ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Spam Email Detection using Deep learning

Jajjara Bhargav [1], Siddi Ajay Bhaskar[2], Kalluri Yuva Ramya[3], Sk mastanvali[4], Peetha Anitha[5]

[1] HOD& Assistant Professor, Department of Computer Science Engineering, Chalapathi Institute of Engineering and Technology, Chalapathi Rd, Nagar, Lam, Guntur, Andhra Pradesh- 522034

[2,3,4,5] Students, Department of Computer Science Engineering, Chalapathi Institute of Engineering and Technology, Chalapathi Rd, Nagar, Lam, Guntur, Andhra Pradesh- 522034

Email id: bhargavchalapathi@gmail.com[1], ajaybhaskar004@gmail.com[2], yuvaramya204@gmail.com[3], vmastan998@gmail.com[4], anithapeetha7200@gmail.com[5]

Abstract:

This paper delves into the growing threat of Business Email Compromise (BEC) phishing attacks, a form of cybercrime that continues to evolve, challenging traditional detection systems. BEC attacks often bypass conventional filters by lacking typical payloads, making them harder to identify using static feature extraction techniques. As a result, there is an increasing need for advanced detection methods that leverage machine learning (ML) models. This study compares the performance of three prominent ML techniques—Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and Bidirectional Long Short-Term Memory (Bi-LSTM)—in identifying BEC phishing attempts. The findings reveal that the ANN model achieves an accuracy of 89%, while the RNN model outperforms it with an accuracy of 97%. Additionally, the Bi-LSTM model further improves the detection accuracy, achieving an impressive 98%. The superior performance of RNNs and Bi-LSTM models is attributed to their ability to process sequential data and capture contextual relationships in email content, a crucial feature for identifying sophisticated phishing tactics. The study further investigates the strengths and limitations of each model, providing insights into the most effective techniques for phishing detection. Additionally, it discusses the importance of dataset quality, feature engineering, and the need for continuous adaptation to emerging phishing tactics. This work underscores the significance of adopting advanced ML models like Bi-LSTMs for more accurate and reliable BEC phishing detection, ultimately strengthening organizational cybersecurity defenses against evolving email-based threats.

Keywords: Business Email Compromise (BEC), Phishing Detection, Machine Learning, Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), Sequential Data Analysis, Email Security, Cybersecurity, Phishing Attack Mitigation.

1.Introduction

In today's digital age, email remains one of the most widely used forms of communication. However, with the increasing volume of emails exchanged daily, spam emails have become a significant issue, causing inconvenience, wasting time, and even posing security threats through phishing and malware attacks. Traditional spam detection techniques, such as rule-based filters and classical machine learning models, have shown effectiveness but often struggle to adapt to the constantly evolving tactics used by spammers.

With the advent of deep learning, more robust and intelligent models have been developed to tackle complex problems like spam detection. Deep learning models, such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN), can automatically learn and extract meaningful features from email text without the need for extensive manual preprocessing or feature engineering. These models excel at handling large datasets and capturing intricate patterns within the data, making them well-suited for spam classification tasks.

This project focuses on leveraging deep learning techniques to build an efficient and accurate spam email detection system. By training models on labeled email datasets, the system learns to distinguish between spam and legitimate (ham) emails based on the content and structure of the messages. The goal is to achieve high accuracy, precision, and recall in identifying spam, thereby improving email security and user experience.

2. Related work:

Valecha et al. [1] proposed a method for detecting phishing emails using persuasion cues. The research specifically focuses on gain and loss persuasion cues. It creates three machine learning models using these cues: one with relevant gain persuasion cues, one with relevant loss persuasion cues, and one with a combination of gain and loss persuasion cues. In their paper, the authors of [**2**] developed a spam filter that combined an integrated distribution-based balancing approach with an N-gram tf-idf feature selection and a deep multilayer perceptron neural network with rectified linear units. This filter accurately detected spam emails in the Enron and Spam Assassin benchmark datasets, even when many different features and additional layers were present in their work, the authors of [3] presented a phishing email detection model called THEMIS that utilized an improved recurrent convolutional neural network (RCNN) model with multilevel vectors and attention mechanisms. This model could simultaneously model email headers, words, email body, and characters, allowing it to identify phishing emails effectively. Alhogail and Alsabih [4] proposed a phishing email detection model that utilized deep learning and natural language processing on the email body to extract features and improve phishing detection. The model was based on a convolutional network (GCN) and was developed as a supervised learning model. The model was trained and tested on a publicly available fraud dataset, including phishing and legitimate emails In their work, Yao et al. [5] explored using graph convolutional networks (GCN) for text classification. The authors proposed a GCN-based model for text classification and evaluated its performance on several benchmark datasets. The results showed that the proposed model achieved competitive performance compared to other state-of-the-art models and demonstrated the potential of using GCN for text classification tasks. Overall, the authors of] presented a promising approach for text classification using GCN and highlighted the potential of this technique in various natural language processing tasks.

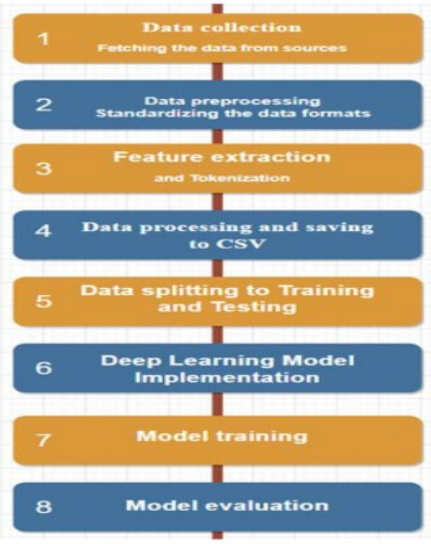
3.Methodology

Spam emails pose significant security and productivity concerns in the digital age. Traditional filtering systems rely heavily on manually crafted rules and classical machine learning techniques, which often struggle to keep up with the evolving nature of spam. In contrast, deep learning offers a powerful and adaptive solution for detecting spam emails with higher

accuracy. This article explores how deep learning techniques such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN) are used to enhance the accuracy and efficiency of spam detection systems.

Proposed Model

This section presents the steps to implement the proposed model, including collecting, preparing, and utilizing a dataset for training and testing deep learning models for detecting phishing emails. Figure illustrates the general framework followed throughout the research; it includes dataset acquisition, data preparation, feature extraction, and training and testing of various deep learning approaches. The following subsections describe the research methodology.



Dataset

Commonly used datasets for spam email detection include:

- Enron Email Dataset
- SpamAssassin Public Corpus
- SMS Spam Collection

These datasets typically contain a collection of labeled emails or messages with categories such as "spam" or "ham".

Preprocessing

Before feeding text data into a deep learning model, it must be cleaned and tokenized. Common preprocessing steps include:

- Lowercasing all text
- Removing punctuation, stopwords, and special characters
- Tokenization (splitting text into words or subwords)
- Converting tokens to sequences using techniques like **word embeddings** (Word2Vec, GloVe) or **TF-IDF**
- Padding sequences to a uniform length

Deep Learning Models for Spam Detection

1. Recurrent Neural Networks (RNN)

RNNs are capable of handling sequential data and are particularly effective in modeling text. However, standard RNNs suffer from vanishing gradients and struggle with long-term dependencies.

2. Long Short-Term Memory (LSTM)

LSTM networks, a type of RNN, address the limitations of traditional RNNs by using gating mechanisms. They are highly effective in capturing context in text data, making them suitable for spam detection.

3. Convolutional Neural Networks (CNN)

Though traditionally used for image data, CNNs can also extract spatial features from text when used with word embeddings. They are faster to train and can be effective in recognizing spam-specific patterns.

4. Hybrid Models (CNN + LSTM)

Combining CNNs and LSTMs can leverage both spatial and sequential features in text data, often resulting in improved performance.

4. Results and Discussion

Deep learning models, particularly LSTM-based architectures, tend to outperform traditional machine learning methods such as Naive Bayes or SVM. With proper tuning and sufficient training data, accuracies above 95% can be achieved. The key advantages include:

- Ability to learn from raw text
- Flexibility to adapt to new spam patterns
- Less reliance on feature engineering

Convolutional Neural Networks (CNNs)

The CNN-based deep learning classifier was implemented in Python and was trained and tested on the provided dataset. The testing accuracy was 98.74%, precision was 98.96%, recall was 98.78%, and F1-score was 98.87% after 50 epochs The figure shows the accuracy and loss plots for the convolutional neural network model during training and validation. The training accuracy starts around 90% and steadily increases with each epoch, reaching over 99% by the 50th epoch. This indicates that the model could fit the training data better with each iteration and minimize errors on the samples it was trained on. The validation accuracy follows a similar trend, starting near 90% and increasing to around 98% by the 50th epoch. However, it is slightly below the training accuracy throughout, indicating some overfitting. The training loss starts around 0.3 and decays rapidly in the first 10 epochs, plateauing under 0.1 by the 30th epoch.

Table: CNN classification report

	Precision	Recall	F1-Score	Support
0	0.98	1.0	0.99	3081
1	0.99	0.98	0.99	2331
micro avg	0.99	0.99	0.99	5412
macro avg	0.99	0.99	0.99	5412
weighted avg	0.99	0.99	0.99	5412
samples avg	0.99	0.99	0.99	5412

The CNN model demonstrated strong performance for phishing email detection, achieving an overall accuracy of 98.74% on the test set. This high accuracy indicates that the model correctly classified most phishing and legitimate emails. The precision of 98.96% shows that of all emails classified by the model as phishing, only a small fraction was mis labeled. The recall of 98.78% means the model could correctly detect almost all the actual phishing emails in the test set, with very few phishing emails missed. Finally, the F1-score of 98.87% reflects the excellent balance between precision and recall attained by the model. Overall, these metrics validate that the CNN model was highly proficient at distinguishing phishing and legitimate emails. The combination of high precision and recall underscores the model's reliability in flagging phishing emails while minimizing false alarms on legitimate emails.

Convolutional Neural Networks (CNNs)

Table illustrates a comparison between the research results convolutional neural networks (CNNs). was used in the comparison because it proposes a deep learning model based on CNN for detecting phishing emails.

Table: CNN comparison

Accuracy	98%	98.74%
Precision	98.50%	98.96%
Recall	98%	98.78%
F1-score	98%	98.87%

Our proposed model achieved an accuracy of 98.74%, which indicates that the model correctly classified 98.74% of emails as either phishing or legitimate. It also achieved a precision of 98.96%, which indicates that when the model classified an email as phishing, it was correct 98.96% of the time. The recall of the proposed model was 98.78%, which indicates that the model correctly identified 98.78% of the phishing emails. The F1-score was 98.87%, which indicates that the proposed model has a high level of accuracy in identifying phishing emails.

Conclusion

Deep learning has revolutionized the way we approach spam detection. Its ability to learn and adapt from data makes it highly effective in a constantly evolving threat landscape. While challenges like data imbalance and model interpretability remain, ongoing advancements in natural language processing and model optimization continue to enhance the robustness and accuracy of spam filters. Additionally, adaptive phishing email filtering needs to be studied so that the system can automatically learn, adapt, and identify phishing emails based on their behaviors. Transformers, which are a type of deep learning model, have shown significant potential in natural language processing tasks such as text classification, machine translation, and text generation. Given the nature of phishing emails, which rely on language-based deception to trick recipients into taking unwanted actions, transformers are a promising avenue for improving the accuracy of phishing email detection. One possible direction for future research in this area is the development of transformer-based models for phishing email detection.

References:

1. Valecha, R.; Mandaokar, P.; Rao, H.R. Phishing Email Detection using Persuasion Cues. *IEEE Trans. Depend. Secure Comput.* **2021**, *19*, 747–756
2. Barushka, A.; Hajek, P. Spam filtering using integrated distribution-based balancing approach and regularized deep neural networks. *Appl. Intell.* **2018**, *48*, 3538–3556.
3. Fang, Y.; Zhang, C.; Huang, C.; Liu, L.; Yang, Y. Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism). *IEEE Access Pract. Innov. Open Solut.* **2019**, *7*, 56329–56340.
4. Alhogail, A.; Alsabih, A. Applying machine learning and natural language processing to detect phishing email. *Comput. Secur.* **2021**, *110*, 102414.
5. kaddoura, S.; Alfandi, O.; Dahmani, N. A spam email detection mechanism for English language text emails using deep learning approach. In Proceedings of the 2020 IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), Bayonne, France, 10–13 September 2020.
6. Singh, S.; Singh, M.P.; Pandey, R. Phishing detection from URLs using deep learning approach. In Proceedings of the 2020 5th International Conference on Computing, Communication and Security (ICCCS), Patna, India, 14–16 October 2020
7. Saha, I.; Sarma, D.; Chakma, R.J.; Alam, M.N.; Sultana, A.; Hossain, S. Phishing attacks detection using deep learning approach. In Proceedings of the 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 20–22 August 2020.
8. McGinley, C.; Monroy SA, S. Convolutional neural network optimization for phishing email classification. In Proceedings of the 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 15–18 December 2021; IEEE: New York, NY, USA, 2021; pp. 5609–5613