Image Spam detection in E-mails using Grasshoppers optimization technique

[1] Deepika Mallampati
Research Scholar, Dept. of CSE, Osmania University, Hyderabad, Telangana, India.
mokshhyd@gmail.com

[2] Dr. Nagaratna P Hegde
Professor, Dept of CSE, Vasavi College of Engineering, Hyderabad, Telangana, India
nagaratnaph@gmail.com

2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3) | 979-8-3503-3800-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/IC2E357697.2023.10262507

***Abstract*— one of the most frequent types of digital communication is email. Spam is unsolicited mass mail, whereas image spam includes spam text incorporated in images. This type of spam threatens email communications because spammers use it to avoid text spam filters. This paper provides a support vector machine (SVM) based classification algorithm and a biology optimization strategy to categorize emails as spam or ham. In this study, we examine image spam detection strategies using different combinations of image processing and machine learning algorithms. The grasshopper optimization algorithm (GOA) is an approach that statistically models and simulates the behavior of natural locust swarms. We investigate and analyze the performance of Grasshopper optimizations using several evaluation measures, such as accuracy, precision, recall, f1-score, and convergence rate to the global optimum solution.**

Keywords Machine learning, Spam, Grasshopper optimization, Support Vector Machine

I. INTRODUCTION

Image spam is a confusing approach in which the content of a message is stored in an image file (e.g., gif, jpeg, jpg, etc.) and shown in an email [1]. This hinders text spam filters from detecting and blocking spam communications [2]. This type of spam typically comprises meaningless computer-generated text that irritates the reader. Spam text embedded inside an image can be an effective method to evade text-based detection [3]. According to a recent report from Symantec [4], spam now accounts for 90.4% of all email. Nevertheless, a novel method is to use an image in the first frame that does not include plain text or to deform the shape of the letters in the image (like in CAPTCHA) to prevent it from being identified by OCR systems. Spam images were formerly thought to be plain text turned into images. This graphic spam has been detected using optical character recognition (OCR). An optical recognition system extracts text from images and applies text-based detection algorithms. Spammers have formed strategies to hide spam images in answer to OCR-based detection. OCR cannot read the text contained in images due to obfuscation [5]. Instead of using optical character identification techniques to detect spam in images, try a more straightforward approach using the features of the images directly. This technique for image processing is examined in this work in conjunction with algorithms for machine learning. Researchers finally combined numerous algorithms to create techniques such as the Machine Learning (DL) algorithm [6-8]. This technique improves a neural network that is now used with other algorithms to combat the highly complicated image-based spam trend.

This algorithm is being utilized in concert with other algorithms to tackle the intricate image-based spam trend. Learning algorithms are a category of Deep - learning (DL) algorithms replicating human brain neural functions. The DL technique was created to compensate for the shortcomings of previous ML algorithms. The most notable limitation of other ML techniques is the well-known "Curse of Dimensionality," [9] in which the algorithm gets less important as the number of characteristics to analyse grows. Although DL is a superior alternative for analyzing information with highly complex features, it generally necessitates learning with a massive quantity of data to achieve accuracy levels equal to those achieved by its predecessor's ML techniques. Several solutions have been proposed to address this issue, including methods for correctly classifying spam from emails. Spammers are developing new tactics to keep one step ahead of spam detection in the war between viruses and antivirus. Image spam is the most modern tactic, in which the main content of a spamming message is sent as an inline image [10]. Most spam images have no text or merely fictitious content. Plain text spam detection is incapable of detecting and blocking it. The bulk of spam that gets through personal antispam filters is visual spam. Spam graphics are more intriguing and defined than simple text at this time.

II. RELATED WORKS AND BACKGROUND

In [11], the authors suggested a text detection algorithm that overcomes the issues. The strategy relied on non-machine learning approaches and basic computations. This study's contribution is divided into two parts: a) a novel edge operator that can be used mainly to detect text edges; and b) a text detection approach for image spam detection that can identify obfuscated text. Accumulated Text Extraction (ATE) is a suggested approach for identifying vertical and horizontal lines and crossing them, after which criteria are used to define the text area and eliminate the non-text area. ATE produces promising findings that may be effectively employed in image spam filtering. Apart from its resistance to obfuscating strategies in image spam, ATE performs well for scene text recognition.

In [12], the authors addressed the issue of standard spam information filtering algorithms seeing a substantial performance drop or even failure while filtering spam image information. This study provides a way to increase data samples based on the clustering method, dramatically enhancing the number of high-quality training instances and fitting the demands of model training. Finally, they developed the convolutional neural network designed to identify SPAM in real-time using the expanded data samples. The experimental findings demonstrate that applying the data augmentation strategy increases the model's accuracy by more than 14%. Compared to other data pre-processing strategies, it can improve model accuracy by 6%. The efficiency of the spam detection model is 7-11% greater than that of traditional approaches when combined with neural network convolution and the suggested Data Augmentation method.

In [13], the authors introduced a particle swarm optimization (PSO) approach that leverages based on chaotic maps to minimize feature dimensionality and increase spam

email classification accuracy. For each particle, the function is shown in binary form. The sigmoid function is used to transform the process to binary. The fit function is used to pick features, which depends on the support vector machine's output (SVM) accuracy. For evaluating the behavior of a Chaotic Binary PSO (CBPSO) using the Spam-Base dataset, the classifier's efficiency and the dimension of the chosen feature extraction as a classification input are considered.

In [14], the authors evaluated and categorized image-based spam emails, and they used classification techniques in conjunction with Feature extraction methods. The resultant process improves on previous Algorithms that suffer from noise created by spam text hidden in photos. Regarding categorization and efficiency, they demonstrate that the hybridization system beats OCR-based systems.

In [15], the authors suggested 123DNet, a basic convolutional neural network (CNN) model trained with 28,929 image s collected from two publicly available datasets and a Personally created dataset. They indicated that spam filtering systems incorporate a system that performs well with new spam image s that didn't previously exist. The model was optimized to contain one input layer, a Convolution layer as a hidden neuron, and three neural network layers. The model was evaluated using 4,339 images from the three dataset samples, followed by a second batch of 1,200 images to assess performance on new images. Using the confusion matrix, a Classification Performance study was performed. Performance measures such as Quality, Accuracy, True Negative Correctness, Sensitivity, Specificity, and F1-score were computed to assess the model's performance. For a common dataset's test sample, the Models obtained a Score of 97% and an F2-Score of 88%.

III. Materials and Methods

Figure 1 depicts the image spam detection model's training and testing procedures. Then, we separate the ham and spam images into two groups: training and testing. Trains and testing sets are mutually exclusive. In other words, there is no overlapping between them. The dataset's features are then extracted one by one. Next, using the scaled training data, train an SVM classifier. The SVM classifier is used to search the test set. A feature selection is also incorporated to reduce dimension according to the feature technique in the training process.

Feature Extraction. Extraction of features is an algorithm for image processing used to minimize a images dimensionality. This helps reduce the resources necessary to explain massive data collection. Estimates containing many variables generally need a lot of processing and memory resources. A broad word for how groupings of variables are structured to overcome these challenges by precisely characterizing the data. We worked with two types of features here: text and texture. Five crucial metadata elements enable fast image retrieval at an inexpensive computational cost. Size of the image, width, height, pixel size, and image file format are examples of these characteristics.

Five crucial metadata elements enable fast image retrieval at an inexpensive computational cost. Size of the image, width, height, pixel size, and image file format are examples of these characteristics. Based on these essential characteristics, create a 10-dimensional feature vector. The image type procedures (f7, f8, & f9) were Boolean procedures that return a value of 1 if the files are of the type and a value of 0 otherwise—data with sufficient precision.

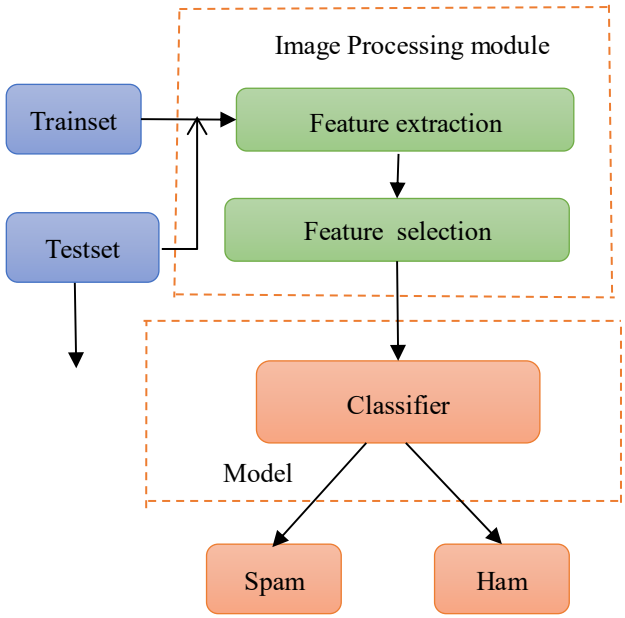

Fig. 1. Image spam detection model

A histogram is a visual representation of an image based on statistics. The grey histogram for grayscale images displays statistics on how various colors of grey appear. The equation includes the description of a global grey histogram

$$H(k) = \frac{r_k}{N} \qquad (1)$$

A texture is a response to an image in the zone of gray-level pixel spatial distribution, and the unique qualities of the pixel structure inside that space may be directly tied to the image. The time of a grey histogram is the easiest method to explain the demography of a texture. The grey value of the grey image is k, and the pixel count with the grey value k is r_k. The total amount of pixels in an image is denoted by N.

$$\mu(r) = \sum_{i=1}^{L} (r_i - m)\, nH(r_i) \qquad (2)$$

The texture features of the paper are μ2, μ3, and μ4. L *is the* histogram dimension, *m* is the mean histogram, and *n* is closely connected to the form of *H(r). However*, these moments have nothing to do with the texturing of the exact position in space. Variance μ2 with in histogram of *n*-order features is a measurement of grey scale variance, a representation of the curve compared to the solution's mean. It denotes the extent of a generally smooth histogram, which shows the degree of dispersion of greyscale images; μ3 indicates skewness, representing the curve compared to the mean symmetric. It represents the amount of histogram skewed, and it's the case if asymmetry histogram points exist; μ4 may be termed kurtosis, which indicates the histogram's relative flatness. It is a histogram distribution of points

gathered around or close to the median at both extremities of the scenario, describing gray-scale image texture differences.

Feature Selection. The cost of translating input data to a higher-dimensional space rises in multiple input spaces. The objective of choosing features is to limit the number of dimensions. The cost of expanding the input vector and performing kernel transformations is substantial. To address optimization problems, GOA mathematically analyzes and duplicates the behaviors of locust swarms in nature. The mathematical equation (3) below replicates the behaviors of a locust swarm.

$$X_i = H_i + W_i + G_i \qquad (3)$$

where Xi is the ith grasshopper's location, G_i is gravity, W_i represents wind advection, and H_i is social interaction. The equation represents it:

$$G_i = -g\hat{e}_g \qquad (4)$$

$$W_i = w\hat{e}_w \qquad (5)$$

$$H_i = \sum_{j=1, j\neq i}^{N} h\left(d_{i,j}\right)\hat{d}_{ij} \qquad (6)$$

where g is the gravitational constant and $\hat{e_g}$ is the unit vector heading to the gravity in the equation. (4) w in the formula (5) is the continuous drift, and $\hat{e_w}$ is the wind direction unit vector. n. In equation (6), d_{ij} is the Distance measure here between *the ith and jth* locusts given by equation (1), and $\hat{d_{ij}}$ is a unit vector represented by equation (5). N denotes the number of grasshoppers.

$$\hat{d}_{ij} = \frac{x_j - x_i}{d_{i,j}} \qquad (7)$$

Eq. (6) gives the h function, which defines the intensity of social forces, where I represent the degree of attraction and r is the attracting length scale.

$$h\left(d_{i,j}\right) = Ier^{\frac{-d_{ij}}{r}} - e^{-d_{ij}} \qquad (8)$$

The above equation is a technique that is not appropriate for tackling optimization problems. Stochastic methods are required for exploration and extraction to obtain a reliable approximation to the global optimum. This is accomplished by introducing a new variable, as illustrated in the equation. (9).

$$X_i^n = c\left[\sum_{j=1, j\neq i}^{N} c\left(\frac{u_n + l_n}{2}\right)h\left(d_{ij}\right)\hat{d}_{ij}\right] + \hat{T}_n \qquad (9)$$

This equation indicates that the locust's future location is determined by its present position, the positions of all other locusts, and the target's place. Where u_n is the *nth* dimension's upper bound and *ln* is its lower bound. *T n* is the value of the object's *nth* dimension that has been determined to be the best answer thus far, and c is the reduction factor for personal space reduction, repulsive force, and suction force. c should be adjusted according to the number of iterations to assist extraction as the number of iterations grows to mix the two processes. The equation is used to compute the value of c. (10).

$$c = c_{max} - s\left(\frac{c_{max} - c_{min}}{L}\right) \qquad (10)$$

where c_{max} and c_{min} represent the highest and lowest values, s indicates the present iteration, and L represents the greatest number of iterations. When balancing exploratory and exploitative inclinations away from local optima, the proposed model yields encouraging results, underlining the role of GOAs in calculating class representations. Despite the problem of slow convergence in its conventional version, it has uses in feature selection.

SVM Model

SVM is a popular classification supervised learning technique. SVMs commonly identify email spam and image spam. The SVM constructs a separating hyperplane during the training phase. During the learning phase, equations for separating hyperplanes are generated. This is accomplished by addressing the Lagrangian dualism problem. Given test instances $X_0, X_1 \ldots, X_n$ and labels $z_0, z_1 \ldots, z_n$, where $z_i \in \{-1, 1\}$, solve the Lagrangian dualism problem in the training step by selecting the kernel functions K and C as

$$Maxinize\ L(\lambda) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j z_i z_j \left(X_i, X_j\right) \qquad (11)$$

Subject to $\sum_{i=1}^{n}\lambda_i z_i = 0$ and $C \geq \lambda_i \geq 0$ for i=1,2…n.

We categorise points in the test phase by identifying whichever part of the hyperplane there are on.

IV. Results and Discussions

The algorithm was learned with three datasets in two simultaneous phases, each with 15% of the spam data and 1200 images with non-spam file as one sample set and a balancing standardized test. Tested. A dataset including 600 images of hams and 600 images of non-spam.

Table 1. Summary of test outcomes for each testset in the dataset before to feature selection

Performance Metric	DataSets		
	Dredze	ISH	PERS-G
Accuracy	96.21%	91.54%	89.54%
Precision	96.24%	91.78%	89.57%
Recall	97.87%	92.47%	90.24%
F1-score	98.21%	92.97%	93.47%

From Figure 2. it can be concluded that the Dredze dataset is having highest accuracy of 96.21%, the precision of 96.24%, recall of 97.87%, and F1-score of 98.21%. Similarly, the PERS-G dataset needs better performance before applying feature selection.

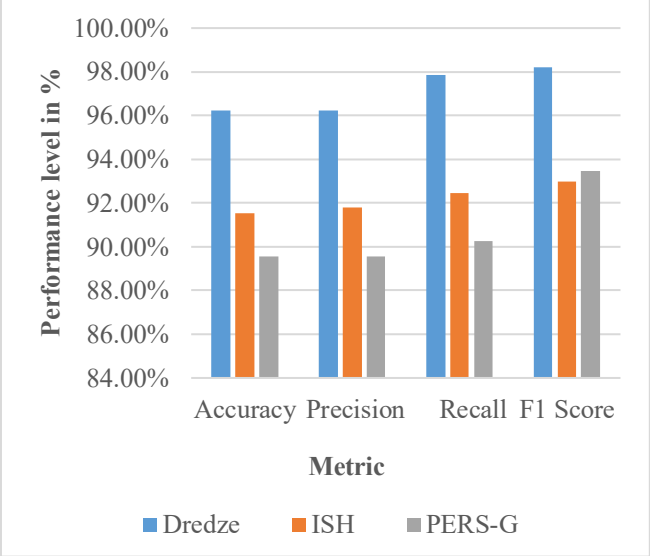

Fig 2. Data set test set test results before GOA

Table 2. Summary of dataset testset test results after feature selection

Performance Metric	DataSets		
	Dredze	ISH	PERS-G
Accuracy	97.64%	92.10%	91.27%
Precision	97.64%	92.41%	90.77%
Recall	98.17%	93.91%	91.80%
F1-score	98.74%	93.31%	95.87%

From Figure 3. it can be concluded that the Dredze dataset is having highest accuracy of 97.64%, the precision of 97.64%, recall of 98.17%, and F1-score of 98.74%. Similarly, the PERS-G dataset needs better performance after applying feature selection.

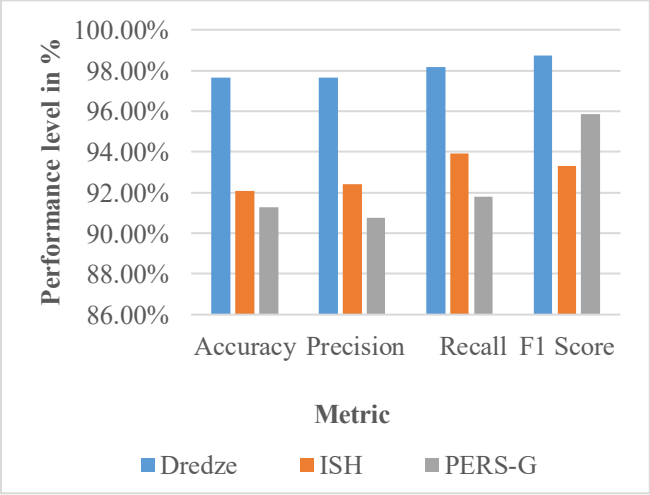

Fig. 3. Test results for testset dataset after GOA.

From Table 3, it can be concluded that the Dredze dataset has a good F1 score of 98.2%, which shows the strength of the feature selection operation before being fed to the classifier.

Table 3. Class labels and F1-score for the test data set.

Dataset	Class Distribution		% of Dataset in the total	F1-score
	Spam	Ham		
Dredze	*91.2%*	*8.8%*	*71.56*	*98.2%*
ISH	*55.3%*	*44.7%*	*7.34*	*92.14%*
PERS-G	*72.4%*	*27.6%*	*25.56*	*91.57%*

CONCLUSION

Thanks to improved spamming strategies, spammers may now avoid classic spam detection systems, including such content-based detection and recognition of optical characters. This has paved the way for image processing tools to detect image spam. A combination of deep learning and image processing techniques might be employed to construct strong classifiers. We created a comparable classifier with just an SVM and image properties as a set of features, and it performed well in image phishing detection on three public datasets. GOA has the best average accuracy of 97.64%.

REFERENCES

[1] Onova, Christopher & Omotehinwa, Temidayo Oluwatosin. (2021). Development of a Machine Learning Model for Image-based Email Spam Detection. 6. 336. 10.46792/fuoyejet.v6i4.718.

[2] Hemalatha, M & Katta, Sriharsha & Santosh, R & Priyanka, Priyanka. (2022). E-MAIL SPAM DETECTION. International Journal of Computer Science and Mobile Computing. 11. 36-44. 10.47760/ijcsmc.2022.v11i01.006.

[3] Kraidia, Insaf & Ghenai, Afifa & Zeghib, Nadia. (2023). HST-Detector: A Multimodal Deep Learning System for Twitter Spam Detection. 10.1007/978-3-031-27099-4_8.

[4] Zhang, Zhibo & Damiani, Ernesto & Al Hamadi, Hussam & Yeun, Chan & Taher, Dr. Fatma. (2022). A Late Multi-Modal Fusion Model for Detecting Hybrid Spam E-mail. 10.48550/arXiv.2210.14616.

[5] S. Assassin, "The apache spamassasin project," Aug 2005. [Online]. Available: http:/spamassasin.apache.orgs

[6] M. Dredze, R. Gevaryahu, A. E.Bachrach. "Learning Fast Classifiers for Image Spam". In Fourth Conference on Email and Anti-Spam, August 2-3, 2007, Mountain View, California USA.

[7] H Aradhye, G. Myers, and J. Herson. "Image Analysis for Efficient Categorization of Image-based spam Email". Proc. IEEE Conf. Document Analysis and Recognition (ICDAR 05), IEEE Press, Aug 2005 pp.914- 918, doi: 1O.1109/lCDAR2005.135.

[8] Mallampati, Deepika, and Nagaratna P. Hegde. "A machine learning based email spam classification framework model: related challenges and issues." International Journal of Innovative Technology and Exploring Engineering 9.4 (2020): 3137-3144..

[9] Pulabaigari, Viswanath & Murty, M. & Bhatnagar, Shalabh. (2023). A pattern synthesis technique to reduce the curse of dimensionality effect.

[10] Salih, Ahmad & Dhannoon, Ban N.. (2021). Weighted k-Nearest Neighbour for Image Spam Classification Iraqi Journal of Science. . .ijs.2021.62.3.32/10.24996. 1036-1045. 62

[11] Hazza, Zubaidah & Aziz, N.A. (2015). A New Efficient Text Detection Method for Image Spam Filtering. International Review on Computers and Software (IRECOS). 10. 10.15866/irecos.v10i1.5111.

[12] Aiwan, Fan & Zhaofeng, Yang. (2018). Image spam filtering using convolutional neural networks. Personal and Ubiquitous Computing. 22. 10.1007/s00779-018-1168-8.

[13] Saleh, Hadeel & Ali Alheeti, Khattab M. & Saad, Saif & Assaf, Omer & Jassam, Noor. (2019). An Enhanced Particle Swarm Optimization algorithm for E-mail Spam Filtering. AUS. 26. 245-251. 10.4206/aus.2019.n26.2.31.

[14] Mallampati, Deepika; Hegde, Nagaratna P. "Feature Extraction and Classification of Email Spam Detection Using IMTF-IDF+Skip-Thought Vectors" Ingénierie des Systèmes d'Information . Dec2022, Vol. 27 Issue 6, p941-948. 8p.

[15] Onova, Christopher & Omotehinwa, Temidayo Oluwatosin. (2021). Development of a Machine Learning Model for Image-based Email Spam Detection. 6. 336. 10.46792/fuoyejet.v6i4.718.)