ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

A Comparative Study of Deep Learning Approaches for Spam Detection

M. P. Sudha[1], Dr. M. Ramesh Kumar[2]

[1]Research scholar, PG and Research Department of Computer Science, [1,2] Government Arts College (Autonomous), Nandanam, Chennai.

[2]Associate Professor and Head, PG and Research Department of Computer Science, [1,2] Government Arts College (Autonomous), Nandanam, Chennai.

DOI: https://doi.org/10.51583/IJLTEMAS.2026.15020000069

Received: 15 February 2026; Accepted: 21 February 2026; Published: 16 March 2026

ABSTRACT

Spam identification is critical in Morden digital communication systems such as email, SMS, social media, and online platforms. The increasing proliferation of unwanted and hazardous messages such as spam, phishing, and scam material poses a severe threat to user privacy and cybersecurity. Deep learning (DL) algorithms can identify automatically learn intricate representations from large-scale various data sources, including reviews, SMS, and email data. They have become powerful alternatives that reflect modern spam qualities across several platforms and languages are sparse. This article presents an exhaustive review of deep learning-based spam detection approaches.

The dataset The lack of dynamic, multilingual, and real-world data limits adaptability and the ability to deal with evolving spam. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory (LSTM) networks, hybrid models, and transformer-based techniques are among the most commonly used designs discussed. Explainable AI (XAI) approaches are not well-integrated to provide clear and understandable explanations for spam detection options. In order to give future research paths for constructing reliable and intelligent spam detection systems, the study discusses datasets, assessment metrics, obstacles, and unfilled research gaps.

The goal of this survey is to provide an organized summary of deep learning techniques. The design of an effective, flexible, and userfriendly deep learning-based spam detection framework that can precisely identify a variety of changing spam messages in real-world communication systems while maintaining low figuring overhead and high robustness is the main issue in this field of study.

Keywords: Spam Detection, Deep Learning, CNN, LSTM, XAI, RNN, Transformer

INTRODUCTION

Spam communications across a variety of platforms, including email, SMS, social media, and instant messaging systems, have increased drastically as digital communication has grown exponentially. Furthermore, these interactions are commonly used for financial fraud, identity theft, phishing attacks, malware distribution, and spreading misleading information. Spam poses major cybersecurity and privacy risks, as well as a negative impact on productivity for both individuals and enterprises.

Traditional spam filtering technologies, including rule-based systems and machine learning models like Support Vector Machines and Naïve Bayes, are ineffective in detecting complex and dynamic spam patternsThey rely heavily on feature engineering and are unable to adjust to a variety of dynamic spam content by enabling autonomous feature learning and collecting comprehensive semantic and contextual information. By automatically creating hierarchical and semantic representations from raw text input, deep learning (DL) has revolutionized spam detection and outperformed conventional techniques.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

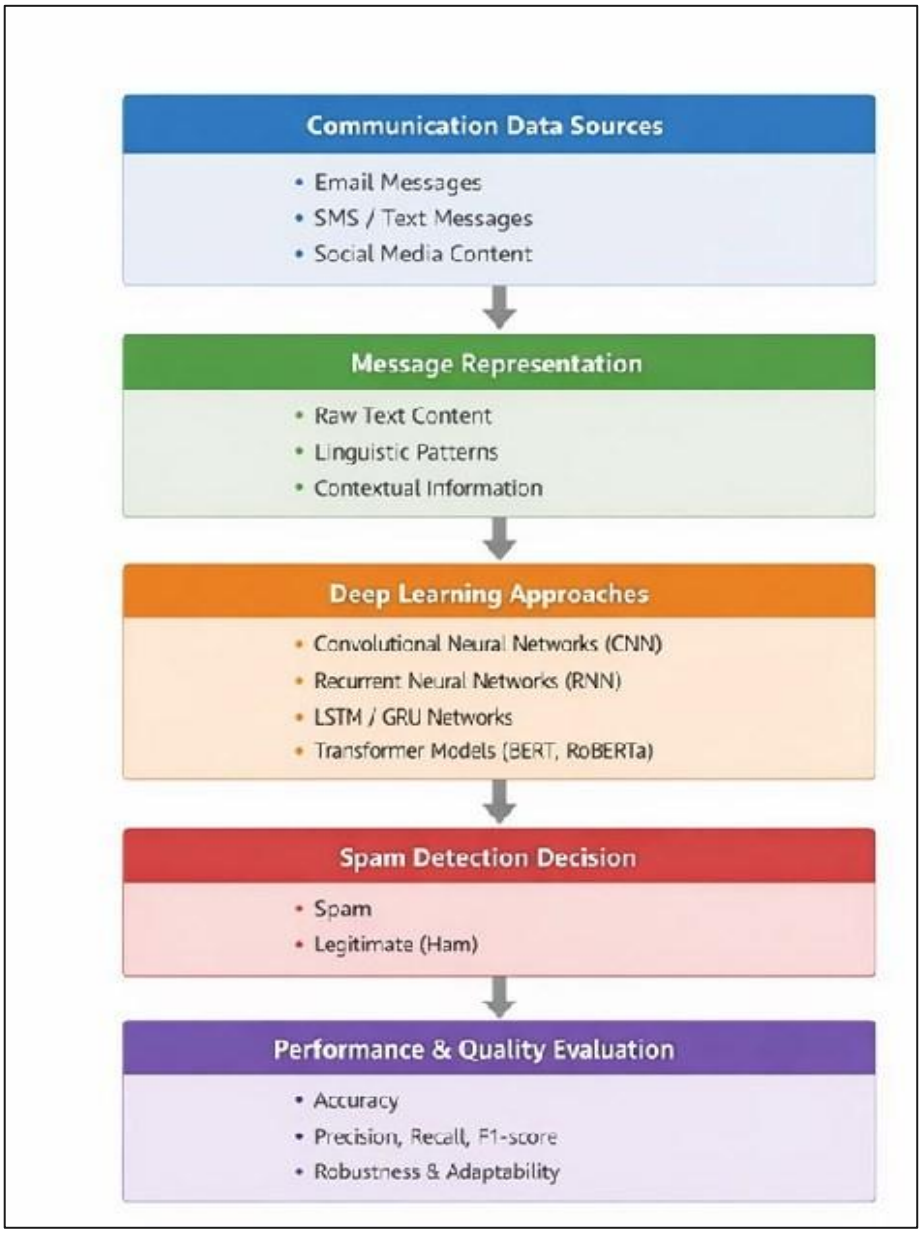

Figure 1 Deep Learning Based Spam Detection

Figure 1 shows the data sources, message properties, deep learning models, classification, and evaluation challenges. Convolutional Neural Networks (CNNs) capture local lexical patterns, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks model sequential dependencies, and Transformerbased architectures like BERT and RoBERTa provide contextualized embedding's and superior generalization. These models have demonstrated significant improvements in accuracy, recall, and robustness when compared to traditional machine learning techniques. Managing class imbalance, integrating multimodal features like URLs and metadata, handling shifting spam patterns (concept drift), ensuring model explain ability, and reducing computational complexity for real-time deployment are some of the disadvantages of deep learning-based spam detection, despite its advantages.. These problems must be resolved in order to create spam detection systems that are dependable, adaptable, and scalable. more recent improvements in performance through the use of selfattention procedures through transfer learning and contextual embedding's. Despite the success of deep learning models, explain ability, computational cost, class imbalance, and adaptability to real-world dynamic spam environments remain problems.

The remainder of the study is structured as follows: A overview of the literature on spam detection is included in Section 2. The proposed deep learning architecture is described in Section 3; the dataset and evaluation matrix is described in Section 4; and collusion and future research directions are described in Section 5.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

Related Work

Primary research on spam detection frequently employed machine learning and arithmetic techniques with physically constructed features such as word frequency, n-grams, and metadata. With the introduction of RNNs and LSTMs to imitate sequential dependencies, researchers began exploring deep learning for automated feature extraction.

CNNs were initially employed to gather local text with URL patterns. Considering that transformerbased models such as BERT and RoBERTa have recently demonstrated cutting-edge performance. Table 1 lists several survey studies that have examined spam detection techniques, as well as current advancements and a comprehensive and updated analysis of deep learning mechanisms. due to the contextualized representations in the call for the identification of spam.

Table 1: Literature Review on Deep learning Spam detection

Paper Ref. No.	Researcher(s)	Dataset Used	Model(s)	RESULTS & Accuracy
1	A. Karim, S. Azam, B. Shanmugam, K. Kannoorpatti, M. Alazab	Public Email Spam	Deep Learning / Hybrid ML Models	Achieved high accuracy (≈95–97%), detection and improved feature representation
2	P. K. Roy, J. P. Singh, S. Banerjee	SMS Spam	Deep Learning (CNN, LSTM)	Achieved Accuracy: 96.2%, Precision: 95.1%, Recall: 96.5%
3	G. M. Shahariar, S. Biswas, F. M. Shah, S. Binte Hassan	Online Review	Deep Learning (CNN, LSTM)	spam using text features, effectively 94%detected review
4	E. S. Rahman, S. Ullah	Email Spam	Hybrid Model	better handling of sequential patterns Achieved Accuracy: 95.7%.
5	M. Popovac, M. Karanovic, S. Sladojevic, M. Arsenovic, A. Anderla	SMS Spam	CNN-based Model	local feature extraction in SMS messages Achieved Accuracy: 94.5%.
6	N. Govil, K. Agarwal, A. Bansal, A. Varshney	SMS & Email D	ML-based Approach	Basic machine learning classifier with text good for accuracy 92%.
7	B. Kim, S. Abuadbba, H. Kim	Image Spam	CNN + Data Augmentation	Image-based spam detection demonstrated capability for Accuracy: 96.8%,
8	M. A. Shaaban, Y. F. Hassan, S. K. Guirguis	Text	Deep Convolutional Forest	Dynamic ensemble detection robustness improved Accuracy: 97%.
9	M. Amaz Uddin, M. N. Islam, L. Maglaras, H. Janicke, I. H. Sarker	SMS Spam	Transformer-based (BERT)	Contextual embeddings improved Explainable model, classification Accuracy: 98%.
10	H. C. Altunay, Z. Albayrak	SMS Spam Dataset	Deep Learning Architectures	Robust SMS spam detection using deep LSTM and CNN

Lastly, the reviewed studies demonstrate that spam identification research has progressed beyond text-based data and image-based spam detection employing CNNs, Transformer-based models, such as the BERT model, which achieved accuracy above 96%.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

Deep Learning Architectures for Spam Detection

The recommended method for deep learning architecture for spam detection is outlined in this section. Data collection, message representation, text preprocessing methods, and the architecture for building deep learning models.

Data Collection

The dataset is often gathered from anonymized real-world systems, research institutes, communication platforms, and public repositories. It includes messages from a variety of communication channels, such as SMS, Social Media Spam Datasets, Review Spam, email, and image-based spam datasets, which enable repeatable research.

Data Preprocessing

Tokenizing messages, converting to lowercase, eliminating special characters and punctuation, eliminating stop words, lemmatizing or stemming, and removing noise are all examples of data pretreatment in deep learning. The processed text is then converted into fixed-length numerical representations (embeddings) to provide effective learning and accurate spam classification.

Message Representation

Word embedding approaches for numerical representations extract words using inks from pre-processed messages prepared for deep learning models. Contextual embeddings can be created using Transformer-based models like BERT or pre-trained models like Word2Vec and GloVe, as well as the architecture. Figure 2 depicts the processed architecture for spam identification.

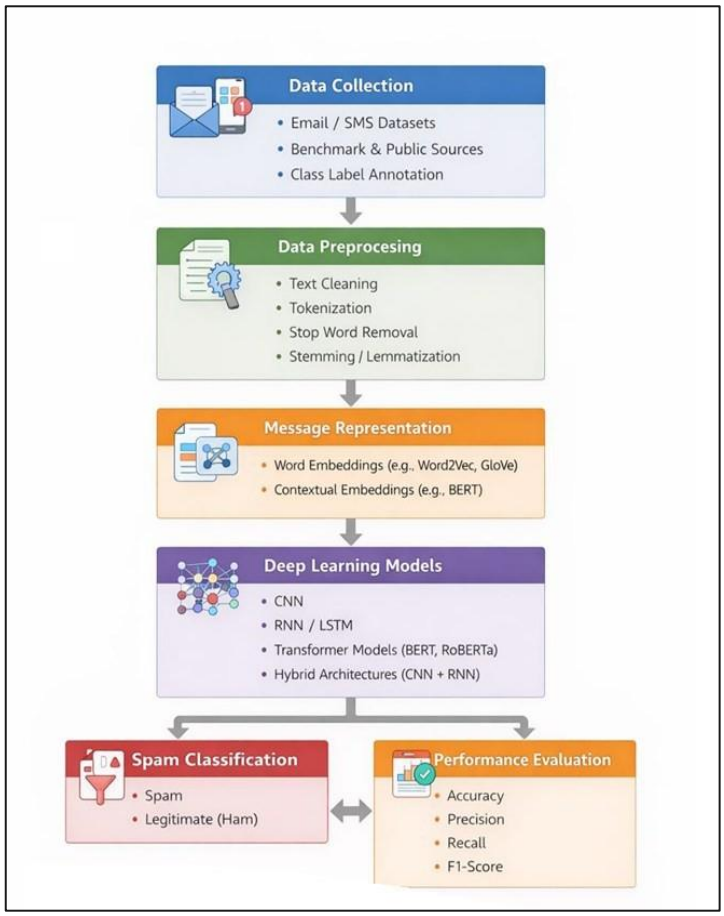

Figure 2 Deep Learning Framework

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

Deep Learning Architectures

Several deep learning architectures are investigated in order to assess how well they detect spam:

Convolutional Neural Networks (CNN): CNNs successfully capture n-gram information and may extract local features such as spam-indicative patterns and key phrases, hence they are commonly used for text classification. By using convolutional filters. CNN-based spam detectors have performed well in email and SMS spam classification tasks. Convolutional filters on word embeddings are used to extract both spatial and local textual information.

Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM): RNNs and LSTMs are designed to process data sequentially from message text and maintain contextual memory. LSTM networks are simulated using long-These models are good in detecting long-message and email spam, as well as sequential relationships.

Transformer-Based Models: Compared to traditional sequential models, systems like BERT and Roberta provide dependable categorization in spam detection by using self-attention techniques to better grasp semantic and contextual linkages in text. By understanding word meaning in context, spam messages that are dynamic and complex can be detected more accurately.

• All deep learning architectures are implemented in a controlled environment using standard frameworks. to ensure that the models are evaluated uniformly under identical conditions and that the outcomes are repeatable. The model's performance is evaluated using common classification measures, such as accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of detection abilities, particularly in cases of class imbalance. Comparative analysis is used to identify the best architecture.

RESULT AND DISCUSSION

Most studies employ static benchmark datasets that don't fairly reflect the variety of spam in the real world. Multilingual, multimodal, large-scale datasets are needed. The idea that spam strategies evolve over time is a problem for current models. Strategies for online and ongoing learning are still not thoroughly researched in terms of their ability to adapt to changing spam. Deep learning models usually lack transparency. The integration of explainable AI systems to generate human-intelligible explanations is one significant research gap noted in the survey. Additionally, the computational expense of transformer-based models necessitates lightweight designs. Currently, most efforts are focused on text-only categorization rather than multimodal classification. Combining text, metadata, behavioral, and visual data can significantly improve detection performance.

Comparative Analysis

The comparison research shows that hybrid and advanced deep learning models outperform standalone architectures. Transformer-based and CNN-LSTM models demonstrated superior generalization and adaptability across a range of spam content categories. Conventional machine learning methods were less accurate and had higher false positive rates.

Table 2: Comparison Result in deep learning model for spam detection

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	94.3	93.8	92.9	93.3
RNN	95.1	94.6	94.0	94.3
LSTM	96.0	95.5	95.2	95.3
CNN–LSTM	96.8	96.2	95.9	96.0
Transformer (BERT)	98.1	97.8	97.6	97.7

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

Transformer-based architectures outperform CNN and RNN-based models on all metrics, according to Table 2's experimental evaluation. However, their higher processing cost limits its real-time deployment. performance. Accuracy and complexity are balanced in hybrid deep learning architectures.

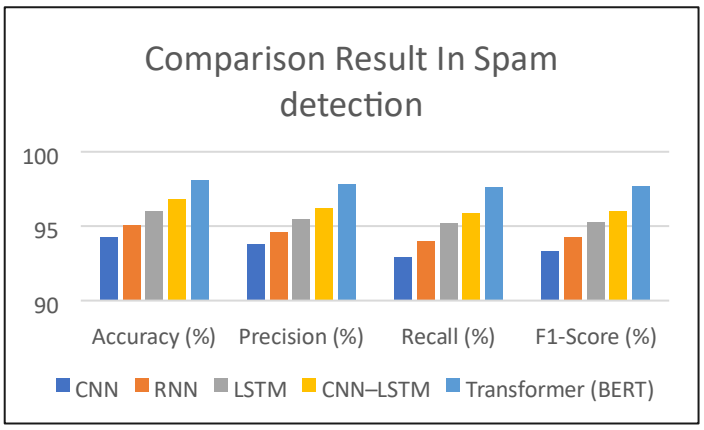

Figure 3: Deep learning Model Comparison results

Figure 3 Shows that Visual representation of various Deep learning model evaluation metrics in Spam detection comparison Analysis in exiting researcher carron work in Different kind of textual data set.

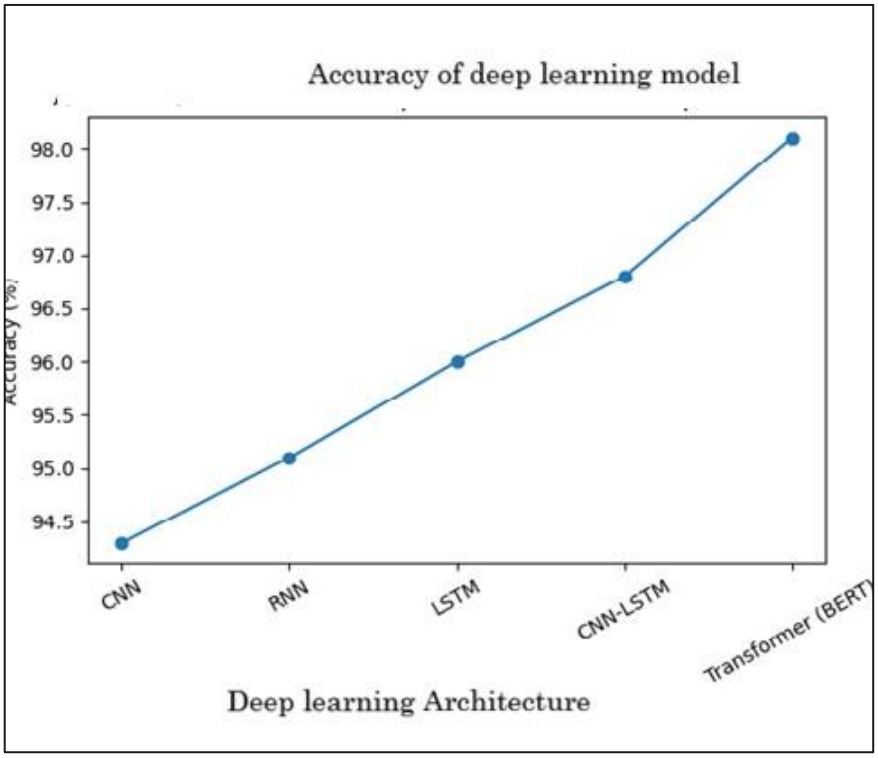

Figure 4: Accuracy of Deep learning model

The accuracy level of deep learning models, such as CNN, RNN, LSTM, and Transformer models, varies in spam text. It is clear from Figure 4 that performance has progressively improved from elderly to advanced. The comparison reveals deep learning architectures for spam detection to enhance practicality.

CONCLUSION

This article included a comparative analysis of deep learning methods for spam detection. Textual representations of spam detection accuracy have significantly enhanced thanks to CNNs, RNNs, hybrid models, and Transformer-based architectures. However, there are still unresolved research hurdles, including changing spam tactics, explainability issues, and computing limitations. In order to develop dependable next-generation spam

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

detection, future research will concentrate on adaptive, explainable, and efficient deep learning models in conjunction with multimodal data sources.. Despite this research, there are still significant problems with deep learning models, including class imbalance, data scarcity, high computational costs, explainability problems, and vulnerability to hostile attacks. To solve these problems, a reliable spam detection system must also be put in place. Multimodal spam detection, which incorporates text, URLs, metadata, and behavioral elements in sophisticated deep learning models, is currently lacking in research.

REFERENCES

1. Karim, S. Azam, B. Shanmugam, K. Kannoorpatti, and M. Alazab, "A Comprehensive Survey for Intelligent Spam Email Detection," IEEE, vol. 7, pp. 168261–168295, 2019. Doi: 10.1109/ACCESS.2019.2954791.
2. P. K. Roy, J. P. Singh, and S. Banerjee, "Deep learning to filter SMS spam," Future Generation Computer Systems, vol. 102, pp. 524–533, 2020. Doi: 10.1016/j.future.2019.09.001.
3. G. M. Shahariar, S. Biswas, F. M. Shah and S. B. Hassan, "Spam review detection using deep learning," in Proc. 2019 IEEE 10th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON), Vancouver, BC, Canada, pp. 01-07,2019. Doi: 10.1109/IEMCON.2019.8936148.
4. E. S. Rahman and S. Ullah, "Email spam detection using bidirectional long short-term memory with convolutional neural network, "in Proc. IEEE Region 10 Symp. (TENSYMP), Dhaka, Bangladesh, pp. 1307–1311,2020.
5. M. Popovac, M. Karanovic, S. Sladojevic, M. Arsenovic, and A. Anderla, "Convolutional neural network based SMS spam detection," 2018 Telecommunication Forum (TELFOR), 2018.
6. N. Govil, K. Agarwal, A. Bansal, and A. Varshney,"A machine learning based spam detection mechanism," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), pp. 954–957, 2020.
7. Kim, S. Abuadbba, and H. Kim,"DeepCapture: Image Spam Detection Using Deep Learning and Data Augmentation," arXiv:2006.08885, 2020. https://arxiv.org/abs/2006.08885
8. M. A. Shaaban, Y. F. Hassan, and S. K. Guirguis, "Deep convolutional forest: A dynamic deep ensemble approach for spam detection in text," Complex & Intelligent Systems, 2021. https://arxiv.org/abs/2110.15718
9. M. A. Uddin, M. N. Islam, L. Maglaras, H. Janicke, and I. H. Sarker, "ExplainableDetector: Exploring transformer-based language modeling approach for SMS spam detection with explainability analysis," arXiv preprint arXiv:2405.08026, 2024.
10. H. C. Altunay and Z. Albayrak,"SMS Spam Detection System Based on Deep Learning Architectures," Applied Sciences, vol. 14, no. 24, 11804, 2024. https://doi.org/10.3390/app142411804