam detection on social media using semantic convolutional neural network," *International Journal of Knowledge Discovery in Bioinformatics (IJKDB)*, vol. 8, no. 1, pp. 12–26, 2018.

[58] ——, "Optimizing semantic lstm for spam detection," *International Journal of Information Technology*, vol. 11, no. 2, pp. 239–250, 2019.

[59] S. Sriram, R. Vinayakumar, V. Sowmya, M. Krichen, D. B. Noureddine, A. Shashank, and K. Soman, "Deep convolutional neural networks for image spam classification," *(2020)*, vol. hal-02510594, 2020.

[60] T. Sharmin, F. Di Troia, K. Potika, and M. Stamp, "Convolutional neural networks for image spam detection," *Information Security Journal: A Global Perspective*, vol. 29, no. 3, pp. 103–117, 2020.

[61] K. Gandhimathi and M. Vijaya, "Identifying similar web pages based on automated and user preference value using scoring methods," *International Journal of Data Mining & Knowledge Management Process*, vol. 3, no. 6, p. 41, 2013.