ISSN 2347 - 3983
Volume 8. No. 5, May 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter83852020.pdf
https://doi.org/10.30534/ijeter/2020/83852020

Botnet Spam E-Mail Detection Using Deep Recurrent Neural Network

MOHAMMAD ALAUTHMAN
Department of Internet Technology, Faculty of information technology, Zarqa University, Zarqa, Jordan.
malauthman@zu.edu.jo

ABSTRACT

The significant amount of SPAM emails that are derived from various botnets worldwide affect the limited capacity of mailboxes. They affect the security of personal mail and the space-loss from the communication. They affect the time required for identifying spam emails and addressing them. Till today, the email spam detection is still considered a challenging process. That is because the email spam is still happening a lot. It is because the detection still needs much improvement. Therefore, the researcher of this study develops a Gated Recurrent Unit Recurrent Neural Network (GRU-RNN) with SVM for Bot Spam email detection. The developed approach got tested by employing the Spambase dataset. The approach shows an accuracy of 98.7%. Through conducting extensive experiments, the researcher concludes that the proposed approach shows an excellent capability of detecting spam email.

Key words: Recurrent Neural Network, Deep learning, Botnet, Spam-email detection, Network security.

1. INTRODUCTION

Nowadays, one receives many emails every day. About 92% of those email messages s are deemed spam[1]. Regarding the spam message, it refers to a message that was sent to one person or several ones repeatedly. From the recipient's perspective, it is undesired. It includes viruses. It includes spyware. It causes damage to the system of the recipient [2-8]. It may be sent through using an electronic mean. Such means include email, messenger, social media, newsgroups, blogs, etc. It can affect the social media platform itself. It can be used for committing crimes, such as a threat. It can be used as a false ad for getting illegal profits [9-13].

Botnets may be used for generating profits through carrying out an attack, such as the (DDoS) attacks. They may be used for financial fraud, Search Engine Optimization (SEO) poisoning, Bitcoin Mining, Corporate and Industrial Espionage and Spamming [14-21].

During the earlier time, the term botnets were not associated with activities causing harm. The IRC bot that is known for the first time is named Eggdrop. In 1993, Eggdrop was published for assisting IRC channel operators [22-25]. Later on, the IRC bots that are malicious emerged. These bots were set to attack IRC users or all the servers. Later on, such bots were employed for launching (DoS) and (DDoS) attacks. The GT-BOT was released in April 1998. It is the first Bot deemed as malicious [26-28].

Regarding spammers, they respond to employing a specific strategy; "Whack-a-Mole". Such response is represented in employing a new IP address each time and shutting down the old one that got observed for the first time in 1996 and. Another innovation related to spam appeared. It has named botnet. Regarding the first spamming botnets, they emerged in 2003. Spam email messages are derived from thousands of IP addresses that frequently change. Today, spammers employ botnets to send a significant amount of Email spam. They adopt spam-sending patterns which are slow and low. That turns all the Blacklisting and filtering techniques are deemed ineffectual [29, 30]. Based on Ref No. [31], the total daily volume of spam in 2008 is more significant than 120 billion messages/ day. Based on the 2010's Symantec report, about 89% of the Email messages on the web are deemed spam messages. About eighty-eight % of spam messages are delivered via botnets [32]. Popular email spam-sending botnets include Storm, Cutwail Bobax, Waledac, Kelihos, Kraken, MeagD, Grum Lethic, Festi, Srizbi, Pushdo, Ozdok, Rustock.

The existent methods for detecting email spam are not very accurate. Hence, raising the performance of those methods is needed. Several data mining methods are deemed effective in raising the performance of the classification. Such methods include the Decision Tree (DT). However, DT includes a weakness. It is represented in being over-sensitive to the noise instance or data, training set, and irrelevant attribute. This weakness reduces the performance of the DT [33]. Some instances are contradictory and noisy existing within the training set. That may lead to the suffering of the created decision tree from overfitting and reducing its accuracy [34].

The researcher believes that recurrent neural networks (RNNs) that are deep may provide an effective solution for implementing a system for spam email detection. A Gated Recurrent Unit Recurrent Neural Network (GRU-RNN) is developed for the detection that's an anomaly. The GRU-RNN is an effective method that is capable of representing the relationship between the current events and the previous ones and improving the rate of the detection that is an anomaly.

To sum up, the significance of the present study are displayed below:

- The researcher introduced a spam email detection system by using GRU-SVM. As far as the researcher knows, it is considered the 1st attempt to use this method for detecting spam email messages.
- The GRU-SVM approach in this study shows a detection rate of 98.7% by employing the minimum amount of features.

The study's design is as follows: Regarding the second part, it displays the relevant literature. Regarding the third part, it displays the proposed technique and the results of the experiment. Regarding the fourth part, it displays the conclusion

2. LITERATURE REVIEW

Many studies were conducted for detecting email spam. The machine learning classifier is mainly used as the primary algorithm. It is combined with optimization of the parameter [35-37]. It is combined with the selection of attributes [38] and the scheme of the threshold [39]. It enhances the outcomes of the detection process. Till today, the accuracy of the detection process is still considered challenging. That is because the best accuracy of the evaluation has not been reached yet.

[39] developed a tri-dimensional cost-sensitive approach; (thresholds, Bayesian & probability). He developed it to filter the spam emails messages. This approach aims at decreasing the error rate derived from miss classifying email messages as spams and non-spams. It aims to show a higher level of performance in cost-sensitive areas. The latter reference divided the dataset into two main parts. These parts are training & testing parts. They show a composition rate of 80% and 20% respectively. The dataset used consists of 3 datasets. These datasets are: Spambase from the UCI Machine Learning Repository, PU1 corpus and Ling-Spam corpus [40] with accuracy 89.88%, 86.35% and 93.94% respectively.

[41] added the differential evolution (DE) to the selection algorithm that is negative (NSA). DE utilized the random generation stage detector distance NSA. Regarding the outcomes, they may get maximized. As for the detector's overlapping, it may get minimized. Regarding the DE, it is used for improving the way of creating detectors during the phase of NSA. As for the local outlier factor (LOF), it is employed as a function of fitness. The developed technique uses Spambase dataset. This dataset got selected from the learning repository of the UCI machine. The accuracy rate of the outcomes is 83.06%.

In [38], the researchers develop a binary search strategy subset by PSO with an operator of a mutation. This operator uses an approach for selecting a wrapper-based feature. This approach extracts the features along with the parameters of weighting. It extracts the classifier of the decision tree (C4.5). The developed technique uses several instances with Spambase. It uses the same standard format for the spambase dataset. The latter researcher collected 6000 email messages during 2012. The rate of accuracy is 94.27%.

[35] developed a model. This model seeks to enhance the random creation of a detector in NSA. That is done by utilizing the stochastic distribution for modelling the data point through the optimization of particle swarm (PSO) was implemented. LOF is displayed as being the function of fitness for identifying the (Pbest) of the candidate detector that offers the best solution. The Spambase dataset derived from the learning repository of the UCI machine, it is used as the dataset. The developed method show accuracy of 91.22%.

[36] used another PSO for enhancing the generation of the random detector in the NSA. Regarding the algorithm, it offers detectors during the generation stage of the algorithm of the negative selection. The NSA-PSO uses (LOF) as the function of fitness for generating detectors. A measure of distance and a value of threshold are used for enhancing the distinctiveness that is existing between the spam detectors and non-spam ones after having the detectors generated. The dataset of the Spambase is employed. The method show accuracy of 83.20%.

Joao Gama, Jerzy Stefanowski [40] employed several classifiers (e.g. KNN, Naive Bayes, Ripper, Decision Tree and SVM. SVM show the highest rate of accuracy (i.e. 98.3%) [42, 43]. They shed light on concept drift. Regarding the concept, it is considered the class being targeted. Regarding the drift, it is the patterns of a stream of data. This method raises the performance level and operates better when having datasets that are small and contain the email dataset that is spam.

[44] used a technique that's employed with the bag of words for representation. He used a classifier of Naive Bayes shall operate better on the probability of classes. The latter researchers employed a few numeric features. He employed a few non-numeric ones which are derived from the dataset. After using the technique, the results are considered promising.

[45]Employed the Naïve Bayes algorithm and Boosting one. Regarding boosting one, it enhances the way of predicting classifiers. The latter scholar employed AdaBoost along with estimators of the decision tree. In Naive Bayes, the selection of feature is carried out based on class.

Through [46], Decision Tree classifier is employed for classifying email messages. Decision tree offers support to the categorical attributes. The critical attribute gets chosen by employing the maximum ratio of gain. Recall, F measure and Precision were employed as a matrix of accuracy.

Skilled spammers started through employing botnets to send a high percentage of Email messages that are spam. That has attributed to the characteristics of botnets which are advantageous [32, 47] as following:

- A botnet offers an infrastructure that is expedient to send out significant volumes of Email messages that are spam since the massive computing distributed network with the bandwidth that's [22, 48]. A botmaster is capable of sending millions of Email messages during a few hours. That is done through leveraging thousands of machines that are infected. Spammers find it easy to get machines infected. They find it easy to enlist these machines into a botnet as new members.
- Botnets work in a manner that ensures getting all the tasks distributed among all the machines that are enlisted and infected. That shall reduce the number of resources that are required for the botmaster. That shall increase the effective throughput.
- The sources or the IP addresses of the machines that are infected change in a consistent manner. Therefore, the majority of the botnets show a specific level of diversity which is geographic. That shall enable the spammer to evade the method used for filtering spam and the method of IP blacklisting regardless of how often they get updated

Rapid growth occurred to the traffic of spam email. That requires a big amount of memory. It requires a big amount of bandwidth. It requires dedicating much time by users. It causes a loss of financial nature. It requires dedicating much time to delete and sort unwanted email messages. He/she requires much time to get rid of spam email [49]. It consumes much time. It is difficult to identify whether the email message is a spam or not through having the content of the mail or its subject read. Thus, a need exists for having a classifier of spams which can be used for detecting and classifying the spam email correctly [50, 51].

3. PROPOSED METHOD

In this part, the researcher proposes the method used for detecting email spam. The researcher selected the RNN-GRU as the essence of the module based on a detector that is an anomaly. The following module loads a trained model. It receives the statistics of the network. It determines whether an email message is considered an anomaly message or not. Regarding the proposed method, it is assessed by employing dataset of Spambase that is derived from the learning repository of the UCI machine [52].

For detecting the spam email messages through using the neural network, there must be 2 stages; the training and testing stages. Figure 1 presents the process carried out for phishing emails using RNN and detecting spam. The proposed models consist of 3 phases. These phases include the features selection, RNN-GRU with SVM.

3.1. Features selection

In the present study, feature reduction aims at selecting a suitable subset of features. That shall enhance the neural network performance. It shall reduce the complexity level of a model of classification without reducing the rates of accuracy significantly. In the present study, the CART algorithm [53-55] has been employed as part of the feature reduction approach. This approach is adopted for eliminating worthless features. It aims at reducing the amount of data that is needed for obtaining higher rates of neural network learning. It aims at doing that for having higher rates of classification accuracy. In the experiment, fifteen 15 features were chosen based on the CART algorithm

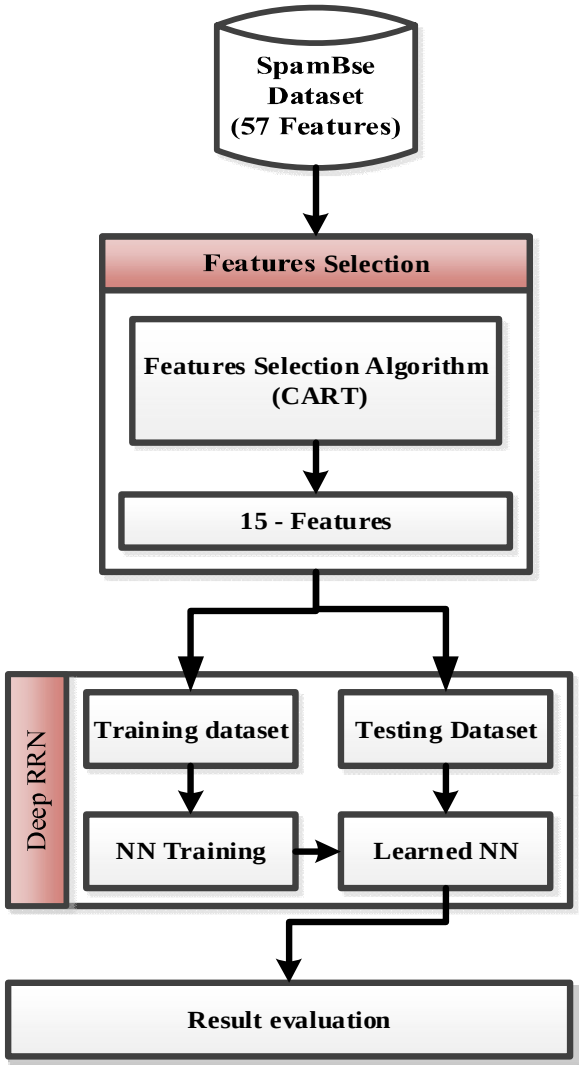

Figure 1: Botnet Spam E-mail Detection Model

3.2. Deep Recurrent Neural Networks

Deep learning is a development in the machine learning domain. It allows one to create models that have discriminative capabilities that surpass any statistical method. The primary algorithms of deep learning are (DNNs). They operate across layers that are connected. The layers are connected in a manner that sees each forward layer taking inputs that are derived from the previous layer and having those inputs modified in a manner that's hidden. Those algorithms can extract the features that are discriminative from the data existing hierarchically, without resorting to handcrafting.

Regarding the RNN, it serves an extension for a feed-forward neural network that is conventional. It utilizes sequential information. The RNNs are called recurrent. That is because they carry out the same task for each element in the sequence, with having the output relying on the previous computations.

The hidden states of the RNN are computed as:

$$h_t = \sigma\left(W_{x_t} + Uh_{t-1} + b_h\right), for\ t = T, \dots, 1. \quad (1)$$

Regarding the hidden states of the RNN, they got computed as follows:

- σ represent a nonlinearity function,
- x_t represent an input vector at time t
- h_t represent a hidden state vector at time t,
- W represent an input to hidden weight matrix,
- U represent a hidden to hidden weight matrix,
- b_h represent a bias term.

The Backpropagation through Time (BPTT) algorithm is used for training the RNN.

3.3. The GRU-SVM Network

The RNN that is traditional encounters gradient problems that are vanishing/exploding [56]. Long Short Term Memory (LSTM) [57] networks and Gated Recurrent Units (GRUs) [58] have been developed for solving that problem. GRUs were chosen in this study because they are not complicated and show a training phase that is faster when having them compared with LSTMs [58]. Figure 2 presents the main elements of GRU.

For this study, there are 15 features employed as the model input. After that, the parameters shall be learnt through using the GRU mechanism of gating of [58] (Equations (2) to (5)).

$$r_t = \sigma(x_t W_r + h_{t-1} U_r) \quad (2)$$

$$z_t = \sigma(x_t W_z + h_{t-1} U_z) \quad (3)$$

$$h_t = (1 - z_t) h_{t-1} + z_t \tilde{h}_t \quad (4)$$

$$\tilde{h}_t = \tanh(x_t W_h + (h_{t-1} \odot r_t) U_h) \quad (5)$$

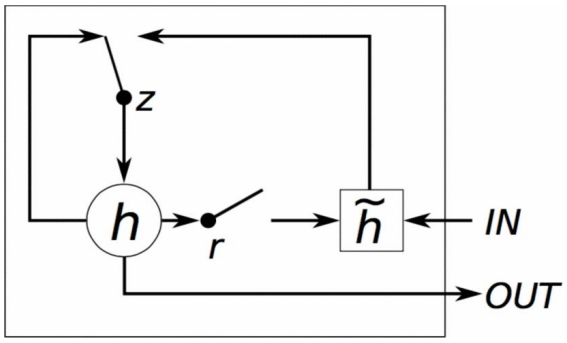

Figure 2: Gated recurrent unit structure[58]

Where r_t is the reset gate, z_t is the update gate, h_t is the activation function and h_t is the candidate activation. $\odot$ is an element-wise multiplication, and σ is the logistic sigmoid function. $W*$ and $U*$ are denoted as learned weight matrices.

The present paper suggests that SVM can be used as the classifier in a neural network architecture. At the final step, the prediction of the model shall be computed by employing the decision function of SVM: $f(x) = sign\,(wx + b)$. A show on in Figure 3. The loss of the neural network is computed using the following equation:

$$min\frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{p} max(0, 1 - y_i'(\mathbf{wx_i} + \mathbf{b}))^2 \quad (5)$$

For minimizing the neural network loss, an optimization algorithm is employed (for this study, the Adam optimizer [59] was used). The optimization algorithm adjusts the weights and biases. That is done based on the computed loss.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Dataset

For analyzing and quantifying the anomalies, the researcher used the SPAMBASE dataset, which includes 57 data attributes that are related with the frequency of some words in the content of the email message. The dataset includes fifty-seven attributes and 4601 messages with 1813 (39%) as a message marked as spam emails. The message marked as non-spam shall show 2788 (61%).

4.2. Experimental Setup

In this paper, we use the Tensorflow [60] as the backend of the KERAS [61], we construct a neural network model, which mainly applies GRU using deep learning KERAS framework with ADAM optimizing function [59]. Moreover, The cost function is estimated based on a binary cross-entropy [62].All experiments in this study were conducted on a laptop computer with Intel Core(TM) i7-3632QM CPU @ 2.20GHz, 8GBand NVIDIA GeForce GTX 4GB GPU. The hyper-parameters used in our experiment is shown in Table 1.

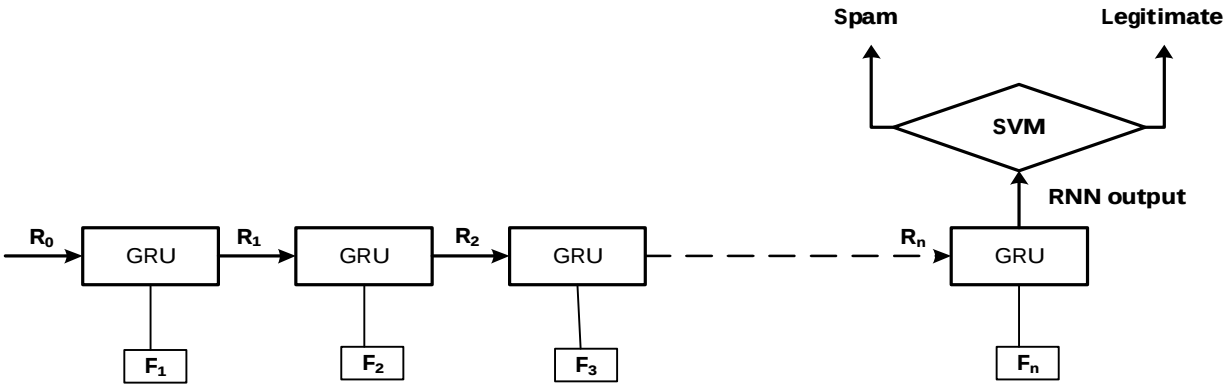

Figure 3: The proposed GRU-SVM structure model and SVM as classifier.

Table 1: List of Hyper-parameters

Parameters	GRU-SVM
Batched	256
Size of Cell	256
Rate of Dropout	0.8
Epochs	10
Learning Rates	10^{-4}
SVM-C	0.5

4.3. Performance evaluation and results

For the selected datasets, the AUC-ROC curve, true positive rate (TPR), true negative rate (TNR), false-positive rate (FPR), average validation accuracy and the F-Score were computed. The results show that the technique archives the highest accuracy and detection rate with the deep recurrent neural network at around 98.65% as presented in Table 2. Moreover, it can be seen that the proposed scheme gives the highest F- score rates, AUC and TPR of about 97.93%, 98.61% and 98.65% respectively.

Table 2:GRU-SVM Results

Measure	GRU-SVM
ACC	0.9865
TPR	0.9836
FPR	0.0153
F1-Score	0.9793
AUC-ROC	0.9861
RMSE	0.0164

We conduct our experiments using classical machine learning algorithms that used for Botnet detection such as Random forest [63], support vector machine (SVM) [64], logistic regression [65] and Gaussian Naive Bayes[66]. As shown in Table 3. our proposed approach based on GRU-SVMperforms better than machine learning algorithm.

Table 3: Comparison of our results with the machine-learning algorithm.

Algorithms	Accuracy	RMSE
Logistic regression	0.849	0.150
Gaussian NB	0.763	0.236
SVM	0.944	0.055
Random Forest	0.957	0.023
Our approach	**0.9865**	**0.0164**

5. CONCLUSION

Spam mails are becoming a severe problem for the networks and the productivity of the users. Through this work, the objective which was to analyze and evaluate the performance of deep RRN on spam email detection. In this research, we apply a gated recurrent unit network and with SVM as an output, layer to predict the email status the legitimate and spam. It's been proved that the proposed method is effective. That's proved based on the experimental results above. There is a need for conducting more studies for exploring other multiclass classifiers. That is needed for improving the impact of the GRU model.

ACKNOWLEDGEMENTS

This research is funded by the Deanship of Research and Graduate Studies in Zarqa University/Jordan.

REFERENCES

[1] D. DeBarr and H. Wechsler, "Spam detection using random boost," *Pattern Recognition Letters*, vol. 33, no. 10, pp. 1237-1244, 2012. https://doi.org/10.1016/j.patrec.2012.03.012

[2] S. M. Lee, D. S. Kim, J. H. Kim, and J. S. Park, "Spam detection using feature selection and parameters optimization," in *2010 International Conference on Complex, Intelligent and Software Intensive Systems*, 2010: IEEE, pp. 883-888. https://doi.org/10.1109/CISIS.2010.116

[3] A. Almomani, B. Gupta, T.-C. Wan, A. Altaher, and S. Manickam, "Phishing dynamic evolving neural fuzzy framework for online detection zero-day phishing email," *arXiv preprint arXiv:1302.0629*, 2013.

[4] K. Alieyan, A. ALmomani, A. Manasrah, and M. M. Kadhum, "A survey of botnet detection based on DNS," *Neural Computing and Applications*, vol. 28, no. 7, pp. 1541-1558, 2017. https://doi.org/10.1007/s00521-015-2128-0

[5] D. A. Almomani *et al.*, "Evolving Fuzzy Neural Network for Phishing Emails Detection," *Journal of computer science*, vol. 8, pp. 1099-1107, 07/01 2012, doi: 10.3844/jcssp.2012.1099.1107.

[6] A. Al-Momani *et al.*, "An online model on evolving phishing e-mail detection and classification method," *journal of applied science*, vol. 11, no. 18, pp. 3301-3307, 2011. https://doi.org/10.3923/jas.2011.3301.3307

[7] M. Almseidin, A. A. Zuraiq, M. Al-kasassbeh, and N. Alnidami, "Phishing Detection Based on Machine Learning and Feature Selection Methods," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 13, no. 12, pp. 171-183, 2019.

[8] A. Niranjan, V. K. Sakhamuri, P. Deepa Shenoy, and K. R. Venugopal, "ERCRFS: Ensemble of random committee and random forest using stackingc for phishing classification," *International Journal of Emerging Trends in Engineering Research*, Article vol. 8, no. 1, pp. 79-86, 2020, Art no. 13, doi: 10.30534/ijeter/2020/13812020.

[9] M. Chakraborty, S. Pal, R. Pramanik, and C. R. Chowdary, "Recent developments in social spam detection and combating techniques: A survey," *Information Processing & Management*, vol. 52, no. 6, pp. 1053-1073, 2016.

[10] A. Almomani, T. Wan, A. Manasrah, A. Altaher, M. Baklizi, and S. Ramadass, "An enhanced online phishing e-mail detection framework based on evolving connectionist system," *International Journal of Innovative Computing, Information and Control (IJICIC)*, vol. 9, no. 3, pp. 169-175, 2013.

[11] W. Xiao-Lin and Cloete, "Learning to classify email: a survey," in *2005 International Conference on Machine Learning and Cybernetics*, 18-21 Aug. 2005 2005, vol. 9, pp. 5716-5719 Vol. 9, doi: 10.1109/ICMLC.2005.1527956.

[12] A. Almomani, A. Obeidat, K. Alsaedi, M. A.-H. Obaida, and M. Al-Betar, "Spam e-mail filtering using ECOS algorithms," *Indian Journal of Science and Technology*, vol. 8, no. S9, pp. 260-272, 2015. https://doi.org/10.17485/ijst/2015/v8iS9/55320

[13] P. K. Kollu and R. S. Prasad, "Intrusion detection system using recurrent neural networks and attention mechanism," *International Journal of Emerging Trends in Engineering Research*, Article vol. 7, no. 8, pp. 178-182, 2019, doi: 10.30534/ijeter/2019/12782019.

[14] M. Alauthman, "An efficient approach to online bot detection based on a reinforcement learning technique," PhD, Northumbria University, UK, 2016. [Online]. Available: http://nrl.northumbria.ac.uk/id/eprint/29617

[15] M. Alauthman, N. Aslam, M. Al-kasassbeh, S. Khan, A. Al-Qerem, and K.-K. Raymond Choo, "An efficient reinforcement learning-based Botnet detection approach," *Journal of Network and Computer Applications*, vol. 150, p. 102479, 2020/01/15/ 2020, doi: https://doi.org/10.1016/j.jnca.2019.102479.

[16] M. Alauthaman, N. Aslam, L. Zhang, R. Alasem, and M. A. Hossain, "A P2P Botnet detection scheme based on decision tree and adaptive multilayer neural networks," *Neural Computing and Applications*, vol. 29, no. 11, pp. 991-1004, 2018/06/01 2018, doi: 10.1007/s00521-016-2564-5.

[17] A. Alnawasrah, A. Alnomani, F. Meziane, and M. Alauthman, "Fast flux botnet detection framework using Adaptive dynamic evolving spiking neural network algorithm," in *2018 9th International Conference on Information and Communication Systems (ICICS)*, 3-5 April 2018 2018.

[18] M. Al-kasassbeh and T. Khairallah, "Winning tactics with DNS tunnelling," *Network Security*, vol. 2019, no. 12, pp. 12-19, 2019.

[19] M. Al-Kasassbeh, "Network intrusion detection with wiener filter-based agent," *World Appl. Sci. J*, vol. 13, no. 11, pp. 2372-2384, 2011.

[20] G. Al-Naymat, M. Al-Kasassbeh, and E. Al-Harwari, "Using machine learning methods for detecting network anomalies within SNMP-MIB dataset," *International Journal of Wireless and Mobile Computing*, vol. 15, no. 1, pp. 67-76, 2018. https://doi.org/10.1504/IJWMC.2018.10015860

[21] N. Chandra Sekhar Reddy, P. C. R. Vemuri, and A. Govardhan, "An emperical study on support vector machines for intrusion detection," *International Journal of Emerging Trends in Engineering Research*, Article vol. 7, no. 10, pp. 383-387, 2019, doi: 10.30534/ijeter/2019/037102019.

[22] R. S. C. Silva, R. M. P. Silva, R. C. G. Pinto, and R. M. Salles, "Botnets: A survey," *Comput. Netw.*, vol. 57, no. 2, pp. 378-403, 2013, doi: doi:10.1016/j.comnet.2012.07.021.

[23] A. Almomani, "Fast-flux hunter: a system for filtering online fast-flux botnet," *Neural Computing and Applications*, vol. 29, no. 7, pp. 483-493, 2018.

[24] K. Alieyan, M. Anbar, A. Almomani, R. Abdullah, and M. Alauthman, "Botnets Detecting Attack Based on DNS Features," in *2018 International Arab Conference on Information Technology (ACIT)*, 2018: IEEE, pp. 1-4. https://doi.org/10.1109/ACIT.2018.8672582

[25] K. Alieyan, A. Almomani, R. Abdullah, and M. Anbar, "A Rule-based Approach to Detect Botnets based on DNS," in *2018 8th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2018: IEEE, pp. 115-120.

[26] L. Chao, J. Wei, and Z. Xin, "Botnet: Survey and Case Study," in *Innovative Computing, Information and Control (ICICIC), Fourth International Conference on*, 7-9 Dec. 2009 2009, pp. 1184-1187, doi: doi: 10.1109/icicic.2009.127.

[27] A. Almomani, O. M. Dorgham, M. Alauthman, M. Al-Refai, and N. Aslam, "Botnet Behavior and Detection Techniques: A Review," *Computer and Cyber Security: Principles, Algorithm, Applications, and Perspectives*, p. 223, 2018. https://doi.org/10.1201/9780429424878-9

[28] K. Alieyan, A. Almomani, M. Anbar, M. Alauthman, R. Abdullah, and B. Gupta, "DNS rule-based schema to botnet detection," *Enterprise Information Systems*, pp. 1-20, 2019.

[29] M. Alauthman, A. Almomani, M. Alweshah, W. Omoushd, and K. Alieyane, "Machine Learning for phishing Detection and Mitigation," *Machine Learning for Computer and Cyber Security: Principle, Algorithms, and Practices*, p. 26, 2019.

[30] A. Almomani, M. Alauthman, O. Almomani, and F. Albalas, "A proposed framework for Botnet Spam-email Filtering using Neucube," in *18th The International Arab Conference on Information Technology (ACIT)*, Yassmine Hammamet, Tunisia, 22-24 December 2017 2017: IEEE.

[31] C. Kreibich *et al.*, "On the Spam Campaign Trail," *LEET*, vol. 8, no. 2008, pp. 1-9, 2008.

[32] B. Stone-Gross, T. Holz, G. Stringhini, and G. Vigna, "The Underground Economy of Spam: A Botmaster's Perspective of Coordinating Large-Scale Spam Campaigns," *LEET*, vol. 11, pp. 4-4, 2011.

[33] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.

[34] D. M. Farid, L. Zhang, C. M. Rahman, M. A. Hossain, and R. Strachan, "Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks," *Expert systems with applications*, vol. 41, no. 4, pp. 1937-1946, 2014. https://doi.org/10.1016/j.eswa.2013.08.089

[35] I. Idris and A. Selamat, "Improved email spam detection model with negative selection algorithm and particle swarm optimization," *Applied Soft Computing*, vol. 22, pp. 11-27, 2014.

[36] I. Idris *et al.*, "A combined negative selection algorithm–particle swarm optimization for an email spam detection system," *Engineering Applications of Artificial Intelligence*, vol. 39, pp. 33-44, 2015.

[37] M. Alkasassbeh, "A Novel Hybrid Method for Network Anomaly Detection Based on Traffic Prediction and Change Point Detection," *arXiv preprint arXiv:1801.05309*, 2018. https://doi.org/10.3844/jcssp.2018.153.162

[38] Y. Zhang, S. Wang, P. Phillips, and G. Ji, "Binary PSO with mutation operator for feature selection using decision tree applied to spam detection," *Knowledge-Based Systems*, vol. 64, pp. 22-31, 2014.

[39] B. Zhou, Y. Yao, and J. Luo, "Cost-sensitive three-way email spam filtering," *Journal of Intelligent Information Systems*, vol. 42, no. 1, pp. 19-45, 2014.

[40] B. Krawczyk, L. L. Minku, J. Gama, J. Stefanowski, and M. Woźniak, "Ensemble learning for data stream analysis: A survey," *Information Fusion*, vol. 37, pp. 132-156, 2017. https://doi.org/10.1016/j.inffus.2017.02.004

[41] I. Idris, A. Selamat, and S. Omatu, "Hybrid email spam detection model with negative selection algorithm and differential evolution," *Engineering Applications of Artificial Intelligence*, vol. 28, pp. 97-110, 2014.

[42] I. A. Lawal and S. A. Abdulkarim, "Adaptive SVM for data stream classification," *South African Computer Journal*, vol. 29, no. 1, pp. 27-42, 2017.

[43] H.-L. Nguyen, Y.-K. Woon, and W.-K. Ng, "A survey on data stream clustering and classification," *Knowledge and information systems*, vol. 45, no. 3, pp. 535-569, 2015.

[44] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, "A Bayesian approach to filtering junk e-mail," in *Learning for Text Categorization: Papers from the 1998 workshop*, 1998, vol. 62: Madison, Wisconsin, pp. 98-105.

[45] S. K. Trivedi and S. Dey, "Interplay between probabilistic classifiers and boosting algorithms for detecting complex unsolicited emails," *Journal of Advances in Computer Networks*, vol. 1, no. 2, pp. 132-136, 2013. https://doi.org/10.7763/JACN.2013.V1.27

[46] J.-J. Sheu, K.-T. Chu, N.-F. Li, and C.-C. Lee, "An efficient incremental learning mechanism for tracking concept drift in spam filtering," *PloS one*, vol. 12, no. 2, 2017.

[47] G. Stringhini, T. Holz, B. Stone-Gross, C. Kruegel, and G. Vigna, "BOTMAGNIFIER: Locating Spambots on the Internet," in *USENIX security symposium*, 2011, pp. 1-32.

[48] S. García, A. Zunino, and M. Campo, "Survey on network-based botnet detection methods," *Security and Communication Networks*, vol. 7, no. 5, pp. 878-903, 2014, doi: 10.1002/sec.800.

[49] P. Parveen and P. Halse, "Spam mail detection using classification," *vol*, vol. 5, pp. 347-349, 2016.

[50] M. A. Shafi'I *et al.*, "A review on mobile SMS spam filtering techniques," *IEEE Access*, vol. 5, pp. 15650-15666, 2017. https://doi.org/10.1109/ACCESS.2017.2666785

[51] S. K. Trivedi and S. Dey, "A study of ensemble based evolutionary classifiers for detecting unsolicited emails," in *Proceedings of the 2014 conference on research in adaptive and convergent systems*, 2014, pp. 46-51.

[52] M. Lichman, "UCI machine learning repository," ed: Irvine, CA, 2013.

[53] M. Al-Kasassbeh, S. Mohammed, M. Alauthman, and A. Almomani, "Feature Selection Using a Machine Learning to Classify a Malware," in *Handbook of Computer Networks and Cyber Security: Principles and Paradigms*, B. B. Gupta, G. M. Perez, D. P. Agrawal, and D. Gupta Eds. Cham: Springer International Publishing, 2020, pp. 889-904.

[54] M. ALKASASSBEH *et al.*, "AN EMPIRICAL EVALUATION FOR THE INTRUSION DETECTION FEATURES BASED ON MACHINE LEARNING AND FEATURE SELECTION METHODS," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 22, 2017.

[55] M. Alkasassbeh and M. Almseidin, "Machine Learning Methods for Network Intrusion Detection," *International Journal of Computer and Information Engineering*, vol. 12, no. 8, pp. 614-619, 2018.

[56] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," ed: A field guide to dynamical recurrent neural networks. IEEE Press, 2001.

[57] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735-1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735

[58] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[59] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[60] T. T. H. Le, J. Kim, and H. Kim, "An Effective Intrusion Detection Classifier Using Long Short-Term Memory with Gradient Descent Optimization," in *2017 International Conference on Platform Technology and Service (PlatCon)*, 13-15 Feb. 2017 2017, pp. 1-6, doi: 10.1109/PlatCon.2017.7883684.

[61] J. Kim, J. Kim, and H. Kim, "An Approach to Build an Efficient Intrusion Detection Classifier," *Journal of Platform Technology*, vol. 3, no. 4, pp. 43-52, 2015.

[62] D. E. Booth, "The Cross-Entropy Method," ed: Taylor & Francis, 2008.

[63] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001. https://doi.org/10.1023/A:1010933404324

[64] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[65] M. Schmidt, N. Le Roux, and F. Bach, "Minimizing finite sums with the stochastic average gradient," *Mathematical Programming*, vol. 162, no. 1, pp. 83-112, 2017/03/01 2017, doi: 10.1007/s10107-016-1030-6.

[66] C. Robert, *Machine learning, a probabilistic perspective*. Taylor & Francis, 2014. https://doi.org/10.1080/09332480.2014.914768