Int. J. Sci. R. Tech., 2025 2(7)
A Multidisciplinary peer-reviewed Journal
www.ijsrtjournal.com [ISSN: 2394-7063]

Original Article

Email Spam Detection

Dheeraj Iti*, Rachana Chavan

Department of Computer Applications JSPM University Pune, India

ABSTRACT

The exponential increase in spam emails poses a significant challenge to digital communication and cybersecurity. Traditional machine learning techniques such as Naïve Bayes and Support Vector Machines (SVM) have been widely employed for spam detection but often fall short in handling the evolving nature and complexity of spam content. This study presents an advanced email spam detection model based on Long Short-Term Memory (LSTM) networks, which are known for their strength in processing sequential data and capturing contextual dependencies in textual information. The research adopts a quantitative experimental approach, utilizing a Kaggle dataset comprising 5,572 email samples. Comprehensive preprocessing, including stop-word removal, tokenization, and word embeddings, was employed to enhance data quality. The LSTM model was trained and evaluated using accuracy, precision, recall, and F1-score metrics and compared with SVM and Naïve Bayes classifiers. Experimental results revealed that the proposed LSTM model significantly outperformed traditional models, achieving 98.74% accuracy, 97.50% precision, and 98.11% F1-score. These findings highlight the model's ability to reduce false positives and adapt to dynamic spam patterns. This research demonstrates that LSTM networks provide a scalable and effective solution for real-time spam detection, with potential applications in email filtering, enterprise communication security, and automated message categorization systems.

Keywords: Deep Learning, Email Classification, LSTM, Machine Learning, Spam Detection, Text Preprocessing

INTRODUCTION

Electronic mail (email) continues to be one of the most widely used forms of communication globally, with more than 4.6 billion active accounts. Despite its convenience and efficiency, email is frequently exploited to distribute spam—unsolicited, irrelevant, or malicious messages. These spam emails not only frustrate users but also pose substantial cybersecurity threats, including phishing attacks, malware dissemination, and the unauthorized collection of sensitive data. Reports indicate that over 14 billion spam messages are sent daily, and even minimal user interaction can yield substantial financial gain for cybercriminals. Traditional spam detection systems primarily utilize rule-based filters and classical machine learning algorithms such as Naïve Bayes (NB) and Support Vector Machines (SVM). While these approaches have achieved moderate success, they often fail to adapt to evolving spam tactics such as contextual manipulation, obfuscated content, and embedded multimedia. These limitations have driven research toward more intelligent and adaptive models.

This study proposes the use of Long Short-Term Memory (LSTM) networks—a form of Recurrent Neural Network (RNN)—to detect spam emails. LSTM models are well-suited for this task as they are designed to analyze sequential data and capture long-term dependencies, which are essential for identifying spam patterns in email text.

The objectives of this research are:

1. To develop an LSTM-based model capable of accurately classifying spam and non-spam emails.
2. To compare its performance with conventional methods like SVM and NB.
3. To assess the effectiveness of text representation methods such as TF-IDF and word embeddings.
4. To evaluate the model using standard classification metrics including accuracy, precision, recall, and F1-score.

MATERIALS AND METHODS

2.1 Dataset Description

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The dataset used in this study was obtained from Kaggle and titled *"Email Spam Classification Dataset"*. It consists of **5,572 labeled email samples**, of which **747 are spam** and **4,825 are non-spam (ham)**. Each email message is classified as either spam or ham, allowing for supervised learning. The dataset was randomly partitioned into **80% for training** and **20% for testing** to ensure unbiased performance evaluation.

2.2 Tools and Frameworks

The following software tools and libraries were utilized to implement and evaluate the spam detection system:

Tool/Library	Purpose
Python 3.8+	Programming environment
TensorFlow & Keras	LSTM model development and training
Scikit-learn	Feature extraction (TF-IDF) and evaluation metrics
NLTK (Natural Language Toolkit)	Text cleaning and tokenization
Pandas & NumPy	Data manipulation and numerical operations
Matplotlib & Seaborn	Visualization of accuracy, loss, and other metrics

2.3 Preprocessing Steps

To enhance data quality, the email texts underwent several preprocessing steps:

- **Lowercasing** all text
- **Removing stop-words**, special characters, and URLs
- **Tokenizing** the cleaned text
- **Transforming** tokens into numerical sequences using word embeddings and padding to fixed lengths

2.4 Model Architecture

The architecture of the LSTM model includes:

- **Embedding Layer**: Converts words into 128-dimensional vectors.
- **LSTM Layers**: Two stacked LSTM layers (128 and 64 units) for sequential learning.
- **Dropout Layers**: Applied after each LSTM to prevent overfitting (rate: 0.2).
- **Dense Layer**: Fully connected layer with ReLU activation.
- **Output Layer**: Sigmoid activation for binary classification.

2.5 Training Configuration

- **Batch Size**: 64
- **Epochs**: 5
- **Loss Function**: Binary Cross-Entropy
- **Optimizer**: Adam
- **Evaluation Metrics**: Accuracy, Precision, Recall, F1-Score

The model was trained on a high-performance machine equipped with an NVIDIA RTX 3090 GPU, which significantly accelerated the training process.

RESULTS AND DISCUSSION

3.1 Performance Evaluation

The Long Short-Term Memory (LSTM) model was trained and tested using the pre-processed dataset. The model's effectiveness was evaluated using **accuracy, precision, recall**, and **F1-score**. These metrics provide a comprehensive view of the classification performance. The results were compared with two widely used machine learning models—**Naïve Bayes (NB)** and **Support Vector Machine (SVM)**. Table 1 summarizes the comparative results.

Table 1: Model Performance Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Naïve Bayes (NB)	91.23	89.76	87.92	88.83
Support Vector Machine	94.56	92.81	91.50	92.15
LSTM (Proposed)	**98.74**	**97.50**	**98.74**	**98.11**

These results clearly show that the proposed LSTM model outperforms traditional classifiers across all evaluation metrics.

3.2 Visual Analysis

The model's training history was visualized using Matplotlib. The following trends were observed:

- **Training Accuracy and Validation Accuracy** steadily increased over the epochs.
- **Loss values** for both training and validation decreased, indicating effective learning and low overfitting.

3.3 Comparative Study with Previous Research

A comparison with recent LSTM-based studies further validates the strength of this approach:

Table 2: Comparison with Previous Research

Study	Model	Accuracy (%)	Remarks
Wijaya et al. (2022)	LSTM	95.60	Dataset: 5,000 emails
Isik et al. (2020)	LSTM + Feature Selection	100.00	Dataset: Turkish spam emails
Lekhya et al. (2024)	LSTM (Email & SMS)	96.19	Hybrid approach for multi-channel spam
This Study (2025)	**LSTM**	**98.74**	Improved preprocessing and optimized architecture

The proposed model's enhanced performance is attributed to **robust preprocessing**, **hyperparameter tuning**, and a **specialized focus on email-only spam**.

DISCUSSION

Why LSTM is More Effective

- Captures **long-term dependencies** in email text
- Learns **contextual relationships** and spam patterns
- Uses dropout layers to prevent overfitting
- Applies the **Adam optimizer** for efficient convergence

Implications

- **Fewer false positives** ensure minimal disruption to users.
- The LSTM model is adaptable to **new spam patterns**.
- Can be scaled and deployed in **real-time filtering systems**.

4. Conclusion and Future Work

This study has demonstrated the effectiveness of **Long Short-Term Memory (LSTM)** networks in detecting spam emails with significantly greater accuracy than traditional machine learning techniques such as **Naïve Bayes** and **Support Vector Machines (SVM)**. Through the use of advanced preprocessing techniques and optimized model architecture, the proposed LSTM model achieved **98.74% accuracy**, **97.50% precision**, and an **F1-score of 98.11%**. The experimental results confirm that LSTM networks are particularly suitable for handling **sequential and contextual data**, which are critical in distinguishing between legitimate and spam emails. Additionally, the model's robustness in reducing false positives and false negatives indicates its practical applicability in real-world scenarios such as **enterprise email systems**, **webmail platforms**, and **messaging applications**. Despite its high accuracy, the model is not without limitations. The dataset used is relatively small and text-centric, which does not reflect the

complexity of **multi-modal spam** (e.g., spam containing images, attachments, or embedded URLs). Furthermore, the **computational demands** of LSTM architectures may restrict deployment in **resource-constrained environments**.

Future Work Recommendations

To enhance the model's effectiveness and scalability, the following directions are recommended:

- Expand the dataset to include **larger and more diverse email sources**.
- Explore **lightweight LSTM variants** (e.g., Distilled LSTM, Edge-LSTM) for mobile deployment.
- Integrate **hybrid models** combining CNN, Attention Mechanisms, or Transformer-based architectures.
- Develop **multi-lingual and multi-modal spam detection systems** to increase generalizability.
- Focus on **adversarial robustness** to resist evolving spam attack strategies.

ACKNOWLEDGEMENTS

- The Dheeraj Iti (author) would like to express sincere gratitude to the faculty and research coordinators of Department of Computer Applications, **JSPM University**, for their valuable support and guidance throughout this study. Special thanks are extended to the mentors and reviewers who provided constructive feedback during the development and evaluation phases of this research.

- Appreciation is also due to the creators and maintainers of the public datasets and open-source libraries that were instrumental in the implementation of the LSTM-based spam detection system. Lastly, the authors acknowledge the contribution of peers and collaborators who offered technical assistance and moral support during the completion of this work.

REFERENCE

1. S. Cahyadi et al., "LSTM sentiment analysis: Use in social media text," International Journal of Analytics and Data Science, vol. xx, no. xx, pp. xx–xx, 2020.
2. M. Isik et al., "Analysis of LSTM-based feature selection techniques for spam detection," in Proc. Decision Aid Sciences and Applications Int. Conf. (DASA), 2020.
3. C. Lekhya et al., "Enhanced spam identification using an LSTM-based method," in Proc. 3rd Int. Conf. on Intelligent Techniques in Signal Processing, Control, and Optimization (INCOS), 2024.
4. R. Mubarikah, "Comparison of SVM and LSTM for classifying email spam," IEEE Cybersecurity Trans., vol. xx, no. xx, pp. xx–xx, 2021.
5. A. Ozcan et al., "A DNN–LSTM hybrid model for phishing URL identification," Applications of Neural Computing, vol. xx, no. xx, pp. xx–xx, 2021.
6. H. Wijaya et al., "Long Spam: A spam detection model based on LSTM," Applications of Machine Learning J., vol. xx, no. xx, pp. xx–xx, 2022.

HOW TO CITE: Dheeraj Iti*, Rachana Chavan, Email Spam Detection, Int. J. Sci. R. Tech., 2025, 2 (7), 322-325. https://doi.org/10.5281/zenodo.16017716