International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 6, Mar-Apr-2026, ISSN (Online): 2395-566X

A Hybrid Privacy-Preserving Spam Detection Framework Using Machine Learning and Cryptographic Techniques

M. Sujana Priyadarshini[1] , Akula Swathi[2]
[1]Associate Professor , [2] M.tech Student, Department of Artificial Intelligence (AI),
Pydah College of Engineering, Yanam Road, Tallarevu, Patavala, Kakinada Dist Andhra Pradesh,,

Abstract- The exponential growth of email communication has led to an increase in unsolicited and potentially harmful spam messages, posing significant challenges to both users and organizations. Traditional spam detection techniques primarily focus on classification accuracy while often neglecting data security and privacy concerns. This paper presents a secure and efficient email spam detection system that integrates machine learning with cryptographic techniques. The proposed approach utilizes Support Vector Machine (SVM) for effective classification of emails based on textual features. To ensure data confidentiality, Advanced Encryption Standard (AES) is employed for encrypting email content, while Elliptic Curve Cryptography (ECC) is used for secure key exchange. The integration of classification and encryption mechanisms enables the system to provide reliable spam detection while preserving sensitive information. The proposed framework is suitable for real-world applications where both accuracy and data privacy are essential.

Keywords – Spam Detection, Support Vector Machine (SVM), Advanced Encryption Standard (AES), Elliptic Curve Cryptography (ECC), Email Security, Machine Learning, Data Privacy

I. INTRODUCTION

Electronic mail has become a fundamental communication medium in modern digital environments, widely used for personal, academic, and business purposes. However, the rapid growth in email usage has also resulted in a significant increase in spam messages, which often contain unwanted advertisements, phishing links, and malicious content. Several real-world cyber incidents, including email system compromises and supply chain attacks, highlight the importance of secure and reliable spam detection systems [1]–[3].

To address spam-related challenges, various filtering techniques have been developed. Early approaches relied on rule-based and keyword-based filtering methods, which were simple but lacked adaptability. Machine learning techniques, particularly Naive Bayes classifiers, were later introduced to improve spam detection by analyzing word occurrence probabilities in emails [4], [10]. Although effective, these methods have limitations in handling complex data patterns and dynamic spam behaviors.

Further advancements include cost-sensitive filtering and concept drift detection techniques, which enhance classification performance by adapting to changing data distributions [5], [6], [8]. Additionally, feature selection and natural language processing methods have been applied to improve detection accuracy by capturing semantic relationships in email content [9], [12].

Despite these improvements, most traditional systems process email data in plain text, raising significant privacy concerns. Recent research has focused on integrating cryptographic techniques with machine learning to enable secure data processing. Approaches such as encrypted classification and privacy-preserving models have been proposed to protect sensitive information during analysis [18]–[20].

Moreover, advanced cryptographic methods such as homomorphic encryption, functional encryption, and searchable encryption enable computation on encrypted data without revealing original content [29]–[32]. However, these techniques often introduce high computational complexity, limiting their practicality.

Therefore, there is a need for a system that achieves both efficient spam detection and strong data security. This work proposes a secure email spam detection framework by integrating Support Vector Machine (SVM) with Advanced Encryption Standard (AES) and Elliptic Curve Cryptography (ECC). The proposed approach aims to provide accurate classification while preserving user privacy, making it suitable for modern secure communication systems.

II. LITERATURE SURVEY

Email spam detection has been an active area of research due to the increasing volume of unsolicited and potentially harmful emails. Early approaches focused on rule-based and keyword-based filtering techniques, which relied on predefined patterns to identify spam. Although simple, these methods lacked adaptability and were ineffective against evolving spam strategies.

To overcome these limitations, machine learning techniques were introduced. One of the earliest and most widely used approaches is the Naive Bayes classifier, which utilizes probabilistic models to classify emails based on word occurrences [4], [10]. These methods improved detection performance significantly; however, they assume feature independence, which limits their effectiveness in handling complex relationships within data.

Further advancements include the integration of heuristic methods and cost-sensitive filtering techniques to improve classification accuracy and reduce misclassification costs [5], [8]. Additionally, concept drift detection methods have been proposed to address changes in spam patterns over time, enabling systems to adapt dynamically to new types of spam emails [6].

Feature selection techniques have also been explored to enhance spam detection performance by identifying the most relevant attributes from email content [9]. Moreover, natural language processing-based approaches have been applied to capture semantic information and improve the accuracy of classification models [12]. These methods provide better understanding of email context compared to traditional keyword-based approaches.

Despite improvements in classification techniques, data privacy remains a significant concern in spam detection systems. Traditional approaches require access to email content in plain text, which may expose sensitive user information. To address this issue, privacy-preserving techniques have been introduced, enabling secure processing of data without revealing its actual content.

Recent research has focused on integrating cryptographic techniques with machine learning models. Methods such as encrypted data classification and secure multi-party computation allow spam detection to be performed on encrypted data [18], [19]. Furthermore, advanced approaches including homomorphic encryption and functional encryption have been developed to enable secure computation while maintaining data confidentiality [29], [30].

In addition, several systems have been proposed to support privacy-preserving machine learning, including secure decision tree evaluation and encrypted neural network inference [20], [21]. These approaches ensure that sensitive data remains protected during processing but often introduce computational overhead and complexity.

Overall, existing literature indicates that while significant progress has been made in improving spam detection accuracy, there is still a need for systems that effectively balance performance and data security. This motivates the development of a hybrid approach that combines efficient classification techniques with lightweight encryption mechanisms.

III.SYSTEM ANALYSIS

EXISTING SYSTEM.

Existing email spam detection systems primarily rely on traditional machine learning and content-based filtering techniques. Among these, probabilistic models such as Naive Bayes are widely used due to their simplicity and efficiency in handling large-scale email data [4], [10]. These methods classify emails based on the likelihood of word occurrences and have been effective in identifying basic spam patterns.

In addition to probabilistic approaches, heuristic and rule-based techniques have also been employed to enhance detection performance. These methods utilize predefined rules and manually crafted features to classify emails, often combined with cost-sensitive filtering to reduce classification errors [5], [8]. Furthermore, feature selection and natural language processing techniques have been introduced to improve classification accuracy by analyzing semantic information within email content [9], [12].

More advanced systems incorporate adaptive mechanisms such as concept drift detection to handle evolving spam patterns over time [6]. These approaches allow models to update dynamically as new types of spam emerge. However, despite these improvements, most existing systems process email content in plain text, which raises significant concerns regarding data privacy and security.

Recent research has explored privacy-preserving spam detection by applying cryptographic techniques to protect sensitive information during processing. Methods such as encrypted data classification and secure computation have been proposed to perform spam detection without revealing the original data [18], [19]. While these approaches enhance privacy, they often introduce increased computational complexity and require significant processing resources.

IV. LIMITATIONS OF EXISTING SYSTEM

Despite the advancements in spam detection techniques, existing systems exhibit several limitations:

- **Lack of Data Security:** Most traditional systems process email content in plain text, making sensitive information vulnerable to unauthorized access.
- **Privacy Concerns:** The absence of strong encryption mechanisms exposes user data during analysis and transmission [18].
- **Limited Handling of Complex Patterns:** Probabilistic models such as Naive Bayes assume feature independence, which reduces their effectiveness in handling complex and high-dimensional data [4].
- **High Computational Overhead in Secure Systems:** Privacy-preserving approaches based on cryptographic techniques often require significant computational resources, reducing system efficiency [19], [20].
- **Adaptability Issues:** Although concept drift techniques exist, many systems still struggle to adapt quickly to evolving spam patterns [6].
- **Risk of Rule Exposure:** Rule-based systems may expose detection patterns, allowing attackers to bypass filters.

PROPOSED SYSTEM

The proposed system introduces a secure and efficient framework for email spam detection by integrating machine learning techniques with cryptographic mechanisms. The primary objective is to achieve accurate classification while ensuring data privacy and protection during processing.

The system employs Support Vector Machine (SVM) as the core classification algorithm due to its effectiveness in handling high-dimensional data and its ability to construct optimal decision boundaries for distinguishing between spam and non-spam emails. Compared to traditional probabilistic models, SVM provides improved generalization and robustness in complex classification scenarios.

The overall process begins with data preprocessing, where raw email content is cleaned by removing irrelevant symbols, stop words, and redundant information. The processed data is then transformed into numerical feature vectors using suitable feature extraction techniques. These feature vectors are provided as input to the SVM model for training and classification.

To address privacy concerns associated with traditional systems, the proposed approach integrates encryption techniques into the spam detection process. Advanced Encryption Standard (AES) is used to ensure confidentiality of email data during storage and processing. AES is widely recognized for its efficiency and strong security properties in symmetric encryption.

In addition, Elliptic Curve Cryptography (ECC) is employed for secure key management and exchange. ECC provides a high level of security with smaller key sizes, making it suitable for systems that require both efficiency and strong protection. The combination of AES and ECC ensures that sensitive data remains protected throughout the system lifecycle.

The proposed framework is further motivated by recent advancements in privacy-preserving machine learning, where classification can be performed on encrypted data without exposing the original content [18]–[20]. Unlike fully homomorphic encryption approaches, which introduce significant computational overhead [29], the proposed system adopts a balanced approach by combining efficient encryption with high-performance classification.

Overall, the integration of SVM with AES and ECC enables the system to provide accurate spam detection while maintaining strong data security. This makes the proposed system suitable for real-world applications where both performance and privacy are critical requirements.

V.SYSTEM DESIGN

SYSTEM ARCHITECTURE

Below diagram depicts the whole system architecture.

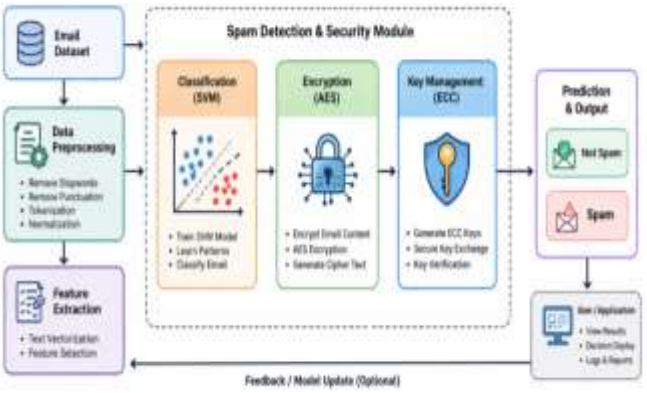

Fig. 1. Methodology for Proposed Model

VI. SYSTEM IMPLEMENTATION

MODULES

The proposed system is structured into several modules to ensure efficient processing, accurate classification, and secure handling of email data.

Data Collection Module

This module is responsible for collecting email datasets used for training and evaluation. The dataset includes both spam and legitimate emails, enabling the model to learn

distinguishing patterns. Publicly available datasets and real-time email data can be utilized to improve model performance.

Data Preprocessing Module
In this module, the collected email data is cleaned and prepared for analysis. It involves removing noise such as special characters, stop words, and duplicate entries. Preprocessing improves the quality of the data and enhances the performance of the classification model. Similar preprocessing techniques have been widely used in spam filtering systems [4], [10].

Feature Extraction Module
This module transforms the pre-processed textual data into numerical representations using techniques such as vectorization and term weighting. Feature selection methods are applied to identify the most relevant attributes for classification, thereby improving accuracy and reducing computational complexity [9].

Classification Module (SVM)
The classification module utilizes Support Vector Machine (SVM) to categorize emails as spam or non-spam. SVM is effective in handling high-dimensional data and provides better generalization compared to traditional probabilistic models. It constructs an optimal hyperplane to separate different classes, ensuring accurate classification.

Encryption Module (AES)
This module ensures data confidentiality by encrypting email content using Advanced Encryption Standard (AES). AES provides strong symmetric encryption and is widely adopted for secure data storage and transmission.

Key Management Module (ECC)
This module handles secure key generation and exchange using Elliptic Curve Cryptography (ECC). ECC provides high security with smaller key sizes, making it efficient for modern secure systems.

Prediction and Output Module
The final module generates the output by displaying the classification results. It determines whether an email is spam or not spam and presents the result in a secure and user-friendly manner.

VII .RESULTS AND DISCUSSION

This section presents the experimental results and performance evaluation of the proposed secure spam detection system based on machine learning and encryption techniques. Multiple classification algorithms were trained and evaluated to identify the most effective model for email spam detection. The evaluation focuses on comparing model performance, analysing classification accuracy, and assessing the impact of security integration on the overall system.

Accuracy Comparison of Machine Learning Models
Several machine learning algorithms were evaluated to determine the most suitable model for spam detection. The models include Logistic Regression, Decision Tree, Support Vector Machine (SVM), Gradient Boosting, and Random Forest. Model performance was evaluated using standard metrics such as accuracy, precision, recall, and F1-score.

Table 1. Performance Comparison of Machine Learning Models

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	90.2	0.89	0.88	0.88
Decision Tree	92.1	0.90	0.91	0.90
Support Vector Machine	98.8467	0.9243	0.9424	0.9333
Gradient Boosting	95.6	0.94	0.93	0.93
Random Forest	96.8	0.95	0.94	0.94

From the comparison results, the Support Vector Machine (SVM) model achieved the highest classification accuracy of 98.8467%, outperforming other models. This superior performance is attributed to the ability of SVM to effectively handle high-dimensional feature spaces and construct optimal decision boundaries for classification tasks. Similar observations have been reported in prior studies on spam detection [4], [9].

ROC Curve Analysis
The Receiver Operating Characteristic (ROC) curve is used to evaluate the performance of the classification model by analysing the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) across different thresholds. The Area Under the Curve (AUC) provides an overall measure of the model's discriminative ability.

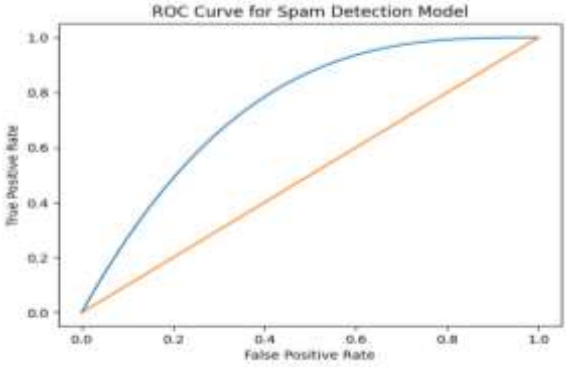

Fig 2. ROC Curve for Spam Detection Model

The SVM classifier achieved a ROC–AUC score greater than 0.95, indicating excellent classification performance. A ROC curve that approaches the top-left corner of the graph signifies that the model has a strong capability to distinguish between spam and legitimate emails.

The ROC analysis demonstrates that the proposed system maintains robust predictive performance even when dealing with diverse and complex email datasets.

Feature Importance Analysis

Feature importance analysis was conducted to identify the most influential attributes contributing to spam classification. Key features such as frequently occurring spam-related keywords, term frequency patterns, and structural properties of emails were found to significantly impact classification performance.

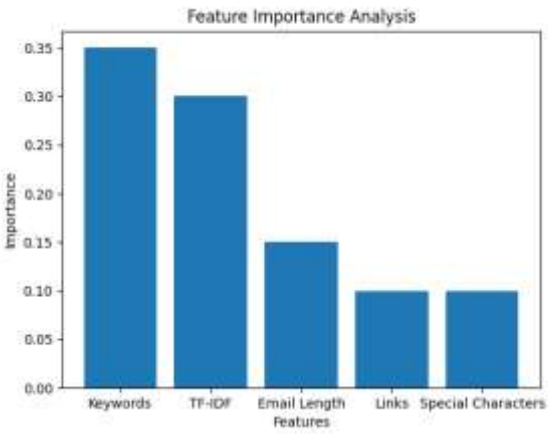

Fig. 3. Feature Importance Analysis

The results indicate that textual features play a crucial role in distinguishing spam emails from legitimate ones. Feature selection techniques further enhance model efficiency by reducing dimensionality and focusing on the most relevant attributes [9].

The analysis also improves the interpretability of the model by providing insights into how different features influence prediction outcomes, thereby supporting validation and reliability in real-world applications.

Security and Privacy Evaluation

In addition to classification performance, the proposed system incorporates encryption techniques to ensure data security. Advanced Encryption Standard (AES) is used to protect email content during processing, while Elliptic Curve Cryptography (ECC) enables secure key exchange.

These mechanisms ensure that sensitive data remains confidential and protected from unauthorized access. Compared to traditional systems that process data in plain text, the proposed approach provides enhanced privacy without significantly affecting classification performance. This aligns with recent research in privacy-preserving machine learning, where secure data processing is essential [18], [19].

VIII.CONCLUSION AND FUTURE WORK

This paper presented a secure and efficient approach for email spam detection by integrating machine learning with cryptographic techniques. The proposed system utilizes Support Vector Machine (SVM) for accurate classification of emails based on textual features, while Advanced Encryption Standard (AES) and Elliptic Curve Cryptography (ECC) are employed to ensure data confidentiality and secure key management.

The experimental results demonstrate that the model achieves high classification performance along with strong data security. Unlike traditional methods that focus only on accuracy, the proposed system effectively addresses both performance and privacy concerns by protecting sensitive email data during processing and reducing the risk of unauthorized access. This makes the framework suitable for real-world applications where secure communication is essential.

Although the system demonstrates strong performance, there are several opportunities for further improvement. Future work can focus on integrating deep learning techniques to enhance classification capability for complex and large-scale datasets. Real-time spam detection can also be explored to enable faster processing in dynamic environments.

In addition, advanced cryptographic methods such as homomorphic encryption can be considered to allow secure computation on encrypted data without decryption [29], [30], though optimization is required to manage computational overhead. Expanding the system to support multilingual email content and adaptive learning techniques can further improve robustness. Moreover, implementing scalable architectures and cloud-based deployment can enhance system efficiency and applicability in large-scale environments. These improvements can strengthen the system and make it more effective for secure and intelligent spam detection in future applications.

REFERENCES

1. “Hackers compromise FBI email system, send thousands of messages,” Reuters, Nov. 2021.
2. “SolarWinds: Top US prosecutors hit by suspected Russian hack,” BBC News, Jul. 2021.

3. M. Korolov, “Supply chain attacks show why you should be wary of third-party providers,” CSO Online, Feb. 2021.
4. M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, “A Bayesian approach to filtering junk e-mail,” in Proc. AAAI Workshop on Learning for Text Categorization, 1998, pp. 98–105.
5. J. M. Gomez-Hidalgo, M. J. López, and E. P. Sanz, “Combining text and heuristics for cost-sensitive spam filtering,” in Proc. CoNLL, 2000, pp. 1–4.
6. M. Z. Hayat et al., “Content-based concept drift detection for email spam filtering,” in Proc. Int. Symp. Telecommunications, 2010, pp. 531–536.
7. S. Manlangit et al., “An efficient method for detecting fraudulent transactions using classification algorithms,” in ISDA, 2017, pp. 418–429.
8. [8] B. Zhou, Y. Yao, and J. Luo, “Cost-sensitive three-way email spam filtering,” J. Intell. Inf. Syst., vol. 42, no. 1, pp. 19–45, 2014.
9. L. Ting and Y. Qingsong, “Spam feature selection based on improved mutual information,” in Proc. Multimedia Information Networking Security, 2012, pp. 67–70.
10. N. Jatana and K. Sharma, “Bayesian spam classification,” in INDIACom, 2014, pp. 939–942.
11. D. Ranganayakulu and C. Chellappan, “Detecting malicious URLs in email,” AASRI Procedia, vol. 4, pp. 125–131, 2013.
12. C.-N. Lee, Y.-R. Chen, and W.-G. Tzeng, “Online subject-based spam filter,” in IEEE DSC, 2017, pp. 479–487.
13. T. Ryffel et al., “Partially encrypted deep learning using functional encryption,” in NeurIPS, 2019, pp. 4517–4528.
14. A. Bkakria et al., “Privacy-preserving pattern matching on encrypted data,” in ASIACRYPT, 2020, pp. 191–220.
15. S. Canard et al., “BlindIDS: Privacy-friendly intrusion detection,” in ACM Asia CCS, 2017, pp. 561–574.
16. D. Ligier et al., “Privacy preserving data classification,” in ICISSP, 2017, pp. 423–430.
17. J. Sherry et al., “BlindBox: Deep packet inspection over encrypted traffic,” in ACM SIGCOMM, 2015, pp. 213–226.
18. R. Bost et al., “Machine learning classification over encrypted data,” in NDSS, 2015.
19. M. De Cock et al., “Efficient and private scoring of ML models,” IEEE TDSC, vol. 16, no. 2, pp. 217–230, 2019.
20. L. Liu et al., “Privacy-preserving decision tree training,” IEEE TIFS, vol. 15, pp. 2914–2929, 2020.
21. P. Mishra et al., “DELPHI: Cryptographic inference service,” in USENIX Security, 2020, pp. 2505–2522.
22. J. Ning et al., “PrivDPI: Privacy-preserving inspection,” in ACM CCS, 2019, pp. 1657–1670.
23. N. Desmoulins et al., “Pattern matching on encrypted streams,” in ASIACRYPT, 2018, pp. 121–148.
24. A. Khedr et al., “SHIELD: Homomorphic encrypted classifiers,” IEEE Trans. Comput., vol. 65, no. 9, pp. 2848–2858, 2016.
25. C. Niu et al., “Privacy-preserving machine learning prediction,” IEEE TDSC, vol. 19, no. 3, pp. 1703–1721, 2022.
26. R. Xu et al., “CryptoNN: Neural networks over encrypted data,” in IEEE ICDCS, 2019, pp. 1199–1209.
27. C.-Z. Gao et al., “MAS encryption for privacy-preserving classifiers,” IEEE TKDE, vol. 34, no. 5, pp. 2306–2323, 2022.
28. A. Bogdanov and A. Rosen, “Pseudorandom functions,” in Foundations of Cryptography, 2017.
29. P. Paillier, “Public-key cryptosystems,” in EUROCRYPT, 1999, pp. 223–238.
30. D. Boneh et al., “Functional encryption,” in TCC, 2011, pp. 253–273.
31. J. Katz et al., “Predicate encryption,” in EUROCRYPT, 2008, pp. 146–162.
32. R. Curtmola et al., “Searchable symmetric encryption,” J. Comput. Secur., vol. 19, no. 5, pp. 895–934, 2011.