Deep Convolutional Neural Networks for Image Spam Classification

S. Sriram, R. Vinayakumar, V. Sowmya, Moez Krichen, Dhouha Ben Noureddine, A. Shashank, K.P. Soman

Deep Convolutional Neural Networks for Image Spam Classification

Sriram S*, Vinayakumar R†, Sowmya V*, Moez Krichen‡§, Dhouha Ben Noureddine¶‖, Shashank A*, Soman KP*
*Amrita School of Engineering, Amrita Vishwa Vidyapeetham Coimbatore, Tamil Nadu, India.
sri27395ram@gmail.com , v_sowmya@cb.amrita.edu, shashankanivilla@gmail.com
†Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Centre, Cincinnati, OH, United States
Vinayakumar.Ravi@cchmc.org, vinayakumarr77@gmail.com
‡FCSIT, Al-Baha University, Al Baha, KSA.
§ReDCAD, University of Sfax, Sfax, Tunisia.
moez.krichen@redcad.org
¶LISI, INSAT, University of Carthage, Tunis, Tunisia.
‖FST, University of El Manar, Tunis, Tunisia.
dhouha.bennoureddine@fst.utm.tn

***Abstract*—With the tremendous growth of the internet, cyberspace is facing several threats from the attackers. Threats like spam emails account for 55% of total emails according to the Symantec monthly threat report. Over time, the attackers moved on to image spam to evade the text-based spam filters. To deal with this, the researchers have several machine learning and deep learning approaches that use various features like metadata, color, shape, texture features. But the Deep Convolutional Neural Network (DCNN) and transfer learning-based pre-trained CNN models are not explored much for Image spam classification. Therefore, in this work, 2 DCNN models along with few pre-trained ImageNet architectures like VGG19, Xception are trained on 3 different datasets. The effect of employing a Cost-sensitive learning approach to handle data imbalance is also studied. Some of the proposed models in this work achieves an accuracy up to 99% with zero false positive rate in best case.**

***Index Terms*—Image spam, Deep learning, Convolutional neural network, Transfer learning, Cost-sensitive learning, Cyber security**

I. Introduction

The Internet has become widely popular nowadays and many people are dependent on it for their social interactions, financial transactions, and communication. But the Internet is not completely safe from cyber criminals. The attackers always try to exploit internet users by employing techniques like phishing, spamming, impersonating, etc. According to the report released by Symantec [1], email spam accounted for approximately 60% of emails in mining, finance, insurance, and real estate industries and techniques like spam filters are essential for safe and secure email communication. Internet of Things (IoT) technologies are growing very rapidly and they are used in applications like Smart cities [2]–[6]. One of the disadvantages of IoT devices is that they are low powered devices with limited resources. They are not built with security in mind. Therefore, many hackers are exploiting the IoT Bot network (the network of compromised IoT devices) for conducting various cyber attacks like DDOS [7]. In [8], the attackers have exploited an IoT Botnet for sending Email Spam. They have used 4500 bots and have sent 1.8 million messages per day.

Spam from emails was initially in the form of text. Several Machine learning (ML) based spam detectors are developed for classification. ML models like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naïve Bayes (NB), etc. are used for filtering email spam with 95% accuracy in [9]. Over time, the attackers came up with new ways like Image spam which is shown in the Fig. 1 to trick the existing spam filters. Image spam attack contains images with text embedded to it and they are used by attackers to evade text-based spam filters. These images trick the user to click on it which might cause redirection to unsafe websites or causes malware infection.

Fig. 1. Sample spam images.

Several techniques are developed over the years to detect image spam. Optical Character Recognition (OCR) techniques are used for extracting textual content which is used for the detection of image spam [10]. Alternative to the content

978-1-7281-2747-7/20/$31 ©2020 IEEE

based approach, several ML and DL based techniques are proposed for effective classification of image spam. Many researchers are proposing DL based methods for various cyber security applications [11] like malware detection [12], [13], malicious domain detection [14], intrusion detection [15], etc. DL techniques can also be used for detecting image spam which may leverage the performance of the existing methods. Therefore, in this work, several Convolutional Neural Network (CNN) models are used for image spam classification.

The main contributions of this work are the following; Firstly, two Deep Convolutional Neural Networks (DCNN) models are designed and their effectiveness is also studied for image spam detection using three different datasets. Secondly, the effects of the transfer learning technique is studied by using pre-trained CNN models like VGG19, Xception, etc. Further, cost-sensitive learning is employed to deal with data imbalance. Lastly, the effectiveness of various ML classifiers that are trained on features extracted by CNN models are studied. The rest of the sections of this work are organized as follows; Section II presents the related works. Section III contains the dataset description. Section IV presents the proposed methodologies. Section V includes experimental results and discussion. Finally, the conclusion is placed in Section VI.

II. Related work

Image spam detection can be based on textual content. In [10], OCR is used to obtain the textual content from the spam image which is then analyzed by various text filters to detect spam. Alternative to OCR based approach, ML-based image spam categorization model is proposed in [16]. The textual content is extracted and then feature extraction is performed. The features are further fed in SVM to classify the images. Similar to [16], ML-based approaches are utilized in [17]. In this work, features based on metadata and file properties are extracted from the spam images which are then fed into ML classifiers such as Maximum Entropy (ME), Decision Tree (DT) and Naive Bayes (NB) to detect the spam. A Probabilistic Boosting Tree (PBT) classifier based spam detection model is proposed in [18]. The global image features including gradient and color histograms are extracted and fed into the classifier. The proposed classifier performed better than the SVM based model. [19] proposes a comprehensive solution for image spam detection in both server and client side using cluster analysis and classifiers like SVM. In [20], the performance of PCA and SVM based image spam classifiers are studied on two datasets in which the Linear and Radial Basis Function SVM models achieved better results. The first dataset is developed by [18] and the second improved dataset is developed by [21] which cannot be detected properly by PCA and SVM based approach.

ML-based approaches require manual feature engineering which can be averted by using Deep Learning (DL) techniques. In [22], the performance of several CNN based models is studied for image spam recognition. CNN models like VGG, Spatial Pyramid Pooling (SPP) network, Weighted Spatial Pyramid (WSP) network, etc. WSP network performed better than other models in terms of accuracy. Similar to [22], in [23], SPP Net is used for image spam detection. In [24], CNN based models are used for Instagram image spam detection. Models such as three and five-level CNNs, VGG-16 and AlexNet are trained on images obtained from Instagram using a web crawler. VGG-16 performed well with 84% accuracy. In [25], a CNN based image spam classifier which is trained on the dataset from [18] is proposed. It achieves 92% accuracy. In [26], CNN based image spam classification model is proposed as part of situational awareness framework for analysing email and url data.

This work studies the effectiveness of two DCNN models and few pre-trained ImageNet architectures like VGG19, Xception, etc. This work uses datasets that are used in [17], [18] and [20].

III. Dataset Description

The following are the three benchmark datasets that are used in this work.

A. Image Spam Hunter Dataset (ISH)

This dataset [18] contains both spam and ham images in JPEG format which are collected from original emails. It is a publicly available dataset that can be found in the Northwestern University website [27]. There are 810 ham and 929 spam images in total. The number of unique spam and ham images is 879 and 810 respectively.

B. Improved Dataset

This dataset is developed as a challenge dataset in work [21] in order to test the performance of image spam models with more advanced spam images. It contains a total of 1,029 spam images that are generated by embedding spam text in ham images as shown in Fig. 2 to trick the existing models. The number of unique images in this dataset is 975. It is available at [28].

C. Dredze ImageSpam Dataset

This dataset [17] contains 3 sets of images. Personal Ham (PHam) has 2,021 images in which there are 1,517 unique images. Personal Spam (PSpam) has 3,298 images in which there are 1,274 unique images. Finally, the Spam Archive (SpamArch) has 16,028 files of various formats like JPEG, PNG, GIF, etc in which there are 3,039 unique images. It is available at [29].

IV. Methodology

A. Pre-processing

The three datasets that are utilized in this work has a lot of duplicate images and corrupt files. Firstly the corrupt files are omitted and then In order to avoid the duplicate files, each image converted into a hash and stored. In this way, when a duplicate image is read, its hash will be matched with existing ones. If the match is found, then the image will be omitted. Finally, all unique images are normalized and resized into the required sizes.

Fig. 2. Sample improved spam images.

B. Deep Convolutional Neural Network (DCNN) models

In this work, two DCNN models are designed. The first CNN1 model has 3 convolution layers of filter size 32, 64 and 128. Each convolutional layer is immediately followed by ReLU activation and max pooling layer of pooling size 2. After the convolution layers, dropout regularization is used and the output is flattened and passed to a Dense layer which contains 128 neurons. This layer is followed by ReLU activation and dropout regularization. Finally, a dense layer of a single neuron is used with a sigmoid activation function. Hybrid models are also utilized in this work where the features are extracted from the last hidden layer of the CNN1 model and passed on to many ML classifiers as shown in Fig. 3. The ML classifiers used in this work are Linear Support Vector Machine (LSVM), Random Forest (RF), AdaBoost (AB), and K-Nearest Neighbor (KNN). The pseudo-code for image spam classification is given below.

Algorithm 1: Image Spam Classification

Input: A set of images extracted from emails $im_1, im_2, .., im_n$

Output: Labels $y_1, y_2, .., y_n$ (0: Legitimate or 1: Spam)

Pre-processing: Images are resized into required size

for *every extracted image* **do**
- Pass the extracted image into the DL model in order to extract optimal feature vector v_i
- Compute $d_i = DenseLayer(v_i)$
- Calculate $y_i = Sigmoid(d_i)$

The Second CNN2 model also has 3 convolution layers of filter size 128, 128 and 256. Each convolutional layer is immediately followed by the ReLU activation and max pooling layer of pooling size 4,3, and 2. After the convolution layers, dropout regularization is used and the output is flattened and passed to a Dense layer which contains 128 neurons. This layer is followed by ReLU activation and dropout regularization. Finally, a dense layer of a single neuron is used with a sigmoid activation function. The structure of the CNN2 model is represented with layer details in Table I.

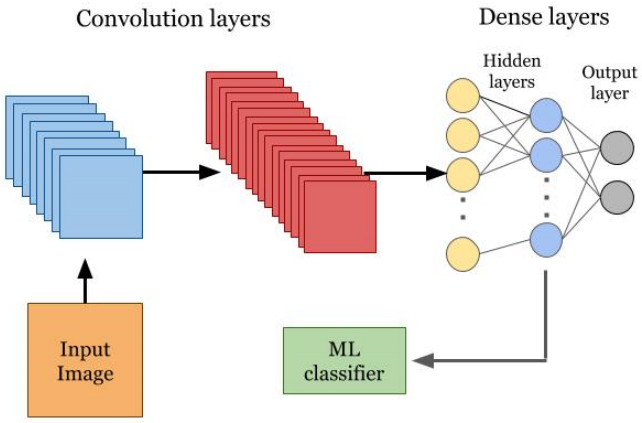

Fig. 3. Overview of Hybrid model.

TABLE I
STRUCTURE OF CNN2 MODEL

Layer Type	Output shape	Parameter count
Conv2D	(-, 156, 156, 128)	13,952
MaxPooling2D	(-, 39, 39, 128)	-
Conv2D	(-, 39, 39, 128)	262,272
MaxPooling2D	(-, 13, 13, 128)	-
Conv2D	(-, 13, 13, 256)	295,168
MaxPooling2D	(-, 6, 6, 256)	-
Dropout	(-, 6, 6, 256)	-
Flatten	(-, 9216)	-
Dense	(-, 128)	1,179,776
Dropout	(-, 128)	-
Dense	(-, 1)	129
Total number of parameter: 1,751,297		

C. Cost-sensitive Learning and Transfer Learning Models

The datasets used in this work are imbalanced dataset and they are a bit skewed towards spam class. Therefore, the cost-sensitive learning approach is used. In this approach, balanced class weights are calculated and passed to the model while fitting process so that the model will penalize the prediction mistakes of minority class proportionally based on how under-represented it is. This approach is employed in this work to the previously mentioned models and they are referred to as CS-CNN1 and CS-CNN2 in this work.

Transfer learning is also employed using the pre-trained ImageNet model such as VGG19, DenseNet201, ResNet152V2, and Xception. The default last dense layer is omitted and

all layers are frozen to facilitate transfer learning. Further, 3 Dense layers of neuron 1024, 512 and 1 are added at the end. Table II indicates the parameter count of all the models mentioned previously. It can be observed that ResNet152V2 and CNN1 has the most number of parameters.

TABLE II
TRAINABLE AND NON-TRAINABLE PARAMETERS OF THE PROPOSED MODELS.

Model	Number of parameters
CNN1	Trainable: 25,013,569
CNN2	Trainable: 1,751,297
VGG19	Trainable: 1,050,625 Non-trainable: 20,024,384
DenseNet201	Trainable: 2,492,417 Non-trainable: 18,321,984
Xception	Trainable: 2,623,489 Non-trainable: 20,861,480
ResNet152V2	Trainable: 2,623,489 Non-trainable: 58,331,648

D. Statistical Metrics

In this work, the standard metrics such as accuracy, precision, recall, f1-score, False Positive Rate (FPR) and False Negative Rate (FNR) are utilized. These metrics can be computed using the terms such as True Positive (TP), False Negative (FN), False Positive (FP), True Negative (TN) that are found in the confusion matrix. TP indicates the number of spam images that are accurately predicted as spam. FN indicates the number of spam images that are wrongly predicted as normal. FP indicates the number of normal images that are wrongly predicted as spam. TN indicates the number of normal images that are accurately predicted as spam.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

TABLE III
PERFORMANCE OF EXISTING STATE-OF-THE-ART IMAGE SPAM MODELS

Reference	Model	Dataset used	Accuracy
[17]	ME	PHam+PSpam	0.98
		PHam+SpamArch	0.89
		PHam+PSpam+SpamArch	0.91
	DT	PHam+PSpam	0.97
		PHam+SpamArch	0.85
		PHam+PSpam+SpamArch	0.87
[20]	LSVM	ISH	0.97
		ISH + Improved dataset	0.7
[30]	NN	ISH	0.99
		ISH + Improved dataset	0.98
		PHam+PSpam	0.98
		PHam+SpamArch	0.96
[25]	CNN	ISH	0.92
[26]	CNN	ISH	0.99

In this work, the proposed models are implemented using Keras and Scikit-learn python library and the source code is publicly available in GitHub. The binary cross-entropy loss function and Adam optimizer used in this work. Dropout regularization is employed in order to avoid over-fitting. The dataset used in this work is divided into 70:30 training and testing sets. From the literature review, the state-of-the-art image spam models are identified and its performance is shown in Table III for the purpose of comparison. For training CNN1 and CNN2 models, the images are resized into 156x156 resolution which is decided after training and testing the model in several input sizes.

The CNN1 and CS-CNN1 models are trained and tested on the image spam hunter and improved dataset for 100 epochs. Hybrid models are also used which extracts the features from the last hidden dense layer of the CNN1 and CS-CNN1 models in order to enhance the performance. The performance of these models is presented in Table IV. It can be observed from Table IV and III that, for the ISH dataset, all of the proposed models performed better than existing LSVM [20] and CNN [25] models. The performance of CS-CNN1-KNN and CS-CNN1-LSVM models is slightly lower when compared to the performance of the existing NN [30] and CNN [26] models. But it is essential to note that unlike [17], the works [20], [25], [30], and [26] did not remove the duplicate images which might have inflated the performance of their models.

In image spam classification, FNR is an important metric as it represents the fraction of spam images that are incorrectly classified as normal. therefore, in terms of FNR, the CS-CNN1-KNN model performed well compared to the rest of the proposed models in this work. For the improved dataset, the proposed models have shown superior performance when compared to existing LSVM [20] and NN [30] models. It could be observed from Table IV that CNN1-RF is better than the rest of the proposed models in this work as its FNR is zero and its accuracy is 0.998.

The CNN2 and CS-CNN2 models are trained on three different combinations of Dredze ImageSpam dataset for 100 epochs as shown in the Table V. For the Dredze PSpam and PHam data, both CNN2 and CS-CNN2 models have achieved an accuracy of 0.97 which is very similar to the performance of DT [17]. But the other ME [17] and NN [30] models have achieved an accuracy of 0.98. For the other two combinations (SpamArch+PHam and PSpam+PHam+SpamArch), the performance of the proposed models are lower when compared to the performance of existing models. It might be due to a couple of reasons. Firstly, in the works [17] and [30], the models are trained on various features like metadata, color, texture, shape and noise features that are extracted during the pre-processing stage whereas the proposed models extract optimal features automatically during the training stage. Secondly, the work [30] did not omit the duplicate images.

The pre-trained architectures that are shown in Table VI are trained on the entire Dredze dataset (PSpam+PHam+SpamArch) for 100 epochs. The images are resized into the resolution 224 x 224. It can be observed from the table that the performance of VGG19 is better than the DT classifier and very close to the ME classifier.

TABLE IV
PERFORMANCE OF CNN1, CS-CNN1 AND HYBRID MODELS

Model	Accuracy	Precision	Recall	F1-score	TP	FN	FP	TN	FNR	FPR
Image spam hunter dataset										
CNN1	0.971	0.981	0.963	0.972	260	10	5	238	0.037	0.021
CNN1-AB	0.975	0.978	0.974	0.976	263	7	6	237	0.026	0.025
CS-CNN1	0.975	0.967	0.985	0.976	266	4	9	234	0.015	0.037
CS-CNN1-RF	0.979	0.974	0.985	0.979	266	4	7	236	0.015	0.029
CS-CNN1-AB	0.979	0.978	0.981	0.979	265	5	6	237	0.019	0.025
CS-CNN1-KNN	0.981	0.978	0.985	0.981	266	4	6	237	0.015	0.025
CS-CNN1-LSVM	0.981	0.981	0.981	0.981	265	5	5	238	0.019	0.021
Image spam hunter Ham + Improved spam dataset										
CNN1	0.996	0.993	1	0.996	293	0	2	241	0	0.008
CNN1-RF	0.998	0.997	1	0.998	293	0	1	242	0	0.004
CS-CNN1	0.998	1	0.997	0.998	292	1	0	243	0.003	0

TABLE V
PERFORMANCE OF CNN2 AND CS-CNN2 MODELS

Model	Accuracy	Precision	Recall	F1-score	TP	FN	FP	TN	FNR	FPR
Dredze PSpam and PHam										
CNN2	0.973	0.981	0.958	0.969	367	16	7	448	0.042	0.015
CS-CNN2	0.974	0.981	0.961	0.971	368	15	7	448	0.039	0.015
Dredze SpamArch and PHam										
CNN2	0.825	0.873	0.864	0.868	788	124	115	339	0.136	0.253
CS-CNN2	0.834	0.874	0.877	0.875	800	112	115	339	0.123	0.253
Dredze PSpam, PHam and SpamArch										
CNN2	0.864	0.911	0.904	0.907	1170	124	114	340	0.096	0.251
CS-CNN2	0.863	0.896	0.922	0.909	1193	101	139	315	0.078	0.306

TABLE VI
PERFORMANCE OF PRE-TRAINED CNN ARCHITECTURES

Model	Accuracy	Precision	Recall	F1-score	TP	FN	FP	TN	FNR	FPR
DenseNet201	0.785	0.796	0.955	0.868	1235	58	317	138	0.045	0.697
Xception	0.811	0.836	0.927	0.879	1198	95	235	220	0.073	0.516
ResNet152V2	0.826	0.836	0.952	0.89	1231	62	242	213	0.048	0.532
VGG19	0.904	0.941	0.93	0.935	1202	91	76	379	0.07	0.167

The proposed CNN2 model performed better than all the pre-trained models except VGG19 in terms of acuracy. The VGG19 model obtained better results than CNN2 model even though it has less number of trainable parameter than CNN2 model. It may be because of the sharing of pre-trained weights as part of transfer learning. The rest of the pre-trained models performed very poorly possibly due to overfitting.

VI. CONCLUSION

In this work, the effectiveness of two Deep Convolutional Neural Networks and hybrid models are studied for image spam classification using 3 different datasets. The effects of employing cost-sensitive learning are studied by assigning balanced class weights and transfer learning is also studied by using several pre-trained CNN architectures like VGG19, Xception, etc. Some of the proposed models performed better than existing works and some of them did not. It can be inferred that in order to build a better image spam classifier, additional information like metadata should also be incorporated into the model training. In future works, the effects of adversarial samples, which are capable of tricking the model to make an incorrect prediction, can be studied.

ACKNOWLEDGMENT

This work was in part supported by Paramount Computer Systems and Lakhshya Cyber Security Labs. We are grateful to NVIDIA India, for the GPU hardware support to the research grant. We are also grateful to the center of Computational Engineering and Networking, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore for encouraging the research.

REFERENCES

[1] "Symantec monthly threat report," accessed: 08 Nov 2019. [Online]. Available: https://www.symantec.com/security-center/publications/monthlythreatreport#Spam

[2] M. Krichen, M. Lahami, O. Cheikhrouhou, R. Alroobaea, and A. J. Maâlej, "Security testing of internet of things for smart city applications: A formal approach," in *Smart Infrastructure and Applications*. Springer, Cham, 2020, pp. 629–653.

[3] M. Krichen and M. Lahami, "Towards a runtime testing framework for dynamically adaptable internet of things networks in smart cities," in *Smart Infrastructure and Applications*. Springer, Cham, 2020, pp. 589–607.

[4] M. Krichen, "Improving formal verification and testing techniques for internet of things and smart cities," *Mobile Networks and Applications*, pp. 1–12, 2019.

[5] M. Krichen and R. Alroobaea, "A new model-based framework for testing security of iot systems in smart cities using attack trees and price timed automata," in *14th International Conference on Evaluation of Novel Approaches to Software Engineering - ENASE 2019*, 2019.

[6] M. Krichen, O. Cheikhrouhou, M. Lahami, R. Alroobaea, and A. J. Maâlej, "Towards a model-based testing framework for the security of internet of things for smart city applications," in *International Conference on Smart Cities, Infrastructure, Technologies and Applications*. Springer, Cham, 2017, pp. 360–365.

[7] S. Akarsh, S. Sriram, P. Poornachandran, V. K. Menon, and K. Soman, "Deep learning framework for domain generation algorithms prediction using long short-term memory," in *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*. IEEE, 2019, pp. 666–671.

[8] C. Cimpanu, "Iot botnet retooled to send email spam," Sep 2017, accessed: 08 Nov 2019. [Online]. Available: https://www.bleepingcomputer.com/news/security/iot-botnet-retooled-to-send-email-spam/

[9] C.-C. Lai and M.-C. Tsai, "An empirical performance comparison of machine learning methods for spam e-mail categorization," in *Fourth International Conference on Hybrid Intelligent Systems (HIS'04)*. IEEE, 2004, pp. 44–48.

[10] "Apache spamassassin - open source anti-spam platform," accessed: 08 Nov 2019. [Online]. Available: http://spamassassin.apache.org/

[11] R. Vinayakumar, K. Soman, P. Poornachandran, and S. Akarsh, "Application of deep learning architectures for cyber security," in *Cybersecurity and Secure Information Systems*. Springer, 2019, pp. 125–160.

[12] S. Venkatraman, M. Alazab, and R. Vinayakumar, "A hybrid deep learning image-based analysis for effective malware detection," *Journal of Information Security and Applications*, vol. 47, pp. 377–389, 2019.

[13] R. Vinayakumar, M. Alazab, K. Soman, P. Poornachandran, and S. Venkatraman, "Robust intelligent malware detection using deep learning," *IEEE Access*, vol. 7, pp. 46 717–46 738, 2019.

[14] V. S. Mohan, R. Vinayakumar, K. Soman, and P. Poornachandran, "Spoof net: syntactic patterns for identification of ominous online factors," in *2018 IEEE Security and Privacy Workshops (SPW)*. IEEE, 2018, pp. 258–263.

[15] R. Vinayakumar, M. Alazab, K. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, pp. 41 525–41 550, 2019.

[16] H. B. Aradhye, G. K. Myers, and J. A. Herson, "Image analysis for efficient categorization of image-based spam e-mail," in *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*. IEEE, 2005, pp. 914–918.

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

[18] Y. Gao, M. Yang, X. Zhao, B. Pardo, Y. Wu, T. N. Pappas, and A. Choudhary, "Image spam hunter," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 1765–1768.

[19] 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.

[20] 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.

[21] A. Chavda, K. Potika, F. Di Troia, and M. Stamp, "Support vector machines for image spam analysis." in *ICETE (1)*, 2018, pp. 597–607.

[22] F. Aiwan and Y. Zhaofeng, "Image spam filtering using convolutional neural networks," *Personal and Ubiquitous Computing*, vol. 22, no. 5-6, pp. 1029–1037, 2018.

[23] Q. Meng, X. Zhu, and L. Gu, "Image spam filtering using weighted spatial pyramid networks," in *Recent Developments in Intelligent Computing, Communication and Devices*. Springer, 2019, pp. 43–52.

[24] C. Fatichah, W. F. Lazuardi, D. A. Navastara, N. Suciati, and A. Munif, "Image spam detection on instagram using convolutional neural network," in *Intelligent and Interactive Computing*. Springer, 2019, pp. 295–303.

[25] A. D. Kumar, S. KP *et al.*, "Deepimagespam: Deep learning based image spam detection," *arXiv preprint arXiv:1810.03977*, 2018.

[26] R. Vinayakumar, K. Soman, P. Poornachandran, S. Akarsh, and M. Elhoseny, "Deep learning framework for cyber threat situational awareness based on email and url data analysis," in *Cybersecurity and Secure Information Systems*. Springer, 2019, pp. 87–124.

[27] Y. Gao, "Image spam hunter dataset," accessed: 12 Sep 2019. [Online]. Available: http://www.cs.northwestern.edu/ yga751/ML/ISH.htm

[28] A. Annadatha, "Improved spam image dataset," accessed: 12 Sep 2019. [Online]. Available: https://www.dropbox.com/s/7zh7r9dopuh554e/New_Spam.zip?dl=0

[29] M. Dredze, "Image spam dataset 2007," accessed: 12 Sep 2019. [Online]. Available: http://www.cs.jhu.edu/ mdredze/datasets/image_spam/

[30] A. P. Singh, "Image spam classification using deep learning," *Master's Projects. 641. SJSU scholarworks*, 2018.