International Journal of Scientific Research & Engineering Trends
Volume 5, Issue5, Sep-Oct-2019, ISSN (Online): 2395-566X

E-Mail Spam Classification Using Long Short-Term Memory Method

M.E. Scholar (CSE) Shripriya Dongre
Jawaharlal Institute of Technology (JIT) Borawan
Shridongre11@gmail.com

HOD (CSE) Prof. Kamlesh Patidar
Jawaharlal Institute of Technology (JIT) Borawan
Kt125412@gmail.com

Abstract - **Inrecentyears, thesingle-modalspamfilteringsystemshavehadahighdetectionratefor text spamming. To avoid detection based on the single-modal spam filtering systems, spammers in jet junk information into the multi-modality part of an email and combine them toreducetherecognitionrateofthesingle-modalspamfilteringsystems,therebyimplementingthe purposeofevadingdetection.Inviewofthissituation,anewmodelcalledtext based dataset module architecture based on model fusion (MMA-MF) is proposed, which use a text based dataset fusion method to ensure it could effectively filter spam whether it is hidden in the text. The model fuses a Convolution Neural Network (CNN) model and a Long Short-Term Memory (LSTM) model to filter spam. Using the LSTM model and the CNN model to process the text parts of an email separately to obtain two classification probability values, then the two classification probability values are incorporated into a fusion mode lt identify whether the email is spam or not. For the hyper parameters of the MMA-MF model, we use a grid search optimization method to get the most suitable hyper parameters for it, and employ a k-fold cross-validation method to evaluate the performance of this model. Our experimental results show that this model is superior to the traditional spam filtering systems and can achieve accuracies in the range of92.64–98.48%.**

Keywords - **Spam Filtering System; Multi-Modal; MMA-MF; Fusion Model; LSTM; CNN.**

I. INTRODUCTION

Spam can be defined as an email which contains sun solicited mail [1].With the rapid development of the Internet, Internet users are increasingly using emails to communicate. At the same time, the issue of spam is getting worse, in which the purpose of most spam is to solicit the recipients for money. In order to achieve this, the products they provide claim to miraculously cure health problems such as diabetes, obesity and hair loss. They may be of any nature, whether it is an advertisement, a text email, an image email or a email that contains text and image data. According to the spam analysis report of Kaspersky Lab, a well-known organization in the security field, the average proportion of global spam in total emails were as high as 56.63% or more in 2017 [2].

This phenomenon indicates that spam is flooding the entire network, which brings inconvenience to cybercitizens. For text spam or image spam, the single-modal spam filtering systems have a high detection rate, while, in order to escape detection, spammers may insert junk information into the multi-modal part of an email, which we call it hybrid spam, to reduce the detection rate of the single-modal spam filtering systems, ultimately achieving the purpose of evading detection. For hybrid spam, it is more harmful than traditional spam because it contains more information than traditional spam, and it requires more network band width and storage space for forwarding and delivery of the mail box servers. Moreover, viruses or unsolicited information carried by hybrid spam are more difficult to detect, which brings tremendous information security risks to people's communication. Therefore, it is extremely important to learn how to effectively identify hybrid spam. In machine learning and cyber security communities, anti-spam methods have been studied for many years [3–15]. These methods roughly are classified into three categories-

- Text-based spam detection
- Image-based spam detection.
- Multi-modal spam detection.

The first and second categories primarily use the textual content or image content of an email to filter spam, respectively. However, the last category processes both the textual and image content of an email to filter spam.

The proposed method the text in an email, so it can efficiently filter spam whether the junk information is hidden in the text. That is, the advantage of the MMA-MF model is that it can not only filter hybrid spam, but also filter spam with only text data. The experimental results indicate that our method is better than other methods significantly. The main contribution is that we apply the CNN and LSTM model to handle the text data

International Journal of Scientific Research & Engineering Trends
Volume 5, Issue5, Sep-Oct-2019, ISSN (Online): 2395-566X

in an email, and combine them into a fusion model by the logistic regression method. To our best knowledge, we firstly shed light on this approach in the email filtering systems. The rest of this paper is organized as follows:Section2 describes the architecture of the MMA-MF model, we present the design frame work of the CNN,LSTM and fusion model, the brief categorization algorithm for text spam.Section3 presents evaluation metrics and validation schemes. Section4 is about experimental results and discussion. In the end, conclusions are given in Section5.

II. MMA-MF MODELARCHITECTURES

Essentially, the spam filtering system is a binary classification problem. In order to make our model not only filter hybrid spam but also filter spam with only text data, we propose a kind of spam filtering framework called MMA-MF. This framework shows in Figure1.

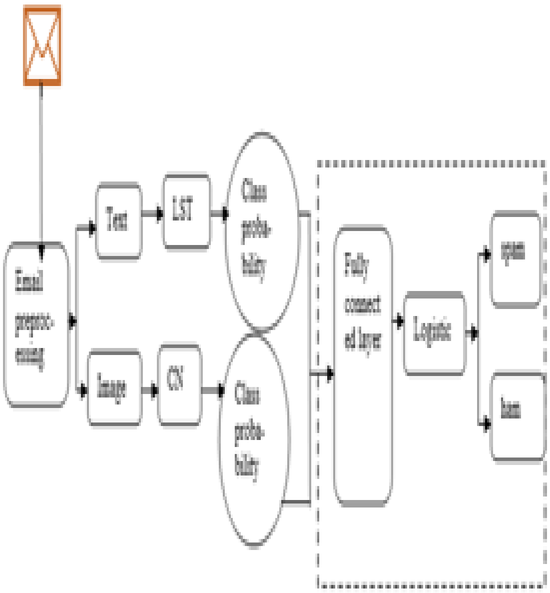

Fig.1. MMA-MF Model Architecture.

The specific steps of the MMA-MF model to identify spam are described as follows:

- Email pre-processing: separate the text data from an email to obtain the text dataset.
- Obtaining the optimal classifiers: the text dataset is used to train and optimize the LST M model and the CNN model, respectively—finally getting the optimal LSTM model and the optimal CNN model.
- Obtaining the classification probability values: the is re-entered into the optimal CNN model to obtain the classification probability values of the as spam. Similarly, the text data set is re-entered in to the optimal LST M model to obtain the classification probability values of the text dataset as spam. For an email that only has text data, we use dropout ideology to set the corresponding model output probability value p =0.5.
- Obtaining the optimal fusion model: the two classification probability values are fed into the fusion model to train and optimize it, ultimately getting the optimal fusion model.

In the above descriptions, through by steps 1, 3 and 4, we can get the classification probability value of a new email as spam, whether the new email is a hybrid email or a single-modal email. In conclusion, we give the overall framework of the MMF-MA model and the brief steps for obtaining the classification probability value of an email as spam. Next, we will introduce the internal structure of the LSTM model, the CNN model and the fusion model, and the selection of the optimal hyper parameter values for the three models in detail.

1.Text Classification Model: LSTM Model

The structure of the LSTM model is roughly shown in Figure 2. It is composed of a one word embedded layer, two LSTM layers and one fully connected (FC) layer. The steps of handling the text portion of an email to obtain the classification probability value of the email are as follows: firstly using the preprocessing technique to acquire the text data of an email, then using the word embedding technique to get its word vector representation. In this paper, we select the word2vec toolkit to get word vector representation. After that, we use the designed two LSTM layers to automatically extract features from the text data. Finally, we apply the FC layer with Soft max activation function to obtain the classification probability value of the text data as spam, and the LSTM model is trained and optimized by using the log-likelihood function to minimize the loss function [22].

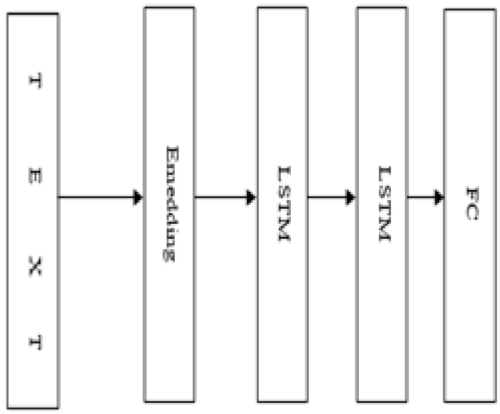

Fig.2. LSTM model framework.

For the hyper parameters of the LSTM model, we use the grid search optimization algorithm to select the optimal values for the five hyper parameters, which are learning rate, batch size, epochs, dropout rate and optimization algorithm. The range and optimal values of these hyper parameters selected by the LSTM model are shown in Table 1.

Table I: The range and optimal values of hyper parameters for LSTM.

Hyperparameter	Range	Optimal Value
learning rate	[0.001, 0.01, 0.1, 0.2]	0.001
batch size	[8, 16, 32]	32
epochs	[10, 20, 30]	30
dropout rate	[0.2, 0.3, 0.4]	0.3
optimization algorithm	[SGD [23], RMSprop [24], Adam [25]]	Adam

WemakeabriefpseudocodedescriptionherefortheLSTMmodel.Foradetailedalgorithm about the LSTM unit, please see the literature [10,26].Let T denote the text data of an email. Input T into the embedding step to convert T into becoming award vector x, x=(x1,x2,•••,xl),where xi∈Rnis then-dimensional word vectors for the i th word in the document T and matrix x∈R lxn denote the document T,where l is the max length of and l≤500.At time-step t, the memory ct and the hidden state ht are updated with the following equations:

$$\begin{bmatrix} i_t \\ f_t \\ o_t \\ \hat{c}_t \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ tanh \end{bmatrix} \left| W \cdot [h_{t-1}, x_t], \right. \quad (1)$$

$$c_t = f_t \odot c_{(t-1)} + i_t \odot \hat{(C_t)} \quad (2)$$

$$h_t = 0_t \odot tanh(c_{(t)}) \quad (3)$$

Where xt is the input at the current time-step, i, f and o is the input gate activation, forget gate activation and output gate activation, respectively, cˆt is the current cell state, σ denotes the logistic sigmoid function and Ⓢ denotes element-wise multiplication. Through training and optimizing the LSTM model, we could obtain the classification probability value of the text part as spam. The entire process of text spam classification algorithm is described in Algorithm 1.

2. Algorithm 1 Text Spam Classification Algorithm.

Input: Text Document T

Output: Text spam classification probability value e

1. Input T into the word2vec toolkit to get the word vector x, x=(x1,x2, ,xl).
2. For the first LSTM layer (64 LSTM units), input x at time t and complete the following calculations:

$$\begin{bmatrix} i_t \\ f_t \\ o_t \\ \hat{c}_t \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ tanh \end{bmatrix} \left| W \cdot [h_{t-1}, x_t], \right.$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \widehat{C}_t$$
$$h_t = 0_t \odot tanh(c_{t)}$$

3. By the first LSTM layer, getting the text feature vector $h=(h_1,h_2, \dots$
4. For the second LSTM layer(32 LSTM units), input *h* at time t and do the same as Equations (1)–(3).
5. Finally, getting more abstract text feature vector *k*, $k=(k_1,k_2, \dots$
6. Input *k* to FC layer and using Softmax activation function to gain the text classification probability value *e*;
7. return ;

The sequences of input (sentences) are fed into the LSTM unit along with the output of the previous LSTM unit. This is repeated with each input sentence and in this way the LSTM units keep on saving the important features.The number of LSTM units save the most important features.Hence, through the LSTM layer, FC layer and Soft max activation function, we can gain the classification probability value *e* of the text part as spam.

3. Fusion Model

The structure of the fusion model is shown in Figure 3. The aim is to fuse the classification probability value of an email text part with the classification probability value of the same email text part to obtain the most accurate classification probability value of the email as spam. The overall steps are as follows:

- Combining the two classification probability values of the LSTM and CNN models to get a feature vector q, q∈R1×4;
- In putting q into the FC layer with 64 neurons to get a comprehensive feature vector;
- Inputting the comprehensive feature vector to the logistic layer, which includes two neurons and chooses the logistic regression function as the activation function to get the most accurate classification probability value of the email as spam. Taking into account the efficiency of our machine, we only use the grid search optimization algorithm to select the optimal values for the four hyper parameters, which are learning rate, batch size, epochs and optimization algorithm, the

best hyper parameter for learning rate is equal to 0.01,batch size is equal to 16,epochs is equal to 30and the optimization algorithm is the SGD algorithm.

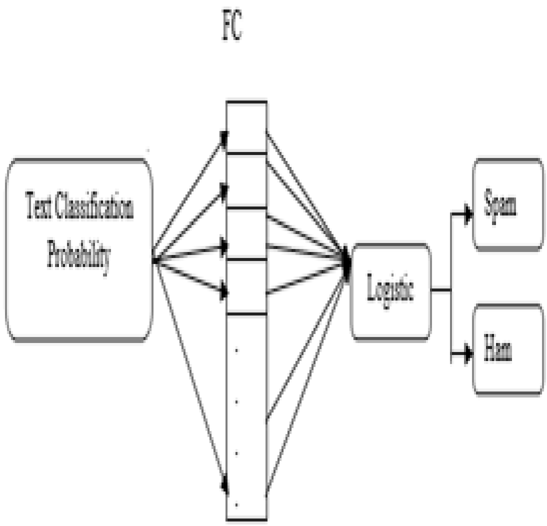

Fig.3. Fusion model structure.

Suppose that the classification probability dataset input to the fusion model is $D=\{(q_1,y_1),(q_2,y_2),\cdots,(q_v,y_v)\}$, $q_i \in R^{1\times 4}$, $y_i \in \{0,1\}$, in which the conditional probability distribution of the logistic regression function is as follows:

$$P(Y=1|q)=\pi(q)\frac{e^{-w^T.q}}{1+e^{-w^T.q}} \quad (4)$$

$$P(Y=0|q)=1-\pi(q)\frac{1}{1+e^{-w^T.q}} \quad (5)$$

We choose the log-likelihood function as the loss function, and the formula is as follows:

$$L(w)=\sum_{i=1}^{v}[y_i \log \pi(q_i)+(1-y_i)\log(1-\pi(q_i))]$$

$$\sum_{i=1}^{v}\left[y_i \, log \frac{\pi(q_i)}{1-\pi(q_i)} \log(1-\pi(q_i))\right]$$

$$\sum_{i=1}^{v}[y_i(w\cdot q_i)-\log(1+e^{(w\cdot q_i)})] \quad (6)$$

The maximum value of $L(w)$is obtained by the Adam algorithm. In addition, the optimal estimate value of the parameter w can be obtained by optimizing $L(w)$. If p >0.5, it means that the email is spam; otherwise, it is a normal email.

III. EVALUATION METRICS AND VALIDATION SCHEME

1. Evaluation Metrics

In order to assess the effectiveness of the proposed method, different evaluation indicators have been used, including accuracy, recall, precision and f1-score, which are defined as follows:

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN}, \quad (7)$$

$$Recall=\frac{TP}{TP+FN}, \quad (8)$$

$$Precision=\frac{TP}{tp+FP}, \quad (9)$$

$$F1-Score=\frac{2*(Precision\ *Recall\ )}{precision\ +Recall} \quad (10)$$

The specific meanings of FP, FN, TP and TN are defined as follows:

- False Positive (FP): The number of legitimate emails (Ham) that are misclassified;
- False Negative (FN): The number of misclassified spam;
- True Positive (TP): The number of spam that are correctly classified;
- True Negative (TN): The number of legitimate emails (Ham) that are correctly classified.

For spam detection, the evaluation metrics about accuracy, recall, precision and f1-score are mainly based on the confusion matrix, which shows in Table 3:

Table II: Confusion matrix.

Prediction	Actual	
	Spam	Ham
Spam	TP	FN
Ham	FP	TN

2. Validation Scheme

In previous studies, a rejection verification scheme has been employed to evaluate the effectiveness of the built spam filtering system. Different studies use different training-test split percentages for data distribution, in which the training dataset is used to evaluate the performance of a model; the testing data set is used to obtain the accuracy of the selected optimal model. The easiest and most straight forward way is to divide the data set into two parts, one for training and the other for testing, which is called the hold out method. The short coming is that the evaluation depends largely on which samples end up in which collection. Another way to reduce the variance of the hold out method is the k-fold cross-validation method, in the k-fold cross-validation method, the data set M is divided into k mutuallyexclusiveparts,and$M_1,M_2,\cdots,M_k$.The inducer is trained on M_i/M and tested againstM_i. This is repeated k times with different i,i=1,2,$\cdots$,k. For a k-fold test, the accuracy, recall, precision and f1-score are defined as follows:

$$Accuracy=\sum_{i=1}^{n}Accuracy_i \quad (11)$$

$$Recall=\sum_{i=1}^{k}Recall_i \quad (12)$$

$$Precision=\sum_{i=1}^{k}Precision_i \quad (13)$$

$$F1-Score=\sum_{i=1}^{k}F1-Score_i \quad (14)$$

International Journal of Scientific Research & Engineering Trends
Volume 5, Issue5, Sep-Oct-2019, ISSN (Online): 2395-566X

where Accuracy, Recall, Precision and F1 − Score are the accuracy, recall, precision and f1-score for each of the k tests. Considering the performance of our computer, we choosea5-foldcross-validation method throughout the experiments.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

1.Corpus

In this paper, we choose three types of email datasets for our experiments: the dataset only contained text, the dataset only contained text and the dataset that contains text data. The dataset only containing text comes from the Indian corpus [29], and we only choose 6000 text emails (4500 Spam, 1500 Ham) by removing duplicates and randomly selecting from 33,645 text emails. The dataset only containing text is composed of Personal text Ham, dataset 1. The dataset details used in the experiments are shown in Table 4below.

Table III: Datasets used in Experiments.

Type	Original Dataset	Before Remove Duplicates	After Remove Duplicates
Text	Indian Ham	17,108	1500
	Indian Spam	16,537	4500

For the mixed dataset 1, the number of text dataset, which contains 600 Spam (text Spam 600) and 600 Ham (text Ham 600 and tex Ham 600 are formed into 600 Ham email).

Table V: Training and Testing Dataset Size.

Type	Training Dataset Size	Testing Dataset Size
Text Dataset 1	5000	1000
Text Dataset 2	960	240

V. RESULTS AND DISCUSSION

In this section, we show our evaluation results on text spam classification, text spam classification and the mixed spam classification. Moreover, we give some analysis and discussions for the experimental results. We use 5-fold cross-validation method to verify the performance of the MMA-MF model on the text dataset, and the mixed datasets1, and obtain the experimental results of the MMA-MF model on the four datasets, as shown in Table 6, in which u means the average value of Accuracy, Recall, F1-Score or Precision after using the 5-fold cross-validation method.

Table VI: Experimental results in 5-fold cross-validation for the MMA-MF model.

Fold	Accuracy	Recall	F1-Score	Precision
MMA-MF Model for Text Dataset 1				
1	98.42	97.84	97.24	98.5
2	98.67	98.15	97.47	98.5
3	98.67	98.19	97.65	99
4	98.25	97.71	97.27	98
5	98.42	97.89	97.53	98.5
MMA-MF Model for Text Dataset 2				
1	93.35	92.64	92.89	90.5
2	92.56	92.63	92.75	90.01
3	91.5	92.33	91.83	93.5
4	92.35	92.83	92.97	92
5	93.44	92.72	92.71	92.5

Fig.4. Fold Cross-Validation Chart for Text Dataset 1.

From Table 6, we can conclude that the MMA-MF model designed in this paper implements the filtering function of spam, whether it is hidden in the text or hidden in the text we are all able to handle it and filter it out pretty well. In conclusion, we have the following observations: for the MMA-MF model, it not only filters well mixed emails, but also filters text emails

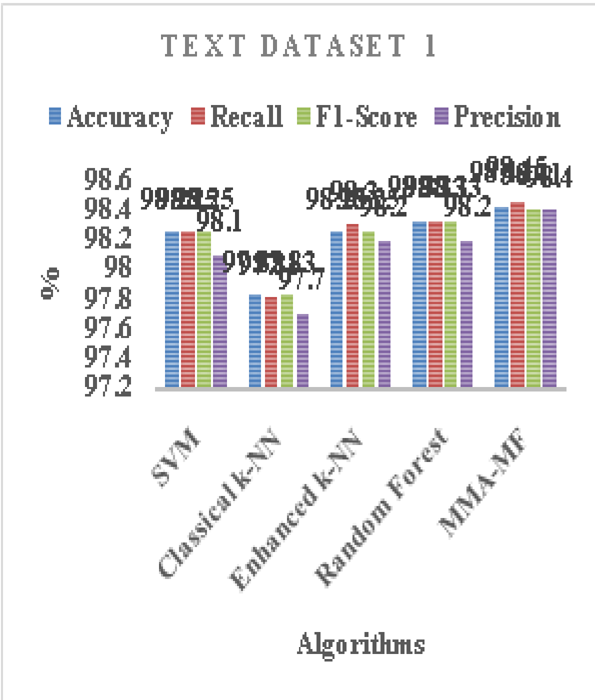

Fig.5. Fold Cross-Validation Chart for Text Dataset 1.

VI. CONCLUSION

We mainly introduce the multi-modal fusion architecture based on model fusion, which we called MMF-MF. The model combines the Convolution Neural Network (CNN), Long Short-Term Memory (LSTM) network and fuses the two models by the logistic regression method to implement spam detection in a variety of email formats to improve spam detection rate. The advantage of the model is that it cannot only filter hybrid spam, but also filter spam with only text data, while other models can only handle text-based spam.

However, we have two issues that need to be solved in the future work.(1)From Table 5,there is no imbalance in our experimental data set. However, in practical applications, spam detection datasets have a large discrepancy between the number of spam emails and non-spam emails. The solutions like one-class classification, few-shot learning and generative adversarial network methods should be proposed to solve the imbalance between the positive and negative samples in the training dataset;(2)Owing to the fact that there is no real mixed email dataset for public use, the mixed email data set is collected by splicing.

In the future, we hope to use the new technique just like the one-class classification method and a few-shot learning method to solve the problem of discrepancy between the number of spam emails and non-spam emails, and we will continue to collect more realistic mixed email data sets to improve the network structure of our model so that the model can get better spam detection performance.

REFERENCES

[1]. Seth, S.; Biswas, S. Multimodal Spam Classification Using Deep Learning Techniques. In Proceedings of the 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Jaipur, India, 4–7 December 2017; pp.346–349.
[2]. Bettencourt,J.Kaspersky Lab Spam and Phishing Report: FIFA 2018 and Bit coin among 2017’s Most Luring Topics. Available online: https://usa.kaspersky.com/about/press-releases/2018_fifa-2018-and-bitcoin- among-2017-most-luring-topics (accessed on 15 February2018).
[3]. Carreras, X.; Marquez, L. Boosting trees for anti-spam email filtering. arXiv 2001, arXiv:cs/0109015.
[4]. Androutsopoulos, I.; Paliouras, G.; Michelakis, E. Learning to Filter Unsolicited Commercial E-Mail; DEMOKRITOS; National Center for Scientific Research: Paris, French, 2014.
[5]. Sahami, M.; Dumais, S.; Heckerman, D.; Horvitz, E. A Bayesian approach to filtering junk e-mail. In Learning for Text Categorization: Papers from the 1998 Workshop; AAAI Technical Report WS-98-05; Monona Terrace Convention Center: Madison, WI, USA, 1998; Volume 62, pp. 98–105.
[6]. Anayat, S.; Ali, A.; Ahmad, H.F. Using a probable weight based Bayesian approach for spam filtering. In Proceedings of the 8th International Multitopic Conference 2004, Lahore, Pakistan, 24–26 December 2004; pp. 340–345.
[7]. Kim, H.J.; Shrestha, J.; Kim, H.N.; Jo, G.S. User action based adaptive learning with weighted bayesian classification for filtering spam mail. In Australasian Joint Conference on Artificial Intelligence; Springer: Berlin/Heidelberg, Germany, 2006; pp. 790–798.
[8]. Yang, Z.; Nie, X.; Xu, W.; Guo, J. An approach to spam detection by naive Bayes ensemble based on decision induction. In Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA’06), Jinan, China, 16–18 October 2006; Volume 2, pp. 861–866.
[9]. Androutsopoulos, I.; Paliouras, G.; Karkaletsis, V.; Sakkis, G.; Spyropoulos, C.D.; Stamatopoulos, P. Learning to filter spam e-mail: A comparison of a naive bayesian and a memory-based approach. arXiv 2000, arXiv:cs/0009009
[10]. Jain, G.; Sharma, M.; Agarwal, B. Optimizing semantic LSTM for spam detection. Int. J. Inf. Technol. 2019, 11, 239–250. [CrossRef]
[11]. Abi-Haidar, A.; Rocha, L.M. Adaptive spam detection inspired by a cross-regulation model of immune dynamics: A study of concept drift. In International Conference on Artificial Immune Systems; Springer: Berlin/Heidelberg, Germany, 2008; pp. 36–47.
[12]. Shang, E.X.; Zhang, H.G. Image spam classification based on con volutional neural network. In Proceedings of the 2016 International Conference on Machine Learning and Cybernetics (ICMLC), Jeju, South Korea, 10–13 July 2016; Volume 1, pp. 398–403.
[13]. Wang, Z.; Josephson, W.K.; Lv, Q.; Charikar, M.; Li, K. Filtering Image Spam with Near-Duplicate Detection; CEAS: Mountain View, CA, USA, 2007.
[14]. Kumar, P.; Biswas, M. SVM with Gaussian kernel-based image spam detection on textual features. In Proceedings of the 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, India, 9–10 February 2017; pp. 1–6.
[15]. Xu, C.; Chiew, K.; Chen, Y.; Liu, J. Fusion of text and image features: A new approach to image spam filtering. In Practical Applications of Intelligent Systems; Springer: Berlin/Heidelberg, Germany, 2011; pp. 129–140.

[16]. Huamin, F.; Xinghua, Y.; Biao, L.; Chao, J. A spam filtering method based on multi-modal features fusion. In Proceedings of the 2011 Seventh International Conference on Computational Intelligence and Security, Hainan, China, 3–4 December 2011; pp. 421–426.
[17]. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems; Curran Associates, Inc.: Red Hook, NY, USA, 2012; pp. 1097–1105.
[18]. Graves, A. Long short-term memory. In Supervised Sequence Labelling with Recurrent Neural Networks; Studies in Computational Intelligence; Springer: Berlin/Heidelberg, Germany, 2012; pp. 37–45.
[19]. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 2014, 15, 1929–1958.
[20]. Sanville, E.; Kenny, S.D.; Smith, R.; Henkelman, G. Improved grid-based algorithm for Bader charge allocation. J. Comput. Chem. 2007, 28, 899–908. [CrossRef] [PubMed]
[21]. Wiens, T.S.; Dale, B.C.; Boyce, M.S.; Kershaw, G.P. Three way k-fold cross-validation of resource selection functions. Ecol. Model. 2008, 212, 244–255. [CrossRef]
[22]. Pinheiro, J.C.; Bates, D.M. Approximations to the log-likelihood function in the nonlinear mixed-effects model. J. Comput. Graph. Stat. 1995, 4, 12–35.
[23]. Bottou, L. Large-scale machine learning with stochastic gradient descent. In Proceedings of COMPSTAT'2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 177–186.
[24]. Tieleman, T.; Hinton, G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. Coursera Neural Netw. Mach. Learn. 2012, 4, 26–31.
[25]. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. arXiv 2014, arXiv:1412.6980.
[26]. Zhou, C.; Sun, C.; Liu, Z.; Lau, F. A C-LSTMneural network for text classification. Ar Xiv 2015, arXiv:1511.08630.
[27]. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv 2015, arXiv:1502.03167.
[28]. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv 2015, arXiv:1502.03167.
[29]. Klimt, B.; Yang, Y. The Indian corpus: A new dataset for email classification research. In European Conference on Machine Learning; Springer: Berlin/Heidelberg, Germany, 2004; pp. 217–226.
[30]. Dredze, M.; Gevaryahu, R.; Elias-Bachrach, A. Learning Fast Classifiers for Image Spam; CEAS: Mountain View, CA, USA, 2007; p. 2007-487.
[31]. Gao, J.; Lanchantin, J.; Soffa, M.L.; Qi, Y. Black-box Generation of Adversarial Text Sequences to Evade Deep Learning Classifiers. arXiv 2018, arXiv:1801.04354
[32]. Nguyen, B.P.; Tay, W.L.; Chui, C.K. Robust biometric recognition from palm depth images for gloved hands. IEEE Trans. Hum. Mach. Syst. 2015, 45, 799–804. [CrossRef]