Alkadhim Journal for Computer Science, Vol. 4, No. 2 (2024)

Open Access ISSN: 3007-1437 (Online), ISSN: 3007-1429 (Print)

IRAQI Academic Scientific Journals

Alkadhim Journal for Computer Science (KJCS)

Journal Homepage: https://alkadhum-col.edu.iq/JKCEAS

KJCS ALKADHIM JOURNAL FOR COMPUTER SCIENCE

Survey of SMS Spam Detection Techniques: A Taxonomy

[1]Hussein Alaa Al-kaabi*, [2]Ali Darroudi , [3]Ali Kadhim Jasim

[1]Ministry of Education Iraq, General Direction of Vocational Education, Al-Najaf – Iraq

[2]Department of Electrical Engineering, Sadjad University of Technology, Mashhad– Iran

[3]Department of Computer Technology Engineering, Imam Ja'afar Al-Sadiq University, Maysan - Iraq

Article information

Article history:

Received: Nov, 10, 2024
Accepted: Dec, 11, 2024
Available online: Dec, 25, 2024

Keywords:
SMS spam detection,
NLP,
Text classification
Deep learning
Machine learning

****Corresponding Author:***
Hussein Alaa Al-kaabi
Hussain.njf7@gmail.com

DOI:
https://doi.org/10.53523/ijoirVolxIxIDxx

Abstract

Short Message Service (SMS) spam remains a significant threat to users and businesses, with spammers constantly adopting more sophisticated techniques. This paper comprehensively surveys SMS spam detection methods, categorizing existing approaches into five primary groups: rule-based methods, traditional machine learning techniques, deep learning models, hybrid models, and ensemble methods. Each category is examined in detail, highlighting its strengths, limitations, and evolution. Rule-based methods, though historically significant, are limited by their inability to handle new or evolving spam tactics. Traditional machine learning techniques, such as Naive Bayes and support vector machines (SVM), offer improved accuracy but depend on handcrafted features. In contrast, deep learning models, including recurrent neural networks (RNN) and convolutional neural networks (CNN), excel in feature extraction and adaptability yet face challenges with model complexity and the need for large labeled datasets. Hybrid and ensemble methods combine the benefits of various models to improve performance, reduce bias, and enhance robustness. This review aims to provide a structured overview of the state of SMS spam detection, identify emerging trends, and suggest future research directions, including improving generalization, reducing data dependency, and exploring the integration of contextual information. The findings underscore the need for continued innovation to address the evolving landscape of SMS spam.

1. Introduction

SMS remains one of the most widely used communication channels globally, with billions of messages sent daily for personal, commercial, and transactional purposes. However, the widespread adoption of SMS has also led to a surge in unsolicited and harmful messages, commonly called SMS spam. SMS spam poses significant challenges to both users and businesses, ranging from financial fraud and data theft to the erosion of trust in mobile communication systems. For businesses, the proliferation of spam threatens customer relationships and undermines legitimate SMS marketing efforts [1]. The increasing sophistication of spam tactics, including deceptive content and advanced evasion techniques, has made traditional detection methods inadequate. As the landscape of SMS communication continues to evolve, there is a critical need for robust and adaptive spam detection mechanisms. Given these challenges, a comprehensive survey of existing SMS spam detection methods is essential to identify current trends, evaluate the effectiveness of various approaches, and uncover gaps in the literature that need further exploration. The primary objective of this survey is to provide a systematic and detailed review of the state-of-the-art methods in SMS spam detection. This includes an examination of both traditional approaches, such as rule-based and machine learning techniques, and more recent advancements in deep learning and hybrid models. Additionally, this survey seeks to categorize these methods into a coherent taxonomy, offering a structured framework for understanding the evolution and diversity of approaches in the field. By doing so, the survey aims to identify prevailing trends, highlight the strengths and limitations of existing methods, and propose potential directions for future research. Ultimately, this paper aspires to serve as a comprehensive resource for researchers and practitioners, guiding the development of more effective and scalable solutions for SMS spam detection. This survey covers many SMS spam detection methods documented in the literature over the past decade. The review focuses on research papers published between 2015 and 2024, encompassing journal articles and conference proceedings. The methods examined include rule-based, machine learning, deep learning, and hybrid approaches, with particular attention given to those that have demonstrated significant empirical results. Additionally, this survey considers various feature extraction techniques, datasets, and evaluation metrics employed in these studies. The scope also extends to identifying the challenges these methods face, such as handling imbalanced datasets, adapting to new types of spam, and ensuring real-time detection capabilities. The remainder of this paper is organized as follows. Section 2 is the related work shown in the survey papers that deal with SMS spam detection challenges. Section 3 provides a background on SMS spam, including its definition, evolution, and the challenges associated with its detection. Section 4 proposes a taxonomy for categorizing SMS spam detection methods into rule-based, machine-learning, deep-learning, hybrid, and ensemble approaches. Section 5 reviews the datasets commonly used for training and evaluating SMS spam detection models and the evaluation metrics that measure their effectiveness. Section 6 offers a comparative analysis of the methods discussed. Section 7 addresses the challenges and future directions for research in SMS spam detection. Finally, Section 8 concludes the paper with a summary of key findings and their implications for academic research and practical applications.

2. Related work

Several survey papers have explored various approaches to SMS spam detection. Al Saidat et al. [2] adopted a systematic review approach to evaluate the effectiveness of Natural Language Processing (NLP) and Machine Learning (ML) techniques. These investigations provide a comprehensive overview of advancements in detection methodologies. The literature consistently highlights the significant advantages of integrating ML classifiers with NLP for achieving more accurate and robust detection outcomes. Qazi et al. [3] emphasize that numerous SMS spam filtering strategies are regarded as leading solutions. In their study, they develop a classification framework of current methodologies, highlighting the widespread adoption of existing SMS anti-spam applications within the literature. Hanif et al. [4] comprehensively review machine learning and deep learning techniques for detecting, classifying, and filtering SMS spam in their study. The review uses various databases, including ResearchGate, Elsevier, Applied Sciences, and IEEE, to identify relevant studies. As SMS remains a more frequently used communication medium than email, this study offers an overview of ML and DL methods, graphical representation approaches, and automated spam filtering techniques previously implemented on Android platforms for SMS spam detection. The primary objective is to identify existing studies' limitations and propose future research directions. Also, Sajedi et al. [5] present a review that examines various machine learning and hybrid algorithms for detecting SMS spam, focusing on their accuracy. A total of 44 articles from sources such as ScienceDirect, Google Scholar, IEEE Explorer, and the ACM library were analyzed, identifying 28 methods, of which 15 were compared based on accuracy, strengths, and weaknesses using the Tiago spam dataset. The findings highlight the DCA algorithm, large cellular network method, and graph-based KNN as the most accurate techniques. Additionally, hybrid approaches for SMS spam detection are discussed. Kaddoura et al. [6] present a new review paper about the spam content on social media platforms, such as Facebook, Twitter, YouTube, SMS, and email, has made spam detection crucial. With the rise in social media usage, especially during the pandemic, users receive numerous messages, often struggling to identify spam. These messages may contain malicious links, fake accounts, news, reviews, and rumors. To enhance social media security, detecting and controlling spam is essential. Their paper provides an extensive survey of recent advancements in spam text detection and classification on social media. All the previous works have not thoroughly reviewed all the techniques used for spam message detection, and at the same time, they focused on short time periods. In our paper, we have provided an extensive review of all the techniques used over the past 10 years.

2. Background

SMS spam, commonly called "text spam," encompasses unsolicited and unwanted messages sent to many recipients via SMS. These messages typically aim to advertise products, services, or fraudulent schemes, often without the recipient's consent [7]. SMS spam can be categorized into several types, including promotional messages, phishing attempts (also known as "smishing"), and messages containing links to malware. The persistent nature of SMS spam poses significant risks to users, ranging from minor annoyances to severe threats such as financial fraud, identity theft, and privacy violations [8]. Unlike email spam, which can often be filtered out with relative ease, SMS spam is particularly problematic because of the personal and direct nature of text messaging. The ubiquity of mobile phones and the high open rates of SMS messages make users more vulnerable to these attacks, making SMS spam a critical issue that requires effective detection and mitigation strategies.

Since the inception of SMS in the early 1990s, spam messages have evolved from simple promotional content to highly sophisticated schemes. Initially, SMS spam largely comprised mass-marketing messages sent indiscriminately to a broad audience. However, as mobile phone usage increased and spam detection methods became more practical, spammers adapted by developing more targeted and deceptive tactics. Phishing, or "smishing," emerged as a prominent form of SMS spam, where attackers impersonate legitimate entities to trick users into revealing sensitive information [9]. Over time, spammers have exploited technological advancements, such as automated message generation and anonymized or spoofed sender information, to evade detection. The increasing use of link-based scams, which direct users to malicious websites, further complicates the landscape of SMS spam. This evolution underscores the need for advanced detection methods capable of adapting to the continually changing tactics employed by spammers. Detecting SMS spam presents several key challenges that stem from the evolving nature of spam tactics and the inherent limitations of detection methods. One of the primary challenges is the adaptability of spammers, who frequently modify their strategies to bypass existing filters. This includes altering message content, using obfuscated or misleading language, and leveraging new technologies to evade detection. Additionally, the balance between accuracy and complexity is a significant concern. While rule-based systems are simple and efficient, they often fail to capture modern spam's nuanced and dynamic nature. On the other hand, advanced machine learning and deep learning models offer higher accuracy but require substantial computational resources and large datasets for training, which may not always be feasible. Furthermore, the real-time detection of SMS spam is crucial, as delays in filtering can expose users to threats. Addressing these challenges requires a multi-faceted approach that combines the strengths of different detection methods while remaining flexible enough to counter the ever-evolving tactics of spammers.

3. Taxonomy of SMS Spam Detection Methods

SMS spam detection methods have evolved from basic rule-based approaches to advanced machine learning, deep learning, and hybrid models. As illustrated in Figure 1, each category addresses different aspects of the spam detection challenge.

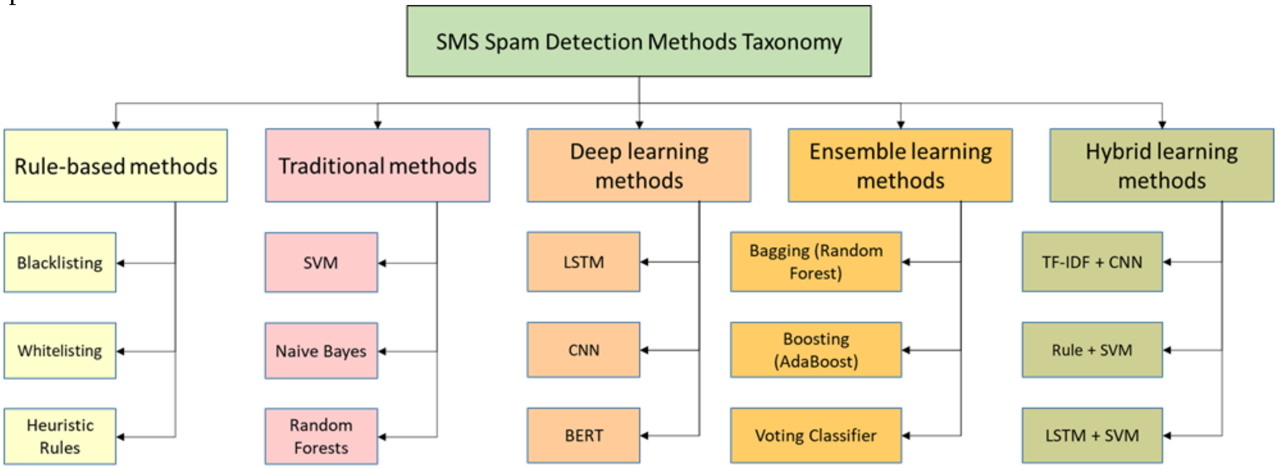

Figure 1. SMS spam detection approaches taxonomy

3.1 Rule-Based and Keyword-Based Approaches

Early methods for SMS spam detection predominantly employed rule-based and keyword-based approaches, which relied on predefined rules or keywords to identify spam messages. These systems flagged messages containing standard spam-related terms, such as "win," "free," or "urgent," as potential spam [10]. While these methods were straightforward and easy to implement, they lacked adaptability to evolving spam tactics [11] and often produced high false favorable rates, misclassifying legitimate messages as spam [12].

3.2 Machine Learning models

As SMS spam tactics became increasingly sophisticated, traditional rule-based approaches were insufficient, prompting the adoption of machine learning techniques. These techniques demonstrated remarkable advancements across various domains, including healthcare and communications [13-14]. The ML models improved spam detection by learning from data and identifying patterns that rule-based systems could not capture [15]. Techniques like Naive Bayes, Support Vector Machines (SVM), and Random Forests gained popularity [16-18], often in conjunction with Term Frequency-Inverse Document Frequency (TF-IDF) for feature extraction [19]. TF-IDF helps quantify the importance of words in a message, allowing these models to better distinguish between spam and non-spam content. Naive Bayes, for instance, uses probabilistic models to classify messages based on features like word frequency [20], while SVM and Random Forests provide more robust classification by finding optimal boundaries between spam and non-spam messages. These approaches significantly enhanced detection accuracy and reduced false positives, though they required substantial labeled data and faced challenges with imbalanced datasets [21].

3.3 Deep Learning Models

In recent years, advances in deep learning have further transformed SMS spam detection by enabling models to capture complex patterns and contextual information within messages [22]. Deep learning techniques, such as the Recurrent neural network (RNN) [23] and its derivatives, such as Long Short-Term Memory (LSTM) networks [24], Bidirectional Long Short-Term Memory (LSTM) networks [25], Convolutional Neural Networks (CNN) [26], and transformer-based models like BERT [27] RoBERTa [28], have shown exceptional performance in text classification tasks, including spam detection. The RNN and its derivatives are particularly effective at handling sequential data, making them well-suited for analyzing the order and context of words in a message. The CNN has also been adapted to capture local patterns in text, such as the presence of specific word combinations indicative of spam; transformer-based models represent the cutting edge of deep learning for text processing [29-30]. These models excel at understanding nuanced language and can adapt to new spam tactics with minimal manual intervention. However, the high computational cost and need for large datasets pose challenges for their deployment in real-time spam detection systems.

3.4 Hybrid models

Hybrid models offer powerful enhancements in SMS spam detection by combining multiple methodologies to achieve both high efficiency and accuracy. These models typically integrate rule-based techniques with machine learning or deep learning approaches, balancing quick filtering and adaptive learning. For instance, rule-based filters can swiftly eliminate obvious spam messages by identifying specific keywords or patterns commonly associated with spam, such as phrases like "free offer" or excessive exclamation marks. Following this initial filtering, a machine learning model like Naive Bayes, SVM, or deep learning models like LSTM or CNN can analyze the remaining ambiguous cases, providing a more nuanced and context-aware classification [31-33]. A hybrid approach might use TF-IDF for feature extraction, where important words or phrases are identified and weighted, combined with rule-based filters to pre-process the data. The final classification is then handled by a deep learning model, which adapts to new and evolving spam tactics. This layered approach ensures that both simple, well-known spam patterns and more sophisticated, evolving spam attempts are effectively detected, striking a balance between speed and adaptability [34-37].

3.5 Ensemble models

Ensemble models further combine strengths by integrating the outputs of multiple classifiers to improve overall robustness and accuracy. These models leverage bagging, boosting, and stacking techniques to enhance prediction performance. In bagging, multiple models are trained on different subsets of data, and their predictions are averaged to reduce variance and improve accuracy [38-39]. This method ensures that the model does not overfit any particular data subset, resulting in a more generalized spam detection system. Boosting, used in algorithms like AdaBoost, sequentially trains models where each new model focuses on correcting the errors made by the previous ones, thus progressively enhancing the model's accuracy. For instance, in SMS spam detection, a boosting approach might first identify the most straightforward spam cases and then iteratively refine its ability to detect more complex or deceptive messages [40]. Stacking, a more sophisticated ensemble method, involves training multiple different types of models (e.g., Naive Bayes, SVM, neural networks) and using a meta-classifier to combine their predictions. This layered approach allows the system to benefit from the unique strengths of each model, whether it's the speed of Naive Bayes, the precision of SVM, or the deep contextual understanding of neural networks [41-42]. By leveraging these diverse models, ensemble techniques offer a powerful solution to the ever-evolving nature of spam, making them particularly well-suited for the complex challenge of SMS spam detection [43].

Figure 2 illustrates the progression of SMS spam detection accuracy from 2015 to 2024 across various methodologies. Over time, there has been a noticeable improvement in accuracy, with earlier methods, such as Rule-Based + Naive Bayes, showing moderate success in 2015 (88.7%). As the years progressed, more advanced techniques like Random Forest, AdaBoost, and Gradient Boosting were introduced, leading to significant gains in accuracy by 2020. From 2021 onwards, the adoption of Transformer-based and BERT-based models marked a shift towards deep learning techniques, achieving accuracies above 97%. By 2023 and 2024, hybrid methods emerged as the top performers, with more than 99% accuracy. These hybrid models, which combine deep learning architectures and feature extraction techniques, demonstrate superior performance in addressing the complexities of SMS spam detection, solidifying their place as the most effective approaches in recent years.

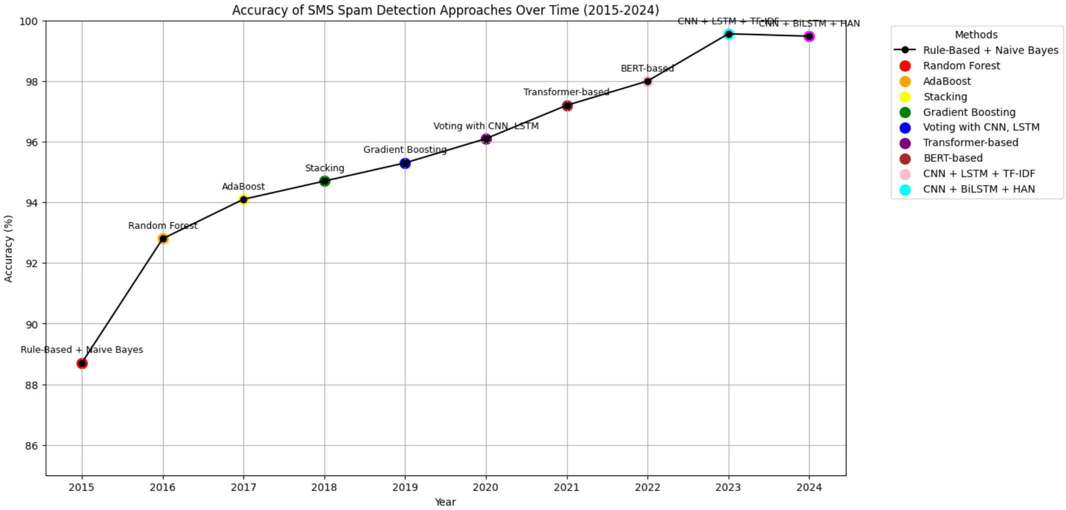

Figure 2. Accuracy of SMS Spam Detection Approaches over Time (2015-2024).

4. Datasets and Benchmarks

4.1 Public Datasets

Public datasets play a crucial role in developing and evaluating SMS spam detection models, providing a standardized basis for comparing different methods. Several publicly available datasets have been widely used in research:

- **UCI SMS Spam Collection**: This dataset is most commonly used in SMS spam detection research. It contains 5,574 English SMS messages labeled as "ham" (non-spam) or "spam." The dataset is imbalanced, with fewer spam messages, making it hard to train and evaluate machine learning models [44].
- **NUS SMS Corpus**: Developed by the National University of Singapore, this dataset contains 10,000 SMS messages, including spam and non-spam [45]. The messages are labeled with various categories, allowing for a more granular analysis of spam types.

- **SMS Spam Corpus v.0.1 Big**: This dataset contains over 10,000 SMS messages, with a significant portion labeled as spam [46]. It is frequently used to evaluate the performance of different machine learning and deep learning models in SMS spam detection.

Figure 3 presents the class distribution of spam and ham messages across four prominent SMS datasets used in spam detection research: UCI SMS Spam Collection, NUS SMS Corpus, and SMS Spam Corpus v.0.1 Big. Each dataset exhibits varying degrees of imbalance between spam and ham messages, which can have significant implications for developing and evaluating SMS spam detection models. For instance, the UCI SMS Spam Collection is highly imbalanced, with only 13.4% of its messages labeled as spam, compared to 86.6% as ham. In contrast, while still imbalanced, the SMS Spam Corpus v.0.1 Big contains a higher proportion of spam messages at 24.3%. The NUS SMS Corpus also demonstrates notable imbalances, with spam comprising 18.0% of its datasets. Such imbalances underscore the challenges of training models that can accurately detect spam while minimizing false positives, emphasizing the need for robust methods capable of handling skewed data distributions.

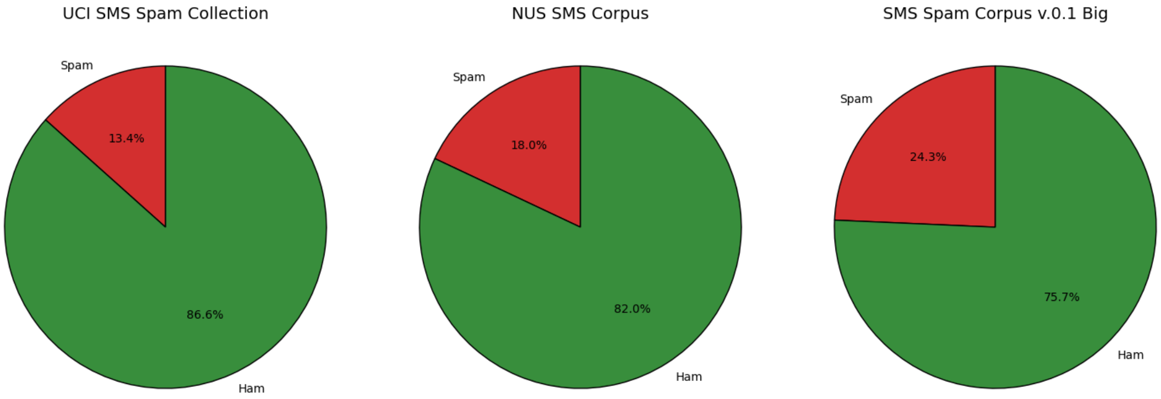

Figure 3. Distribution of spam and ham messages across different datasets

4.2 Proprietary Datasets

While public datasets are vital for research, companies often use proprietary datasets for SMS spam detection, collected from live systems with diverse messages and languages. These datasets, maintained by telecommunications companies and messaging platforms, capture spam patterns specific to their user bases. Though not publicly available, they offer valuable insights into evolving spam tactics. Researchers typically need industry collaborations to access them. Many SMS spam datasets are private or use local datasets, as seen in references [47-50].

4.3 Benchmarking

Benchmarking is critical for evaluating and comparing the performance of SMS spam detection methods. Key benchmarks include:

- **Accuracy**: The proportion of correctly classified messages (spam and non-spam) out of the total number of messages. While accuracy is straightforward, it may be misleading in imbalanced datasets, where non-spam messages vastly outnumber spam [51].
- **Precision, Recall, and F1-Score**: Precision measures the proportion of correctly identified spam messages out of all messages classified as spam. In contrast, recall measures the proportion of actual spam messages correctly identified by the model. The F1-score is the harmonic mean of precision and recall, providing a balanced metric for false positives and negatives [52].
- **Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC)**: The ROC curve plots the true positive rate against the false positive rate at various threshold settings, while the AUC quantifies the model's overall performance across all thresholds. A higher AUC indicates better discrimination between spam and non-spam messages [53].

- **Confusion Matrix**: A detailed breakdown of the model's performance, showing the number of true positives, true negatives, false positives, and false negatives. The confusion matrix provides valuable insights into the types of errors made by the model, guiding further improvements [54].
- **Computational Efficiency**: Given the real-time nature of SMS communication, computational efficiency is an important benchmark. This includes the time taken to process and classify messages, as well as the resource requirements of the model [55]. Models that achieve high accuracy but are too slow for real-time deployment may be less practical in real-world applications.

By employing these benchmarks, researchers can objectively assess the strengths and weaknesses of different SMS spam detection methods, facilitating the development of more effective and scalable solutions.

4.4 Data Splitting Strategies

Data splitting is critical in developing and evaluating SMS spam detection models, ensuring that the models are trained and tested on separate data to avoid overfitting and provide an unbiased performance review. Several strategies are commonly employed:

- **Training, Validation, and Test Split**: This is the most commonly used data splitting strategy [56]. The dataset is typically divided into three subsets :
  - **Training Set (60-70%)**: Used to train the model, allowing it to learn from labeled examples.
  - **Validation Set (15-20%)**: Used to tune hyper-parameters and select the best model configuration, preventing overfitting on the training data [57].
  - **Test Set (15-20%)**: Used to evaluate the final model's performance on unseen data, providing an unbiased assessment of its generalization ability.
- **K-Fold Cross-Validation**: This approach divides the dataset into K equal-sized folds. The model is trained and validated K times, each using a different fold as the validation set and the remaining K-1 folds as the training set [58]. The results are averaged to provide a more robust estimate of model performance. This method is beneficial when the dataset is small, as it maximizes the use of available data.

5. Comparative Analysis

The performance of SMS spam detection methods varies by techniques, datasets, and metrics. Traditional models like Naive Bayes and SVM perform well, with Naive Bayes excelling in recall but prone to false positives, while SVM balances precision and recall. Deep learning models like LSTM, CNN, and BERT outperform traditional methods by capturing complex patterns, with BERT excelling on large datasets. Ensemble methods boost accuracy by combining models but require substantial resources and careful tuning to avoid overfitting.

5.1 Comparative Analysis of SMS Spam Detection Methods

Each SMS spam detection method exhibits distinct strengths and weaknesses, making them suitable for specific scenarios. Rule-based and keyword-based approaches are simple and quick but falter against evolving spam strategies. Machine learning methods balance accuracy and efficiency but face difficulties with complex data. Deep learning techniques, while excellent for capturing intricate patterns, are resource-intensive. Hybrid models combine multiple methods for robust performance, though they demand significant computational power. Ensemble methods achieve high accuracy by integrating diverse classifiers but may not be practical for real-time applications due to latency and resource constraints. Table 1 provides a detailed comparison of these approaches.

Table 1. Strengths and Weaknesses of SMS Spam Detection Methods

Detection Method	Strengths	Weaknesses
Rule-based & Keyword-based	Simple and fast to implement. Effective for straightforward spam patterns.	Poor adaptability to new spam tactics. High false positive rate.
Machine Learning	Balances accuracy and efficiency. Can identify patterns beyond predefined rules.	Requires labeled data. Struggles with high-dimensional or complex data.
Deep Learning	Excels at identifying complex patterns. Minimal feature engineering is required.	Computationally intensive. Needs large labeled datasets.
Hybrid Methods	Combines strengths of different techniques. Enhances accuracy and robustness.	High resource and computational demands. It may not suit resource-limited environments.
Ensemble Methods	High accuracy. Reduces overfitting. Handles diverse spam patterns effectively.	Computationally expensive. High latency, unsuitable for real-time detection.

This table clearly outlines the trade-offs in selecting an SMS spam detection approach, aiding researchers and practitioners in choosing the most appropriate technique for their needs.

5.2 Trends and Insights

Recent trends in SMS spam detection highlight a shift towards deep learning models, especially transformer-based ones like BERT and GPT [59], which excel in handling large, diverse datasets and detecting complex text patterns. Hybrid and ensemble methods are also gaining traction, combining multiple techniques to improve detection performance and adapt to evolving spam tactics. Real-time detection is a growing focus, with efforts to optimize deep learning models for faster processing and deploy lightweight models on mobile devices. Additionally, integrating metadata and contextual features alongside textual data is becoming more common, enhancing the detection system's ability to address multifaceted spam.

6. Challenges and Future Directions

As SMS spam detection evolves, it faces numerous challenges that must be addressed to enhance effectiveness and adaptability. This section explores the obstacles hindering progress and highlights potential research directions to overcome these limitations and improve spam detection methodologies.

6.1 Current Challenges

Despite progress, several challenges persist in SMS spam detection. Data scarcity remains a significant issue, as many available datasets lack diversity and representativeness, limiting model generalization [60]. Model interpretability is another challenge, as deep learning models often function as "black boxes," hindering transparency and trust. Adversarial attacks, where spammers manipulate content to evade detection, pose a growing threat. The trade-off between accuracy and computational efficiency also persists, as resource-intensive models are hard to deploy in real-time or on devices with limited processing power. Ethical concerns, especially regarding privacy and data security, must be addressed.

6.2 Future Research Directions

Future research could focus on multilingual models that handle different languages and regions by creating more diverse datasets. Transfer learning techniques can help models adapt across domains, reducing the need for large labeled datasets. Interpretable AI is another key area, focusing on models that offer transparency without sacrificing accuracy. Research should also address adversarial resilience, using methods like adversarial training or hybrid systems to improve security. Finally, lightweight models optimized for mobile devices are crucial to ensure practical and efficient spam detection, especially in regions with limited internet access.

7. Conclusion

This paper comprehensively reviews the evolution of SMS spam detection methodologies, spanning from traditional rule-based approaches and machine learning models to advanced deep learning and hybrid frameworks. This survey aims to provide researchers and practitioners with a clear understanding of each method's strengths, limitations, and practical applications. Traditional models, such as Naive Bayes and SVM, have demonstrated efficacy on well-curated datasets but face challenges when dealing with the increasing sophistication of modern spam tactics, including evolving linguistic patterns and obfuscation strategies. In contrast, deep learning models, such as LSTM and BERT, have shown significant promise by capturing complex patterns, contextual nuances, and semantic relationships in SMS text data. Furthermore, hybrid and ensemble methods, which integrate multiple techniques, have emerged as robust solutions to enhance detection accuracy and adaptability in real-world scenarios. This survey aims to serve as a taxonomy and a guide for researchers in natural language processing (NLP) and mobile security, helping them navigate the diverse range of available approaches. By organizing and analyzing existing methods, this work seeks to assist in identifying the most suitable techniques for addressing specific challenges in SMS spam detection. Additionally, the survey highlights trends, gaps, and opportunities for future research, contributing to the advancement of secure and efficient spam detection systems in the ever-evolving digital communication landscape.

Conflict of Interest: The authors declare no conflicts of interest related to this research. No financial or personal relationships could influence the results or interpretation of the findings.

References

1. S. J. Delany, M. Buckley, and D. Greene, "SMS spam filtering: Methods and data," Expert Syst. Appl., vol. 39, no. 10, pp. 9899-9908, Aug. 2012.
2. S. M. R. A. Saidat, S. Y. Yerima, and K. Shaalan, "Advancements of SMS Spam Detection: A Comprehensive Survey of NLP and ML Techniques," *Procedia Computer Science*, vol. 244, pp. 248-259, 2023. doi: 10.1016/j.procs.2024.10.198.
3. A. Qazi, N. Hasan, R. Mao, M. Elhag Mohamed Abo, S. Kumar Dey and G. Hardaker, "Machine Learning-Based Opinion Spam Detection: A Systematic Literature Review," in *IEEE Access*, vol. 12, pp. 143485-143499, 2024, doi: 10.1109/ACCESS.2024.3399264.
4. K. Hanif and H. Ghous, "Detection of SMS Spam and Filtering by Using Data Mining Methods: Literature Review," *Irjmets.com*, vol. 1, pp. 874-886, 2021.
5. H. Sajedi, G. Z. Parast, and F. Akbari, "SMS Spam Filtering Using Machine Learning Techniques: A Survey," *Machine Learning Research*, vol. 1, no. 1, pp. 1-14, 2016.
6. S. Kaddoura, G. Chandrasekaran, D. E. Popescu, and J. H. Duraisamy, "A Systematic Literature Review on Spam Content Detection and Classification," *PeerJ Computer Science*, vol. 8, e830, 2022.
7. S. Ali, "SMS spam identification based on message duplication detection by cuckoo filters," J. Kerbala Univ., vol. 10, pp. 48-55, 2014.
8. H. A. Al-Kabbi, M. -R. Feizi-Derakhshi and S. Pashazadeh, "Multi-Type Feature Extraction and Early Fusion Framework for SMS Spam Detection," in IEEE Access, vol. 11, pp. 123756-123765, 2023, doi: 10.1109/ACCESS.2023.3327897.
9. J. W. Joo et al., "S-Detector: An enhanced security model for detecting smishing attack for mobile computing," Telecommun. Syst., vol. 66, pp. 29-38, 2017.
10. J. M. Gómez Hidalgo, G. Cajigas Bringas, E. Puertas Sánz, and F. Carrero García, "Content-based SMS spam filtering," in *Proc. 2006 ACM Symp. Document Engineering (DocEng)*, 2006, pp. 107-114.
11. A. G. West, A. J. Aviv, J. Chang, and I. Lee, "Spam mitigation using spatiotemporal reputations from blocklist history," in Proc. 26th Annu. Comput. Security Appl. Conf., Dec. 2010, pp. 161-170.
12. T. Almeida and A. Yamakami, "Content-based SMS spam filtering," in *Proc. 2010 Int. Symp. Multimedia and Security (ISMS)*, 2010, pp. 123-130.
13. A. N. Kamber, "A variety of electrocardiogram (ECG) signal processing," *Solid State Technol.*, vol. 63, no. 3, pp. 5370-5377, 2020.
14. N. B. Hassan, A. K. Nawar, S. A. Jebur, and I. Tareq, "Internet of Things: Architecture, technologies, applications, and challenges," Alkadhim J. Comput. Sci., vol. 2, no. 1, 2024.

15. S. Gupta, S. D. Saha, and S. K. Das, "SMS spam detection using machine learning," J. Phys.: Conf. Ser., vol. 1797, no. 1, p. 012017, 2021.
16. D. D. Arifin and M. A. Bijaksana, "Enhancing spam detection on mobile phone Short Message Service (SMS) performance using FP-growth and Naive Bayes Classifier," in Proc. 2016 IEEE Asia Pacific Conf. Wireless Mobile (APWiMob), 2016, pp. 80-84.
17. Tekerek, Adem. "Support vector machine-based spam SMS detection." *Politeknik Dergisi* 22, no. 3 (2019): 779-784.
18. N. N. A. Sjarif, N. F. M. Azmi, S. Chuprat, H. M. Sarkan, Y. Yahya, and S. M. Sam, "SMS spam message detection using term frequency-inverse document frequency and random forest algorithm," *Procedia Comput. Sci.*, vol. 161, pp. 509-515, 2019.
19. S. M. Hossain, K. M. Kamal, A. Sen, and I. H. Sarker, "TF-IDF feature-based spam filtering of mobile SMS using a machine learning approach," in *Applied Intelligence for Industry 4.0*, 2023, pp. 162-175.
20. G. Ubale and S. Gaikwad, "SMS Spam Detection Using TFIDF and Voting Classifier," *2022 International Mobile and Embedded Technology Conference (MECON)*, Noida, India, 2022, pp. 363-366, doi: 10.1109/MECON53876.2022.9752078.
21. O. Abayomi-Alli, S. Misra, A. Abayomi-Alli and M. Odusami, "A review of soft techniques for SMS spam classification: Methods approaches and applications," *Eng. Appl. Artif. Intell.*, vol. 86, pp. 197-212, Nov. 2019.
22. W. H. Gomaa, "The impact of deep learning techniques on SMS spam filtering," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 1, 2020.
23. V. V. Kalyani, M. R. Sundari, S. Neelima, P. S. S. Prasad, P. P. Mohan, and A. Lakshmanarao, "SMS Spam Detection using NLP and Deep Learning Recurrent Neural Network Variants," in *Proc. 2024 Int. Conf. Cogn. Robot. Intell. Syst. (ICC-ROBINS)*, Apr. 2024, pp. 92-96.
24. C. L. Sri, D. D. Lakshmi, K. Ravali, V. Kukreja, and S. Hariharan, "Improved spam detection through LSTM-based approach," in *Proc. 2024 Third Int. Conf. Intelligent Techniques Control, Optim. Signal Process. (INCOS)*, 2024, pp. 1-6.
25. A. L. Rosewelt, N. D. Raju, and S. Ganapathy, "An effective spam message detection model using feature engineering and bi-LSTM," in *Proc. 2022 Int. Conf. Adv. Comput., Commun. Appl. Inform. (ACCAI)*, Jan. 2022, pp. 1-6.
26. T. Huang, "A CNN model for SMS spam detection," in *Proc. 2019 4th Int. Conf. Mechanical, Control Comput. Eng. (ICMCCE)*, 2019, pp. 851-85110.
27. C. Oswald, S. E. Simon, and A. Bhattacharya, "Spotspam: Intention analysis–driven SMS spam detection using BERT embeddings," *ACM Trans. Web (TWEB)*, vol. 16, no. 3, pp. 1-27, 2022.
28. H. A. Al-Kabbi, M.-R. Feizi-Derakhshi, and S. Pashazadeh, "A hierarchical two-level feature fusion approach for SMS spam filtering," *Intell. Autom. Soft Comput.*, vol. 39, no. 4, 2024.
29. X. Liu, H. Lu, and A. Nayak, "A spam transformer model for SMS spam detection," *IEEE Access*, vol. 9, pp. 80253-80263, 2021.
30. A. Ghourabi and M. Alohaly, "Enhancing spam message classification and detection using transformer-based embedding and ensemble learning," *Sensors*, vol. 23, no. 8, p. 3861, 2023.
31. H. Baaqeel and R. Zagrouba, "Hybrid SMS spam filtering system using machine learning techniques," in *Proc. 2020 21st Int. Arab Conf. Inf. Technol. (ACIT)*, 2020, pp. 1-8.
32. A. Ghourabi, M. A. Mahmood, and Q. M. Alzubi, "A hybrid CNN-LSTM model for SMS spam detection in Arabic and English messages," *Future Internet*, vol. 12, no. 9, p. 156, 2020.
33. M. R. F. Derakhshi, E. Zafarani-Moattar, H. A. Al-Kabi, and A. H. J. Almarashy, "PCLF: Parallel CNN-LSTM fusion model for SMS spam filtering," in *BIO Web Conf.*, vol. 97, p. 00136, 2024.
34. E. Ramanujam, K. Shankar, and A. Sharma, "Multi-lingual spam SMS detection using a hybrid deep learning technique," in *Proc. 2022 IEEE Silchar Subsection Conf. (SILCON)*, 2022, pp. 1-6.
35. V. Gupta, A. Mehta, A. Goel, U. Dixit, and A. C. Pandey, "Spam detection using ensemble learning," in *Harmony Search and Nature Inspired Optimization Algorithms: Theory and Applications, ICHSA 2018*, Springer Singapore, 2019, pp. 661-668.
36. R. Hajahan and P. L. Lekshmy, "Hybrid Learning Approach for E-mail Spam Detection and Classification," in *Intelligent Cyber Physical Systems and Internet of Things. ICoICI 2022. Engineering Cyber-Physical Systems and Critical Infrastructures*, vol. 3, J. Hemanth, D. Pelusi, and J. I. Z. Chen, Eds. Cham: Springer, 2023.

37. T. Xia, X. Chen, J. Wang, and F. Qiu, "A Hybrid Model with New Word Weighting for Fast Filtering Spam Short Texts," *Sensors*, vol. 23, no. 21, p. 8975, 2023.
38. S. Hosseinpour and H. Shakibian, "An ensemble learning approach for SMS spam detection," in *Proc. 2023 9th Int. Conf. Web Res. (ICWR)*, 2023, pp. 125-128.
39. A. Al Maruf, A. Al Numan, M. M. Haque, T. T. Jidney, and Z. Aung, "Ensemble approach to classifying spam SMS from Bengali text," in Proc. Int. Conf. Adv. Comput. Data Sci., Cham: Springer Nature Switzerland, Apr. 2023, pp. 440-453.
40. F. Akbari and H. Sajedi, "SMS spam detection using selected text features and boosting classifiers," in *Proc. 2015 7th Conf. Inf. Knowl. Technol. (IKT)*, May 2015, pp. 1-5.
41. C. Ulus, Z. Wang, S. M. Iqbal, K. M. S. Khan, and X. Zhu, "Transfer Naïve Bayes Learning using Augmentation and Stacking for SMS Spam Detection," in *Proc. 2022 IEEE Int. Conf. Knowl. Graph (ICKG)*, Nov. 2022, pp. 275-282.
42. W. Saeed, "Comparison of automated machine learning tools for SMS spam message filtering," in *Advances in Cyber Security: Third Int. Conf. ACeS 2021, Penang, Malaysia, Aug. 24–25, 2021, Revised Selected Papers 3*, Singapore: Springer, 2021, pp. 307-316.
43. G. Airlangga, "Optimizing SMS spam detection using machine learning: A comparative analysis of ensemble and traditional classifiers," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 6, no. 4, pp. 1942-1951, 2024.
44. T. A. Almeida, J. M. G. Hidalgo, and A. Yamakami, "Contributions to the study of SMS spam filtering," *Proc. 11th ACM Symp. Document Eng.*, pp. 259-262, Sep. 2011.
45. T. Chen and M.-Y. Kan, "Creating a live, public short message service corpus: the NUS SMS corpus," *Language Resour. Eval.*, vol. 47, pp. 299-335, 2013.
46. M. A. Shafi'I, M. S. Abd Latiff, H. Chiroma, O. Osho, G. Abdul-Salaam, A. I. Abubakar, and T. Herawan, "A review on mobile SMS spam filtering techniques," *IEEE Access*, vol. 5, pp. 15650-15666, 2017.
47. I. S. Mambina, J. D. Ndibwile, D. Uwimpuhwe, and K. F. Michael, "Uncovering SMS spam in Swahili text using deep learning approaches," *IEEE Access*, 2024.
48. U. Maqsood, S. Ur Rehman, T. Ali, K. Mahmood, T. Alsaedi, and M. Kundi, "An intelligent framework based on deep learning for SMS and e-mail spam detection," *Appl. Comput. Intell. Soft Comput.*, vol. 2023, no. 1, Art. no. 6648970, 2023.
49. T. Xia and X. Chen, "A discrete hidden Markov model for SMS spam detection," *Appl. Sci.*, vol. 10, no. 14, p. 5011, 2020.
50. A. Theodorus, T. K. Prasetyo, R. Hartono, and D. Suhartono, "Short message service (SMS) spam filtering using machine learning in Bahasa Indonesia," in *Proc. 2021 3rd East Indonesia Conf. Comput. Inf. Technol. (EIConCIT)*, Apr. 2021, pp. 199-203.
51. A. H. J. Almarashy, M. -R. Feizi-Derakhshi and P. Salehpour, "Enhancing Fake News Detection by Multi-Feature Classification," in *IEEE Access*, vol. 11, pp. 139601-139613, 2023, doi: 10.1109/ACCESS.2023.3339621.
52. A. K. Jasim, J. Tanha, and M. A. Balafar, "Neighborhood information based semi-supervised fuzzy C-means employing feature-weight and cluster-weight learning," *Chaos Solitons Fractals*, vol. 181, p. 114670, 2024.
53. Z. H. Ali, H. M. Salman, and A. H. Harif, "SMS spam detection using multiple linear regression and extreme learning machines," *Iraqi J. Sci.*, 2023, pp. 6342-6351.
54. A. I. Jabbooree, L. M. Khanli, P. Salehpour, and S. Pourbahrami, "Geometrical facial expression recognition approach based on fusion CNN-SVM," *Int. J. Intell. Eng. Syst.*, vol. 17, no. 1, 2024.
55. F. Wei and T. Nguyen, "A lightweight deep neural model for SMS spam detection," in *Proc. 2020 Int. Symp. Networks, Comput. Commun. (ISNCC)*, 2020, pp. 1-6.
56. Y. Xu and R. Goodacre, "On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning," *J. Anal. Test.*, vol. 2, no. 3, pp. 249-262, 2018.
57. M. Ramezani, M.-R. Feizi-Derakhshi, M. A. Balafar, et al., "Automatic personality prediction: an enhanced method using ensemble modeling," *Neural Comput. Appl.*, vol. 34, pp. 18369–18389, 2022, doi: 10.1007/s00521-022-07444-6.
58. M. I. Khaleel, Z. A. Taha, I. A. Murad, and M. Zahawii, "ChatGPT and the Crisis of Academic Honesty," *AlKadhim Journal for Computer Science*, vol. 2, no. 1, pp. 28–35, 2024.

59. D. Salman and N. Sulaiman, "A Review of Encryption Algorithms for Enhancing Data Security in Cloud Computing," *AlKadhim Journal for Computer Science*, vol. 2, no. 1, pp. 53–71, 2024.
60. H. K. Hoomod, A. J. Al-Mousawi, and J. R. Naif, "New Complex Hybrid Security Algorithm (CHSA) for Network Applications," in *Inventive Communication and Computational Technologies: Proceedings of ICICCT 2019*, Singapore: Springer, 2020, pp. 87–103.