electronics

MDPI

Article

Hybrid Features by Combining Visual and Text Information to Improve Spam Filtering Performance

Seong-Guk Nam, Yonghun Jang, Dong-Gun Lee and Yeong-Seok Seo *

Department of Computer Engineering, Yeungnam University, Gyeongsan 38541, Korea; sd05031@yu.ac.kr (S.-G.N.); killerwise@ynu.ac.kr (Y.J.); dklee77@ynu.ac.kr (D.-G.L.)
* Correspondence: ysseo@yu.ac.kr; Tel.: +82-53-810-3534

Abstract: The development of information and communication technology has created many positive outcomes, including convenience for people; however, cases of unsolicited communication, such as spam, also occur frequently. Spam is the indiscriminate transmission of unwanted information by anonymous users, called spammers. Spam content is indiscriminately transmitted to users in various forms, such as SMS, e-mail, and social network service posts, causing negative experiences for users of the service, while also creating costs, such as unnecessarily large amounts of network traffic. In addition, spam content includes phishing, hype or false advertising, and illegal content. Recently, spammers have also used images that contain stimulating content to effectively attract users' curiosity and attention. Image spam contains more complex information than text, making it more difficult to analyze and to generalize its properties compared to text. Therefore, existing text-based spam detectors are vulnerable to spam image attacks, resulting in a decline in service quality. In this paper, a "hybrid features by combining visual and text information to improve spam filtering performance" method is proposed to reduce the occurrence of misclassification. The proposed method employs three sub-models to extract features from spam images and a classifier model to output the results using the features. Each sub-model extracts topic-, word-, and image-embedding-based features from spam images. In addition, the sub-models use optical character recognition, latent Dirichlet allocation, and word2Vec techniques to extract features from images. To evaluate spam image classification performance, the spam classifiers were trained using the extracted features and the results were measured using a confusion matrix. Our model achieved an accuracy of 0.9814 and a macro-F1 score of 0.9813. In addition, the application of OCR evasion techniques resulted in a decrease in recognition performance. Using the proposed model, a mean macro-F1 score of 0.9607 was obtained.

Keywords: multimodal; deep learning; CNN; image processing; natural language processing; computer vision; spam identification

check for updates

Citation: Nam, S.-G.; Jang, Y.; Lee, D.-G.; Seo, Y.-S. Hybrid Features by Combining Visual and Text Information to Improve Spam Filtering Performance. *Electronics* **2022**, *11*, 2053. https://doi.org/10.3390/electronics11132053

Academic Editor: Arkaitz Zubiaga

Received: 29 April 2022
Accepted: 6 June 2022
Published: 30 June 2022

1. Introduction

In modern society, many people benefit from the development of information and communication technology (ICT) [1–6]. Users who own desktops and mobile devices can connect to the network anytime and anywhere to communicate with diverse people through social network services (SNS) or enjoy leisure activities through various streaming services.

However, technological advances do not always have positive effects. Simultaneously, as technology advances, people or organizations use technology for malicious purposes [7–12]. For example, spammers (who distribute spam content) can cause various problems by including spam content in comments or posts in services that many users use or by including spam content in email or personal messages. Spam content promotes products, disseminates harmful information to teenagers, such as regarding adult products, and provides inducements to access gambling, drug, and phishing sites [13]. If users are exposed to this content, various social problems may occur. Financial problems may arise due to downloading of viruses and phishing, and teenagers can be indiscriminately exposed to harmful information, such as adult products and drugs [14,15]. Service managers must protect users from indiscriminate spam content by considering many factors, such as network traffic, storage, and security, to maintain service quality [16].

Recently, various services, such as SNS and streaming, have provided avenues for commentary. In the past, only text was available for commenting; however, now, comments can be created using icons, images, and videos. Spammers also distribute spam content using images or videos to effectively engage users' curiosity and attention [13]. Many studies have sought to find methods to filter spam images using traditional classification algorithms and artificial intelligence (AI) techniques for image processing. However, spammers evade spam image detection using various techniques that interfere with image recognition [17]. Spam content composed of text can be detected using rule-based techniques, such as specific keyword checking or the use of certain expressions. Unlike text, images are complex in composition, making them much more difficult to analyze [18]. Moreover, spammers use different techniques, such as the inclusion of optical illusions, noise, resizing, and repositioning, to circumvent spam image filtering systems. A common approach to filtering spam images is to use a classification model; however, this is considered complicated to implement. An increase in the ratio of general images misclassified as spam images causes inconvenience to service users [19]. Although available spam image datasets are unchanged and limited, spammers are constantly generating new spam content.

In this paper, a method that harnesses hybrid features by combining visual and text information to improve spam filtering performance is proposed to analyze the features of various components included in the image to improve the classification of spam images that can contain various anomaly patterns. The goal is to reduce the misclassification rate of generic or ham (i.e., non-spam) [20] images. The proposed method employs three sub-models to extract subject- and word-embedding-based text features from spam images in addition to a spam classifier model based on artificial intelligence. Optical character recognition (OCR) [21] is used to extract the text contained in an image, and latent Dirichlet allocation (LDA) [22–25] and word2Vec [26] are employed to extract subject-related features from the text. A recurrent neural network (RNN) further extracts word embeddings from text and a convolutional neural network (CNN) is used for extracting visual features from images. Finally, a classifier model performs spam classification of the images using the text and visual features extracted by the three sub-models.

In this report, spam-detection-related studies are introduced in Section 2. The proposed method is explained in Section 3. The experimental results of the performance evaluation of the proposed method are presented in Section 4, and Section 5 concludes the paper.

Research contribution: The main contributions of the study are as follows:

- A method is proposed that employs three sub-models to extract features from spam images and a classifier model to output the results using the features.
- The proposed method showed significant performance enhancement compared to existing techniques.
- An ablation study was performed to analyze the impacts of each sub-model.
- Optimal combinations of sub-models for detecting spam images applying OCR evasion techniques are presented.

2. Related Work

In this section, related studies on the classification of various types of spam content are investigated.

Network and storage technologies were not as advanced in the early 2000s as they are now, and neither were information and communication environments, causing spam to unnecessarily consume network traffic and server storage [16]. Many businesses operating these services suffered from economic losses [14]. Spam problems have occurred in various areas of the Internet, such as email, social media, and reviews, along with the development of technology. Services with frequent spam problems mainly use text to communicate effectively. Researchers have conducted studies on effective spam filtering to counter spam problems, with most studies considering approaches which analyze and filter text-based spam content. In early spam classification studies, spam filtering was performed using traditional machine learning techniques, such as support vector machine (SVM) and k-nearest neighbor (kNN) [27,28]. Other studies have improved spam detection accuracy by combining metadata provided by the service with text analysis. Murugavel et al. [29] classified the type of spam thread using spam keywords using a content text analysis method and extracted frequently occurring spam threads to provide better solutions in terms of handling ethical hacking from spammers. Alom et al. [30] trained a deep learning (DL)-based model to detect spammers' accounts using the text of writers and metadata of accounts, such as the age of the account, number of followings, and number of followers. In addition, recently, various studies have been conducted to detect spam reviews [31]. Spam filtering may have been effective for a short period.

However, to avoid text-based spam filtering services, many spammers have utilized the sending of spam mail, including images [32,33]. The method for distributing spam content has evolved continuously with the development of technology. Researchers have also analyzed image spam content to counteract spam trends. Barbar et al. [34] described domain authentication, which is a protocol for verifying sender authentication used in domain keys identified mail (DKIM) documents, to perform spam email filtering. They suggested adding a sender policy framework (SPF) record to identify whether an email is able to be sent in SPF for authentication in email services where DKIM is unavailable. They proposed a FENOMAA (feature extraction neural network with OCR enhanced by mail authentication and analyzer of context) technique to filter spam mail. In general, researchers have analyzed spam images using CNN-based models, which have shown excellent performance in image analysis [35–37].

Unfortunately, existing studies have not considered the text and image features of spam images together. Because many spammers use images to distribute spam content, it is crucial to analyze the images by considering the image and text features together. Therefore, this paper proposes a technique for classifying spam images using image and text features extracted from images to overcome these problems. Table 1 briefly shows a comparison of related studies in terms of approach, target service, target content, model, and dataset.

Table 1. A comparison with related studies.

Research	**Approaches**	**Target Service**			**Target Contents**		**Model**	**Dataset**
		SMS	**Email**	**SNS**	**Text**	**Image**		
Şahin et al. [27]	Using text mining technologies	O	O		O		TF-IDF + kNN	Enron, Ling-Spam, SMS-Spam-Collection
Zamil et al. [28]	Combining kNN and SVM		O			O	kNN + SVM	Dredze [38]
Murugavel et al. [29]	Detecting keywords and threads in email spam corpus		O		O		Multi-split spam corpus algorithm	Email Dataset
Alom et al. [30]	Classifying spam text and spam account with metadata of twitter accounts			O	O		Deep learning model	Twitter Social Honeypot dataset, Twitter 1KS − 10KN dataset
Barbar et al. [34]	Proposing complete solution model with authentication of domain and enhanced OCR		O			O	FENOMAA	-
Sharmin et al. [35]	Classifying spam images with CNN model		O	O		O	CNN	ISH [39], Advanced Dredze, Challenge dataset

Table 1. *Cont.*

Research	Approaches	Target Service			Target Contents		Model	Dataset
		SMS	Email	SNS	Text	Image		
Fatichah et al. [36]	Classifying spam images with CNN models of 3 and 5 layers, AlexNet, VGG16			O		O	CNN	8000 captured images in Instagram
Srinivasan et al. [37]	Training VGG19 and Xception models using transfer learning.		O	O		O	CNN	ISH, Improved dataset, Dredze
The proposed method	Classifying spam images using image and text features	O	O	O	O	O	CNN, Word-embedding, LDA, word2vec	ISH, Dredze

Recently, artificial intelligence (AI) techniques have demonstrated high performance in various fields. In particular, computer vision techniques been shown to exhibit excellent performance in fields, such as medical disease prediction, image synthesis, emotion recognition, motion recognition, object detection, and image summary [40–43]. Many studies have been conducted to classify spam images using AI techniques. People focus on images and videos more easily than text because of the characteristics of the media, making images very effective in highlighting the desired information [13]. Spammers can easily use images on various platforms.

However, it is difficult to accurately analyze the information in images compared to text, and, although many studies have been conducted to classify spam content, several problems still remain [18]. First, the number of public datasets is insufficient. Spammers constantly produce and distribute new spam content, but collecting data is generally quite difficult because of user privacy issues. Therefore, many researchers have continued to conduct research using only a few spam image datasets. However, machine-learning-based spam classification models require additional datasets for higher classification performance because classification performance can vary depending on the amount and characteristics of the trained datasets. Existing spam filters can misclassify new types of spam content with a high probability. Deep-learning-based techniques that have evolved rapidly in recent years have achieved excellent performance through training and testing in many fields; however, they are still not widely used in the industry. AI is being used cautiously in fields such as medicine, to avoid human accidents. Many industrial fields find it difficult to use AI techniques because of their low reliability [44–46]. When AI techniques fail, financial losses can be very large [19]. In this study, the objective was to train high-accuracy spam image classifiers using high-level features extracted from images.

3. Overall Approach

This section describes the proposed method for more accurate spam filtering of images. Figure 1 shows the overall approach of the proposed method in detail. The proposed method extracts various features from images using AI-based text and image processing techniques. Our model uses three sub-models to extract three features from images, and each sub-model extracts topic-, word-embedding-, and convolution-based features.

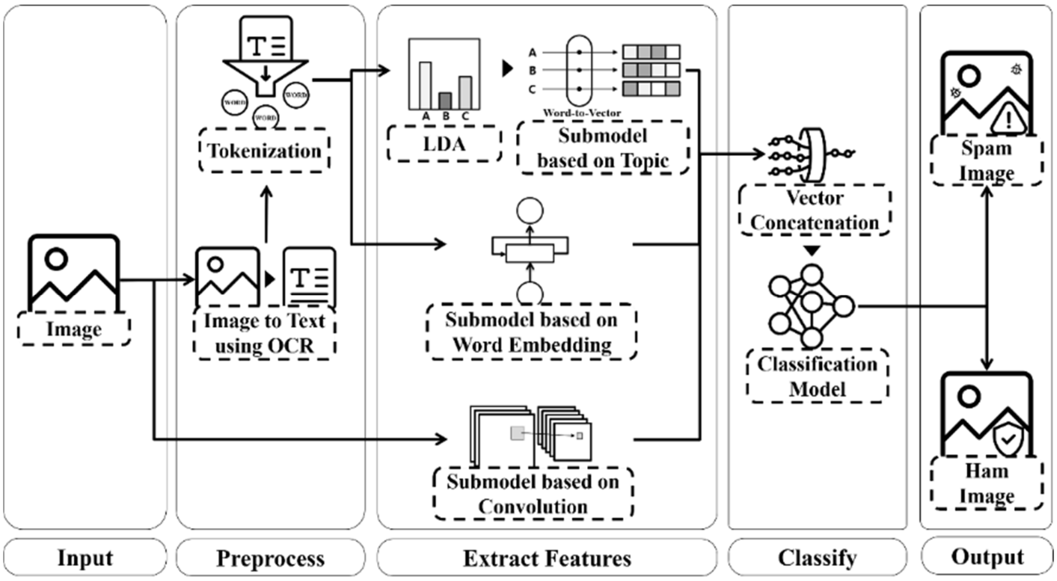

Figure 1. Overall approach.

The sub-models for text processing extract topic-based and word-embedding-based features using the text contained in the images. The OCR technique is used for preprocessing to extract the text contained in images [21]. In addition, the latent Dirichlet allocation technique, LDA is used to extract topic words from the text [22–25]. Each extracted topic word is converted into an embedding vector using a pretrained word-embedding model. Word-embedding-based features are extracted from text using the word-embedding-based sub-model and convolution-based features from images using the convolution-based sub-model. Finally, text and image features extracted from each sub-model are input into the classifier model in the form of a single one-dimensional vector, whereby the model outputs the classification of the input image as spam or ham.

Spammers can include expensive brand products, gambling (money fraud), and adult products in their images to attract a user's curiosity and attention. They can also include SNS IDs, contacts, and links to webpages to induce users to purchase products/services or contact them. The proposed model for extracting and analyzing text from images to filter spam effectively is discussed in Sections 3.1 and 3.2, considering the characteristics of the spam content. The images were preprocessed into text using an OCR technique to analyze the text content in the images.

A neural network-based analysis is described in Sections 3.1 and 3.2 which identifies features that are generally difficult to determine through human analytic power in texts and images [47,48]. A word-embedding-based sub-model was used to analyze text and a convolution-based sub-model to analyze images.

3.1. Sub-Model Based on Topic (Topic Sub-Model)

In this section, the extraction of topic-based features from texts for determining whether the texts contained in the images are spam-related is discussed. First, LDA, which is a stochastic topic model based on the Dirichlet distribution, is used to extract topic words from text. The flowchart in Figure 2 outlines the details of the word-filtering process. Word-filtering is performed on words extracted from the LDA model. The word-filtering process checks whether the extracted words exist in the word list of the pretrained word2vec sub-model that contains many commonly used words. Therefore, words that do not exist in the word2vec sub-model are recognized as errors by the OCR model. In contrast, words that are not misrecognized, but do not exist in the word list, are either checked by converting them into lowercase during the word filtering process or into verbs in their base forms to obtain similar topic-based features. If the modified word is included in the word list, the word filtering process returns the changed word. Finally, word-embedding techniques convert the filtered words into vectors corresponding to each word. Each vector is an embedding vector containing the characteristics of words. The embedding vector is obtained from a pre-trained model that learns various relationships between words from many documents. The vectors can perform various calculations instead of words, and the calculated vectors can be converted back into the most similar words. For example, vector (“Seoul”) − vector (“Korea”) + vector (“Japan”) = vector (“Tokyo”) illustrates a possible calculation. Spammers use images that contain words related to spam to distribute spam content. We used the embedding vector to extract the features of frequently appearing words in spam images.

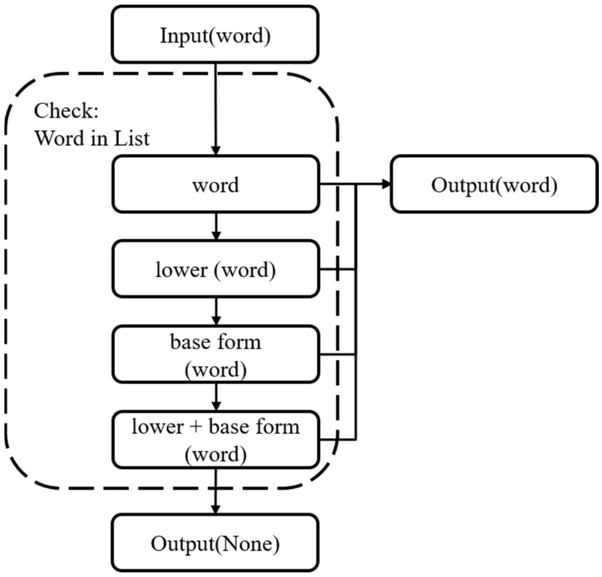

Figure 2. Process of word filtering.

3.2. Sub-Model Based on Word Embedding (Text Sub-Model)

In this section, we introduce a word-embedding-based sub-model. Figure 3 shows the process of the sub-model in detail. The sub-model converts the input words into one feature vector. For the efficient neural network use of natural languages, a vocabulary model consisting of the words used in the dataset is generated. The vocabulary set stores the information obtained by matching index values, which are integers for each word, and converts the input word into index values. If words that are not contained in the vocabulary set are entered, the words are set to zero. An embedding bag model is used to convert the index values into an embedding vector. The word embedding model receives a list of index values. Unfortunately, natural language sentences are not of a fixed length. In other words, the input size is unknown. Therefore, the model receives a list of indexes and sentence lengths. Then, it repeatedly receives the inputs based on the input size. The model converts the input indexes into trained vectors of a fixed size. An embedding bag model outputs the average value of the converted vectors that include features of the input sentence. Thus, the model always outputs a fixed-size vector through the corresponding process. Both the topic-based and the word-embedding-based sub-models convert natural language into trained embedding vectors. The topic-based sub-model aims to identify the features of the key words contained in spam images. However, the word embedding sub-model aims to identify the features of sentences contained in spam images.

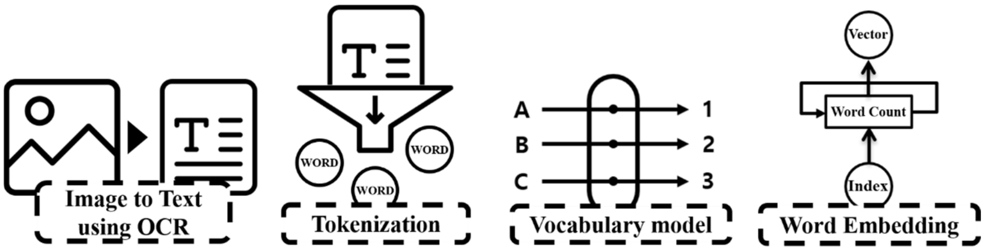

Figure 3. Process of the sub-model based on word embedding.

3.3. Sub-Model Based on Convoltion (Image Sub-Model)

In this section, we introduce the convolution-based sub-model. The sub-model uses a CNN model to effectively extract various features from images that cannot be identified through human analytical power. The CNN model is used to extract feature maps, including edges, texture, and color in images, called convolution-based feature vectors in this paper. The results of the CNN modules, consisting of a convolution layer, rectified linear unit (ReLU) activation function, and pooling layer, are represented by a feature map in multi-dimensional vector form. The feature maps are converted into a single one-dimensional form through a flattening process for use as input in a neural network.

CNN-based models require a fixed-size input. Thus, we resize the input images before using them as inputs of the model to optimize and change to a fixed size. In addition, resizing in our method aims to make the contents of the images recognizable to the human eye. Spammers use images to distribute spam content; images frequently appearing in spam content are as follows: (1) images containing text on a monochrome background; (2) images containing eye-catching objects; and (3) images containing trademarks of companies. The sub-model based on convolution uses multiple layers of CNN modules to extract feature maps that contain various features from large- to small-area units of the images. Feature maps are used for image analysis and spam image filtering.

3.4. Classifier Model

Studies that classify spam images using image or text features already exist; however, there are several problems with these approaches. For example, models that use textual features are OCR-dependent. If the spammer uses OCR evasion techniques, it is difficult for a model using textual features to perform properly. In addition, image data characteristics are difficult to analyze because the pixel value changes significantly even with a slight modification of the image. For example, image changes, such as resizing, repositioning, rotation, brightness, and noise, do not affect human vision significantly, but the values of the data change significantly. Image-based systems are very sensitive to these changes; therefore, when attacks target these changes, performance can significantly degrade. The proposed method reduces the misclassification rate of existing sub-models and can improve classification performance in identifying spam and ham images.

Each model has different advantages; thus, a classifier model using all these features is proposed so that the advantages can complement each other. The process of extracting features from images using the three sub-models presented in Sections 3.1–3.3 is as follows. The three features extracted from the images are used as inputs to the fully connected (FC) classifier model to determine whether the image is spam or ham. The proposed method extracts three feature vectors from images using topic-, word-embedding- and convolution-based sub-models. The extracted vectors consist of one vector for use as an input to the FC layer of the classifier that outputs the image classification as spam or ham.

4. Evaluation

In this section, the dataset, preprocessing, tools, libraries, and frameworks used to conduct the experiment are described, along with details of the experiment conducted to evaluate the spam classification performance of the proposed model.

4.1. Dataset and Preprocessing Methods

To evaluate the spam image classification performance of the proposed model, open image datasets provided by Image Spam Hunter (ISH) [39] and Dredze [38] were used, as in many previous spam image studies. The ISH dataset consists of 930 spam images and 810 ham (non-spam) images. The ISH dataset used for the training and testing of the proposed model consists of files that did not generate errors during preprocessing, as shown in Table 2. The Dredze dataset consists of 3299 spam images and 2021 ham images. The Dredze dataset was used to validate the model trained on the ISH dataset. As shown in Table 3, the Dredze dataset consists of files that did not generate errors during preprocessing. Tables 2 and 3 display the number of images configured in the source and dataset after preprocessing, as well as the difference in the number of sources before and after preprocessing.

Table 2. Components of ISH datasets.

-	SPAM Image	HAM Image
Source dataset	930	810
After pre-processing	921	798
Number of differences	9	12

Table 3. Components of Dredze datasets.

-	SPAM Image	HAM Image
Source dataset	3299	2021
After pre-processing	3265	1892
Number of differences	34	129

Two preprocessing steps were performed to extract text- and convolution-based features from one image. First, the size and channel of the image were changed for use as input to the sub-model to enable the extraction of convolution-based features. A convolution-based sub-model receives only fixed-size data. Therefore, the images of the three channels representing red, green, and blue (RGB) colors were changed to a width and height of 256 pixels each. Next, the characters contained in the image were transformed into text for use as input into the sub-models to perform text feature extraction. The Tesseract OCR [49] tool was used to extract text from images. The extracted text was then tokenized using the Natural Language ToolKit (NLTK) library. In this process, special characters (e.g., @, #, !, ~, etc.) contained in the words were removed.

4.2. Experimental Settings

In this study, three sub-models were proposed for extracting features from images and a classifier model for classifying spam images. The three sub-models consisted of topic- and word-embedding modules for extracting textual features, and a convolution-based module for extracting image features. Finally, the fully connected classifier model received the combined feature vector from the three sub-models and output the image classification as spam or ham.

In our experiments, we prepared seven models including the existing proposed models to evaluate the performance. The existing models were used for classifying spam text (Baseline-1) and images (Baseline-2), respectively [37,50]. Other models were combined using the two baseline models and the topic-based sub-model. Table 4 shows the features used as input by the models in seven modes to classify spam images. In Table 4, the circle represents the sub-model used in the corresponding mode. In addition, Table 5 shows the input type, output size, preprocessing techniques, and processes used by the proposed method in extracting the features of each model.

Table 4. Combination of sub-model usage.

Sub-Models \ Models	Baseline-1 [50]	Baseline-2 [37]	Model 1	Model 2	Model 3	Model 4	Model 5 (Proposed)
Text (word-embedding)	O			O	O		O
Image (Convolution)		O		O		O	O
Topic (LDA/word2vec)			O		O	O	O

Table 5. Details of each model.

Model	Pre-Process	Input Type	Process for Extracting Features from Images	Output (Size)
Topic	OCR, Tokenize	Words [1]	Latent Dirichlet Allocation (LDA), Pretrained word2Vec	Vector (1500)
Text		Words [1]	Word Embedding	Vector (64)
Image	Resize Image Vectorization	3 Channels [2] 256 × 256 Image	Convolution Neural Network (CNN), Flattening	Vector (21,632) (13^2 × 128)
Classifier	Vector Concatenation	Combined Vector [3]	Fully Connected Layer, Sigmoid	True/False

[1] List of tokenized words from input text, [2] Red, green and blue (RGB) channel, [3] 23,196 size vector (=topic(1500) + text(64) + image(21,632)).

To verify the performance of the proposed method, the following three research questions were defined.

- RQ1. Can the proposed method perform classification in an environment in which the existing techniques for filtering spam images are advantageous?
- RQ2. Can the proposed method perform classification in an environment in which the existing techniques for filtering spam images are disadvantageous?
- RQ3. Can the proposed model perform classification on a new spam image dataset that is not used for RQ1 and RQ2 for the model validation?

In this study, the environment in which the existing techniques for spam filtering are advantageous refers to the case in which spam image-filtering models easily classify images because the image quality is good. By contrast, in disadvantageous environments, the various image filtering models applied to images have difficulty in classifying the spam images created by spammers. The spam image classification performance of the proposed model (overall approach) was analyzed in the first experiment to obtain an answer to RQ1. The classification performance results of the proposed model are presented using a confusion matrix. The second experiment analyzed whether the features extracted from the three sub-models contributed to improving image classification performance, to obtain an answer to RQ1. The spam image classification performance of models was analyzed according to feature usage by combining three sub-models in seven ways, as shown in Table 4. The classification performance of the models was further compared based on accuracy and the macro-F1 scores.

To obtain an answer to RQ2, OCR evasion techniques were applied to images to analyze classification performance using the proposed model. Spammers apply various techniques to images to bypass image-based filtering systems. Initially, the OCR evasion techniques used by spammers in the past to evade spam image filter systems were reproduced. OCR evasion techniques were applied to the dataset, as shown in Table 6. A list of the techniques, options, and values used to evade OCR is displayed in Table 6. In addition, none of the techniques and options listed in Table 6 prevents people from recognizing the content. The rate of decrease in text recognition was analyzed through differences in the text recognized in the original images and images to which OCR evasion techniques were applied. Subsequently, images to which the various techniques were applied were classified using the proposed model. The sub-models were further combined in seven ways, as in the second experiment of RQ1, and their spam image classification performance was compared.

Table 6. Techniques and examples to circumvent the OCR systems.

Technique	Option	Value
Gaussian Noise	Variance	0.01~0.03
Salt and Pepper Noise	Amount	5%
	Salt vs. Pepper	50% vs. 50%
Gaussian Blurring	Sigma	1
Median Blurring	Filter size	3 × 3
Rotation	Clockwise	90 degree
Flipping	Vertical (*X*-axis)	None
CAPTCHA *	Random	30%

* Completely Automated Public Turing test to tell Computers and Humans Apart.

In all the experiments, the learning rate scheduler technique was applied to improve the accuracy of the classifier model and an early stopping technique was used to prevent overfitting of the classifier model during training. If the learning rate did not decrease within a certain number of iterations, the learning rate scheduler decreased the learning rate by a factor and early stopping terminated the training. In addition, the hyperparameters used in the training process for the classifier model were configured, as listed in Table 7.

Table 7. Hyperparameter settings for model training.

Hyperparameter	Option
Cross Validation	Shuffled 10 Folds (Training:9, Validation:1)
Start Learning Rate	0.001
Learning Rate Scheduler	Reduce learning rate on plateau Monitor: validation loss, Patience: 3, Factor: 0.9 *
Optimizer	Stochastic gradient descent
Early Stopping	Monitor: validation loss, Patience: 15
Save Option	Monitor: validation loss, Lowest~5th Model
Drop Out	20%

* Factor by which the learning rate will be reduced (learning rate = learning rate × factor).

4.3. Experimental Results

In Section 4.3, the explored performance of the proposed model in the environment defined in Section 4.2 to obtain answers to the two research questions is presented. The accuracy and macro-F1 scores of the models were used to evaluate performance.

4.3.1. Answer for RQ1

The first experiment was conducted to determine the RQ1. The classification results using the proposed model (overall approach model) and 10-fold cross-validation are presented in Table 8. The confusion matrix of the classification results is presented in Tables 9 and 10. Table 8 shows the number of images in the classified results and the ratio of each item. Tables 9 and 10 display the accuracy, precision, recall, F1 and macro-F1 scores. Accuracy in classification performance is an intuitive indicator of performance; however, it is difficult to trust the results if the composition of the dataset is unbalanced. Therefore, precision, recall, and the F1 scores were used to compensate for these problems and the results were analyzed to differentiate spam and ham. The misclassification of ham images is fatal to the service; therefore, the classification performance of ham images is as important as that of spam images. Finally, the macro-F1 score, which is the average value of the F1 score on the spam and ham sides, was used to compare the performances of the models.

Table 8. Overall result of model classification approaches.

Label / Predict	SPAM	HAM
SPAM	777 (45.20%)	11 (0.64%)
HAM	21 (0.12%)	910 (52.94%)

Table 9. Confusion matrix of classification results.

Name		Value
	Accuracy	0.9814
Spam Side	Precision	0.974
	Recall	0.986
	F1-Score	0.980
Ham Side	Precision	0.988
	Recall	0.977
	F1-Score	0.983
	Macro-F1-Score	0.9813

Table 10. Features and performance indicators of the models used.

	Spam Side			Ham Side			Total	
Models	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Macro-F1-Score**
Baseline-1 (only Text) [50]	0.9724	0.9823	0.9773	0.9848	0.9763	0.9805	0.9791	0.9789
Baseline-2 (only Image) [37]	0.9561	0.9658	0.9610	0.9707	0.9623	0.9665	0.9639	0.9637
Model 1 (only Topic)	0.9712	0.9949	0.9829	0.9957	0.9755	0.9855	0.9843	0.9842
Model 2 (Text + Image)	0.9586	0.9696	0.9641	0.9739	0.9645	0.9692	0.9668	0.9666
Model 3 (Text + Topic)	0.9674	0.9936	0.9803	0.9946	0.9724	0.9834	0.9820	0.9818
Model 4 (Image + Topic)	0.9825	0.9949	0.9887	0.9957	0.9850	0.9903	0.9895	0.9895
Model 5 (All, proposed)	0.9737	0.9860	0.9798	0.9881	0.9774	0.9827	0.9814	0.9813

4.3.2. Answer for RQ2

The second experiment was conducted to determine the answer to RQ2. OCR evasion techniques used by past spammers were applied to images to reproduce situations in which OCR could not perform properly.

The images subjected to OCR evasion techniques exhibited a decrease in text recognition performance, as shown in Table 11. Table 11 shows the extent of the text recognition rate decrease compared to the original for the techniques applied to the image, resulting in a decrease in OCR performance. Each item in Table 11 is a duplicate ratio calculated by recognizing text in units of words in the images. If the same character is not recognized or is recognized as different from the original, it contributes to the decrease in the recognition rate. Furthermore, the classification performance of the models changed, as shown in Table 12, depending on the OCR evasion technique applied, using models trained on the original images. In particular, it was demonstrated that classification performance decreased rapidly for the topic and text sub-models analyzing text. In Table 12, the rows represent the techniques applied to the images and the columns represent the features of the image used as input to the classifier model. In addition, each item in the table has a value between 0 and 1. Models that used only text-based features showed a decline in the overall classification performance and dropped to a maximum of approximately 33 points. On the other hand, models using image-based features were less affected by the decrease in classification performance than the text model, by up to 6.4 points. Finally, the classification performance was compared using the mean macro-F1 score of each model. Our proposed model using all features achieved the highest score of 0.9607, showing the best classification performance in OCR evasion techniques among the seven models. Spammers send spam images to avoid detection by the image-spam filtering system. Furthermore, it was demonstrated that our proposed model involved less performance reduction and was more stable in spam image classification performance compared to other models, even with the OCR evasion techniques applied to the images. Therefore, our proposed model was superior in classification performance on images with regards to OCR evasion techniques.

Table 11. Text recognition rate according to the application of OCR evasion techniques.

Technique	Text Recognition Rate
Gaussian Noise	67.28%
Salt and Pepper Noise	22.86%
Gaussian Blurring	47.06%
Median Blurring	24.12%
Rotation	69.43%
Flipping	3.24%
CAPTCHA	48.06%

Table 12. Performance of spam image classification model with OCR evasion techniques.

Models Dataset	Baseline-1 (Only Text)	Baseline-2 (Only Image)	Model 1 (Only Topic)	Model 2 (Text + Image)	Model 3 (Text + Topic)	Model 4 (Image + Topic)	Model 5 (All)
Original	0.9789	0.9637	0.9842	0.9666	0.9818	0.9895	0.9813
Gaussian Noise	0.9575	0.9673	0.9557	0.9643	0.9574	0.9673	0.9737
Salt and Pepper	0.9203	0.9614	0.9296	0.9626	0.9307	0.9592	0.9696
Gaussian Blur	0.8999	0.9568	0.9255	0.9414	0.9261	0.9662	0.9696
Median Blur	0.7698	0.9643	0.8684	0.9655	0.8643	0.9644	0.9661
Rotation	0.9498	0.9179	0.9441	0.9336	0.9493	0.9279	0.9365
Flipping	0.6404	0.9562	0.6768	0.9597	0.6693	0.9250	0.9435
CAPTCHA	0.9377	0.9579	0.9603	0.9591	0.9603	0.9714	0.9661
Average	**0.8679**	**0.9535**	**0.8943**	**0.9552**	**0.8939**	**0.9545**	**0.9607**

Two research questions were defined, and experiments were conducted to verify the performance of the proposed technique. The first research question was "Can the proposed technique method classification work in an environment in which existing techniques for filtering spam images are advantageous?". Sub-models were combined in seven ways to analyze whether they contributed to spam image classification performance. The macro-F1 score of our proposed model was 0.9813, exhibiting the best performance among the seven models.

To examine the experimental results, a chi-square test [51] was used to compare differences in the proportions of spam image classification between the baseline models and the five models. Tables 13 and 14 show the p-values of the chi-square test of homogeneity. Values in bold indicate statistical significance ($p < 0.05$) between the models. As presented in Table 13, Models 2, 4, and 5, including the image sub-model, showed significant differences compared to Baseline-1 using only the text sub-model. As provided in Table 14, Models 1, 3, and 4 including the topic sub-model showed statistical differences in performance compared to Baseline-2. From the statistical tests, Model 4 generally showed that there were significant differences in performance in both Baseline-1 and Baseline-2 cases. In addition, the proposed methods could be complementary to the existing models.

Table 13. *p*-values of the chi-square test with Baseline-1 on the ISH dataset.

Models Dataset	Model 1 (Only Topic)	Model 2 (Text + Image)	Model 3 (Text + Topic)	Model 4 (Image + Topic)	Model 5 (All)
Original	0.13088	0.17434	0.19919	0.05941	0.94254
Gaussian Noise	0.95252	**0.00001**	0.93610	0.45632	**0.00329**
Salt and Pepper	0.73807	**0.00000**	0.60672	**0.00000**	**0.00000**
Gaussian Blur	0.05261	**0.00000**	0.05122	**0.00000**	**0.00000**
Median Blur	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**
Rotation	0.74227	0.21855	0.90336	**0.00000**	**0.02575**
Flipping	0.12073	**0.00000**	0.28285	**0.00000**	**0.00000**
CAPTCHA	**0.00807**	**0.00005**	**0.01374**	**0.00004**	**0.00023**

Bold indicates the *p*-value less than 0.05.

Table 14. *p*-values of the chi-square test with Baseline-2 on the ISH dataset.

Models Dataset	Model 1 (Only Topic)	Model 2 (Text + Image)	Model 3 (Text + Topic)	Model 4 (Image + Topic)	Model 5 (All)
Original	**0.00018**	0.97038	**0.00079**	**0.00001**	**0.01472**
Gaussian Noise	**0.00096**	0.77645	**0.00195**	**0.00230**	0.07059
Salt and Pepper	**0.00000**	0.95324	**0.00000**	**0.00069**	0.11195
Gaussian Blur	**0.00000**	0.92326	**0.00000**	**0.01034**	0.13302
Median Blur	**0.00000**	0.99828	**0.00000**	**0.00056**	**0.03140**
Rotation	**0.02538**	0.35590	**0.00306**	**0.00019**	**0.04086**
Flipping	**0.00000**	0.95990	**0.00000**	**0.00000**	**0.00001**
CAPTCHA	0.24253	0.99867	0.11910	**0.01258**	0.28896

Bold indicates the *p*-value less than 0.05.

4.3.3. Answer for RQ3

This section presents a verification of the performance of the seven models trained on the ISH dataset using the Dredze dataset [38]. The Dredze dataset is approximately three times larger than the ISH dataset; therefore, it is expected to show more realistic results than the current evaluation results (Table 12). Table 15 shows the macro-F1 score evaluation of the performance of the seven models using the Dredze dataset.

Table 15. Performance of spam image classification model with the Dredze dataset.

Models Dataset	Baseline-1 (Only Text)	Baseline-2 (Only Image)	Model 1 (Only Topic)	Model 2 (Text + Image)	Model 3 (Text + Topic)	Model 4 (Image + Topic)	Model 5 (All)
Original	0.9342	0.8163	0.8823	0.8154	0.8817	0.8945	0.8689
Gaussian Noise	0.9042	0.8138	0.8704	0.8223	0.8721	0.8819	0.8722
Salt and Pepper	0.8721	0.8152	0.8302	0.8179	0.8314	0.8684	0.8583
Gaussian Blur	0.7724	0.7814	0.8212	0.7888	0.8164	0.8759	0.8586
Median Blur	0.6581	0.8170	0.8040	0.8171	0.8005	0.8760	0.8623
Rotation	0.9224	0.7727	0.8813	0.7768	0.8798	0.8406	0.8077
Flipping	0.4389	0.8172	0.5642	0.8194	0.5619	0.8467	0.8464
CAPTCHA	0.8964	0.7765	0.8737	0.7750	0.8766	0.8484	0.8264
Average	**0.7998**	**0.8013**	**0.8159**	**0.8041**	**0.8151**	**0.8666**	**0.8501**

The evaluation result shows a pattern similar to the result shown in Table 12 (ISH dataset), but the overall performance dropped by approximately 7% to 15%. Interestingly, the patterns in Tables 12 and 15 provide certain observations. First, Baseline-1, which uses only the text sub-model, exhibited the most robust performance in the original dataset without any attack. However, Baseline-1 showed the lowest average performance at 0.7998 in the evaluation by applying various image distortions, indicating that it had the weakest ability to respond to OCR evasion attacks. A further interesting observation is that Baseline-1, Model-1, and Model-2 were highly vulnerable to flipping attacks. These models have in common that they omit an image sub-model. Thus, it can be inferred that models composed only of text or topic models based on OCR are very vulnerable to flipping attacks. Flipping is an attack that must be dealt as it does not involve a problem with human perception, as it is simply an image distortion applied with vertical inversion. Finally, it is of note that the models, including the image sub-model, were highly vulnerable to rotation and capture attacks. However, the models including only the text or topic sub-model showed a performance drop of only approximately 1% for the rotation attack.

This experiment demonstrated that the OCR-based (text, topic) model and the image-based model have a complementary relationship with each other, and that the most efficient model is Model-4.

Tables 16 and 17 show the *p*-value of the chi-square test for the analysis of significant differences between the baseline models and the five models in the Dredze dataset. Values in bold indicate statistical significance ($p < 0.05$) between the models. As presented in Table 16, all cases showed significant differences with Baseline-1. In Table 17, all cases except Model 2 showed significant differences with Baseline-2. The models using the topic sub-model were able to complement the performance of the baseline models even for datasets that were not used for the model training.

Table 16. *p*-values of the chi-square test with Baseline 1 for the Dredze dataset.

Models **Dataset**	**Model 1** **(Only Topic)**	**Model 2** **(Text + Image)**	**Model 3** **(Text + Topic)**	**Model 4** **(Image + Topic)**	**Model 5** **(All)**
Original	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**
Gaussian Noise	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**
Salt and Pepper	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**
Gaussian Blur	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**
Median Blur	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**
Rotation	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**
Flipping	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**
CAPTCHA	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**

Bold values indicate less than or equal to 0.05.

Table 17. *p*-value of chi-square test with Baseline 2 for the Dredze dataset.

Models **Dataset**	**Model 1** **(Only Topic)**	**Model 2** **(Text + Image)**	**Model 3** **(Text + Topic)**	**Model 4** **(Image + Topic)**	**Model 5** **(All)**
Original	**0.00000**	0.16052	**0.00000**	**0.00000**	**0.00000**
Gaussian Noise	**0.00000**	0.10389	**0.00000**	**0.00000**	**0.00000**
Salt and Pepper	**0.00000**	0.09382	**0.00000**	**0.00000**	**0.00000**
Gaussian Blur	**0.00000**	0.33331	**0.00000**	**0.00000**	**0.00000**
Median Blur	**0.00000**	0.06803	**0.00000**	**0.00000**	**0.00000**
Rotation	**0.00000**	0.35616	**0.00000**	**0.00000**	**0.00000**
Flipping	**0.00000**	0.00591	**0.00000**	**0.00000**	**0.00000**
CAPTCHA	**0.00000**	0.61324	**0.00000**	**0.00000**	**0.00000**

Bold values indicate less than or equal to 0.05.

5. Conclusions

A hybrid features by combining visual and text information to improve spam filtering performance method was proposed to reduce the occurrence of misclassification. The proposed method uses three sub-models to extract topic-, word-embedding-, and convolution-based features from images. Features extracted from the image are combined into a single one-dimensional vector and used as input to the classifier to determine whether the image is spam or ham. The proposed model was trained and validated using an image dataset provided by ISH to evaluate the performance of spam image classification. Furthermore, a 10-fold cross-validation technique was used to prevent overfitting of the proposed model. However, the images contained in the dataset were not the same as those actually used in spam. The second research question was "Can the proposed method perform classification in an environment in which existing techniques for filtering spam images are disadvantageous?". A disadvantageous environment was constructed for OCR whereby seven OCR evasion techniques (tricks) used by spammers to reduce the performance of the image filtering system were explored. It was demonstrated that text recognition rates in OCR were reduced, although there was no significant difference in the human interpretation of images. Furthermore, the average macro-F1 score of the proposed model, which uses all three sub-models in this environment, was 0.9613, displaying strong performance on images with OCR evasion techniques compared to other models.

In this study, the model was trained and evaluated using only public datasets. However, new spam content and images are constantly being generated and there are more varieties of these than the images used in this study. In future work, a pipeline will be proposed to collect the latest spam images to continuously develop improved models. In addition, we intend to study techniques to perform spam image classification according to the service environment by analyzing the information on the platform and community interest.

Author Contributions: Conceptualization, S.-G.N., D.-G.L. and Y.-S.S.; data curation, S.-G.N.; funding acquisition, Y.-S.S.; investigation, S.-G.N., Y.J., D.-G.L. and Y.-S.S.; methodology, S.-G.N., Y.J., D.-G.L. and Y.-S.S.; project administration, D.-G.L. and Y.-S.S.; resources, Y.J. and D.-G.L.; software, S.-G.N. and D.-G.L.; supervision, Y.J. and Y.-S.S.; validation, S.-G.N. and Y.-S.S.; visualization, S.-G.N. and Y.-S.S.; writing—original draft, S.-G.N., Y.J. and Y.-S.S.; writing—review and editing, Y.-S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2020R1I1A3073313).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Strielkowski, W.; Firsova, I. Effective management of energy consumption during the COVID-19 pandemic: The role of ICT solutions. *Energies* **2021**, *14*, 893. [CrossRef]
2. Gong, D.; Liu, S. Who benefits from online financing? A sharing economy E-tailing platform perspective. *Int. J. Prod. Econ.* **2020**, *222*, 107490. [CrossRef]
3. Cannon, P.; Lumsden, L. An innovative and authentic way of learning how to consult remotely in response to the COVID-19 pandemic. *Educ. Primary Care* **2022**, *33*, 53–58. [CrossRef] [PubMed]
4. Tanwar, S.; Parekh, K. Blockchain-based electronic healthcare record system for healthcare 4.0 applications. *J. Inf. Secur. Appl.* **2020**, *50*, 102407. [CrossRef]
5. Alhaboobi, Z.A.; Yousif, S.T. Intelligent classroom a conceptual model for the effective use of internet of things technique. In Proceedings of the 2019 2nd Scientific Conference of Computer Sciences (SCCS), Baghdad, Iraq, 27–28 March 2019; pp. 116–120.
6. Meqdad, M.N.; Majdi, H.S. Enabling Techniques for 10 Gbps Long-Haul Transmission in Non-Coherent OCDMA Systems. In Proceedings of the 2018 9th International Symposium on Telecommunications (IST), Tehran, Iran, 17–19 December 2018; pp. 457–459.
7. Ilker, K.A.R.A.; Aydos, M. Cyber fraud: Detection and analysis of the crypto-ransomware. In Proceedings of the 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 28–31 October 2020; pp. 764–769.
8. Hayes, D.R.; Cappa, F. A framework for more effective dark web marketplace investigations. *Information* **2018**, *9*, 186. [CrossRef]
9. Datta, P.; Panda, S.N. A technical review report on cyber crimes in India. In Proceedings of the 2020 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 12–14 March 2020; pp. 269–275.
10. Lee, M.; Park, E. Real-time Korean voice phishing detection based on machine learning approaches. *J. Ambient. Intell. Humaniz. Comput.* **2021**, 1–12. [CrossRef]
11. Loukas, G.; Patrikakis, C.Z. Digital deception: Cyber fraud and online misinformation. *IT Prof.* **2020**, *22*, 19–20. [CrossRef]
12. Shambhavee, H.M. Cyber-Stalking: Threat to People or Bane to Technology. *Int. J. Trend Sci. Res. Dev.* **2019**, *3*, 350–355. [CrossRef]
13. Yu, S. Sex in Spam: A Content Analysis. *Int. J. Crim. Justice Sci.* **2014**, *9*, 35.
14. Ukai, Y.; Takemura, T. Spam mails impede economic growth. *Rev. Socionetwork Strateg.* **2007**, *1*, 14–22. [CrossRef]

15. Available online: https://www.weforum.org/agenda/2018/05/its-40-years-since-the-first-spam-email-was-sent-here-are-6-things-you-didnt/ (accessed on 16 May 2022).
16. Fonseca, O.; Fazzion, E. Measuring, characterizing and avoiding spam traffic costs. *IEEE Internet Comput.* **2016**, *20*, 16–24. [CrossRef]
17. Biggio, B.; Fumera, G.; Pillai, I.; Roli, F. Image Spam Filtering by Content Obscuring Detection. In Proceedings of the Fourth Conference on Email and Antispam (CEAS), Mountain View, CA, USA, 2–3 August 2007; pp. 1–5.
18. Bouma-Sims, E.; Reaves, B. A First Look at Scams on YouTube. *arXiv* **2021**, arXiv:2104.06515.
19. Adadi, A.; Berrada, M. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160. [CrossRef]
20. Karim, A.; Azam, S. Efficient clustering of emails into spam and ham: The *foundational* study of a comprehensive unsupervised framework. *IEEE Access* **2020**, *8*, 154759–154788. [CrossRef]
21. Memon, J.; Sami, M. Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR). *IEEE Access* **2020**, *8*, 142642–142668. [CrossRef]
22. Abrigo, A.B.C.; Estuar, M.R.J.E. A comparative analysis of N-Gram deep neural network approach to classifying human perception on Dengvaxia. In Proceedings of the 2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT), Kahului, HI, USA, 14–17 March 2019; pp. 46–51.
23. Anwar, W.; Bajwa, I.S. An empirical study on forensic analysis of Urdu text using LDA-based authorship attribution. *IEEE Access* **2018**, *7*, 3224–3234. [CrossRef]
24. Huang, Y.; Wang, R. Sentiment Classification of Crowdsourcing Participants' Reviews Text Based on LDA Topic Model. *IEEE Access* **2021**, *9*, 108131–108143. [CrossRef]
25. Lee, D.G.; Seo, Y.S. Improving bug report triage performance using artificial intelligence based document generation model. *Hum. -Cent. Comput. Inf. Sci.* **2020**, *10*, 26. [CrossRef]
26. Mikolov, T.; Chen, K. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.
27. Şahin, D.Ö.; Demirci, S. Spam Filtering with KNN: Investigation of the Effect of k Value on Classification Performance. In Proceedings of the 2020 28th Signal Processing and Communications Applications Conference (SIU), Gaziantep, Turkey, 5–7 October 2020; pp. 1–4.
28. Zamil, Y.K.; Ali, S.A. Spam image email filtering using K-NN and SVM. *Int. J. Electr. Comput. Eng.* **2019**, *9*, 2088–8708. [CrossRef]
29. Murugavel, U.; Santhi, R. Detection of spam and threads identification in E-mail spam corpus using content based text ana-lytics method. *Mater. Today Proc.* **2020**, *33*, 3319–3323. [CrossRef]
30. Alom, Z.; Carminati, B. A deep learning model for Twitter spam detection. *Online Soc. Netw. Media* **2020**, *18*, 100079. [CrossRef]
31. Hussain, N.; Turab Mirza, H. Spam review detection techniques: A systematic literature review. *Appl. Sci.* **2019**, *9*, 987. [CrossRef]
32. Wang, D.; Irani, D. A study on evolution of email spam over fifteen years. In Proceedings of the 9th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing, Austin, TX, USA, 20–23 October 2013; p. 9.
33. Annadatha, A.; Stamp, M. Image spam analysis and detection. *J. Comput. Virol. Hacking Tech.* **2018**, *14*, 39–52. [CrossRef]
34. Barbar, A.; Ismail, A. Image Spam Detection Using FENOMAA Technique. In Proceedings of the International Conference on Artificial Intelligence and Applied Mathematics in Engineering (ICAIAME 2019), Antalya, Turkey, 20–22 April 2019.
35. Sharmin, T.; Di Troia, F. Convolutional Neural Networks for Image Spam Detection. *Inf. Secur. J. A Glob. Perspect.* **2020**, *29*, 103–117. [CrossRef]
36. Fatichah, C.; Lazuardi, W.F. Image Spam Detection on Instagram Using Convolution Neural Network. In *Intelligent and Interactive Computing*; Piuri, V., Balas, V., Borah, S., Syed Ahmad, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2019; Volume 67, pp. 295–303.
37. Srinivasan, S.; Ravi, V. Deep Convolutional Neural Network based Image Spam Classification. In Proceedings of the 2020 6th Conference on Data Science and Machine Learning Applications (CDMA), Riyadh, Saudi Arabia, 4–5 March 2020; pp. 112–117.
38. Dredze, M.; Gevaryahu, R. Elias-Bachrach, A. Learning fast classifiers for image spam. In proceedings of the Fourth Conference on Email and Anti-Spam (CEAS), Mountain View, CA, USA, 2–3 August 2007; pp. 487–493.
39. Gao, Y.; Yang, M. Image spam hunter. In Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 1765–1768.
40. Zaidi, S.S.A.; Ansari, M.S. A survey of modern deep learning based object detection models. *Digit. Signal Processing* **2022**, *126*, 103514. [CrossRef]
41. Lee, D.G.; Jang, Y. Intelligent Image Synthesis for Accurate Retinal Diagnosis. *Electronics* **2020**, *9*, 767. [CrossRef]
42. Huh, J.H.; Seo., Y.S. Understanding Edge Computing: Engineering Evolution with Artificial Intelligence. *IEEE Access* **2019**, *7*, 164229–164245. [CrossRef]
43. Kim, S.K.; Huh, J.H. Artificial Neural Network Blockchain Techniques for Healthcare System: Focusing on the Personal Health Records. *Electronics* **2020**, *9*, 763. [CrossRef]
44. Gade, K.; Geyik, S.C. Explainable AI in industry. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, New York, NY, USA, 4–8 August 2019; pp. 3203–3204.
45. Samek, W.; Wiegand, T. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv* **2017**, arXiv:1708.08296.
46. Arrieta, A.B.; Díaz-Rodríguez, N. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 85–115.

47. Shi, C.T. Signal pattern recognition based on fractal features and machine learning. *Appl. Sci.* **2018**, *8*, 1327. [CrossRef]
48. Wang, Z.; Wang, J. An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition. *Neurocomputing* **2018**, *310*, 213–222. [CrossRef]
49. Tesseract OCR. Available online: https://github.com/tesseract-ocr (accessed on 16 May 2022).
50. Alsaffar, D.; Alfahhad, A. Machine and deep learning algorithms for Twitter spam detection. In *International Conference on Advanced Intelligent Systems and Informatics*; Hassanien, A., Shaalan, K., Tolba, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2019; Volume 1058, pp. 483–491.
51. Bolboacă, S.D.; Jäntschi, L.; Sestraş, A.F.; Sestraş, R.E.; Pamfil, D.C. Pearson-Fisher Chi-Square Statistic Revisited. *Information* **2011**, *2*, 528. [CrossRef]