Journal of Machine Learning and Deep Learning (JMLDL)
Volume 3, Issue 2, Date: 31- December– 2026
ISSN: 2978-4336, DOI: 10.64820/AEPJMLDL.32.1.5.122026

Optimized Machine Learning and Deep Learning Approaches for Email Spam Detection

Ramprakash Velmurugan [1, *], Keerthana T [2], Yuvadarshan M [3], Muthupandi G [4], Pavithra S [5]
[1, 2, 3, 4, 5] Department of Computer Science and Engineering, Chennai Institute of Technology, Tamil Nadu, India
Email: [1] ramprakashv.cse20232@citchennai.net, [2] keerthanat@citchennai.net, [3] yuvadarshanmurugann@gmail.com, [4] pg.muthupandi@gmail.com, [5] pavithras@citchennai.net
*Corresponding Author

***Abstract*—A comprehensive review of spam detection techniques emphasizes the growing importance of addressing the widespread and emerging threat of email spam, which has become a major challenge. Spam emails, including malicious content like phishing links and malware, present significant security, economic, and privacy challenges to individuals and organizations. The objective of the study is to examine and evaluate modern machine learning (ML) and deep learning (DL) strategies for detecting spam more accurately and efficiently compared to earlier techniques like blacklists or content-based filtering. This paper identifies significant ML and DL methods commonly used for detecting email spam by evaluating their effectiveness and limitations. It also highlights research gaps in existing systems and suggests integrating AI technology into spam detection. Challenges discussed include false positives, computational efficiency, and the need for adaptive models to counter advanced spam tactics. The review aims to guide future studies and improve spam detection techniques.**

Keywords—Machine learning (ML), Deep learning (DL), Fuzzy systems, Feature selection, Spam detection challenges, Hybrid models.

I. INTRODUCTION

A. Background

Email spam, including phishing and malware, continues to pose a significant threat despite efforts to curb the problem, that have seen in traditional solutions such as signature-based detection have fallen short of tackling modern threats [1]. It is imperative to undertake this study to determine the effectiveness of new ML and DL approaches to detect problems, since they optimize adaptability and scalability. Concerning limitations such as false positives and concept drift, the study proposes a solution that can foster superior protection against emerging and contemporary spam invasions.

B. Need for advanced systems

Spam is a rapidly evolving phenomenon, and the measures presently in use, such as simple rule filters and blacklists, are inefficient. Since the spammers keep changing their ways, there is rising demand for systems that can learn and detect new threats in real time. ML and DL make the process scalable, accurate, and flexible, and provide better detection and moreover, guarantee a high level of protection from advanced spam attacks.

C. Challenges to overcome

This research is inspired by the increasing difficulties posed by emergent properties of email spam that are beyond the capacity of traditional approaches. But as spam becomes smarter day to day, previous phishing and obfuscation techniques prove to be ineffective. The answers to these questions are the focus of this study: to determine whether it is possible to apply machine learning (ML) and deep learning (DL) to overcome such challenges and offer a more adaptive, accurate, and scalable approach to counter emerging spam threats [2]. In this pursuit, the study endeavours to enhance the strength and functionality of the spam detection systems to better protect the users through exploitation of the above advanced technologies.

D. Objective of the system

The aim is to assess the performance of ML and DL methods to identify emails as spam. It is intended to evaluate Naive Bayes, Support Vector Machines (SVM), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks with several issues like false positives and concept shift [3]. The aim is to deal with the problem of identifying emerging email spam, which has drawbacks in classical approaches. This paper focuses on the ability of applying ML and DL into the increased detection rate, capability and flexibility with the hope of boosting the effectiveness of spam detection system.

II. LITERATURE SURVEY

E-mail spam, including phishing scams, malware attacks, and fraudulent messages, remains the biggest threat to cybersecurity. The rule-of-thumb method, which uses filters and blacklists, was the initial way used but has now failed due to advancements in techniques used by spammers. More sophisticated techniques such as machine learning (ML) and deep learning (DL) are also more scalable and adaptable

Received: 10 -12 -2024
Revised: 7- 11 -2026
Published: 31-12-2026

© 2026 The Author(s). This article is published under the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0). This license permits others to share, copy, distribute, remix, adapt, and build upon the material, provided they give appropriate credit to the original authors and source. Any adapted or derivative work must be distributed under the same CC BY-SA 4.0 license.

detection systems [4]. Naïve Bayes and Support Vector Machines (SVM), and Convolutional Neural
Networks (CNN) are other types of algorithms used in filtering out spam depending on statistical attributes and content. Conversely, the following issues are a cause of concern: false positive results, unbalanced datasets, and the potential need for constant updates.
Recently, deep learning models, especially LSTM because of their capability to manage nonlinear sequential data and also explain the features. Hybrid models of CNNs and LSTMs have been researched well, and the results suggest enhanced accuracy, as well as flexibility. However, challenges such as the change of concept and the high computation complexity remain a challenge. There is an attempt to propose novelties such as transfer learning, federated learning, and explainable artificial intelligence (XAI) to overcome these challenges [5]. These innovations are intended to improve the reliability, speed, and security of the spam identification tools, and thus, better protect against the continuing spam threats and contribute to the improvement of the overall security situation.

III. System Design

The system utilizes state-of-the-art AI techniques to detect email spam with high accuracy and adaptability. By integrating advanced machine learning and deep learning methods, it addresses evolving spam tactics while ensuring scalability and robustness. Transparency and interpretability are prioritized, making the system practical for real-world cybersecurity applications. The system architecture, detailing its components for effective spam classification, is shown in Fig. 1 below.

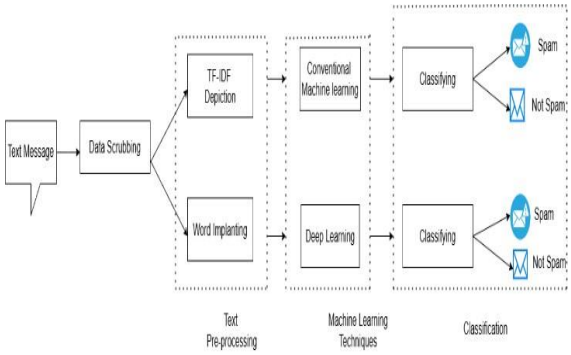


Fig. 1. Architecture of spam detection.

A. Text message

This is the natural language text message, which in most cases will be an email or an SMS that requires classification. It includes all information for spam classification, including the unwanted one and the desired one. The system processes this to determine often spam.

B. Data scrubbing

This step works to standardize the text by eliminating noise in the text the noise could be punctuation, stop words, or special characters. It lowers down the text to be processed and then splits the text into words. It is done to get a standardized input for the next steps.

C. TF-IDF depiction

It converts the text data into numerical values by applying the technique Term Frequency-Inverse Document Frequency (TF-IDF) [6]. It underlines certain words by giving a score of their importance in a document in relation to the given set. These scores are supplied to machine learning analysis algorithms.

D. Conventional machine learning

Naive Bayes or SVM are traditional models where TF-IDF is used to detect spam in the given structure. They use features of words and their joint distribution, which help separate spam from non-spam [7]. Classification is a process where a system uses data learned during training to classify an input signal.

E. Word implanting

Word vectors convert words to vectors into high-dimensional vectors that represent their meaning. They retain connections between words as in the case of offer and discount. These vectors make it possible for the deep learning models to know what context of meaning the text holds.

F. Deep learning

Word embeddings are then fed to more complex neural networks like Recurrent Neural Networks, or even Transformers. The models especially perform well in identifying long-range dependencies and other inconspicuous patterns. They improve spam identification, especially in cases when a message might contain other texts or is ambiguous.

G. Output (spam or not spam)

The system provides the decision-making for filtering or quarantining the spam messages in a message list. Normal messages, on the other hand, are transferred to the user's inbox. This brings efficiency in spam management or message delivery to their recipients.

IV. Methodology

The process of diagnosing spam messages through deep learning and machine learning follows a structured methodology comprising the following steps:

Step 1. Rule based systems

Keyword based filtering

Filters Email with the help of looking for certain words or symbols that are related to spam, for instance "win"/" prize" [8]. As observed, it is easier to implement and effective, especially in dealing with easy spam patterns. Still, it faces issues with spammers who provide misspelled keywords or never use them in the search at all.

Static rule sets

A fixed rule set is used to sort the emails according to their specific traits as defined beforehand, phrases for instance. These systems are quite effective for identification of certain types of spam but cannot adapt to different novel and ever-

changing spam approaches. Therefore, they are applied together with other more flexible current detector techniques.

STEP 2. BLOCKLISTS AND REAL-TIME BLACKHOLE LISTS (RBLS)

Blocking known spam sources

Blocklists, on the other hand, enable one to block corresponding email from a spammer's IP or domain [9]. If used as filters to reidentify new sources, they are useful for combating continual spam but are not helpful for new spam sources. Dynamic IPs even lower the reliability of these techniques.

Real-Time Updates with RBLs

RBLs build on blocklists by automatically updating a database of undesirable sources. This enables real time identification of spam and subsequently have it blocked. However, the difficulty with false positive results and testing at a big scale continue to be problems.

STEP 3. MACHINE LEARNING (ML) TECHNIQUES

Learning patterns from data

The Naïve Bayes model and other ML models operate under the statistical premise and examine the past messages to locate patterns in spam emails [10]. They also can learn from large volumes of data and categorise new emails according to a set of characteristics observed previously, to name a few like word frequencies and the structure of the emails. This can enable the avoidance of new spam tactics that the mechanisms of rule-based systems can miss detecting.

Flexibility in detection

ML techniques can learn new tricks in combating a type of spam and flexibly switch to using them, for example, in combating phishing or polymorphic spam. This flexibility makes them a strong weapon for flexible spam identification. However, the computational demands and overfitting issues are the limitations of the study.

STEP 4. DEEP LEARNING (DL) TECHNIQUES

Structural analysis with CNNs

The Convolutional Neural Networks (CNNs) considers structure and content of the e-mail messages in its pursuit of spam. Bots are especially proficient at analysing URLs and format. Their large number of computations may restrict the degree to which they can be used in smaller systems.

Sequential analysis with LSTMs

Long Short-Term Memory networks (LSTMs) incorporate time series into email text and are effective at identifying phishing and advanced spam [11]. However, they are expensive to conduct, and the results are often difficult to interpret.

STEP 5. FUZZY LOGIC AND FUTURE SELECTION

Fuzzy Logic for spam detection

Fuzzy logic provides flexibility in the membership level, which makes it easier to categorize emails. This reduces VFP and NVFP since there occurs uncertainty in the classification of email content as spam or not spam.

Feature selection methods (TF-IDF, Word2Vec)

TF-IDF represent words' significance and has value difference as Word2Vec shows words' connections. These methods minimize noise features and enhance the effectiveness of given spam detection models.

STEP 6. HYBRID DEEP LEARNING MODEL

Combining CNNs and LSTMs

In a hybrid approach, we use both CNNs and LSTMs to take the best from the structural and sequential methods [12]. This combination improves the outlook in terms of accuracy and versatility and uniquely positions the offering to counter complicated spam strategies. However, these models require considerable computation power, which is an important drawback of these models.

Enhanced Detection with Hybrid Approaches

The hybrid models provide improved performance of complex spam detection since they use various deep learning approaches. They are suitable for phishing and polymorphic attack and yet, they need large amounts of training data and powerful computational strengths.

The various methods of email spam detection are shown in the figure 2 below.

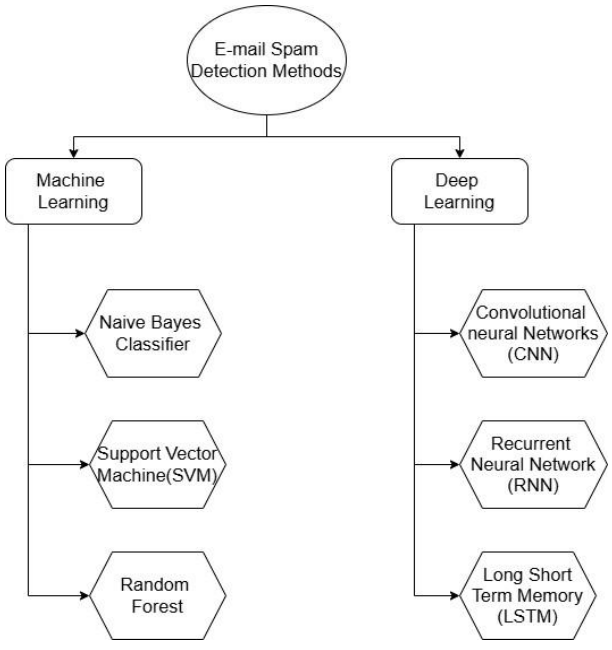


Fig. 2. Spam detection methods

V. FUZZY LOGIC IN SPAM DETECTION

Fuzzy logic offers a reliable approach for dealing with the shroud of vagueness evident in spam categorization. Fuzzy logic does not belong to the binary category as it considers likelihood of the outcomes therefore making it easy for it to

categorize the emails as per the probability aspect. This is especially helpful in spam filtering since most real messages contain words similar to those of spam (for example, 'free' or 'win'). As a result, fuzzy logic can attach these e-mails less degree of sameness and decrease false positives.

A. Methods of Fuzzy Logic in Spam Detection:

Fuzzification

The process starts with converting quantitative crisp input values such as the number of times specific keywords (for example, 'free', 'win') have been used into fuzzy values. These fuzzy values define how much specific terms are related to spam content. For instance, a word is marked as 'urgent,' the sameness degree of this word is greater.

After fuzzifying the inputs, a Fuzzy Inference System (FIS) makes use of fuzzy rules whereby it assesses the extent to which the received email is spam. For instance, with the presence of frequently used words such as 'urgent' and 'win ', the system could easily predict that the email in question is most probably spam. The fuzziness is then removed in making clean-cut decisions such as "spam" or "not spam."

Adapting thresholding in fuzzy systems

The operation of fuzzy logic systems shows that adaptive thresholding can improve the efficiency of spam detection. This technique then adapts thresholds relative to changing email patterns and individual preferences. For example:

In the case where legitimate emails are often wrongly classified as spam (false positives), the system adjusts the weights of fuzzy rules to minimize errors. This improves the correctness of the system and guarantees the possibility of reacting to changing spam strategies, which makes the approach to spam detection more personal and efficient. A fuzzy logic system is flexible in detecting spam patterns, unlike regular filters that only use lists of keywords and patterns. Fuzzy logic compares the degree of match instead of looking for a perfect match; then it will be able to pick variations such as fr€€ or f.re.e. It can also periodically update its rules of detection by using data from large data sets or user feedback. Also, compatibility with streaming threat feed and implement concept drift. For this reason, fuzzy logic systems are very effective against the current modern tactics of spam. The fuzzy logic architecture for spam detection is described in Table 1.

B. How fuzzy logic helps in spam detection

Fuzzy logic improves spam detection because it measures the graduated level of spam probability in emails. Compared to the conventional systems where Emails are categorized as either spam or not-spam can be done using fuzzy logic. This flexibility reduces false positives (when a legitimate message is classified as spam) and false negatives (when a spam message sneaks in), especially at the edge where the classification is most ambiguous. The architecture of the fuzzy logic system is shown in Figure 3 below.

VI. CHALLENGES IN SPAM DETECTION

Some issues include spam attack techniques, changing spam patterns, imbalanced data, and workload. This includes text camouflage, changing URLs, and creating polymorphic spam, which is why spam changes constantly.

Another factor that makes the detection difficult is concept drift; given that spam content and/or patterns change periodically, the model needs frequent updates. Imbalance in datasets generalizes models to classify emails as non-spam because they outnumber spam emails; this inequality increases the false negative rate.

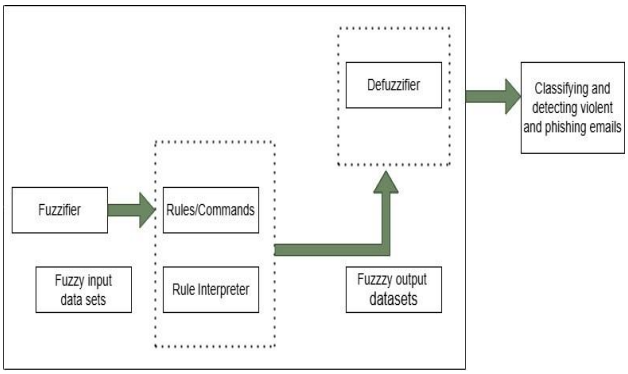


Fig.3. Fuzzy logic system architecture. described in Table 1 below.

Table 1. Fuzzy logic architecture.

Component	Description	Example	Advantage
Input parameters	Selection criteria include keywords, length of the email desired and sender's domain.	This brings terms such as 'win, free' along with sender's domain.	Evaluates the email contents in detail.
Fuzzification	Transforms input values into fuzzy degrees – for example, "urgent" is 0.8 which means high chance of spam.	Percentage of words in the 'urgent' fuzzified means that there is high probability of the message having spam into 0.8.	Maintains flexibility in decision making.
Fuzzy Inference System (FIS)	Fuzzy rules are applied to compute numerical characteristics of email and then to conclude it as spam email or not everything.	If "win" is equal to 0.9 then, "urgent" is equal to 0.8 and so email is 0.85 likely spam.	Can change with spam tactics.
Defuzzification	Translates the ambiguous decision back into a clean-cut decision of defining the email as spam or not.	Email containing 0.85 membership referred as spam	Provides precise decisions.

Also, current deep learning models need high computational power, and many are considered ‘’black boxes’’ thus lacking in interpretability [19]. Such things affect the precision, flexibility, and speed of the antispam measures, which presents the need for the creation of effective, comprehensible, and versatile approaches combating complex spam strategies, altering URLs, or creating polymorphic spam, to bypass detection. Concept drift further complicates detection, as spam content and patterns continuously evolve, requiring frequent model updates. Data imbalance, where legitimate emails far outnumber spam, can bias models toward classifying emails as non-spam, leading to higher false negatives. The challenges faced in spam detection is described in Table 2 below.

Table 2. Challenges in spam detection [20]

Challenge	Description	Impact on spam detection
Evasive strategies by spammers	Spammers employing misspellings, disguises, or random text	The effectiveness of keywords Based filtering is reduced
Concept drift	Spam strategies evolve gradually	Models that are trained on old data become ineffective
Divergence datasets	Large number of official emails, fewer spam content	Reduces the efficiency of Machine learning models

VII. CONCLUSION

This paper focuses on the evolution of techniques in spam detection, highlighting a transition from rule-based systems to advanced ML and DL models, including CNNs and LSTMs. These approaches have brought higher accuracy and adaptability, but challenges like changing tactics in spam, imbalance in data, and computational demands still prevail. This system will be lightweight, adaptive, and transparent to handle emerging spam patterns.

REFERENCES

[1] Tusher, Ekramul Haque, et al. "Email Spam: A Comprehensive Review of Optimize Detection Methods, Challenges, and Open Research Problems." IEEE Access (2024).

[2] Ahmad, Sheikh Muhammad. Spam Classification Using Machine Learning and Deep Learning. Diss. Dublin Business School, 2024.

[3] Abimbola, Bolanle, Enrique de La Cal Marin, and Qing Tan. "Enhancing Legal Sentiment Analysis: A Convolutional Neural Network–Long Short-Term Memory Document-Level Model." Machine Learning and Knowledge Extraction 6.2 (2024): 877-897.

[4] Kocher, Geeta, and Gulshan Kumar. "Machine learning and deep learning methods for intrusion detection systems: recent developments and challenges." Soft Computing 25.15 (2021): 9731-9763.

[5] Wewaldeni Pathirannehelage,Yasintha. Analysis of centralized to federated learning-based anomaly detection in networks with explainable AI (XAI). MS thesis. Y. Wewaldeni Pathirannehelage, 2022.

[6] Alfarizi, Muhammad Ibnu, Lailis Syafaah, and Merinda Lestandy. "Emotional text classification using tf-idf (term frequencyinverse document frequency) and lstm (long short-term memory)." JUITA: Jurnal Informatika 10.2 (2022): 225-232.

[7] Amayri, Ola, and Nizar Bouguila. "A study of spam filtering using support vector machines." Artificial Intelligence Review 34 (2010): 73-108.

[8] Zdziarski, Jonathan A. Ending spam: Bayesian content filtering and the art of statistical language classification. No starch press, 2005.

[9] Ramachandran, Anirudh, Nick Feamster, and Santosh Vempala. "Filtering spam with behavioral blacklisting." Proceedings of the 14th ACM conference on computer and communications security. 2007.

[10] Gomes, Sebastian Romy, et al. "A comparative approach to email classification using Naive Bayes classifier and hidden Markov model." 2017 4th international conference on advances in Electrical Engineering (ICAEE). IEEE, 2017.

[11] Hassan, Mohamed. "Evaluation of Deep Learning Algorithms in Comparison for Phishing Email Identification." Applied Mathematics on Science and Engineering 1.1 (2024): 21-35.

[12] Tasdelen, Abdulkadir, and Baha Sen. "A hybrid CNN-LSTM model for pre-miRNA classification." Scientific reports 11.1 (2021): 14125.