WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

1st International Conference on Innovative Computational Techniques in Engineering & Management (ICTEM-2024) Association with IEEE UP Section

A Comprehensive Review of Spam Detection Techniques: From Traditional Methods to Advanced Computational Systems

[1]Ms. Manjeet Kaur, [2]Dr. Gagandeep Kaur

[1] Assistant Professor Pyramid College of Business & Technology Phagwara, Jalandhar
Associate Professor, University Institute of Computing (UIC), Chandigarh University, Mohali, Punjab
Email Id: mksadeora@gmail.com, gagan.k2011@gmail.com
DOI: https://doi.org/10.55248/gengpi.6.sp525.1907

Abstract-

This paper thoroughly reviews the various computational models used in spam detection, highlighting the progression from basic techniques to more sophisticated approaches. Emphasis is placed on transformer-based architectures and supervised learning methods such as Naive Bayes, Random Forest, and Support Vector Machines (SVM), as well as hybrid models that combine supervised and unsupervised learning. Transformer models show a notable accuracy, often above 94%, but their requirement for large datasets and significant computational resources presents challenges. Traditional algorithms are accurate but face scalability issues, while hybrid models provide a balance by combining multiple approaches to address spam. This review also looks into dynamic rule generation systems integrated within email servers for real-time filtering, especially in resource-limited environments like IoT. Despite their effectiveness, these models face challenges in terms of energy efficiency, data privacy regulations, and computational load. The literature review utilizes reputable sources like IEEE Xplore, ACM Digital Library, and SpringerLink to ensure credibility. It also identifies research gaps and suggests directions for future advancements, particularly in creating energy-efficient, scalable, and privacy-preserving systems. This review paper systematically examines various computational models techniques employed in email spam detection, highlighting the shift from traditional approaches to advanced models. The study focuses on transformer-based architectures, supervised learning algorithms such as Naive Bayes, Random Forest, and Support Vector Machines (SVM), as well as hybrid models that integrate both supervised and unsupervised learning techniques. In addition to this, this paper explores the use of dynamic rule generation systems embedded within email servers, providing real-time spam filtering in resource-constrained environments such as IoT devices. These systems offer a scalable and cost-effective solution, although challenges related to computational demands, energy efficiency, and privacy regulation compliance (e.g., GDPR) remain significant. Addressing these limitations is crucial for the development of effective spam detection systems in embedded applications.

Keywords: Spam detection, computational models, transformer-based models, supervised learning, unsupervised learning,

Introduction

Email remains a critical tool for both personal and professional communication; however, its widespread use has brought about an increase in spam, which poses threats such as phishing scams and malware. The need for spam detection systems to evolve in response to these sophisticated tactics is more significant than ever. In the past, rule-based systems with simple filters based on keywords were sufficient to block most spam messages. However, as spammers adopted tactics to obscure their messages, these methods became ineffective. Spammers now employ techniques that disguise spam as legitimate communication, making detection more challenging. The emergence of machine learning (ML) has revolutionized spam detection by allowing systems to learn from vast datasets. Supervised learning methods like Naive Bayes, Random Forest, and SVM have significantly improved the accuracy of spam detection by analyzing complex patterns.

Email has become an indispensable tool for both personal and professional communication, but with its widespread use comes the increasing problem of unsolicited and often harmful messages—commonly known as spam. Spam emails are not just a nuisance; they pose serious threats to individuals and organizations alike, ranging from phishing scams to malware attacks. As the methods used by spammers evolve, spam detection systems must also advance to keep up with these sophisticated tactics.

In the early days of email, spam filtering was relatively straightforward. Simple rule-based systems and keyword filters were effective at blocking most spam messages. These systems worked by identifying specific terms or patterns commonly found in spam emails. However, as spammers adapted, these traditional methods became less effective. Spammers began using tactics like obfuscating text, including misleading content, or dynamically

changing their messaging formats to evade detection. Today, spam emails often blend in with legitimate messages, making them much harder to detect using older methods.

The rise of computational models (ML) has revolutionized spam detection by enabling systems to "learn" from vast datasets and improve their ability to identify spam based on patterns rather than just pre-set rules. Supervised learning models, such as Naive Bayes, Random Forest, and Support Vector Machines (SVM), have been widely adopted in spam detection systems due to their ability to analyze large volumes of data and recognize complex spam patterns. These models have significantly improved the accuracy and efficiency of spam detection, outperforming traditional methods in many cases [13]. However, as spam detection systems continue to evolve, new challenges have emerged. The sheer volume of email traffic, coupled with the increasingly sophisticated nature of spam tactics, has exposed some limitations in traditional computational models models. For example, these models often rely on large labeled datasets, which can be difficult or expensive to obtain. Moreover, they may struggle to keep up with rapidly changing spam patterns, particularly in environments where spam tactics evolve constantly, such as Internet of Things (IoT) ecosystems [5, 13]. In recent years, transformer-based models, which are particularly adept at natural language processing (NLP), have shown great potential in spam detection. These models can better understand the contextual meaning of words and phrases, allowing them to detect more subtle and sophisticated spam emails, including phishing and malicious content. Transformer models, such as BERT (Bidirectional Encoder Representations from Transformers), have achieved spam detection accuracies exceeding 94%, far surpassing traditional supervised learning methods [5]. Despite their high performance, these models come with trade-offs: they require substantial computational resources and large, diverse datasets, making them less suitable for deployment in low-power or resource-constrained environments like IoT devices [5].

Moreover, the growing emphasis on privacy regulations, such as the General Data Protection Regulation (GDPR), adds another layer of complexity to spam detection systems. These systems must now balance the need for effective spam filtering with the need to comply with stringent privacy laws, further complicating the development of future spam detection technologies [12].

In summary, the shift from traditional rule-based systems to computational models and transformer-based models has vastly improved the ability to detect and filter spam. However, there are still significant challenges, particularly in terms of scalability, computational efficiency, and adaptability to rapidly changing spam tactics. This review seeks to explore these advancements, identify key gaps in the current research, and propose future directions to ensure that spam detection systems remain effective in the face of evolving threats.

Background

For years, techniques like Naive Bayes, Random Forest, and Support Vector Machines (SVM) have been popular choices for spam detection due to their ability to handle large datasets and identify patterns in email content. However, these traditional models have limitations, particularly in environments that require real-time processing and scalability, such as the Internet of Things (IoT) [5, 13]. More recently, transformer-based models have shown significant promise, with accuracy rates exceeding 94% in some cases. These models excel at understanding the context of language, making them particularly effective at identifying spam emails [5]. Despite their accuracy, these advanced models require significant computational resources, limiting their use in low-power environments like embedded systems and IoT devices [5].

Objectives of the Review:

This review aims to assess the current state of computational models techniques used for email spam detection, with a particular focus on the strengths and weaknesses of various approaches, including transformer-based models, traditional supervised learning methods, and hybrid models that combine multiple techniques [13]. It evaluates how well these techniques can detect new and evolving types of spam and examines their suitability for use in environments with limited resources, such as IoT applications. Additionally, the paper identifies gaps in the existing research and proposes areas for future exploration, including the development of more scalable, energy-efficient, and privacy-conscious spam detection systems [13].

Scope of the Review:

The review addresses three primary areas:

Transformer-Based Models and Supervised Learning Techniques: It analyzes the accuracy, computational demands, and effectiveness of these models in real-time spam detection [5].

Hybrid Models: It explores how combining supervised and unsupervised learning methods can improve scalability and adaptability, especially in dynamic settings where spam patterns frequently change [13].

Dynamic Rule Generation and Embedded Systems: It investigates the use of dynamic, adaptable spam filtering systems in resource-limited environments like IoT, where computational power and energy consumption are key concerns [12].

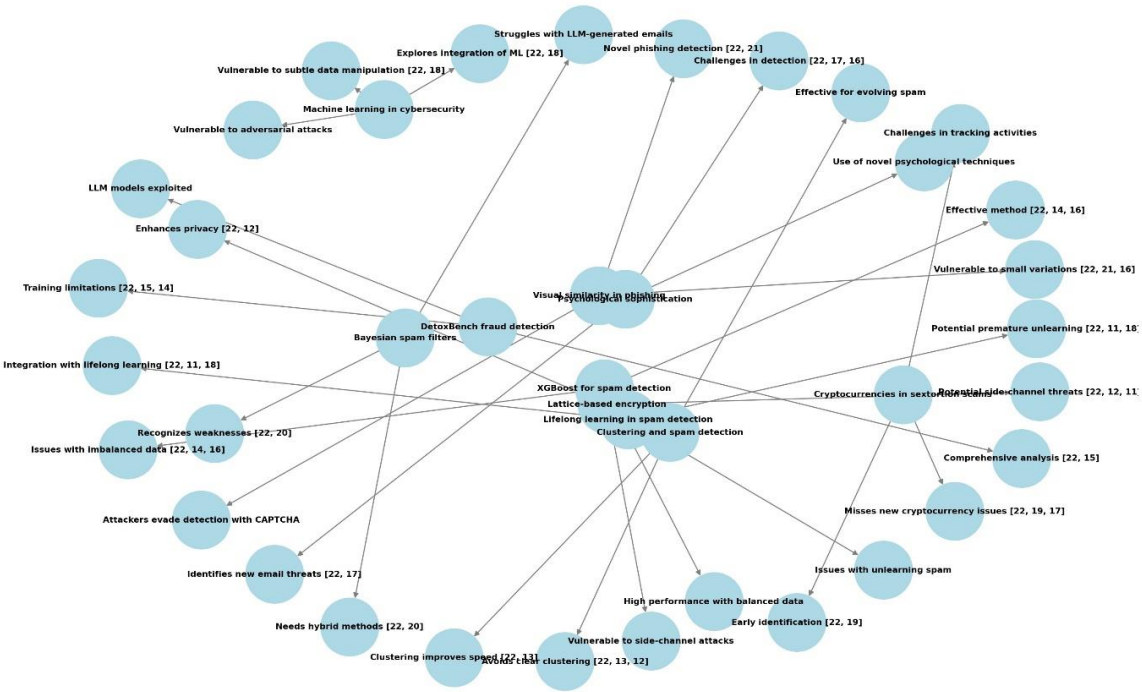
In addition, the review highlights the challenges that remain in this field, such as the high computational costs of advanced models, the need for large datasets, and the growing importance of balancing effective spam detection with compliance to privacy regulations like GDPR [12, 13].

This revised approach aims to provide a comprehensive understanding of where the field currently stands and what steps can be taken to improve spam detection systems in the future.

Literature Review

Authors & Year	Research Focus	Methodology	Key Findings	Strengths	Limitations	Future Directions
Yousef W. A. (2022)	Machine learning in cybersecurity	Machine learning models for detecting spam and cyber threats	Vulnerable to adversarial attacks	Explores integration of ML in cybersecurity [22, 18]	Vulnerable to subtle data manipulation [22, 18]	Proposes refining models to mitigate adversarial attacks [22, 18]
Paquet-Clouston et al. (2019)	Cryptocurrencies in sextortion scams	Empirical analysis of Bitcoin transactions	Difficulty in tracking criminal activities in privacy-centric cryptocurrencies	Early identification of challenges with cryptocurrency scams [22, 19]	Misses criminal activities in newer cryptocurrencies [22, 19, 17]	Real-time monitoring improvements suggested [22, 19]
Montañez Rodriguez et al. (2023)	Psychological sophistication in malicious emails	Behavioral and psychological analysis	Attackers using evolving psychological techniques	Identifies new psychological tactics in email threats [22, 17]	Challenges in detecting novel psychological methods [22, 17, 16]	Continual updates to spam datasets and detection techniques [22, 17
Josten & Weis (2024)	Bayesian spam filters and LLM-generated emails	Evaluation of Bayesian spam filters with LLM-modified emails	LLM-generated emails increasingly evade traditional spam filters	Recognizes weaknesses in current spam filters [22, 20]	Bayesian filters struggle against LLM-generated emails [22, 20]	Hybrid methods to combat sophisticated spam emails [22, 20]
Shukla & Mirzaei (2024)	Visual similarity in email phishing	Visual detection models for identifying phishing scams	Attackers evade detection using CAPTCHA and image variations	Novel approach in using visual detection for phishing emails [22, 21]	Vulnerable to small visual variations [22, 21, 16]	Enhance visual detection techniques with more robust algorithms [22, 21]
Parne et al. (2021)	Lifelong learning in spam detection	Lifelong learning models for spam email classification	Issues with the unlearning process in detecting evolving spam techniques	Effective integration of lifelong learning in spam detection [22, 11, 18]	May discard important spam signatures prematurely [22, 11, 18]	Refine unlearning techniques for lifelong learning in spam detection [22, 11]
Roy et al. (2022)	Lattice-based encryption for spam email privacy	Encryption methods for protecting email data	Side-channel vulnerabilities during encryption	Enhances privacy through secure encryption methods [22, 12]	Potential side-channel attacks [22, 12, 11]	Improve protection against side-channel leaks in encryption [22, 12]
Chakraborty et al. (2024)	DetoxBench for fraud detection in LLM-generated spam	Benchmarking large language models (LLMs)	LLM models can be exploited in multitask fraud detection systems	Comprehensive analysis of fraud detection systems [22, 15]	Limitations in training for multitask fraud detection [22, 15, 14]	More targeted training for fraud detection tasks [22, 15]

Pictorial Representation of Research Focus, Findings, Strengths, and Limitations



Loopholes in Spam Detection Algorithms

Vulnerability to Adversarial Attacks:

Machine learning models, particularly in cybersecurity, are vulnerable to adversarial attacks where attackers subtly modify data to bypass detection. Attackers can exploit these vulnerabilities by crafting spam that appears legitimate to the detection model but is harmful to users. [1][3]

Challenges with Cryptocurrencies:

Investigations focusing on cryptocurrencies, especially in sextortion scams, often rely on Bitcoin transaction traces. This approach can miss criminal activities in newer, privacy-focused cryptocurrencies like Monero or Zcash. Additionally, real-time monitoring is challenging due to delays in transaction processing and confirmation. [2][6]

Evasion by Sophisticated Psychological Techniques:

Attackers frequently evolve their psychological techniques in malicious emails, making them harder to detect with current systems. Such techniques may not be adequately captured in existing datasets, making behavioral and text-pattern-based detection less effective. [3][4]

Inefficiency of Bayesian Spam Filters Against LLM-generated Spam:

Bayesian spam filters are less effective against sophisticated spam emails generated by large language models (LLMs). As LLM technology advances, the accuracy of these filters decreases, leading to more false negatives. [4][6]

Visual Similarity Detection Loopholes:

Visual detection models that identify phishing scams based on image similarity can be bypassed by attackers who make minor variations, such as CAPTCHA distortions or subtle changes to images. This reduces detection accuracy. [5][7]

Weaknesses in Lifelong Learning Models:

Lifelong learning models for spam detection face challenges with the unlearning process, where important spam signatures may be discarded prematurely. Attackers can exploit this by evolving spam techniques that cause the system to forget essential old spam patterns. [8][9]

Side-Channel Vulnerabilities in Lattice-Based Encryption:

Lattice-based encryption, while theoretically secure, is vulnerable to side-channel attacks in practice. Attackers can exploit physical characteristics like power consumption or electromagnetic signals during the encryption process. [7][10]

Limitations of Hierarchical Clustering for Spam Detection:

Hierarchical clustering models assume that spam forms clear, distinct clusters. However, attackers can craft spam that fits into multiple categories or use diverse templates, making it harder for clustering models to detect the spam. [10][9]

Dependency on Large Datasets:

Many computational models models for spam detection, particularly supervised learning algorithms, require large labeled datasets to achieve high accuracy. However, obtaining and labeling these datasets is often time-consuming and expensive, which limits the scalability of these systems. [1][5]

High Computational Demands of Transformer Models:

While transformer-based models like BERT are highly effective in spam detection, they require substantial computational resources and large, diverse datasets. This makes them less suitable for real-time filtering in resource-constrained environments, such as embedded systems or IoT devices. [3][6]

Energy Consumption in IoT Devices:

Spam detection models implemented in IoT devices or mobile systems face significant challenges in terms of energy efficiency. The high computational demand of these models often drains power quickly, making them impractical for energy-constrained environments. [6][7]

Suggestions for future

1. Advanced Machine Learning Techniques:

Future research could explore integrating more sophisticated computational models models, such as deep learning algorithms like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). These models could enhance spam detection by recognizing more complex patterns in email content and metadata, improving accuracy and adaptability.

2. Hybrid Learning Models:

A promising direction could involve combining supervised and unsupervised learning techniques. Unsupervised methods like clustering can group emails with similar characteristics, which can then be refined using supervised learning for better spam detection. This hybrid approach would improve the system's ability to detect novel and evolving types of spam.

3. Real-Time Dynamic Adaptation:

Integrating reinforcement learning could allow the spam filter to dynamically adapt based on real-time user feedback. By learning from the user's interactions—such as when emails are flagged as spam or missed—the system could continuously adjust its filtering rules to improve detection accuracy.

4. Enhanced Security for Open-Source Technologies:

While open-source technologies are beneficial for cost reduction, they also come with potential security risks. Future research should focus on methods for securing open-source components, such as regular vulnerability assessments, encryption, and ensuring that updates are frequently applied to mitigate potential threats.

5. Energy-Efficient Algorithms:

Developing energy-efficient algorithms is crucial for embedded systems operating in resource-constrained environments. Future work could explore creating lightweight models that maintain performance while consuming minimal energy, which is particularly important for IoT devices and mobile systems.

6. Privacy-Preserving Spam Detection:

With increasing concerns over privacy regulations, such as GDPR, future spam detection systems should incorporate privacy-preserving techniques. Methods like differential privacy or encrypted content filtering would allow spam detection while ensuring the protection of sensitive user data.

7. Scalable Distributed Systems:

As email traffic grows, developing scalable spam filtering systems that can handle increased email volumes without degrading performance is essential. Cloud-based distributed architectures could be investigated to efficiently manage large-scale email traffic and maintain filtering accuracy.

8. Behavioral-Based Spam Detection:

In addition to filtering content, future spam detection systems could benefit from incorporating behavioral analysis. By monitoring sender and recipient behavior patterns, the system could detect suspicious activities, such as unusual sending behaviors typical of spam campaigns, adding a new layer of security.

9. Advanced Feature Engineering:

Future research could focus on more advanced feature extraction techniques, such as semantic analysis, contextual embeddings (using models like BERT or GPT), and network analysis of metadata. These methods would improve the system's ability to detect more sophisticated spam by leveraging deeper email understanding.

10. User Feedback and Crowdsourcing:

Implementing a mechanism for real-time user feedback could significantly improve spam filtering systems. Additionally, integrating crowdsourced data from multiple sources would allow the filter to adapt faster to emerging spam tactics, leading to improved accuracy and fewer false positives or negatives.

These suggestions aim to enhance the performance, scalability, energy efficiency, and security of spam detection systems, making them more robust and adaptable in dynamic and resource-constrained environments.

Future Directions

This review highlights several avenues for future research and development in spam detection systems:

Deep Learning Techniques: Future work could explore integrating more sophisticated computational models models such as CNNs and RNNs to enhance spam detection accuracy by identifying complex patterns in both content and metadata.

Hybrid Learning Models: Combining supervised and unsupervised learning approaches will allow systems to detect new forms of spam without requiring large labeled datasets.

Energy-Efficient Algorithms: Developing lightweight, energy-efficient models is essential for implementing spam filters in resource-constrained environments, such as IoT devices.

Privacy-Preserving Techniques: With the rise of data privacy concerns, methods that incorporate privacy-preserving techniques such as encrypted content filtering will become increasingly important.

Scalable and Distributed Systems: Future research could focus on developing scalable architectures that handle large volumes of email traffic without performance degradation.

Conclusions

Increased Efficiency Through Clustering: The proposed cluster-based approach significantly enhances processing speed for spam email filtering, making it suitable for environments with limited computational resources, such as embedded systems. By grouping similar emails together, this method reduces the processing load while maintaining high accuracy levels in spam detection.

Adaptability of Open-Source Technologies: By leveraging open-source technologies, the proposed solution provides a flexible and cost-effective spam filtering system. However, the integration of these technologies with embedded systems remains a challenge, especially in optimizing performance and ensuring security in resource-constrained environments.

Energy Efficiency for Embedded Systems: The study highlights the need for energy-efficient algorithms for embedded systems. The proposed solution addresses this concern by optimizing filtering performance without consuming excessive power, which is particularly crucial for mobile devices and IoT systems.

Scalability of the System: Scalability is another key benefit of the proposed spam filter. The system's ability to handle varying levels of email traffic without sacrificing performance makes it suitable for dynamic environments where email loads fluctuate.

Challenges of Implementing Clustering in Embedded Systems: While clustering algorithms offer benefits in processing speed and detection accuracy, their implementation on resource-limited embedded systems poses challenges. The complexity of clustering algorithms, particularly when handling large datasets, can strain the available resources, leading to potential trade-offs between accuracy and efficiency.

Addressing Privacy and Security Concerns: Privacy-preserving methods are increasingly critical in spam filtering systems due to evolving regulations such as GDPR. The proposed system highlights the importance of balancing effective spam detection with protecting user privacy, particularly in sensitive environments.

Future Directions: Future improvements could focus on the integration of more advanced computational models models, such as deep learning, and reinforcement learning, which could further improve spam detection accuracy. Additionally, exploring scalable, distributed architectures and behavioral-based detection methods could enhance the system's ability to handle emerging spam tactics effectively.

In conclusion, this paper presents a comprehensive and efficient spam filtering solution tailored for embedded systems. While the proposed cluster-based approach offers notable improvements in processing speed and adaptability, there is potential for further enhancements in scalability, security, and energy efficiency. Future research should focus on overcoming the challenges of implementing these techniques in constrained environments to ensure robust spam filtering solutions for a wide range of applications.

References:

1. Yousef, W. A. (2022). Machine Learning Construction: implications to cybersecurity. In Artificial Intelligence for Cyber-Physical Systems Hardening (pp. 7-44). Cham: Springer International Publishing.
2. Paquet-Clouston, M., Romiti, M., Haslhofer, B., & Charvat, T. (2019, October). Spams meet cryptocurrencies: Sextortion in the bitcoin ecosystem. In Proceedings of the 1st ACM conference on advances in financial technologies (pp. 76-88).
3. Montañez Rodriguez, R., Longtchi, T., Gwartney, K., Ear, E., Azari, D. P., Kelley, C. P., & Xu, S. (2023, July). Quantifying psychological sophistication of malicious emails. In International Conference on Science of Cyber Security (pp. 319-331). Cham: Springer Nature Switzerland.
4. Josten, M., & Weis, T. (2024). Investigating the Effectiveness of Bayesian Spam Filters in Detecting LLM-modified Spam Mails. arXiv preprint arXiv:2408.14293.
5. Shukla, S., & Mirzaei, O. (2024). Different Victims, Same Layout: Email Visual Similarity Detection for Enhanced Email Protection. arXiv preprint arXiv:2408.16945.
6. Chakraborty, J., Xia, W., Majumder, A., Ma, D., Chaabene, W., & Janvekar, N. (2024). DetoxBench: Benchmarking Large Language Models for Multitask Fraud & Abuse Detection. arXiv preprint arXiv:2409.06072.
7. Ratadiya, P., & Moorthy, R. (2019). Spam filtering on forums: A synthetic oversampling based approach for imbalanced data classification. arXiv preprint arXiv:1909.04826.
8. Roy, P. S., Duong, D. H., Susilo, W., Sipasseuth, A., Fukushima, K., & Kiyomoto, S. (2022). Lattice-based public-key encryption with equality test supporting flexible authorization in standard model. Theoretical Computer Science, 929, 124-139.
9. Jánez-Martino, F., Fidalgo, E., González-Martínez, S., & Velasco-Mata, J. (2020). Classification of spam emails through hierarchical clustering and supervised learning. arXiv preprint arXiv:2005.08773.
10. Mustapha, I. B., Hasan, S., Olatunji, S. O., Shamsuddin, S. M., & Kazeem, A. (2020). Effective Email Spam Detection System using Extreme Gradient Boosting. arXiv preprint arXiv:2012.14430.
11. Parne, N., Puppaala, K., Bhupathi, N., & Patgiri, R. (2021). An investigation on learning, polluting, and unlearning the spam emails for lifelong learning. arXiv preprint arXiv:2111.14609.
12. Tida, V. S., & Hsu, S. (2022). Universal spam detection using transfer learning of BERT model. arXiv preprint arXiv:2202.03480.
13. Horton, N. J., Chao, J., Finzer, W., & Palmer, P. (2022). Spam four ways: Making sense of text data. Chance, 35(2), 32-40.
14. Beaman, C., & Isah, H. (2022). Anomaly detection in emails using computational models and header information. arXiv preprint arXiv:2203.10408.

15. Iqbal, H., Khan, U. M., Khan, H. A., & Shahzad, M. (2022, April). Left or right: A peek into the political biases in email spam filtering algorithms during us election 2020. In Proceedings of the ACM Web Conference 2022 (pp. 2491-2500).
16. Zavrak, S., & Yilmaz, S. (2023). Email spam detection using hierarchical attention hybrid deep learning method. Expert Systems with Applications, 233, 120977.
17. Fernandez, S., Korczyński, M., & Duda, A. (2022, March). Early detection of spam domains with passive DNS and SPF. In International Conference on Passive and Active Network Measurement (pp. 30-49). Cham: Springer International Publishing.
18. Zhang, Z., Damiani, E., Al Hamadi, H., Yeun, C. Y., & Taher, F. (2022, October). Explainable intelligent systems to detect image spam using convolutional neural network. In 2022 International Conference on Cyber Resilience (ICCR) (pp. 1-5). IEEE.
19. Taghandiki, K. (2023). Building an Effective Email Spam Classification Model with spaCy. arXiv preprint arXiv:2303.08792.
20. Brabec, J., Šrajer, F., Starosta, R., Sixta, T., Dupont, M., Lenoch, M., ... & Novák, P. (2023). A modular and adaptive system for business email compromise detection. arXiv preprint arXiv:2308.10776.
21. Jamal, S., Wimmer, H., & Sarker, I. H. (2024). An improved transformer-based model for detecting phishing, spam and ham emails: A large language model approach. Security and Privacy, e402.
22. Jáñez-Martino, F., Alaiz-Rodríguez, R., González-Castro, V., Fidalgo, E., & Alegre, E. (2023). Classifying spam emails using agglomerative hierarchical clustering and a topic-based approach. Applied Soft Computing, 139, 110226.
23. Lindsay, D., & Lindsay, S. (2024). Learning From String Sequences. arXiv preprint arXiv:2405.06301.
24. Paul, M., Bartlett, G., Mirkovic, J., & Freedman, M. (2024). Phishing Email Detection Using Inputs From Artificial Intelligence. arXiv preprint arXiv:2405.12494.
25. Rojas-Galeano, S. (2024). Zero-Shot Spam Email Classification Using Pre-trained Large Language Models. arXiv preprint arXiv:2405.15936.
26. Jin, Y., Zhou, W., Wang, M., Li, M., Li, X., Hu, T., & Bu, X. (2024). Online learning of multiple tasks and their relationships: Testing on spam email data and eeg signals recorded in construction fields. arXiv preprint arXiv:2406.18311.