Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4827

Submitted : Oct 10, 2024
Accepted : Oct 22, 2024
Published : Oct 31, 2024

A Comparative Analysis of Deep Learning Models for SMS Spam Detection: CNN-LSTM, CNN-GRU, and ResNet Approaches

Gregorius Airlangga[1)*]
[1)] Atma Jaya Catholic University of Indonesia, Indonesia
[1]gregorius.airlangga@atmajaya.ac.id

ABSTRACT

Spam messages have become a growing challenge in mobile communication, threatening user security and data privacy. Traditional spam detection methods, including rule-based and machine learning techniques, are increasingly insufficient due to the evolving sophistication of spam tactics. This research evaluates the effectiveness of advanced deep learning models such as CNN-LSTM, CNN-GRU, and ResNet for SMS spam detection. The dataset used consists of diverse SMS messages labeled as either spam or legitimate (ham), ensuring broad coverage of real-world spam patterns. The study employs a robust ten-fold cross-validation approach to assess the generalization capabilities of the models, measuring performance based on accuracy, precision, recall, and F1 score. The results indicate that ResNet outperformed the other models, achieving an average accuracy of 99.08% and an F1 score of 0.9646, making it the most reliable model for spam detection. CNN-GRU demonstrated competitive performance with a balance between accuracy (98.97%) and computational efficiency, making it suitable for real-time applications. CNN-LSTM, while highly accurate (98.92%), showed a slightly lower recall compared to the other models, indicating a more cautious approach to detecting spam. These findings highlight the potential of hybrid deep learning models in addressing the complexities of SMS spam detection. Future research could focus on optimizing these models for deployment in resource-constrained environments, such as mobile devices, and further exploring the integration of residual connections for more effective spam filtering.

Keywords: SMS Spam Detection; CNN-LSTM; CNN-GRU; ResNet; Deep Learning Models

INTRODUCTION

The importance of short message services for businesses, verification purposes and individuals is directly proportional to the increase in potentially dangerous and focused spam messages (Alkhalil et al., 2021; Kigerl, 2020; Rao et al., 2021). These spam messages, ranging from phishing attempts to fraudulent schemes, not only cause inconvenience but also pose serious security risks (Vijayakumar & Thomas, 2024). Traditional rule-based systems and keyword filters have long been used to detect spam, but these methods are becoming increasingly inadequate in the face of evolving spam techniques (Jain, 2021). Spammers have adapted their tactics, using more complex and diverse patterns to bypass detection (Salman et al., 2024). As a result, the need for more sophisticated methods of spam detection has grown, and machine learning, followed by deep learning, has become a focal point in tackling this issue (Do et al., 2022). The early adoption of machine learning models, such as Naive Bayes, Support Vector Machines (SVM), and Decision Trees, brought improvements in spam detection by leveraging labeled datasets to classify messages (Agarwal et al., 2024). These models, however, struggled with the complexity of unstructured text data, especially in understanding the context and sequence of words (Gasparetto et al., 2022). Ensemble methods like Random Forest and Gradient Boosting improved accuracy by combining multiple models, yet these approaches were still limited in their ability to handle long-term dependencies within text (Kumari & Toshniwal, 2021).

The rise of deep learning has revolutionized text classification tasks, including SMS spam detection. Convolutional Neural Networks (CNNs), originally developed for image processing, proved effective in identifying local patterns in text, such as keywords or phrases common in spam messages (Sharmin et al., 2020). However, CNNs alone cannot capture the sequential nature of text, leading to the adoption of Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), which can preserve context over longer sequences (Oruh et al., 2022). By combining CNNs with LSTMs or GRUs, hybrid models have emerged that exploit the strengths of both architectures, showing remarkable improvements in spam detection performance. In addition to these models, Residual Networks (ResNet), known for their success in deep learning tasks like image classification, have been explored for text classification tasks (Minu & Canessane, 2022). ResNet's use of skip connections enables the development of deeper networks without the risk of performance degradation due to

* Corresponding author

vanishing gradients. This allows the model to capture more complex patterns in SMS data, potentially improving spam detection capabilities (Mehnatkesh et al., 2023). Despite their success in other domains, ResNet models remain underutilized in SMS spam detection, leaving room for further exploration.

Spam detection research has a long history, starting with rule-based systems that relied on manually defined keywords and patterns to identify unwanted messages (Saidani, 2021). These early approaches were effective against static forms of spam but quickly became obsolete as spammers learned to circumvent simple filters (Gaurav et al., 2020). Machine learning models, including Naive Bayes, SVM, and Decision Trees, provided a more adaptive solution by learning from labeled data. However, these models struggled with text's unstructured nature, particularly when spam messages were long or contained subtle differences from legitimate messages (Dey et al., 2024). Ensemble methods like Random Forest and Gradient Boosting further enhanced spam detection by combining multiple weak classifiers to improve robustness. These methods offered better accuracy but lacked the ability to fully capture the contextual relationships between words (Ansari et al., 2022). The introduction of deep learning, particularly CNNs, LSTMs, and GRUs, marked a major shift in text classification approaches. CNNs were able to detect localized features in text data, such as specific phrases, while LSTMs and GRUs captured long-term dependencies in sequential data, allowing for a more nuanced understanding of message content (Islam et al., 2024). Hybrid models combining CNNs with LSTMs or GRUs have shown promise in SMS spam detection. CNN layers are effective at identifying key features within messages, while LSTM or GRU layers capture the sequence of words, helping the model understand context (Ahmadzadeh et al., 2022). This dual approach has led to more accurate spam detection models, capable of adapting to diverse and evolving spam strategies. Although less commonly applied in this domain, ResNet has demonstrated its ability to create deeper networks that retain performance through the use of residual connections. This allows for more complex feature learning without the risk of overfitting or gradient issues, making ResNet a potential candidate for enhancing spam detection models (Durga & Rajesh, 2022). Despite this, ResNet's application to text classification remains limited, particularly in SMS spam detection. Modern SMS spam detection methods predominantly rely on hybrid deep learning architectures that combine CNNs with LSTM or GRU layers. These models capitalize on the strengths of both types of layers: CNNs extract critical local features from the text, while LSTMs or GRUs capture sequential dependencies and context (Daraghmi et al., 2024). This combination has proven effective in handling the diverse and adaptive nature of spam messages, which can vary significantly in structure and content.

Additionally, ResNet's introduction to text classification tasks offers new possibilities for SMS spam detection. By employing skip connections, ResNet allows for deeper networks that can learn more complex patterns, enhancing the model's ability to differentiate between spam and legitimate messages (Nandwani & Verma, 2021). However, ResNet's use in this specific domain has not been widely explored, leaving a gap in current research. Another challenge faced by deep learning models is the computational demand required for training and real-time deployment, particularly in mobile or resource-constrained environments. Thus, there is growing interest in optimizing models for both accuracy and efficiency (Abbas, 2021). While significant strides have been made in SMS spam detection using deep learning models, several gaps persist. Although hybrid models such as CNN-LSTM and CNN-GRU have been widely adopted, there is limited research comparing the performance of these models on the same dataset, making it difficult to assess their relative strengths and weaknesses (Klemm & Vennemann, 2021). Additionally, while ResNet has shown promise in other areas of text classification, its application to SMS spam detection remains underexplored.

Furthermore, much of the existing research focuses on improving model accuracy, often overlooking the need for computational efficiency in real-world applications (Yin et al., 2021). Many of the current models are computationally intensive and may not be suitable for deployment on mobile devices, where processing power is limited (Li et al., 2024). Addressing these gaps requires a more comprehensive evaluation of different models, considering both their performance and their practical feasibility for real-time spam detection in constrained environments (Cai et al., 2022). This study contributes to the field of SMS spam detection by providing a detailed evaluation of hybrid models, including CNN-LSTM, CNN-GRU, and ResNet (Rayan, 2022). Through comparative analysis, this research offers insights into the strengths and weaknesses of each model, highlighting their effectiveness in detecting spam messages. In addition to improving detection accuracy, this study emphasizes the importance of computational efficiency, exploring the trade-offs between model complexity and performance. The introduction of ResNet into SMS spam detection also fills a gap in current literature, offering a novel approach to handling complex text patterns.

The rest of this article is structured as follows: The Methodology section describes the dataset, data preprocessing techniques, model architecture, and evaluation metrics used in this study. The Results section presents the findings from the experiments, including a comparison of the performance of CNN-LSTM, CNN-GRU, and ResNet models. The Discussion section interprets these results in the context of existing research, highlighting the practical

* Corresponding author

Submitted : Oct 10, 2024
Accepted : Oct 22, 2024
Published : Oct 31, 2024

implications of each model's performance. Finally, the Conclusion summarizes the key contributions of the research and suggests potential directions for future work in the field of SMS spam detection.

LITERATURE REVIEW

SMS spam detection has been a critical area of research for over two decades, evolving from simple keyword-based systems to advanced machine learning and deep learning models (Sabeeh et al., 2020). The main challenge in this domain lies in the ability to adapt to the dynamic and evolving nature of spam messages, which often vary in structure and content. Early detection methods primarily relied on rule-based systems, which flagged messages based on predefined keywords or patterns associated with spam (Thakur et al., 2023). These systems were initially effective but quickly became obsolete as spammers adapted their strategies to evade detection. The static nature of rule-based systems could not keep pace with the increasing sophistication of spam techniques, leading to a high number of false negatives as spammers learned to bypass these basic filters (Tubishat et al., 2023). Moreover, rule-based approaches were limited in their ability to handle language variability and failed to generalize across different contexts, making them insufficient for comprehensive spam detection.

As limitations of rule-based recognition methods to machine learning, researchers turned to machine learning models such as Naive Bayes, Support Vector Machines (SVM), and Decision Trees. These models utilized labeled datasets to classify messages as either spam or legitimate based on text features such as word frequency or the presence of specific characters (Das et al., 2023). While Naive Bayes was widely adopted due to its simplicity and efficiency in handling text classification, it, like other traditional machine learning models, struggled with the complexity of unstructured text data. These models treated words as independent entities, neglecting the relationships between them, which are often critical in understanding the overall intent of a message (Sabir et al., 2021). Furthermore, these machine learning models required manual feature extraction, a process that was both time-consuming and prone to inaccuracies. As the volume and complexity of spam messages grew, the ability of traditional machine learning models to effectively distinguish spam from legitimate messages diminished, particularly in cases where the content was more nuanced (Karasoy & Balli, 2022).

Research of Random Forest and Gradient Boosting were introduced to improved robustness and accuracy, ensemble methods such as Random Forest and Gradient Boosting were introduced. These methods combined multiple weak classifiers to generate better predictions, reducing the likelihood of overfitting and improving generalization (Naseem et al., 2021). Although ensemble methods outperformed single-model approaches in many cases, they still relied on manual feature engineering and could not fully capture the sequential nature of text data. Additionally, ensemble models were computationally expensive, making them less suitable for large-scale datasets (Remya et al., 2024). The introduction of deep learning models marked a significant advancement in text classification and spam detection. Unlike traditional machine learning models, deep learning methods such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) could automatically learn features from raw data, eliminating the need for manual feature extraction (Cao et al., 2020; Kernbach & Staartjes, 2022). CNNs, initially developed for image processing, were adapted for text classification tasks by applying convolutional filters to text sequences (Vankdothu & Hameed, 2022). This allowed CNNs to detect local patterns in the data, such as commonly occurring words or phrases in spam messages. However, CNNs were limited in their ability to model long-range dependencies within text, as they primarily focused on local features.

Nowadays, researcher began working for hybrid model to address the limitation of CNNs, researchers began exploring hybrid models that combined CNNs with Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) (Akhter et al., 2020). These hybrid models leveraged the strengths of both architectures: CNNs for local feature extraction and LSTMs or GRUs for capturing the sequential relationships between words. LSTM networks were particularly well-suited for handling long sequences of data, maintaining context over time and preserving important information throughout the message (Mienye et al., 2024). GRUs, a simplified variant of LSTMs, offered similar performance but with reduced computational overhead, making them a popular choice for SMS spam detection tasks where efficiency is critical. Hybrid models such as CNN-LSTM and CNN-GRU demonstrated superior performance in detecting spam messages compared to traditional machine learning models and standalone CNNs (Bukhari et al., 2024). These models effectively handled the complex, sequential nature of text data, allowing them to capture both local and global context within a message. However, despite their improved accuracy, these models presented increased computational demands, particularly in real-time applications. The balance between performance and efficiency remains a key challenge for deploying these models in large-scale or mobile applications.

In addition to hybrid models, Residual Networks (ResNet) have been introduced in the domain of text classification,

* Corresponding author

Submitted : Oct 10, 2024
Accepted : Oct 22, 2024
Published : Oct 31, 2024

offering a new approach to improving spam detection. ResNet, originally developed for image classification tasks, uses skip connections to allow the model to bypass certain layers, addressing the issue of vanishing gradients and enabling the development of deeper networks (Jafari & Byun, 2023). Deeper networks can capture more complex features in the data, potentially improving spam detection accuracy. While ResNet has shown great promise in other deep learning applications, its use in SMS spam detection has been limited. This gap in the research provides an opportunity to explore the potential of ResNet in this domain, particularly when combined with other architectures such as CNNs or LSTMs (Ahmed et al., 2023).

Despite the significant advancements in SMS spam detection, several gaps remain in the current literature. Although hybrid models like CNN-LSTM and CNN-GRU have been widely adopted, there is a lack of comprehensive studies that compare the performance of these models on the same dataset (Dua et al., 2021; Hua et al., 2023; Lu et al., 2022). Such comparative studies are essential for understanding the relative strengths and weaknesses of each approach, especially in terms of their ability to generalize across different types of spam messages (Guo et al., 2023). Our research is conducted by considering this gap, especially in evaluating several deep learning approaches in SMS spam domain.

METHOD

This section outlines the methodology used to develop and evaluate SMS spam detection models, with a focus on addressing current challenges such as improving model generalization, computational efficiency, and accuracy. The study applies advanced deep learning architectures, including CNN-LSTM, CNN-GRU, and ResNet, with the implementation of ten-fold cross-validation to ensure robust evaluation and reliable results.

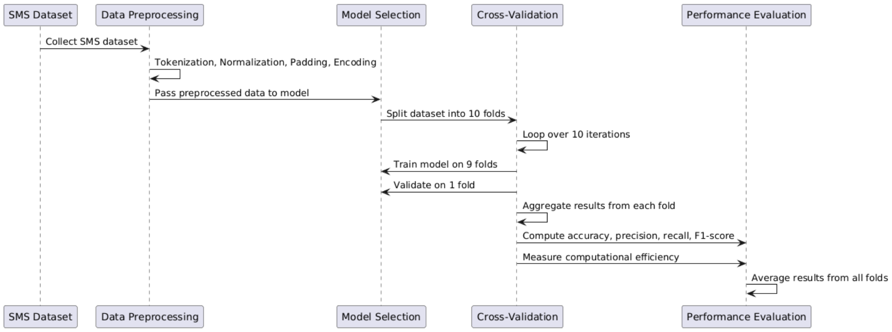

Figure 1. Sequence Diagram of Research of Methodology

Dataset Description

The dataset used in this study consists of SMS messages labeled as either spam or legitimate (ham) and can be download from (Dapat, 2024). These messages were collected from various sources to ensure diversity in language, structure, and content. The dataset provides a binary classification where the messages are labeled as spam (1) or ham (0). It is essential to have a balanced dataset to avoid model bias, ensuring that both types of messages are well-represented. Prior to model development, the dataset was divided into two main subsets: 80% of the data was used for training the models, while 20% was held out for final testing to evaluate the models' performance on unseen data.

Data Preprocessing

In text-based classification tasks, preprocessing is a critical step that significantly impacts model performance. The preprocessing pipeline for this study involved several key stages: tokenization, text normalization, sequence padding, and encoding. Tokenization involves splitting the SMS messages into individual words or tokens, which allows the model to process textual data. In this study, tokenization was carried out using TensorFlow's Keras library, which converted the words into numerical representations. Once the text was tokenized, each token was mapped to an integer, forming a sequence of numbers that the model could process.

Text normalization was applied to ensure consistency across the dataset. This process involved converting all text to lowercase, removing punctuation, and eliminating stopwords that did not contribute to the classification task. These normalization steps helped reduce noise and focus the model on meaningful patterns in the text. Furthermore, padding was applied to ensure that all text sequences were of uniform length. Since SMS messages vary in length, it was necessary to define a maximum sequence length based on the longest message in the dataset. Shorter sequences were padded with zeros to meet the required length, ensuring that the input to the models had a consistent format. Finally,

* Corresponding author

Submitted : Oct 10, 2024
Accepted : Oct 22, 2024
Published : Oct 31, 2024

encoding was performed to transform the preprocessed text into a format suitable for input into the deep learning models. The Keras Tokenizer was used to convert each word into its corresponding numerical representation, creating sequences of integers for the SMS messages. This encoding process enabled the models to process the textual data and learn from the patterns present in the message content.

Model Architecture

Three deep learning models were developed for this study: CNN-LSTM, CNN-GRU, and ResNet. Each model was designed to address the specific challenges of SMS spam detection by leveraging different aspects of deep learning techniques. The CNN-LSTM model integrates Convolutional Neural Networks (CNNs) for local feature extraction with Long Short-Term Memory (LSTM) networks for capturing long-term dependencies in the text. CNN layers are particularly effective at identifying local patterns, such as specific words or phrases that frequently appear in spam messages. The LSTM layers then capture the sequential nature of the data, maintaining context over longer text sequences and ensuring that the order of words is considered. This hybrid approach allows the model to handle both local and global information, making it well-suited for the complex nature of SMS messages.

In the CNN-GRU model, a similar approach is used, but with the LSTM layers replaced by Gated Recurrent Units (GRUs). GRUs are a variant of LSTMs and offer similar performance benefits in capturing sequence dependencies, but with fewer parameters and reduced computational complexity. This makes the CNN-GRU model more computationally efficient, especially in cases where training speed and resource usage are critical considerations. By combining CNN layers for feature extraction and GRU layers for sequence learning, the model is able to capture both the structure of the text and its temporal relationships with less computational overhead compared to the CNN-LSTM model.

The ResNet model introduces the concept of residual connections, which allow the network to bypass certain layers and mitigate the vanishing gradient problem commonly encountered in deep neural networks. Residual connections make it possible to build deeper networks that can capture more complex features in the data without sacrificing performance. The ResNet model applied in this study leverages these deep connections to extract intricate patterns in the SMS messages, which can help improve detection accuracy, particularly for more sophisticated or subtle forms of spam. Although ResNet is traditionally used in image classification tasks, its ability to handle deep feature extraction makes it a promising candidate for text-based tasks like SMS spam detection.

Ten-Fold Cross-Validation

To ensure that the models were evaluated rigorously and to minimize the risk of overfitting, the study employed ten-fold cross-validation. This technique is a robust method for evaluating model performance by dividing the dataset into ten equal parts, or "folds." In each iteration of cross-validation, nine folds were used to train the model, while the remaining fold was used for validation. This process was repeated ten times, with each fold being used as the validation set exactly once. By averaging the results across all ten iterations, a more reliable and generalizable estimate of model performance was obtained. Ten-fold cross-validation offers several advantages over traditional train-test splits. First, it ensures that the model is trained and validated on multiple subsets of the data, which reduces the likelihood of overfitting and provides a better measure of how well the model will perform on unseen data. Second, it helps to account for any variability in the data by ensuring that every data point is used for both training and validation. This is particularly important in SMS spam detection, where the patterns in the data can be subtle, and the risk of overfitting to specific subsets is high.

Evaluation

The models were evaluated using several performance metrics, including accuracy, precision, recall, and F1-score, which are standard measures for classification tasks. Accuracy measures the overall correctness of the model's predictions, providing a general sense of its performance. Precision and recall offer more specific insights, with precision assessing the proportion of predicted spam messages that were correctly classified, and recall measuring the proportion of actual spam messages that were correctly identified. The F1-score provides a harmonic means of precision and recall, offering a balanced view of the model's performance, particularly in cases where there is a trade-off between false positives and false negatives. In addition to these classification metrics, computational efficiency was also a key focus of the evaluation, particularly for the CNN-GRU model, which was designed to reduce computational complexity. Computational efficiency was assessed by measuring the time taken to train each model and the resources required for inference, particularly in terms of memory and processing power.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4827

Submitted : Oct 10, 2024
Accepted : Oct 22, 2024
Published : Oct 31, 2024

RESULT

As presented in the table 1, the performance of the five models CNN-LSTM, CNN-GRU, ResNet, GRU, and BiLSTM was evaluated on the SMS spam detection task. The results for each model were assessed using metrics such as accuracy, F1 score, precision, and recall. The use of ten-fold cross-validation for CNN-LSTM, CNN-GRU, and ResNet ensured a robust evaluation of their generalization capabilities, while the GRU and BiLSTM models were evaluated using standard test accuracy on a separate test set. Below is a detailed discussion of the results for each model. The CNN-LSTM model achieved an average accuracy of 98.92% with a standard deviation of ±0.0047, demonstrating its ability to generalize well across the different folds in the dataset. The average F1 score for the model was 0.9579, reflecting a balanced performance between precision and recall. The precision for CNN-LSTM was 0.9773, which indicates that the model was highly effective in identifying spam messages while minimizing false positives. The recall, however, was slightly lower at 0.9403, suggesting that a small proportion of actual spam messages were misclassified as legitimate.

The results suggest that the CNN-LSTM model effectively captures both local features (via the CNN layers) and the sequential structure of the SMS messages (via the LSTM layers). The relatively high precision indicates that the model is more cautious in predicting spam, potentially at the expense of recall. This makes CNN-LSTM suitable for scenarios where minimizing false positives is critical, such as in security-sensitive applications. The CNN-GRU model performed slightly better than the CNN-LSTM model, with an average accuracy of 98.97% and a smaller standard deviation of ±0.0043. The F1 score was also higher at 0.9596, indicating a slightly better balance between precision and recall. The precision for CNN-GRU was 0.9807, while the recall was 0.9399. These results suggest that CNN-GRU is similar in performance to CNN-LSTM but offers a marginal improvement in terms of overall accuracy and F1 score.

The GRU layers in this model provided similar benefits as the LSTM layers, capturing the sequential dependencies in the SMS messages. However, GRUs tend to be computationally more efficient than LSTMs, which could make CNN-GRU a more attractive option when training time and resource usage are considerations. The model's ability to maintain high precision indicates its effectiveness in correctly identifying spam messages, while its slightly lower recall suggests it may still miss some spam instances. The ResNet model achieved the highest performance among the deep learning models, with an average accuracy of 99.08% and a standard deviation of ±0.0046. The F1 score for ResNet was 0.9646, which is the highest among the evaluated models, indicating that ResNet achieved the best balance between precision and recall. The precision was 0.9854, and the recall was 0.9453, both of which are higher than those of the CNN-based models.

The superior performance of ResNet can be attributed to its deeper architecture, enabled by the residual connections that allow the network to learn more complex patterns without the risk of vanishing gradients. This architecture enables ResNet to capture both local and global features in the SMS messages more effectively than the CNN-LSTM and CNN-GRU models. The high precision and recall suggest that ResNet is both cautious in predicting spam and capable of identifying a greater proportion of actual spam messages, making it a highly reliable model for SMS spam detection.

The standalone GRU model achieved a test accuracy of 98.39%, which is slightly lower than the CNN-based models. The F1 score for GRU was 0.9375, indicating that its overall balance between precision and recall was somewhat lower than that of CNN-GRU and ResNet. The precision for GRU was 0.9783, while the recall was 0.9000, suggesting that the model was highly effective in minimizing false positives but missed a considerable portion of actual spam messages. Although GRU is known for being computationally efficient, its slightly lower performance compared to CNN-GRU highlights the importance of combining convolutional layers with recurrent layers to capture both local and sequential information in text data. The GRU model's high precision makes it useful in scenarios where false positives must be minimized, but the lower recall indicates that it may be less effective in identifying all spam messages.

The BiLSTM model produced the lowest performance among the evaluated models, with a test accuracy of 98.12%. The F1 score for BiLSTM was 0.9268, which is significantly lower than the F1 scores of the other models. The precision for BiLSTM was 0.9708, indicating that it was still effective in minimizing false positives, but the recall was only 0.8867, suggesting that it failed to identify a substantial portion of actual spam messages. BiLSTM is known for its ability to capture both past and future context in sequences, but the results suggest that this capability may not have provided significant advantages in the SMS spam detection task. The lower recall indicates that the BiLSTM model struggled to correctly classify spam messages, potentially because the bidirectional structure introduced additional complexity that was not necessary for this task.

* Corresponding author

DISCUSSIONS

Among the models evaluated, ResNet stood out as the best performer in terms of both accuracy and F1 score, demonstrating its ability to handle complex patterns in SMS data. The CNN-GRU model also performed well, with only marginally lower accuracy and F1 score than ResNet, while offering improved computational efficiency compared to CNN-LSTM. CNN-LSTM, while highly accurate and precise, had slightly lower recall, indicating that it may be more cautious in predicting spam messages, which could lead to missed spam detections. The standalone GRU model, although efficient, underperformed compared to the CNN-based models, particularly in terms of recall, indicating that it struggled to identify all spam messages. The BiLSTM model, while still performing reasonably well, was the weakest in terms of recall and overall accuracy, suggesting that its bidirectional nature did not provide a significant benefit for this specific task. Overall, the results indicate that hybrid models that combine convolutional layers with recurrent layers, particularly CNN-GRU and CNN-LSTM, offer a robust solution for SMS spam detection. ResNet, with its deeper architecture, provides the best performance, making it a highly effective model for this task. However, when computational efficiency is a key concern, CNN-GRU emerges as a strong contender, offering a good balance between performance and resource usage. In summary, ResNet's superior performance in both precision and recall makes it the most reliable model for SMS spam detection, particularly in applications where high accuracy is critical. CNN-GRU offers a competitive alternative with improved efficiency, while CNN-LSTM provides a slightly more cautious approach with strong precision but lower recall. GRU and BiLSTM, though useful in certain contexts, are less effective compared to the hybrid and residual models.

Table 1 The Comparison of Model

Model	Average Accuracy	F1 Score	Precision	Recall
CNN-LSTM	0.9892	0.9579	0.9773	0.9403
CNN-GRU	0.9897	0.9596	0.9807	0.9399
ResNet	0.9908	0.9646	0.9854	0.9453
GRU	0.9839	0.9375	0.9783	0.9
BiLSTM	0.9812	0.9268	0.9708	0.8867

CONCLUSION

In this study, we evaluated the performance of several deep learning models, including CNN-LSTM, CNN-GRU, ResNet, GRU, and BiLSTM, for the task of SMS spam detection. The use of ten-fold cross-validation provided a robust assessment of the models' generalization capabilities, while accuracy, precision, recall, and F1 score were used to measure their effectiveness. Among the models tested, ResNet emerged as the best performer, achieving the highest accuracy (99.08%) and F1 score (0.9646). Its ability to capture complex patterns through deeper architectures, enabled by residual connections, made it particularly effective in detecting spam messages while maintaining both high precision and recall. This suggests that ResNet is the most reliable model for SMS spam detection, especially in environments where minimizing false negatives and false positives is crucial. The CNN-GRU model also performed well, with only marginally lower accuracy and F1 score than ResNet, while offering the advantage of reduced computational complexity. This makes CNN-GRU an attractive option for applications where computational efficiency is a concern, particularly in real-time or mobile-based systems. CNN-LSTM, while highly accurate, demonstrated a slight trade-off in recall, making it more cautious in detecting spam but still highly reliable.

The standalone GRU and BiLSTM models, although effective to some degree, underperformed compared to the CNN-based models. GRU's lower recall indicates that it may miss a significant number of spam messages, while BiLSTM's complexity did not translate into superior performance, leading to the lowest overall results among the models tested. In conclusion, hybrid models such as CNN-GRU and CNN-LSTM provide strong performance for SMS spam detection, with ResNet offering the highest accuracy and reliability. These models represent a significant improvement over traditional machine learning approaches, effectively handling the complexities of SMS data while balancing computational demands. Future work could focus on optimizing these models further for deployment in resource-constrained environments, such as mobile devices, and exploring additional architectures that can improve efficiency without sacrificing performance.

REFERENCES

Abbas, A. M. (2021). Social network analysis using deep learning: applications and schemes. Social Network Analysis and Mining, 11(1), 106.

Submitted : Oct 10, 2024
Accepted : Oct 22, 2024
Published : Oct 31, 2024

Agarwal, R., Dhoot, A., Kant, S., Bisht, V. S., Malik, H., Ansari, M. F., Afthanorhan, A. & Hossaini, M. A. (2024). A novel approach for spam detection using natural language processing with AMALS models. IEEE Access.

Ahmadzadeh, E., Kim, H., Jeong, O., Kim, N. & Moon, I. (2022). A deep bidirectional LSTM-GRU network model for automated ciphertext classification. IEEE Access, 10, 3228–3237.

Ahmed, S. F., Alam, M. S. Bin, Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A. B. M. & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artificial Intelligence Review, 56(11), 13521–13617.

Akhter, M. P., Jiangbin, Z., Naqvi, I. R., Abdelmajeed, M., Mehmood, A. & Sadiq, M. T. (2020). Document-level text classification using single-layer multisize filters convolutional neural network. IEEE Access, 8, 42689–42707.

Alkhalil, Z., Hewage, C., Nawaf, L. & Khan, I. (2021). Phishing attacks: A recent comprehensive study and a new anatomy. Frontiers in Computer Science, 3, 563060.

Ansari, L., Ji, S., Chen, Q. & Cambria, E. (2022). Ensemble hybrid learning methods for automated depression detection. IEEE Transactions on Computational Social Systems, 10(1), 211–219.

Bukhari, S. M. S., Zafar, M. H., Abou Houran, M., Moosavi, S. K. R., Mansoor, M., Muaaz, M. & Sanfilippo, F. (2024). Secure and privacy-preserving intrusion detection in wireless sensor networks: Federated learning with SCNN-Bi-LSTM for enhanced reliability. Ad Hoc Networks, 155, 103407.

Cai, H., Lin, J., Lin, Y., Liu, Z., Tang, H., Wang, H., Zhu, L. & Han, S. (2022). Enable deep learning on mobile devices: Methods, systems, and applications. ACM Transactions on Design Automation of Electronic Systems (TODAES), 27(3), 1–50.

Cao, Y., Geddes, T. A., Yang, J. Y. H. & Yang, P. (2020). Ensemble deep learning in bioinformatics. Nature Machine Intelligence, 2(9), 500–508.

Dapat, V. (2024). SMS Spam Detection Dataset. https://www.kaggle.com/datasets/vishakhdapat/sms-spam-detection-dataset/data

Daraghmi, E. Y., Qadan, S., Daraghmi, Y., Yussuf, R., Cheikhrouhou, O. & Baz, M. (2024). From Text to Insight: An Integrated CNN-BiLSTM-GRU Model for Arabic Cyberbullying Detection. IEEE Access.

Das, S., Mandal, S. & Basak, R. (2023). Spam email detection using a novel multilayer classification-based decision technique. International Journal of Computers and Applications, 45(9), 587–599.

Dey, A., Nayak, S., Kumar, R. & Mohanty, S. N. (2024). How Machine Learning is Innovating Today's World: A Concise Technical Guide. John Wiley \& Sons.

Do, N. Q., Selamat, A., Krejcar, O., Herrera-Viedma, E. & Fujita, H. (2022). Deep learning for phishing detection: Taxonomy, current challenges and future directions. Ieee Access, 10, 36429–36463.

Dua, N., Singh, S. N. & Semwal, V. B. (2021). Multi-input CNN-GRU based human activity recognition using wearable sensors. Computing, 103(7), 1461–1478.

Durga, B. K. & Rajesh, V. (2022). A ResNet deep learning based facial recognition design for future multimedia applications. Computers and Electrical Engineering, 104, 108384.

Gasparetto, A., Marcuzzo, M., Zangari, A. & Albarelli, A. (2022). A survey on text classification algorithms: From text to predictions. Information, 13(2), 83.

Gaurav, D., Tiwari, S. M., Goyal, A., Gandhi, N. & Abraham, A. (2020). Machine intelligence-based algorithms for spam filtering on document labeling. Soft Computing, 24(13), 9625–9638.

Guo, Z., Yang, C., Wang, D. & Liu, H. (2023). A novel deep learning model integrating CNN and GRU to predict particulate matter concentrations. Process Safety and Environmental Protection, 173, 604–613.

Hua, H., Liu, M., Li, Y., Deng, S. & Wang, Q. (2023). An ensemble framework for short-term load forecasting based on parallel CNN and GRU with improved ResNet. Electric Power Systems Research, 216, 109057.

Islam, M. S., Kabir, M. N., Ghani, N. A., Zamli, K. Z., Zulkifli, N. S. A., Rahman, M. M. & Moni, M. A. (2024). Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach. Artificial Intelligence Review, 57(3), 62.

Jafari, S. & Byun, Y.-C. (2023). A CNN-GRU Approach to the Accurate Prediction of Batteries' Remaining Useful Life from Charging Profiles. Computers, 12(11), 219.

Jain, A. (2021). SPAM filtering using artificial intelligence. Artificial Intelligence and Data Mining Approaches in Security Frameworks, 261–291.

Karasoy, O. & Ball\i, S. (2022). Spam SMS detection for Turkish language with deep text analysis and deep learning methods. Arabian Journal for Science and Engineering, 47(8), 9361–9377.

* Corresponding author

Submitted : Oct 10, 2024
Accepted : Oct 22, 2024
Published : Oct 31, 2024

Kernbach, J. M. & Staartjes, V. E. (2022). Foundations of machine learning-based clinical prediction modeling: Part II—Generalization and overfitting. Machine Learning in Clinical Neuroscience: Foundations and Applications, 15–21.

Kigerl, A. (2020). Spam-based scams. The Palgrave Handbook of International Cybercrime and Cyberdeviance, 877–897.

Klemm, C. & Vennemann, P. (2021). Modeling and optimization of multi-energy systems in mixed-use districts: A review of existing methods and approaches. Renewable and Sustainable Energy Reviews, 135, 110206.

Kumari, P. & Toshniwal, D. (2021). Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. Journal of Cleaner Production, 279, 123285.

Li, H., Wang, S. X., Shang, F., Niu, K. & Song, R. (2024). Applications of large language models in cloud computing: An empirical study using real-world data. International Journal of Innovative Research in Computer Science \& Technology, 12(4), 59–69.

Lu, L., Zhang, C., Cao, K., Deng, T. & Yang, Q. (2022). A multichannel CNN-GRU model for human activity recognition. IEEE Access, 10, 66797–66810.

Mehnatkesh, H., Jalali, S. M. J., Khosravi, A. & Nahavandi, S. (2023). An intelligent driven deep residual learning framework for brain tumor classification using MRI images. Expert Systems with Applications, 213, 119087.

Mienye, I. D., Swart, T. G. & Obaido, G. (2024). Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications. Information, 15(9), 517.

Minu, M. S. & Canessane, R. A. (2022). Deep learning-based aerial image classification model using inception with residual network and multilayer perceptron. Microprocessors and Microsystems, 95, 104652.

Nandwani, P. & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. Social Network Analysis and Mining, 11(1), 81.

Naseem, U., Razzak, I., Khan, S. K. & Prasad, M. (2021). A comprehensive survey on word representation models: From classical to state-of-the-art word representation language models. Transactions on Asian and Low-Resource Language Information Processing, 20(5), 1–35.

Oruh, J., Viriri, S. & Adegun, A. (2022). Long short-term memory recurrent neural network for automatic speech recognition. IEEE Access, 10, 30069–30079.

Rao, S., Verma, A. K. & Bhatia, T. (2021). A review on social spam detection: Challenges, open issues, and future directions. Expert Systems with Applications, 186, 115742.

Rayan, A. (2022). Analysis of e-Mail Spam Detection Using a Novel Machine Learning-Based Hybrid Bagging Technique. Computational Intelligence and Neuroscience, 2022(1), 2500772.

Remya, S., Pillai, M. J., Nair, K. K., Subbareddy, S. R. & Cho, Y. Y. (2024). An Effective Detection Approach for Phishing URL Using ResMLP. IEEE Access.

Sabeeh, V., Zohdy, M., Mollah, A. & Al Bashaireh, R. (2020). Fake news detection on social media using deep learning and semantic knowledge sources. International Journal of Computer Science and Information Security (IJCSIS), 18(2), 45–68.

Sabir, B., Ullah, F., Babar, M. A. & Gaire, R. (2021). Machine learning for detecting data exfiltration: A review. ACM Computing Surveys (CSUR), 54(3), 1–47.

Saidani, N. (2021). A learning approach for spam detection using semantic representation. Université du Québec en Outaouais.

Salman, M., Ikram, M. & Kaafar, M. A. (2024). Investigating Evasive Techniques in SMS Spam Filtering: A Comparative Analysis of Machine Learning Models. IEEE Access.

Sharmin, T., Di Troia, F., Potika, K. & Stamp, M. (2020). Convolutional neural networks for image spam detection. Information Security Journal: A Global Perspective, 29(3), 103–117.

Thakur, P., Joshi, K., Jain, S. & Thakral, P. (2023). Spam Detection in Emails using Machine Learning.

Tubishat, M., Al-Obeidat, F., Sadiq, A. S. & Mirjalili, S. (2023). An Improved Dandelion Optimizer Algorithm for Spam Detection: Next-Generation Email Filtering System. Computers, 12(10), 196.

Vankdothu, R. & Hameed, M. A. (2022). Brain tumor MRI images identification and classification based on the recurrent convolutional neural network. Measurement: Sensors, 24, 100412.

Vijayakumar, B. & Thomas, C. (2024). The ethics of envisioning spam free email inboxes. AI and Ethics, 1–24.

Yin, X., Liu, Q., Pan, Y., Huang, X., Wu, J. & Wang, X. (2021). Strength of stacking technique of ensemble learning in rockburst prediction with imbalanced data: Comparison of eight single and ensemble models. Natural Resources Research, 30, 1795–1815.