

Original Article

# Detection of Violent E-mails Using Fuzzy Logic

Victoria Oluwatoyin Oyekunle<sup>1</sup>, Prince Oghenekaro Asagba<sup>2</sup>, Fubara Egbono<sup>2</sup>

<sup>1,2,3</sup> Department of Computer Science, University of Portharcourt,  
Portharcourt, Nigeria.

Received Date: 14 February 2021

Revised Date: 29 March 2021

Accepted Date: 31 March 2021

**Abstract** - People all around the world spend billions of e-mail messages daily, and the use of mobile e-mail (e-mail sent via a mobile device) is growing at an astounding rate. Despite its advantages, one of the biggest threats to an e-mail today is Violent and phishing e-mail. This research improves the detection and filtering of violent and phishing e-mails by implementing a fuzzy Logic detection model that classifies e-mails into classes' violent, phishing and ham and then determines how harmful the classified e-mails are. Incoming e-mails were classified based on how well their features as compared with their rank values satisfied the stated fuzzy rules. From the results, output e-mail classes and their corresponding degrees of threats were provided at high accuracy and improved speed from Moderate to High or Very High.

**Keywords** - E-mail, Fuzzy logic, Ham, Phishing, Violent E-mail, Artificial Intelligence

## I. INTRODUCTION

Electronic mail (frequently called e-mail) is the method of exchanging electronic messages between computers over a network- usually the internet. It is undoubtedly one of the internet applications most commonly used today. With the increased use of the internet, and the number of e-mail users multiplying day by day, it has become one of the finest advertising ways to generate and send messages. The advantage of using e-mails for adverts today has inspired the introduction of unwanted e-mails by spammers.

Scammers use specialized computer programs to gather people's e-mail addresses from social media pages, websites, consumer lists, password dumps, newsgroups, etc., and sell them to other scammers. The program looks at the code of every webpage; it looks for an e-mail address and saves it to the spammers' database of harvested addresses. They send bulk e-mail messages to internet users to obtain a response from a few in order to ensure their profit. Response from a few in order to ensure their profit.

In their article on Violent e-mail scams, techradar.com noted that some violent e-mails come in the form of bomb

scare e-mails where the sender claims to have planted a bomb in a target's home or office that will be triggered if a requested amount is not paid. Some of these e-mails also contain passwords or partial phone numbers associated with the e-mail address the e-mail was sent to, making it appear to the target that the attacker has access to his or her private information, which could be used for blackmail [1]. But most times, many of these user details are usually cleverly guessed and used as bait by scammers, all in a bid to ensure profit off their targets.

Experiences have shown that these violent e-mails, usually sent in dozens to several people, usually contain threats of bodily harm, vulgar threats, threats of sextortion, or making reference to mutilation of female genitalia. They usually contain the foulest of language imaginable. The violent words come both in the subject text and e-mail body of such e-mails. They are usually sent with the sole aim of intimidating and frightening the targets and cause emotional distress, which may make the target succumb to such threats and pay the expected sum. Setting up an e-mail filter that will direct abusive e-mails into a separate or dummy e-mail account will do a lot of good because this will prevent seeing the abuse on a regular basis and to have a place to store them pending the time decision is made as to what to do with them [2] In fact, sending of Violent and phishing scam e-mails constitute e-mail harassment because participating in sending such e-mails violates ethical usage of the computer account, and in some extreme cases may even provoke victims to press criminal charges [3].

Many current e-mail filtering methods do well, but they must be frequently maintained and tuned as the characteristics of unwanted messages change. Some problems emerge from an e-mail filtering model judging a legitimate e-mail to be an unwanted e-mail which is usually far worse than judging an unwanted e-mail to be a legitimate one. Web-based mail systems (for example, Yahoo and Hotmail) have inbox quota limits of a couple of megabytes. These quotas may be exceeded on a daily basis by unwanted e-mail, and legitimate messages will be rejected by the mail



servers because the user's inbox is full. For businesses that depend on e-mail services for income, the loss of legitimate mail could prove very expensive and render the utilization of such e-mail services ineffective as a communication tool [4].

Automatic e-mail filtering of malicious e-mails is an important and popular one among methods created to prevent unwanted messages. It can be described as the automatic classification of unwanted and legitimate or Ham email [4]. Numerous strategies have been recommended to distinguish and group (classify) e-mail messages as legitimate and non-legitimate. It has been discovered that the algorithm success rate of Artificial Intelligence is extremely high. AI algorithms have been utilized in E-mail Classification and Filtering Systems to give better classifications, as seen in the works done by [5] to [22].

Numerous spammers today are progressively employing creative strategies to send unwanted e-mails. They have been successful in outsmarting e-mail filtering systems that use only classification algorithms. Consequently, existing unwanted e-mail filtering methods need to be enhanced.

We used the concept of Fuzzy logic to build an automatic classifier. Fuzzy logic was chosen as the best option for this research because it assumes the boundary between two neighboring classes as the shared area within which an object has partial membership in each class [23]. As stated in [16], in everyday existence, linguistic factors that are closer to human reasoning are more important than assurance. Linguistic conditions are therefore used to model the e-mail detection scheme.

Fuzzy logic allows us to generate models that represent values between 0 and 1. Real-life situations today aren't just 0 and 1 and True or false. We use statements in day-to-day life activities that cannot be represented by ordinary Boolean logic. In Boolean logic, memberships of sets are either full memberships which are represented by '1', or total non-membership, which is represented by '0'. Or 'True' for full membership and 'False' for complete non-membership. Partial memberships are not allowed. In e-mail filtering, degrees of the threat posed by unwanted messages (both violent and phishing scam) differs for each individual, and they fall under unwanted messages as some e-mail messages can be more dangerous than the others. Hence fuzzy logic is best to model these fuzzy boundaries.

The rest of the paper is organized as follows. Section II discusses the related work on spam detection and filtering using fuzzy logic-based systems. Section III describes the fuzzy logic modeling for violent e-mail detection, and Section IV details the results. Conclusion and future work are presented in Section V.

## II. RELATED WORK

While there is no directly related work on the filtering of violent e-mails, many researchers employed Artificial intelligence and fuzzy logic for Spam, phishing, and unwanted e-mail detection.

[18] presented the design and execution of a trainable Fuzzy logic-based e-mail classification scheme that learned the most efficient Fuzzy laws during the training stage and then applied the Fuzzy Control Model to classify invisible texts. Their findings showed that spam filters that are automatically trainable are practically feasible and can have an important impact on spam detection.

In their paper, [25] for all incoming messages identified and implemented blurred rules. Input e-mail was categorized as spam or ham based on the outcome of the different rules against user behavior. Fuzzy rules were designed to deduce spam messages for 5 input parameters, namely Sender's Address, SenderIP, SubjectWords, ContentWords, and Common User Attachment. The suggested simplistic strategy to spam e-mail deduction was conducted more accurately and quicker than the current approaches.

[26] suggested the classification of spam emails with fuzzy word rankings based on content. The work used a word ranking database to classify the messages, and the ranks were used depending on the degree of risk that each word had. The work obtained a better result from ranking and classifying spam words, but it did not classify the spam words in the subject.

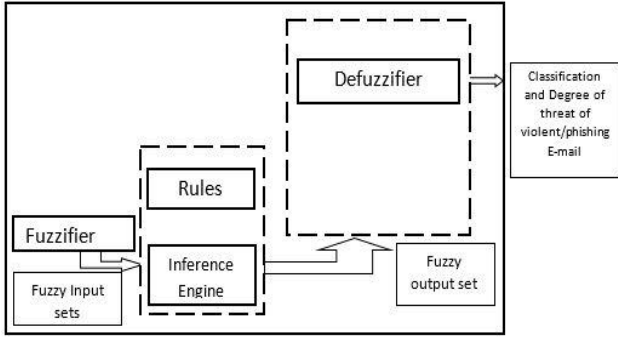
[27] proposed a spam mail classification approach using a spam word ranking database and fuzzy rules to classify spam mails according to spam content. This work categorized e-mails based on the degree of danger each term has in combination with other current techniques. The suggested work used sets of language terms to rank and classify spam mails. This method extracted only four features from an e-mail instead of extracting all the features from the mail.

In their paper, [28] also presented a fuzzy expert system to detect spam. They considered the pre-processing of the subject, content, the sender's e-mail address, and attachments of the e-mail to be ranked by using common spam words list. These ranked items represented the input variables for the proposed system, which classify the e-mail as spam or not. The fuzzy expert system performed well to filter the spam and gave good results in terms of spam recall and precision.

This research proposes a Fuzzy Logic detection model for filtering violent e-mails. It also specifies the degree of threat each violent e-mail poses.

**III. MATERIALS AND METHODS**

From Fig 1, the steps to constructing the fuzzy logic system are;



**Fig. 1. Fuzzy logic System Architecture**

Firstly, the definition of the linguistic variables and terms. Second, construction of membership functions, the knowledge base of rules, and conversion of crisp data into fuzzy data sets using membership functions (fuzzification). Third, evaluation of rules in the rule base (Inference Engine). And lastly, conversion of output data into non-fuzzy values. (Defuzzification)

An internet corpus of SpamAssassin public corpus, which we had gathered and saved over time were used in this research. The subject and body of the input e-mail are segmented into tokens and analyzed. Message blocks that are probable to contain violent (brutal) and recurring phishing phrases are labeled. The extracted phrases were classified according to rank values assigned to them based on their probabilities which are calculated from their word count. The distinct words or phrases generated from the e-mail pre-processing and feature extraction stages were passed as input variables to the fuzzy inference system, which classifies the words and produces output. Membership functions give the degree of membership (from 0 to 1) of an input variable from the Subject text and e-mail body to a specific fuzzy set A. In order to obtain the membership values of each fuzzy linguistic set, a fuzzification operation is processed to compare the input variables with the membership functions on the premise part.

For each e-mail function, linguistic values were allocated as very low, low, moderate, high, and very high. Table 3 shows the Linguistic labels attached to features.

**Table 1. Linguistic Labels attached to parameters**

S/N	Parameters	Linguistic Values
1	Subject Text (ST)	Very High, High, Moderate, Low, Very Low
2	E-mail Body (EB)	Very High, High, Moderate, Low, Very Low

**A. The Fuzzy rules**

Fuzzy rules were given as if-then statements that relate the likelihood of violent e-mail messages in e-mails to various levels of key e-mail indicators. A value is allocated

to the e-mail phrases and classified into five language factors. Very High (VH), High (H), Moderate (M), Very Low (VL), and Low (L), called fuzzy sets, which are mapped to each variable of input (Subject text and e-mail body). For these variables, these fuzzy sets overlap and cover the required ranges. Twenty-five (25) rules were defined.

The rank values assigned to the features and corresponding linguistic values are;

If  $0 \leq ST$  AND  $EB < 0.2$ , Then very low unwanted word (Ham)

ElseIf  $0.2 \leq ST$  AND  $EB < 0.4$  Then low unwanted words (Ham)

ElseIf  $0.4 \leq ST$  AND  $EB < 0.5$  Then Moderate unwanted words (Phishing)

ElseIf  $0.5 \leq ST$  AND  $EB < 0.7$  Then Strong unwanted words (Phishing)

ElseIf  $0.7 \leq ST$  AND  $EB \leq 1$  Then Very strong unwanted word (Violent)

The fuzzy rules defined for the main e-mail filtering system are defined below;

R1: If (SubjectText is very low) and (E-email body is very low) then (Ham is NOT Dangerous)

R2: If (SubjectText is very low) and (E-email body is low) then (Ham is NOT Dangerous)

R3: If (SubjectText is very low) and (E-email body is moderate) then (Ham is NOT Dangerous)

R4: If (SubjectText is very low) and (E-email body is high) then (Ham is NOT Dangerous)

R5: If (SubjectText is low) and (E-email body is very high) then (Ham is NOT Dangerous)

R6: If (SubjectText is low) and (E-email body is very low) then (Ham is NOT Dangerous)

R7: If (SubjectText is low) and (E-email body is low) then (Ham is NOT Dangerous)

R8: If (SubjectText is low) and (E-email body is moderate) then (Ham is NOT Dangerous)

R9: If (SubjectText is low) and (E-email body is high) then (Ham is NOT Dangerous)

R10: If (SubjectText is low) and (E-email body is very high) then (Ham is NOT Dangerous)

R11: If (SubjectText is moderate) and (E-email body is very low) then (Ham is NOT Dangerous)

R12: If (SubjectText is moderate) and (E-email body is low) then (Ham is NOT Dangerous)

R13: If (SubjectText is moderate) and (E-email body is high) then (Phishing is Moderate Dangerous)

R14: If (SubjectText is moderate) and (E-email body is very high) then (Violent is Dangerous)

R15: If (SubjectText is high) and (E-email body is very low) then (Ham is NOT Dangerous)

R16: If (SubjectText is high) and (E-email body is low) then (Ham is NOT Dangerous)

R17: If (SubjectText is high) and (E-email body is moderate) then (Phishing is Moderate Dangerous)

R18: If (SubjectText is high) and (E-email body is high) then (Violent is Dangerous)

R19: If (SubjectText is high) and (E-email body is very high) then (Violent is Dangerous)

R20: If (SubjectText is very high) and (E-email body is very low) then (Ham is NOT Dangerous)

R21: If (SubjectText is very high) and (E-email body is low) then (Ham is NOT Dangerous)

R22: If (SubjectText is very high) and (E-email body is moderate) then (Phishing is Moderate Dangerous)

R23: If (SubjectText is very high) and (E-email body is high) then (Violent is Dangerous)

R24: If (SubjectText is very high) and (E-email body is very high) then (Violent is Most Dangerous)

#### IV. IMPLEMENTATION AND RESULTS

Classification learner was used in training and classification of the e-mail datasets. A fuzzy logic designer in MATLAB (R2017a) was used to implement the proposed fuzzy e-mail filtering model. An interface was also designed to display the results of the classification based on the fuzzy concept. The choice MATLAB was used on the basis that it is highly adaptable for data visualization, has a record of program enhancement, and is helpful for modeling applications. This Fuzzy logic model brings the advantage of controlling at any moment the situation as X "subject text," and Y "e-mail body" predicts Z "violent, phishing or ham." The fuzzy E-mail number can be given as a triangular Membership Function for inputs (subject text and e-mail body) and output (violent, phishing, or ham).

Fig.2.A. to Fig.2.E shows the membership functions of the variables, while Fig 2F shows the rule view with the 25 rules that were defined.

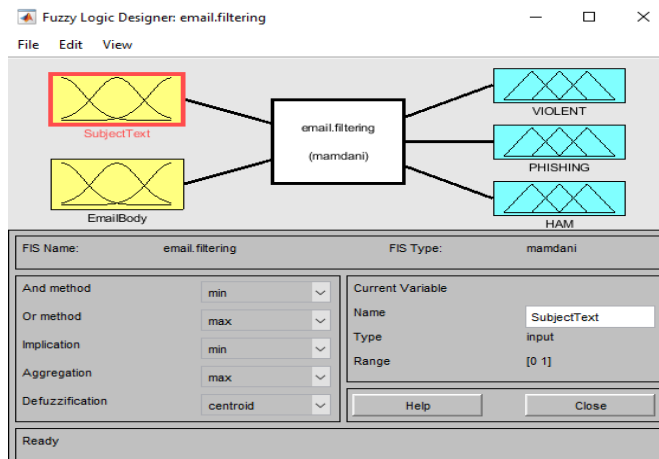


Fig. 2A The FIS was showing the I/O variables.

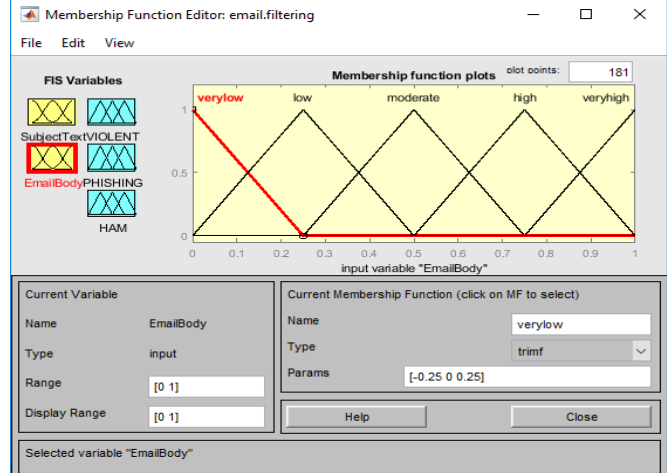


Fig. 2B Triangular membership functions for EB (input) showing the degree of membership of the values to a fuzzy set (Antecedent)

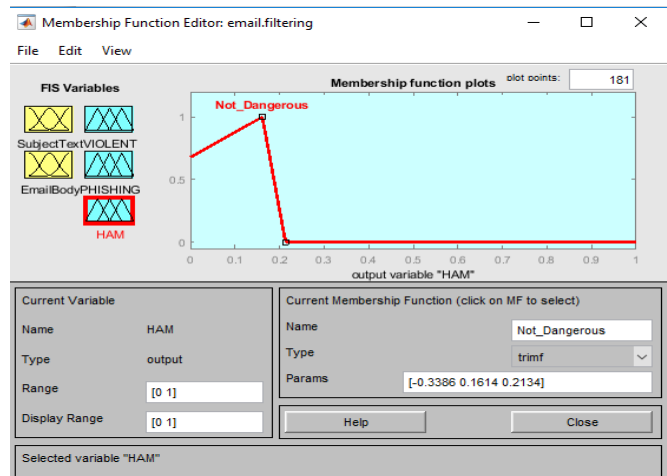


Fig. 2C Triangular membership functions for Ham (output) showing a degree of membership of the values to a fuzzy set (Consequent)

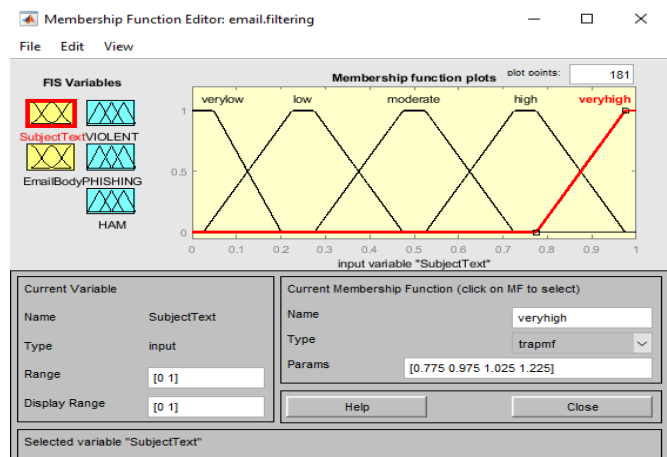


Fig. 2D Trapezoidal membership functions for Subject Text(input) showing the degree of membership of the values to a fuzzy set

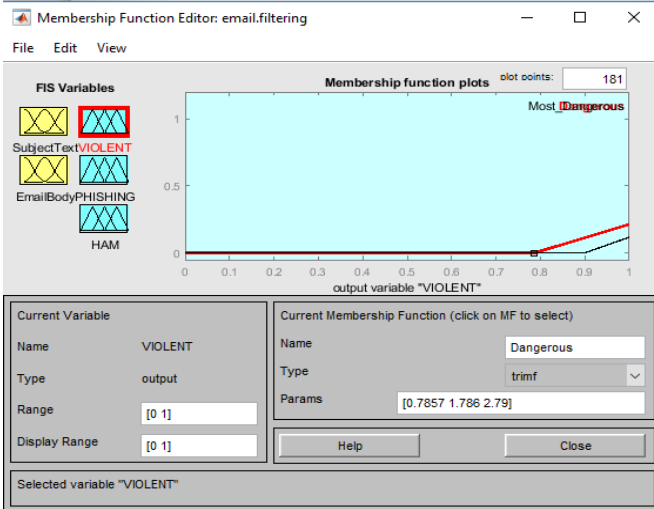


Fig. 2E Triangular membership functions for Violent (output) showing the degree of membership of the values to a fuzzy set (Consequent)

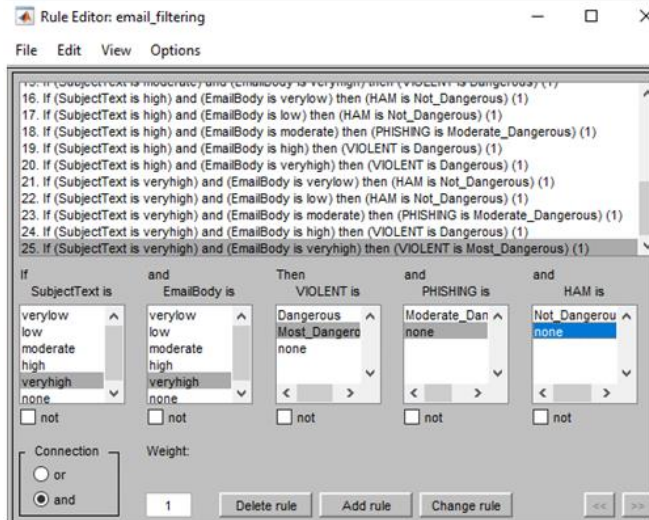


Fig. 2F The rule view was showing the 25 rules that were defined.

Results show that very low in the subject, and email body gives ham, very low in subject text, and low in email body gives ham. However, high in subject text and very high in email body gives violent, and so on. This method extracted the words and applied a fuzzy inference system with fuzzy rules for the classification of violent mails and indicating the degree of violence.

Table II consists of System inputs (subject text and email body) and outputs (violent, phishing, or ham). From (0.1 to 0.49) indicates not ham while (0.5 to 1) indicates unwanted words. Degree outline; from (0.5 to 0.8) indicates moderate phishing and (0.81 to 1) indicates very high violence. From Table III, Degree “Low” signifies Weak and Least dangerous email messages. Degree “Moderate” signifies moderately dangerous email messages. Degree “High” signifies Strong and Dangerous email messages. Very high signifies very strong and most dangerous email messages.

Table 2. Table showing Rules Extraction and their corresponding classified classes.

Subject Text	Email Body	Violent	Phishing	Ham
0.1	0.2	-	-	0.226
0.2	0.4	-	-	0.226
0.2	0.5	-	-	0.232
0.3	0.6	-	0.539	-
0.4	0.7	-	0.529	-
0.5	0.8	-	0.6	-
0.6	0.9	-	0.652	-
0.7	0.8	-	0.68	-
0.8	0.8	-	0.738	-
0.8	0.9	-	0.742	-
0.9	0.9	0.823	-	-
0.9	1	0.922	-	-

Table 3. Table showing degrees of Violent Words and linguistic values that produce them.

Subject Text	Email Body	Violent	Degree	Subject Text
0.1	0.2	0.226	LOW	0.1
0.2	0.4	0.226	LOW	0.2
0.2	0.5	0.232	LOW	0.2
0.3	0.6	0.539	MODERATE	0.3
0.4	0.7	0.529	MODERATE	0.4
0.5	0.8	0.6	MODERATE	0.5
0.6	0.9	0.652	HIGH	0.6
0.7	0.8	0.68	HIGH	0.7
0.8	0.8	0.738	HIGH	0.8
0.8	0.9	0.742	HIGH	0.8
0.9	0.9	0.823	VERY HIGH	0.9
0.9	1	0.922	VERY HIGH	0.9

## VI. CONCLUSION

In this study, the objective was to enhance the identification of incoming violent and phishing e-mails by using fuzzy logic to classify e-mails into violent, phishing, and ham (legitimate) based on some features. To understand their weakness, we checked a current scheme. We also used fuzzy logic to determine the degree of phishing and violence with high accuracy and improved speed from Moderate, High, or Very High. Violent emails with very high degrees of threat were classified as very hazardous; this allowed users the capacity to understand and block such unwanted e-mail levels. Moderate phishing phrases were classified as moderately dangerous, making consumers cautious with such messages. The proposed approach will work only for e-mails having Subject text and E-mail body as plain text. But today, scammers also include multimedia content and HTML links in e-mails sent to exploit users. Our future work aims at detecting and filtering such content.

## REFERENCES

- [1] Violent e-mail scams are on the rise: <https://www.techradar.com/news/violent-e-mail-scams-are-on-the-rise/> retrieved 28 (2020).
- [2] R. E. Sorace, How to handle E-mail harassment: [https://www.huffpost.com/entry/how-to-handle-e-mail-haras\\_b\\_5606031/](https://www.huffpost.com/entry/how-to-handle-e-mail-haras_b_5606031/) retrieved 28 (2020).
- [3] Avoid getting in trouble with e-mail: <https://kb.iu.edu/d/afnf> retrieved 28 (2020).
- [4] M. Sahami, S. Dumais, D. Heckerman, E. Horvitz, A Bayesian approach to filtering junk e-mail, In Proceedings AAAI Workshop on Learning for Text Categorization, (1998)
- [5] Y. Hong, L. Qihe, Z. Shijie, L. Yang, A Spam Filtering Method Based on Multi-Modal Fusion. *Applied Sciences*,(9) (2019) 1152.
- [6] A. Saleh, A. Karim, B. Shanmugam, S. Azam, K. Kannoorpatti, M. Jonkman, F. De Boer, An Intelligent Spam Detection Model Based on Artificial Immune System. *Information* (10) (2019) 209.
- [7] M. .A. Shafi'I, M. S. A. Latiff, H. Chiroma, O. Osho, G. Abdul-Salaam, A. I. Abubakar, T. Herawan., A review on mobile SMS spam filtering techniques. *IEEE Access*, (5) (2017) 15650-15666.
- [8] E. G. Dada, J. S. Bassi, H. Chiroma, A. O. Adetunmbi, O.E. Ajibuwa., Machine learning for e-mail spam filtering: review, approaches and open research problems. *Heliyon*, 5(6) (2019).
- [9] O. Abayomi-Alli, S. Misra, A. Abayomi-Alli, M. Odusami., A review of soft techniques for SMS spam classification: Methods, approaches, and applications. *Engineering Applications of Artificial Intelligence*, (86) (2019) 197-212.
- [10] B. K. Dedeturk, B. Akay., Spam filtering using a logistic regression model trained by an artificial bee colony algorithm. *Applied Soft Computing*, PDCA12-70 data sheet, Opto Speed SA, Mezzovico, Switzerland, (10) (2020) 6229.
- [11] J. R. Méndez, T. R. Cotos-Yañez, D. Ruano-Ordás, A new semantic-based feature selection method for spam filtering. *Applied Soft Computing*, (76) (2019) 89-104.
- [12] A. Fahfouh, J. Riffi, M. A. Mahraz, A. Yahyaouy, H. Tairi, PV-DAE: A hybrid model for deceptive opinion spam based on neural network architectures. *Expert Systems with Applications*, 11(3) (2020) 517.
- [13] N. Andrew, Jeff, Building High-level Features Using Large Scale Unsupervised Learning. *Proceedings of the 29th International Conference on Machine Learning*, Ed-inburgh, Scotland, UK, (2013) 1-13.
- [14] R. Anju., V. Vaidhehi, E-mail Classification Using Machine Learning Algorithms. *International Journal of Engineering and Technology (IJET)*, 9(2) (2017).
- [15] I. Androutopoulos., P. Georgios, K. Vangelis, S. Georgios, D.S. Constantine, S. Panagiotis., Learning to Filter Spam E-Mail: A Comparison of a Naïve Bayesian and a Memory-Based Approach, *Proceedings of the workshop, Machine Learning and Textual Information Access*, H. Zaragoza, P. Gallinari, and M. Rajman (Eds.), 4th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-2000), Lyon, France, (2000) 1-13.
- [16] B. Cetisli., Development of an adaptive neuro-fuzzy classifier using linguistic hedges: Part 1, *Expert Systems with Applications*, 37 (2010) 6093-6101.
- [17] M. Dewan, D. Farid, Z. Li, M. R. Chowdhury, M.A. Hossain, Rebecca S., Hybrid Decision tree and Naïve Bayes classifiers for multi-class classification tasks. *Expert Systems with Applications*, 41 (2014) 1937-1946.
- [18] M. Fuad., D. Debzani, S.M. Hossain., A Trainable Fuzzy Spam Detection System. Department of Computer Science Montana State University Bozeman, Montana, USAfuad@cs.montana.edu, (2008).
- [19] R. Giyanani, M. Desai., Spam Detection using Natural Language Processing. *IOSR Journal of Computer Engineering (IOSR-JCE)* e-ISSN: 2278-0661,p-ISSN: 2278-8727, 16(5) (2014) 116-119.
- [20] I. Ismaila, S. Ali, N. Thanh Nguyen, S.O. Omatu, M. P. KamilKuca., A combined negative selection algorithm-particle swarm optimization for an e-mail Spam detection system. *Engineering Applications of Artificial Intelligence* 39 (2015) 33-44.
- [21] A. Sharma, Anchal, SMS Spam Detection Using Neural Network Classifier. *IEEE*, (2014) 240-244
- [22] S. Seth, S. Biswas., Multimodal Spam Classification Using Deep learning Techniques. *On Signal-Image Technology and Internet-Based Systems, Proceedings of the International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, Jai-pur, India, (2017).
- [23] S. Mehta, U. Eranna, K. Soundararajan., A Fuzzy Technique for Classification of Intercepted Communication, *International Journal of Communication Engineering Applications-IJCEA*, 03(1) (2012).
- [24] M. N. Marson, M. W El-Kharashi, Fayez Gebali., Targeting Spam control on middleboxes: Spam detection based on layer-3 E-mail content classification. *Elsevier Com-puter Networks* 53 (2009) 835-848.
- [25] P. Sudhakar., G. Poonkuzhali, K. Thiagarajan, R.K. Keshav, K. Sarukesi, Fuzzy Logic for E-Mail Spam Deduction, *Proceedings of the International Conference on Applied Computer and Applied Computational Science*, Vernice, Italy, (2011).
- [26] G. Santhi, S. M. Wenisch, Sengutuvan, P., A Content-Based Classification of Spam Mails with Fuzzy Word Ranking, *IJCSI International Journal of Computer Science Issues*, 10 (3) (2013) 2. ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784.
- [27] H. Kotian., K. Gupta, J. Stephy., Using Fuzzy Logic for E-mail Spam Filtering. *International Journal of Advanced Research in Computer Science and Software Engineering* 5(10) (2015) 15-20.
- [28] S. Almasan, W. Qaid, Khalid, A., I. Alqubat., Filtering Spam Using Fuzzy Expert System, *Journal of Emerging Trends in Computing and Information Sciences* 10(12) (2015) 655-660.
- [29] Surendiran,R., and Alagarsamy,K., 2012. An Extensive Survey on Mobile Security and Issues . *SSRG International Journal of Computer & organization Trends (IJCOT)*, 2(1) 39-46.