2021 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/IOP/SCI)

Email Classification and Forensics Analysis using Machine Learning

Maryam Hina*, Mohsan Ali†, Abdul Rehman Javed‡, Gautam Srivastava§,
Thippa Reddy Gadekallu¶, Zunera Jalil‡
* Department of Computer Science, Air University, Islamabad, Pakistan
† National Center for Cyber Security, Air University, Islamabad, Pakistan
‡ Department of Cyber Security, Air University, Islamabad, Pakistan
§ Dept. of Math and Computer Science, Brandon University, Brandon R7A 6A9, Canada
§ School of Information Technology and Engineering, Vellore Institute of Technology, Tamil Nadu, India

Email: {180160@students.au.edu.pk, mohsan.ali@mail.au.edu.pk, abdulrehman.cs@au.edu.pk,
srivastavag@brandonu.ca, thippareddy.g@vit.ac.in, zunera.jalil@mail.au.edu.pk }

***Abstract*—Emails are being used as a reliable, secure, and formal mode of communication for a long time. With fast and secure communication technologies, reliance on Email has increased as well. The massive increase in email data has led to a big challenge in managing emails. Emails so far can be classified and grouped based on sender, size, and date. However, there is a need to detect and classify emails based on the contents contained therein. Several approaches have been used in the past for content-based classification of emails as Spam or Non-Spam Email. In this paper, we propose a multi-label email classification approach to organize emails. An efficient classification method has been proposed for forensic investigations of massive email data (e.g., a disk image of an email server). This method would help the investigator in Email related crimes investigations. A comparative study of machine learning algorithms identified Logistic Regression as a method that achieves the highest accuracy compared to Naive Bayes, Stochastic Gradient Descent, Random Forest, and Support Vector Machine. Experiments conducted on benchmark data sets depicted that logistic Regression performs best, with an accuracy of 91.9% with bi-gram features.**

***Index Terms*—Digital Forensics, Machine Learning, Email Forensics, Fraud Detection, Crime Investigation**

I. Introduction

With the advent of broadband internet, the complexity of cybersecurity problems has increased, which makes identifying, analyzing, and controlling the relevant risk events significantly challenging [1]–[3]. Electronic mail (Email) is a critical communication source and declared as the 2nd most used application on the Internet [1]. According to the statistics in 2019, there are 3.9 billion active users of Email, and 246.5 billion emails had been sent in 2019. Social media is popular amongst people, but social media users are still less in number than email account holders. With the increase in the volume of data, attackers are exploiting user identities and misusing this data. Electronic crimes by cybercriminals affect the security chain by phishing, spoofing, spamming, threatening, harassment, and terrorism [4]–[8].

Many organizations have faced great financial losses due to crimes committed through emails. Enron, an American company, was bankrupted due to the leakage of its confidential emails. According to a study before the 9/11 attack, most of the communication took place by emails [9]. Cybercriminals take advantage of emails to steal confidential information [10]. Email protocols are not effective in capturing attacks in a real-time environment. Different mechanisms like blocklisting and filtering were used for email authentication, but these techniques have certain limitations and are ineffective for forensic investigations since forensic investigators always have a specific crime in mind before initiating an investigation. Digital Forensics is defined as the process of preservation, identification, extraction, and documentation of digital evidence [11]. Investigators have used many forensics solutions in the past for analysis of cybercrimes involving email-related crimes such as Add4Mail, MailXaminer, Email trackerpro, etc.

Digital forensics frameworks are used for header and body analysis of emails to collect evidence, but these tools do not perform any processing based on email contents [12]. In addition, these tools are used for post-crime forensic investigation [13]. Government agencies currently use the NLP to detect the news published on social media platforms or websites. Rumor detection is also an essential application of the NLP for the best of the internet world. Text summarising, text generation, topic clustering, and topic modeling, image captioning, and video captioning are the essential fields where NLP-based solutions are helpful [14].

There is a need for a system to analyze all emails going through the server to identify and detect damaging emails using AI (Artificial Intelligence) algorithms. This can be used as a proactive method of defense against email crimes. Machine learning is used in different fields of computer sciences to resolve issues. Machine learning algorithms have been used for classifying emails into Spam and legitimate emails. With the increase in cybercrimes, a system to make more detailed classification is needed. In this work, we have proposed a system that can classify emails into four different classes. Five different machine learning algorithms have been used in this research work.

[1] https://www.statista.com/statistics/420391/spam-email-traffic-share/

978-1-6654-1236-0/21/$31.00 ©2021 IEEE
DOI 10.1109/SWC50871.2021.00093

This paper makes the following contributions:

- Propose an efficient approach to detect and classify emails as Normal, Harassing, Fraudulent, and Suspicious Email.
- Present a comparative analysis of different machine learning algorithms for this problem.
- Email forensics with four classes and unique combination of TF-IDF features makes this study unique as existing works classify Email into spam and non-spam email
- Evaluate the effectiveness of the proposed approach using different machine learning classifiers and achieve promising accuracy.

The remaining paper is organized as follows: Section II discusses the related research on email classification. Section III explains the proposed email classification approach and discusses its working. In Section IV, we present and discuss experimental results. Finally, Section V summarizes the findings of this work.

II. Prior and related work

In this section, existing work related to email classification techniques and methodologies is presented. [15] proposed a fraudulent email detection model by employing advanced features. The authors used the fraudulent email data having 2500 emails and normal data having 3000 emails and applied a cluster-based classification model to classify emails as normal or fraudulent. They concluded from the experimentation that the selection of feature sets is more important than the classification algorithm to improve accuracy.

Authors in [16] presented a content-based phishing detection approach. The authors used Random forest for the classification of Phishing emails. They classified phishing and ham emails. They extracted relevant features and improved phishing email classifiers with better prediction accuracy. The accuracy can be further improved by optimal kernel parameter selection. [17] adopted a manual approach for Email analysis. The authors spotted spoofed messages sent by SMTP, decoded them, analyzed IP addresses, traced their locations, and made a timeline of all the events. They also checked server logs to ensure the activities mentioned in the timeline, but it is a long and tiring procedure and needed a large quantity of email data for meaningful analysis.

Authors in [18] presented email classification as Spam by using the Fuzzy C means algorithm. The authors implemented a membership threshold value to detect Spam. [19] proposed a hybrid system for email classification. The proposed solution consists of preprocessing, feature extraction & email classification. The authors correctly differentiated spam emails from homogeneous work-related emails. [20] presented a method to classify an email as fraudulent and ham. Fraudulent and Normal Email Dataset is used in this work for email classification. Machine learning techniques are used o classify emails. The accuracy is improved, and misclassification is reduced. The authors in [21] and [22] presented multi-label email classification using clustering techniques. They distributed work into two phases, the first one is generating taxonomic, and the second one is labeling text in different categories. They performed topic modeling using the Latent Dirichlet Algorithm and generated categorical terms. Finally, they mapped emails with their categories. Similarly, the authors in [23], and [24] presented works on spam classification by extracting features from the email body and word and document length. They performed experimentation on four email datasets, SpamAssassin, Enron-Spam, Lingspam, and CSDMC2010. They experimented using different classifiers like NB, RF, SVM, Bagging, and Boosting.

III. Proposed Methodology

The proposed methodology is explained with a block diagram shown in Fig. 1. The methodology starts with the email data collected from the Enron email dataset containing four classes in the dataset. The class distribution of the dataset is shown in Fig. 2. The dataset was imbalanced, but we make the manual data balance to ensure the model training's unbiased decisions. The dataset contains four classes such as fraudulent, harassment, normal and suspicious. The original dataset was based on three classes; we added one more class to classify all the emails into one category.

A. Pre-Processing

The dataset is preprocessed based on the steps mentioned in the block diagram Fig. 1 such as lower case conversion of text, removing unwanted keywords, Reply keyword, forward keyword, links, and tags from the email body. The more the data is cleaned, the more accurately the model will learn the patterns. The data preprocessing is sometimes required to transform the data for the required underlying model input. We apply noise removal techniques at the preprocessing stage. For instance, the stop words are just rules for understanding the sentence's semantics, but they do not contribute to classification problems. The most repeating words, called the stopwords, are removed from the dataset, such as "are," "am," "i," and "we.".

The sender, receiver, BCC, CC, body tags are not helpful to evaluate the email contents. The removal of this information is also part of this study. For instance, if we have many emails in the dataset, each Email has header information and sender-receiver information. A sparse data structure is used to store all these chunks, which is inefficient for large-scale processing. So, removing headers is also mandatory as it does not contribute to the classification of emails. The lower case conversion is helpful, for example, if two words are in the dataset and both of them are the same such as "email" and "EMAIL". The text to real-valued vector will consider these entries differently. So, lower case conversion will help in saving the sparse of vectors by lowercasing the words.

The two more steps in the preprocessing are stemming and lemmatization. They both normalize the words to their base form, but they are different with respect to their normalization process. For example, the words "are", "am" "is" are lemmatized towards the word "be," which is different from the original words. The words "learning", "learnt", and

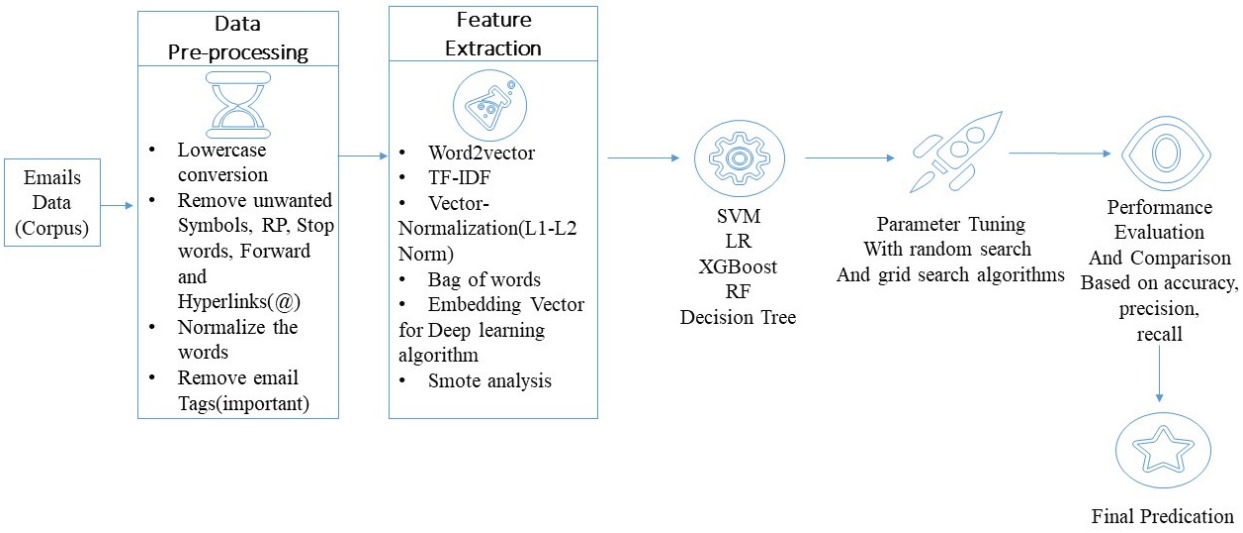

Fig. 1. Illustration of Email Classification Block diagram

"learner" stemmed towards the "learn" which is the origin of individual words in this vocabulary. This step is performed using the natural language processing libraries such as spacy and NLTK. The removal of the specific keyword is performed using the regular expression technique. Regular expression performs pattern matching to remove the unwanted words from the corpus.

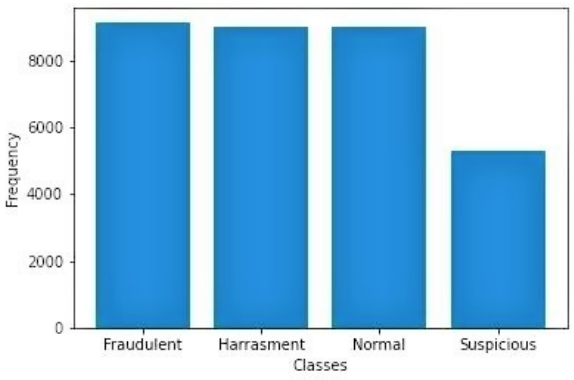

Fig. 2. Class-wise Distribution of Dataset

B. Feature Extraction

The features or characteristics distinguish normal emails from other types of harmful emails. In terms of textual data, the features from the Email body will decide that either an email is normal or harmful for the end-user. There are numerous methods used in NLP for feature extraction with respect to traditional or modern learning. In this study, the bag of words features extraction is used. This technique uses sparse vectors for the representation of the words as numerical vectors. Each word has the same importance as the balls have the same importance in a bag, due to which this method is called the bag of words. The bag of words feature extraction starts with the tokenization of the samples of the textual dataset. Each term in the dataset is assigned a unique number value which will not be assigned to any other term in that experiment. There are predefined priorities to tokenize the sentence. For instance, first, break the sentence with respect to the spaces found in the sentence, follow the quotation marks. The tokenization is used as the input for further processing, such as calculating the TF-IDF vectors.

Secondly, we extract the features of the email dataset using the natural language processing features extraction technique based on sparse vector representations such as TF-IDF, bag-of-words, and N-grams. The mathematical equations related to extracting the TF-IDF features are written in equation 1. The TF explains the words part of the vocabulary, but IDF increases the word's importance, which is less repetitive in the dataset. The denominator of Equation 1 highlights the lower-ranked words in the vocabulary. Lastly, the feature extractions phase output will be the feature vector that contains the tokenized corpus, TF-IDF calculations used to train the machine learning algorithms.

$$TF - IDF = \text{TermFrequency} \times \left(\frac{1}{\text{DocumentFrequency}} \right) \quad (1)$$

$$TF{-}IDF = \text{TermFrequency} \times \text{InverseDocumentFrequency} \quad (2)$$

$$TF - IDF(t, d, D) = TF(t, d) \cdot IDF(t, D) \quad (3)$$

The size of the dataset size is 32427 instances. The extracted features are split into a 75-25% ratio in the training and testing

dataset. The model is trained on the corpus of the data. The trained model learned the data patterns based on the features of the individual class. The results achieved are evaluated using the testing dataset, the twenty percent of the complete data. We tried the shuffling technique to shuffle the data samples to reduce the biases in the data sequence.

C. Machine Learning Models

Sequential and non-sequential machine learning models are part of the literature. Models with the greatest entropy are hidden Markov models. A decision tree, SVM, Naive Bayes, and logistic regression are not sequential models. This study employed non-sequential models, including logistic regression, random forest, SVM, naive Bayes, and stochastic gradient descent.

The retrieved textual characteristics are now numerical vectors. It will be used to train supervised learning models. The broadest machine learning model will perform best on unseen data. Model training is critical since it might lead to over-fitting or under-fitting issues. Many machine learning algorithms utilize the hit-and-try approaches to find the optimum intelligent and generalized trained model. In the realm of machine learning, there are no complicated rules that dictate which model performs better. We utilized ML models in this investigation. [25].

The suggested technique assesses the trained models using accuracy, precision, recall, and F-score. It is based on learning score, scalability, and performance over training samples. Due to its simplicity, Logistic Regression performed the best. Stochastic Gradient Descent(SGD) has the lowest accuracy because of a lack of a pattern-finding log. The operating system, CPU, RAM, GPU, and other language technologies used in the creation of the email categorization system are shown in Table I.

TABLE I
COMPUTING ENVIRONMENT

Parameter	Value
Operating System	Ubuntu 18.04.2 LTS
CPU	Xeon E5/Corei5
RAM	128GB
GPU	NVIDIA GeFroce 1080
Python Version	3.7

IV. PERFORMANCE ANALYSIS

The dataset used for this study comprises four classes, namely Normal Class having 9000 emails, Fraudulent Class having 9001 emails, 9138 Harassing, and 5287 Suspicious emails. This data is taken from well-known dataset repositories, Kaggle and ACL Repository. The following datasets are used in this study. Enron [2], Normal and Fraudulent [26], Hillary Clinton Email Dataset and Hate speech and offensive data [3].

[2]https://www.kaggle.com/wcukierski/enron-email-dataset
[3]https://www.kaggle.com/kaggle/hillary-clinton-emails

We used the accuracy, confusion matrix, precision, recall, and ROC curve for the model evaluation. The general Eq. 4 for the accuracy is given below:

$$\text{Accuracy} = \frac{PredictedCorrectly}{TotalPredictions} \tag{4}$$

The Eq. 5 given below is used for the accuracy calculation because we have the confusion matrix entries.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

The proposed methodology explained in Section III is implemented to get results from the machine learning algorithms. Firstly, we got the dataset words frequency representation as shown in Fig. 3. The word cloud contains all the most frequent words in the dataset. The bigger the font size of a word in the word cloud, the more frequent it is. The different words have different colors.

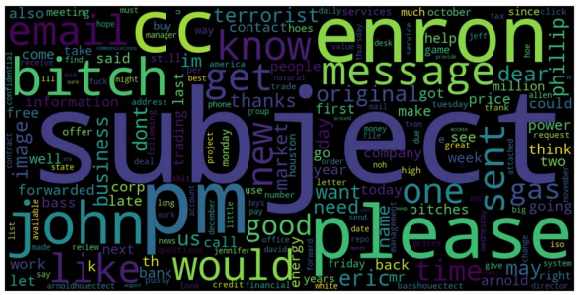

Fig. 3. Wordcloud Representation of dataset

The dataset is split into 75%-25% training and testing ratios, respectively. The model is trained with UnigramUnigram, UnigramUnigram plus bigram, and unigram plus bi gram plus trigram features with L2 normalization. The highest accuracy is achieved using the logistic regression model, which is almost 92% in uni and bigram features with L2 normalization. The results achieved by following the methodology are depicted in Table II.

The vertical chart represents the highest accuracies achieved by individual machine learning algorithms in Fig. 4.

There are two types of data with respect to the available samples during the model's training. one is the balanced dataset, and the other is the imbalance dataset. The accuracy measure is not suitable for the imbalanced datasets, although we have an almost balanced dataset for the Email forensic. In imbalanced datasets, we move towards the other options such as precision, recall, and ROC curve.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

One more parameter which is the harmonic mean of the recall (see Eq. 7) and precision (see Eq. 6) is F-score is calculated as the Eq. 8 given below:

$$F = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

TABLE II
COMPARISON OF TUNED MODELS WITH DIFFERENT N-GRAM FEATURES AND TF-IDF VALUES

N-gram with TF-IDF Norm	Accuracy LR	SVM	SGD	NB	RF
Unigram with L2 norm	0.9191	0.9001	0.8715	0.9045	0.9054
Unigram+bigram with L2 norm	0.9191	0.8990	0.8763	0.9050	0.9050
Unigram+bigram + trigram with L2 norm	0.9179	0.9000	0.8746	0.8950	0.9035

The precision, recall, F-score of the model are depicted in Table III. The F-score of the model is perfect, which is 96% for the fraudulent class. As depicted in Fig. 4, the highest accuracy is almost 92% (0.9191) achieved by logistic regression model. As the email dataset is textual, and length is more than 1000 words in some instances, which is a long sequence to deal with this, the model accuracy is a little bit lower than 100%.

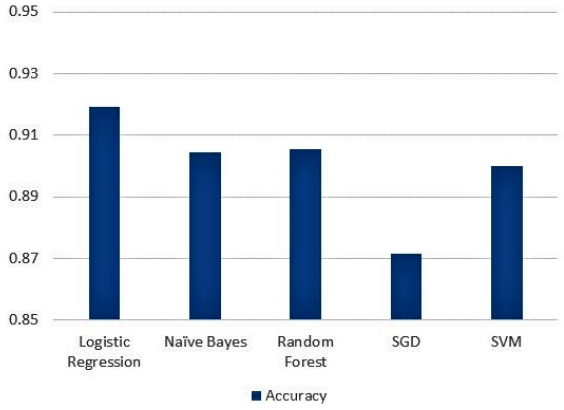

Fig. 4. Accuracy Comparison of Machine Learning Models

TABLE III
FINAL TUNED LOGISTIC REGRESSION (LR) MODEL SCORES ON TEST DATA

	Precision	Recall	F-Score
Normal	0.87	0.99	0.93
Harassment	0.90	0.96	0.93
Fraudulent	0.98	0.95	0.96
Suspicious	0.96	0.68	0.80

Fig. 5 depicts the confusion matrix over the four email classification classes in a heat map-based confusion matrix. The vertical line with number labeling shows the frequency of samples in the matrix. The greater the values in the matrix's diagonal, the accuracy will be high and vice versa. The performance evaluation metric is calculated using this multi-class confusion matrix such as precision-recall and F-score.

The parameters used to get the more efficient email classification models are represented in Table IV. The logistic Regression achieved the best results, 92%, with the parameters of C, mixiter, and njobs. The other parameters that we tuned during the training for models were also mentioned. The best

Fig. 5. Confusion Matrix of Logistic Regression

parameters are searched using the grid-search technique over these features mentioned in the parameter tuning table with 10-fold cross-validation.

TABLE IV
PARAMETER TUNING OF THE MACHINE LEARNING MODELS

Algorithms	Parameters
LR	C=[0.1, 0.001, 1],maxiter= 100, njobs=-1
SVM	C=[0.1, 0.001, 1, 100, 1000]
SGD	gamma=0, learning_rate=0.1
NB	alpha=[0.01, 0.001, 0.00000001]
RF	min_samples_leaf=[5,4], n_estimators=[100,150]

There are different methods to evaluate the supervised learning problem. Fig. 6 depicts the logistic regression receiver operating characteristics (ROC) curve used to calculate each class's independence in the dataset. When the dataset is imbalanced, the receiver operating characteristic curve is one of the evaluations that come into the picture. The ROC is calculated with the help of different probability thresholds. The true positive rate, true negative rate, false-positive rate, and false-negative rate calculations plot a ROC curve.

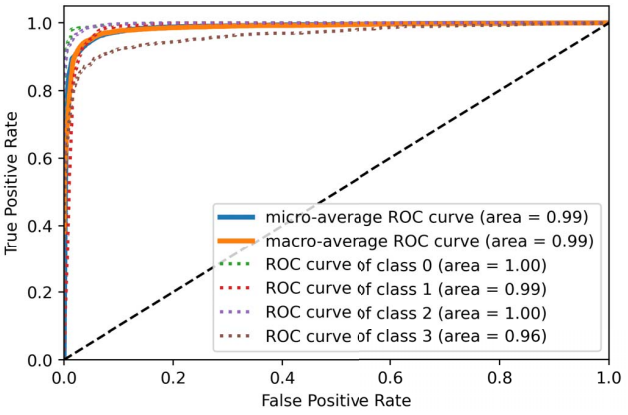

Fig. 6. Receiver Operating Characteristic Curve for Logistic Regression

The values of the TPR and TNR near to one are best for a good model. The area under the curve in case ROC will

be in between 0 and 1. The AUC near to one is the sign of an intelligent model. In this case, if AUC is equal to 0.5, the trained model is inferior. Class 0 means "Fraudulent," class 1 means "Harassment," class 2 means "Normal", and class 3 means "Suspicious" in the ROC curve graph. The ROC curve plotting depends on the true and false prediction rate of the trained model. The more the area under the curve, means more the classes are distinguishable than the other classes. A false-positive rate is mentioned on the x-axis, and on the y-axis, a true positive rate is mentioned.

V. Conclusion

We presented a study about multi-label email classification that is helpful for investigative analysis. We run different machine learning algorithms for this purpose. We obtained a promising Linear Regression accuracy on the Email data set by best parameter tuning from the study and experimental results. A comparative study of machine learning algorithms identified Logistic Regression as a method that achieves the highest accuracy compared to Naive Bayes, Stochastic Gradient Descent, Random Forest, and Support Vector Machine. Experiments conducted on benchmark data sets depicted that logistic Regression performs best, with an accuracy of 91.9% with bi-gram features. Stochastic Gradient Descent showed the lowest accuracy because of perfect parameter learning. This work leads towards automated forensic investigation that is the need of time and efficiently detects incoming emails. In the future, we intend to incorporate blockchain to store and access forensics analysis data [27]–[29].

References

[1] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, "A comprehensive survey of ai-enabled phishing attacks detection techniques," *Telecommunication Systems*, pp. 1–16, 2020.

[2] C. Iwendi, Z. Jalil, A. R. Javed, T. Reddy, R. Kaluri, G. Srivastava, and O. Jo, "Keysplitwatermark: Zero watermarking algorithm for software protection against cyber-attacks," *IEEE Access*, vol. 8, pp. 72 650–72 660, 2020.

[3] C. Iwendi, S. U. Rehman, A. R. Javed, S. Khan, and G. Srivastava, "Sustainable security for the internet of things using artificial intelligence architectures," *ACM Transactions on Internet Technology (TOIT)*, vol. 21, no. 3, pp. 1–22, 2021.

[4] F. Iqbal, R. Batool, B. C. Fung, S. Aleem, A. Abbasi, and A. R. Javed, "Tweet-to-act: Towards tweet-mining framework for extracting terrorist attack-related information and reporting," *IEEE Access*, 2021.

[5] A. Rehman, S. U. Rehman, M. Khan, M. Alazab, and T. Reddy, "Canintelliids: Detecting in-vehicle intrusion attacks on a controller area network using cnn and attention-based gru," *IEEE Transactions on Network Science and Engineering*, 2021.

[6] A. R. Javed, M. O. Beg, M. Asim, T. Baker, and A. H. Al-Bayatti, "Alphalogger: Detecting motion-based side-channel attack using smartphone keystrokes," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020.

[7] A. R. Javed, S. U. Rehman, M. U. Khan, M. Alazab, and H. U. Khan, "Betalogger: Smartphone sensor-based side-channel attack detection and text inference using language modeling and dense multilayer neural network," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 5, pp. 1–17, 2021.

[8] A. Muhammad, M. Asad, and A. R. Javed, "Robust early stage botnet detection using machine learning," in *2020 International Conference on Cyber Warfare and Security (ICCWS)*. IEEE, 2020, pp. 1–6.

[9] A. S. Aski and N. K. Sourati, "Proposed efficient algorithm to filter spam using machine learning techniques," *Pacific Science Review A: Natural Science and Engineering*, vol. 18, no. 2, pp. 145–149, 2016.

[10] A. Basit, M. Zafar, A. R. Javed, and Z. Jalil, "A novel ensemble machine learning method to detect phishing attack," in *2020 IEEE 23rd International Multitopic Conference (INMIC)*. IEEE, 2020, pp. 1–5.

[11] A. R. Javed and Z. Jalil, "Byte-level object identification for forensic investigation of digital images," in *2020 International Conference on Cyber Warfare and Security (ICCWS)*. IEEE, 2020, pp. 1–4.

[12] W. Ahmed, F. Shahzad, A. R. Javed, F. Iqbal, and L. Ali, "Whatsapp network forensics: Discovering the ip addresses of suspects," in *2021 11th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE, 2021, pp. 1–7.

[13] V. K. Devendran, H. Shahriar, and V. Clincy, "A comparative study of email forensic tools," *Journal of Information Security*, vol. 6, no. 2, p. 111, 2015.

[14] G. G. Chowdhury, "Natural language processing," *Annual review of information science and technology*, vol. 37, no. 1, pp. 51–89, 2003.

[15] S. Nizamani, N. Memon, M. Glasdam, and D. D. Nguyen, "Detection of fraudulent emails by employing advanced feature abundance," *Egyptian Informatics Journal*, vol. 15, no. 3, pp. 169–174, 2014.

[16] S. B. Rathod and T. M. Pattewar, "Content based spam detection in email using bayesian classifier," in *2015 International Conference on Communications and Signal Processing (ICCSP)*. IEEE, 2015, pp. 1257–1261.

[17] R. P. Iyer, P. K. Atrey, G. Varshney, and M. Misra, "Email spoofing detection using volatile memory forensics," in *2017 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 2017, pp. 619–625.

[18] A. K. Singh, S. Bhushan, and S. Vij, "Filtering spam messages and mails using fuzzy c means algorithm," in *2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU)*. IEEE, 2019, pp. 1–5.

[19] T. Ayodele, R. Khusainov, and D. Ndzi, "Email classification and summarization: A machine learning approach," in *2007 IET Conference on Wireless, Mobile and Sensor Networks (CCWMSN07)*. IET, 2007, pp. 805–808.

[20] R. S. H. Ali and N. El Gayar, "Sentiment analysis using unlabeled email data," in *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*. IEEE, 2019, pp. 328–333.

[21] A. Sharaff and N. K. Nagwani, "Ml-ec2: An algorithm for multi-label email classification using clustering," *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)*, vol. 15, no. 2, pp. 19–33, 2020.

[22] M. Hina, M. Ali, A. R. Javed, F. Ghabban, L. A. Khan, and Z. Jalil, "Sefaced: Semantic-based forensic analysis and classification of e-mail data using deep learning," *IEEE Access*, vol. 9, pp. 98 398–98 411, 2021.

[23] R. Shams and R. E. Mercer, "Classifying spam emails using text and readability features," in *2013 IEEE 13th international conference on data mining*. IEEE, 2013, pp. 657–666.

[24] T. Sajid, M. Hassan, M. Ali, and R. Gillani, "Roman urdu multi-class offensive text detection using hybrid features and svm," in *2020 IEEE 23rd International Multitopic Conference (INMIC)*, 2020, pp. 1–5.

[25] S. Gilda, "Notice of violation of ieee publication principles: Evaluating machine learning algorithms for fake news detection," in *2017 IEEE 15th Student Conference on Research and Development (SCOReD)*, 2017, pp. 110–115.

[26] D. Radev, "Clair collection of fraud email, acl data and code repository," *ADCR2008T001*, 2008.

[27] P. Kumar, R. Kumar, G. Srivastava, G. P. Gupta, R. Tripathi, T. R. Gadekallu, and N. Xiong, "Ppsf: A privacy-preserving and secure framework using blockchain-based machine-learning for iot-driven smart cities," *IEEE Transactions on Network Science and Engineering*, 2021.

[28] R. Kumar, R. Tripathi, N. Marchang, G. Srivastava, T. R. Gadekallu, and N. N. Xiong, "A secured distributed detection system based on ipfs and blockchain for industrial image and video data security," *Journal of Parallel and Distributed Computing*, vol. 152, pp. 128–143, 2021.

[29] A. Mubashar, K. Asghar, A. R. Javed, M. Rizwan, G. Srivastava, T. R. Gadekallu, D. Wang, and M. Shabbir, "Storage and proximity management for centralized personal health records using an ipfs-based optimization algorithm," *Journal of Circuits, Systems and Computers*, p. 2250010, 2021.