Hybrid NLP Framework for Real-Time Multilingual SMS Spam and Malicious URL Intelligence

JAMMULA RISHITHA[1], Dr.P.JOHN PAUL[2]
[1]Student, Department of Computer Science and Engineering, Ellenki College of Engineering and Technology, Patelguda, Patancharu, Hyderabad, Telangana.
[2]Principal & Professor, Department of CSE, Ellenki College of Engineering and Technology, Patelguda, Patancharu, Hyderabad, Telangana.

To Cite this Article
Jammula Rishitha, Dr. P. John Paul, "Hybrid NLP Framework for Real-Time Multilingual SMS Spam and Malicious URL Intelligence", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 07, July 2026, pp: 917-923, DOI: http://doi.org/10.64771/jsetms.2026.v03.i07.pp917-923
Submitted: 19-06-2026 *Accepted: 24-07-2026* *Published: 30-07-2026*

***Abstract*— The increasing use of mobile messaging applications has resulted in a rapid rise in spam messages and malicious URLs, posing significant threats to user privacy and online security. This paper presents a web-based system that performs multilingual SMS spam detection along with malicious URL classification using machine learning and natural language processing techniques. The proposed approach utilizes the Multilingual Bidirectional Encoder Representations from Transformers (MBERT) model to classify both English and Hindi SMS messages as either spam or legitimate after performing text preprocessing and data cleaning. For URL analysis, a Random Forest classifier is employed to identify whether a submitted web link is safe or malicious. The application consists of secure user authentication, dataset processing, model training, real-time SMS prediction, URL verification, and performance evaluation modules. Experimental results indicate that the MBERT model achieves an accuracy for SMS spam detection, while the Random Forest classifier attains accuracy in malicious URL classification. Performance is evaluated using precision, recall, F1-score, confusion matrices, and graphical analysis. The developed system provides an effective, practical, and scalable solution for protecting users from spam messages and harmful web links in multilingual communication environments.**

***Keywords*— Multilingual SMS spam detection, MBERT, malicious URL classification, Random Forest, natural language processing, text classification, phishing detection, cybersecurity, web-based application, machine learning.**

I. INTRODUCTION

The rapid expansion of digital communication has made **Short Message Service (SMS)** an essential medium for personal, business, and commercial interactions. Along with its widespread usage, unwanted spam messages and malicious web links have become increasingly common. Such messages are frequently used to advertise fake products, spread phishing attacks, or redirect users to harmful websites that compromise personal information. As cyber threats continue to evolve, the need for reliable and automated detection systems has become more important than ever **[3], [9], [15]**.
Conventional spam filtering techniques mainly depend on predefined rules or keyword matching, which often produce inaccurate results when dealing with multilingual text or newly emerging spam patterns. Modern machine learning and natural language processing techniques have significantly improved the ability to understand contextual information within text, making them suitable for identifying complex spam messages. Likewise, machine learning algorithms can effectively analyze URL characteristics to distinguish between legitimate and malicious web addresses **[1], [2], [10], [12], [13], [17]**.

This work presents an integrated system for detecting spam SMS messages and classifying suspicious URLs through a web-based interface. The proposed framework supports both English and Hindi messages, enabling multilingual text analysis using the **Multilingual Bidirectional Encoder Representations from Transformers (MBERT)** model. Before classification, the input data undergoes preprocessing to remove unnecessary symbols and improve prediction quality. In addition, a **Random Forest** classifier is employed to examine URLs and determine whether they are safe or malicious. The application includes user authentication, dataset processing,

model training, prediction modules, and graphical performance evaluation. Experimental observations demonstrate that the proposed framework provides accurate and efficient results, making it a practical solution for improving communication security and reducing the risks associated with spam messages and malicious websites **[2]**.

II. RELATED WORK

N. Mageshkumar, A. Vijayaraj, N. Arunpriya, and A. Sangeetha (2022) presented "Efficient Spam Filtering Through Intelligent Text Modification Detection Using Machine Learning" [1]. The authors investigated the challenges posed by modern spam messages that intentionally modify words and characters to bypass conventional filtering techniques. Their work focused on developing a machine learning framework capable of recognizing manipulated text patterns instead of relying solely on predefined keywords. The proposed approach included text preprocessing, feature extraction, and supervised learning to improve the identification of unwanted messages. The experimental analysis demonstrated that machine learning models can successfully detect spam even when messages contain altered spellings or hidden patterns. The study emphasized that intelligent feature representation plays an important role in reducing false classifications while improving overall detection accuracy. This research highlights the importance of adaptive learning techniques in developing robust spam filtering systems that remain effective against continuously evolving messaging attacks and deceptive communication strategies. **[1]**

N. N. A. Sjarif, N. F. M. Azmi, S. Chuprat, H. M. Sarkan, Y. Yahya, and S. M. Sam (2019) proposed "SMS Spam Message Detection Using Term Frequency–Inverse Document Frequency and Random Forest Algorithm" [2]. The study explored the application of text mining techniques for distinguishing spam messages from legitimate SMS communications. Before classification, the message contents were transformed into numerical representations using the TF-IDF feature extraction method, allowing important words to receive greater significance during model training. A Random Forest classifier was then applied to categorize messages based on these extracted features. The authors reported that combining effective preprocessing with ensemble learning produced reliable classification results while maintaining computational efficiency. Their findings demonstrated that feature weighting methods significantly influence prediction performance and help improve the identification of unwanted messages. This work provides valuable insight into the use of statistical text representation together with machine learning algorithms for practical SMS spam filtering applications. **[2]**

P. K. Roy, J. P. Singh, and S. Banerjee (2020) introduced "Deep Learning to Filter SMS Spam" [12]. The researchers examined the advantages of deep learning methods over traditional machine learning algorithms for spam message classification. Their framework automatically learned meaningful text representations from large collections of SMS data without depending heavily on manually engineered features. Different neural network architectures were evaluated to capture contextual information and improve prediction capability. The experimental results indicated that deep learning models achieved higher accuracy and better generalization when compared with several conventional classifiers. The authors also discussed the importance of handling diverse writing styles and informal language commonly found in mobile messaging. Their work demonstrated that automatic feature learning enables the detection system to adapt to evolving spam patterns more effectively, making deep learning a promising approach for future intelligent messaging security applications. **[12]**

A. Ghourabi and M. Alohaly (2023) proposed "Enhancing Spam Message Classification and Detection Using Transformer-Based Embedding and Ensemble Learning" [13]. The study combined transformer-based language representations with ensemble machine learning techniques to improve spam message classification performance. Transformer embeddings were employed to capture semantic relationships and contextual meaning within text, while ensemble models increased prediction stability by integrating multiple learning strategies. The proposed framework was evaluated on benchmark datasets containing legitimate and spam messages collected from different sources. Experimental findings showed noticeable improvements in classification accuracy and robustness when compared with conventional text representation approaches. The authors concluded that contextual language models are highly effective for multilingual and complex spam detection because they preserve linguistic information more efficiently than traditional feature extraction methods. This research demonstrates the growing significance of transformer architectures in modern text classification and cybersecurity applications. **[13]**

S. Ballı and O. Karasoy (2019) developed "Development of Content-Based SMS Classification Application Using Word2Vec-Based Feature Extraction" [17]. Their research focused on improving SMS classification by employing Word2Vec to generate distributed word representations that capture semantic similarity among words.

Instead of depending on frequency-based features alone, the proposed approach transformed textual information into meaningful vector representations before classification. This allowed the system to recognize relationships between words even when different vocabulary was used to express similar meanings. Various machine learning algorithms were evaluated to determine the effectiveness of the generated features. The results confirmed that semantic word embeddings contributed to higher classification performance compared with conventional text representation methods. The study highlighted the importance of contextual feature learning for spam detection and demonstrated that distributed word representations provide a strong foundation for intelligent SMS filtering systems capable of handling diverse communication patterns. **[17]**

III. DATASET DETAILS

The proposed system utilizes two publicly available datasets to perform multilingual SMS spam detection and malicious URL classification. The first dataset consists of SMS messages written in both **English and Hindi**, where each message is assigned a class label indicating whether it is **Spam** or **Ham (legitimate)**. The collected messages include promotional advertisements, phishing attempts, financial fraud notifications, lottery messages, and normal day-to-day conversations. This combination of multilingual text enables the system to learn language-specific characteristics while maintaining effective performance across different communication styles. Before training, the SMS data undergoes preprocessing steps such as removing special characters, punctuation, unnecessary symbols, and numerical values, followed by text normalization to improve data quality. The processed dataset is then divided into **80% training data** and **20% testing data** for model development and performance evaluation.

The second dataset contains a collection of website links labelled as **Normal** or **Malicious**. These URLs represent a variety of online resources, including trusted websites, phishing pages, suspicious domains, and potentially harmful links. Each record is associated with a corresponding class label that allows the classification model to distinguish between safe and unsafe URLs. During model training, meaningful URL characteristics are extracted and supplied to the **Random Forest** classifier for learning discriminative patterns. The dataset includes sufficient variations in URL structures, domain names, and link formats, enabling the classifier to recognize malicious behaviour with high reliability. Similar to the SMS dataset, the URL dataset is partitioned into **80% training samples** and **20% testing samples** to ensure unbiased evaluation of the prediction model.
The multilingual SMS dataset is used to train the **Multilingual Bidirectional Encoder Representations from Transformers (MBERT)** model, which captures contextual information from both English and Hindi text without requiring separate language-specific models. Unlike conventional machine learning approaches that depend heavily on manually designed features, MBERT automatically learns semantic relationships and contextual representations from the input messages. The trained model is capable of accurately identifying spam messages even when different writing styles, mixed-language content, or modified words are used. The combination of a multilingual text dataset and a labelled URL dataset provides a comprehensive framework for secure message filtering and malicious website detection, supporting reliable real-time prediction through the developed web-based application.

IV. PROPOSED METHODOLOGY

The proposed system introduces an integrated framework for detecting multilingual SMS spam and identifying malicious URLs using advanced machine learning and natural language processing techniques. The objective of the system is to provide accurate classification of suspicious messages and unsafe web links through a single web-based platform. The framework combines the contextual understanding capability of the **Multilingual Bidirectional Encoder Representations from Transformers (MBERT)** model with the classification strength of the **Random Forest** algorithm to improve prediction accuracy while supporting both English and Hindi text.

The first stage of the methodology involves collecting two labelled datasets: a multilingual SMS dataset containing **Spam** and **Ham** messages, and a URL dataset consisting of **Normal** and **Malicious** web links. Before model training, the SMS messages undergo several preprocessing operations, including the removal of unwanted symbols, punctuation marks, numbers, and unnecessary spaces. The cleaned text is then normalized to produce consistent input suitable for multilingual language processing. Similarly, the URL dataset is prepared by extracting meaningful characteristics that help distinguish safe links from malicious ones.

Following preprocessing, the SMS data is supplied to the **MBERT** model for feature learning and classification. Unlike traditional machine learning approaches that depend on manually designed features, MBERT automatically captures semantic relationships and contextual information from multilingual text. This enables the system to recognize spam messages even when different writing styles, mixed-language content, or modified

words are used to avoid detection. The trained model learns complex linguistic patterns that improve classification performance across multiple languages.

For malicious URL detection, the prepared URL dataset is processed using the **Random Forest** classifier. This ensemble learning algorithm constructs multiple decision trees and combines their predictions to determine whether a submitted URL is legitimate or malicious. The use of Random Forest improves prediction stability, minimizes overfitting, and provides reliable classification for various URL formats and domain structures. The datasets are divided into **80% training data** and **20% testing data** to evaluate the performance of both models under unseen conditions.
After successful training, the developed models are integrated into a secure web application that supports user authentication, dataset processing, model execution, and real-time prediction. Users can submit SMS messages in either English or Hindi for spam analysis or enter a website link for security verification. The system immediately returns the prediction result while presenting performance statistics such as accuracy, precision, recall, F1-score, confusion matrices, and comparison graphs. The proposed methodology therefore offers an efficient, scalable, and practical solution for multilingual spam filtering and malicious URL detection, contributing to safer digital communication and enhanced protection against phishing and cyber threats.

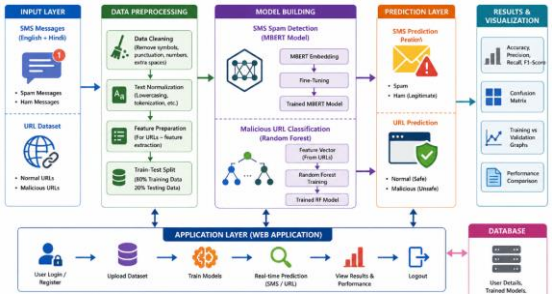

Figure 1: Workflow of the Proposed SMS Spam Detection and Malicious URL Classification System

Figure 1 illustrates the overall workflow of the proposed multilingual **SMS Spam Detection and Malicious URL Classification** system. The process begins with the collection of two input datasets: multilingual SMS messages (English and Hindi) and a labelled URL dataset containing both normal and malicious links. The SMS messages undergo preprocessing, which includes data cleaning, text normalization, feature preparation, and dataset partitioning into training and testing sets. The processed text is then supplied to the **MBERT** model for spam classification, while the extracted URL features are provided to the **Random Forest** classifier for malicious URL detection. After model training, both classifiers are integrated into a web-based application where authenticated users can upload datasets, train the models, and perform real-time SMS and URL prediction. Finally, the application presents the classification results along with performance measures such as accuracy, precision, recall, F1-score, confusion matrix, training-validation graphs, and model comparison charts, enabling effective evaluation of the proposed framework.

V.RESULT AND DISCUSSION

The proposed system was developed to perform multilingual SMS spam detection and malicious URL classification using a combination of **MBERT** and **Random Forest** algorithms. The experimental evaluation was carried out using separate datasets containing English and Hindi SMS messages along with labelled URLs representing both legitimate and malicious websites. Before model training, the datasets were preprocessed to remove unwanted symbols, normalize the text, and prepare meaningful features for classification.
The **MBERT** model was trained to identify spam and legitimate SMS messages by learning contextual information from multilingual text. Similarly, the **Random Forest** classifier was trained to distinguish between safe and malicious URLs using extracted URL features. The trained models were integrated into a web-based application that allows users to perform real-time predictions through a simple interface.

Performance was assessed using standard evaluation measures, including **accuracy, precision, recall, F1-score**, confusion matrices, and comparative performance graphs. The experimental results demonstrate that the **MBERT** model achieved an overall **accuracy of 99.55%** in multilingual SMS spam detection, while the **Random Forest** classifier obtained an **accuracy of 99.46%** for malicious URL classification. The confusion matrix and graphical performance analysis further indicate that both models correctly classified the majority of the testing samples with very few misclassifications. These results confirm that the proposed framework provides reliable and efficient detection of spam messages and malicious web links, making it suitable for practical deployment in applications that require secure digital communication and protection against phishing and online threats.

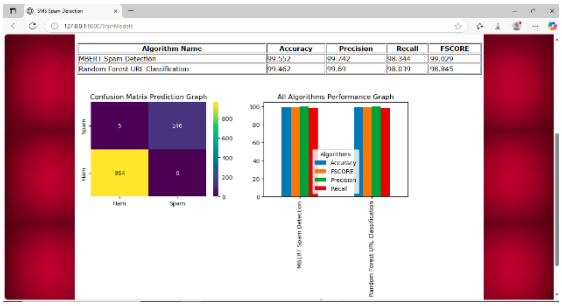

Figure 2: Performance Evaluation of the Proposed SMS Spam Detection and Malicious URL Classification System

Figure 2 presents the performance analysis of the proposed multilingual SMS spam detection and malicious URL classification framework. The evaluation compares the performance of the **MBERT** model for SMS spam detection and the **Random Forest** classifier for malicious URL classification using standard performance metrics. The results show that the **MBERT** model achieved an overall **accuracy of 99.55%**, with a **precision of 99.74%**, **recall of 98.34%**, and an **F1-score of 99.02%** for multilingual SMS classification. Similarly, the **Random Forest** classifier obtained an **accuracy of 99.46%**, **precision of 99.69%**, **recall of 98.03%**, and an **F1-score of 98.85%** for URL classification. The confusion matrix indicates that the majority of SMS messages were correctly classified into **Spam** and **Ham** categories, with only a few misclassified instances. The comparative performance graph further demonstrates that both models consistently achieved high values across all evaluation metrics, confirming the effectiveness and reliability of the proposed framework for detecting spam messages and identifying malicious web links in real-time applications.

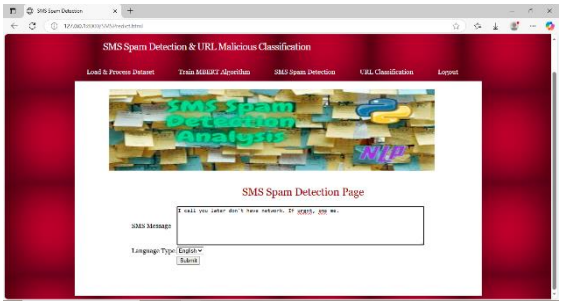

Figure 3: SMS Spam Detection Interface

Figure 3 shows the SMS spam detection module of the proposed web-based application. After successful login and model training, users can access this interface to analyse SMS messages in real time. The user enters the SMS text into the input field and selects the appropriate language, either **English** or **Hindi**, before submitting the message for classification. The application forwards the preprocessed text to the trained **MBERT** model, which analyses the contextual information and predicts whether the message is **Spam** or **Ham (Legitimate)**. The interface provides a simple and user-friendly environment for multilingual spam detection without requiring technical expertise. This module enables fast and accurate identification of suspicious messages, helping users recognize unwanted promotional content, phishing attempts, and fraudulent communications while ensuring secure digital messaging.

DISCUSSION

The proposed framework combines the **MBERT** model for multilingual SMS spam detection with the **Random Forest** algorithm for malicious URL classification to improve the security of digital communication. The system was evaluated using labelled datasets containing English and Hindi SMS messages along with normal and malicious URLs. Prior to training, the data underwent preprocessing, including text cleaning, normalization, feature preparation, and dataset partitioning into **80% training** and **20% testing** samples to ensure reliable model evaluation.

The experimental results demonstrate that both models achieved high classification performance across all evaluation metrics. The **MBERT** model obtained an overall **accuracy of 99.55%** for SMS spam detection, while

the **Random Forest** classifier achieved **99.46% accuracy** in identifying malicious URLs. The corresponding precision, recall, and F1-score values also remained consistently high, indicating that only a few samples were incorrectly classified. The confusion matrix further confirmed the robustness of the proposed approach.
The results indicate that contextual language modelling effectively captures multilingual text patterns, whereas Random Forest efficiently identifies suspicious URL characteristics. The integration of these models into a web-based application provides a practical, accurate, and scalable solution for protecting users against spam messages, phishing attempts, and malicious websites in real-world communication environments.

VI.CONCLUSION

This work presents an effective framework for **multilingual SMS spam detection** and **malicious URL classification** by integrating the **MBERT** model with the **Random Forest** algorithm. The proposed system was developed to improve digital communication security by accurately identifying spam messages written in English and Hindi while also detecting unsafe web links. Through data preprocessing, model training, and performance evaluation, the framework demonstrated reliable classification capability using standard metrics such as accuracy, precision, recall, and F1-score. The experimental results showed that the MBERT model achieved **99.55% accuracy** for SMS spam detection, whereas the Random Forest classifier obtained **99.46% accuracy** for malicious URL identification. These findings confirm that the proposed approach can effectively reduce the risks associated with phishing attacks, fraudulent messages, and harmful websites while providing real-time prediction through a user-friendly web application.

Future work will focus on expanding the multilingual capability by incorporating additional regional and international languages to improve the system's applicability across diverse user communities. The framework can also be enhanced by integrating advanced transformer architectures and deep learning techniques to further improve detection accuracy against newly emerging spam patterns and sophisticated cyber threats. In addition, deploying the proposed model on cloud and mobile platforms, together with continuous model updating using real-world datasets, will improve scalability, adaptability, and practical usability in large-scale communication environments.

REFERENCES

[1] N. Mageshkumar, A. Vijayaraj, N. Arunpriya, and A. Sangeetha, "Efficient spam filtering through intelligent text modification detection using machine learning," *Materials Today: Proceedings*, vol. 64, pp. 848–858, 2022. doi: 10.1016/j.matpr.2022.05.364.

[2] 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 Computer Science*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:213854042

[3] A. Marcus, "Effect of SMS advertising on attitudes of Nigeria GSM phone users," vol. 3, pp. 215–225, Apr. 2019.

[4] N. Hussain, H. Mirza, and I. Hussain, "Detecting spam review through spammer's behavior analysis," *Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 8, no. 2, pp. 61–71, Mar. 2019. doi: 10.14201/ADCAIJ2019826171.

[5] D. Saraswathi and D. Sowmya, "SMS spam classification using PSO-C4.5," in *Lecture Notes in Electrical Engineering*, vol. 967, pp. 41–47, 2023. doi: 10.1007/978-981-19-7169-3_4.

[6] K. Ranjith Reddy and S. Chaudhary, "An efficient text mining technique and its application to SMS spam detection," in *Data Engineering and Intelligent Computing*. Singapore: Springer Nature, 2022, pp. 201–213.

[7] A. Maruf, A. Numan, M. Haque, T. Jidney, and Z. Aung, "Ensemble approach to classify spam SMS from Bengali text," 2023. doi: 10.1007/978-3-031-37940-6_36.

[8] E. I. Hanif *et al.*, "Malay SMS spam detection tool using keyword filtering technique," *Journal of Physics: Conference Series*, vol. 1793, no. 1, 2021. doi: 10.1088/1742-6596/1793/1/012064.

[9] H. Baqeel and R. Zagrouba, "Hybrid SMS spam filtering system using machine learning techniques," in *International Conference on Advanced Communication and Information Technology (ACIT)*, 2020. doi: 10.1109/ACIT50332.2020.9300071.

[10] S. Gupta, S. Saha, and S. K. Das, "SMS spam detection using machine learning," *Journal of Physics: Conference Series*, vol. 1797, no. 1, Feb. 2021. doi: 10.1088/1742-6596/1797/1/012017.

[11] O. Abayomi-Alli, S. Misra, and A. Abayomi-Alli, "A deep learning method for automatic SMS spam classification: Performance of learning algorithms on the indigenous dataset," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 17, 2022. doi: 10.1002/cpe.6989.

[12] 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.

[13] A. Ghourabi and M. Alohaly, "Enhancing spam message classification and detection using transformer-based embedding and ensemble learning," *Sensors*, vol. 23, no. 8, 2023. doi: 10.3390/s23083861.

[14] S. Das Gupta, S. Saha, and S. K. Das, "SMS spam detection using machine learning," *Journal of Physics: Conference Series*, vol. 1797, no. 1, 2021. doi: 10.1088/1742-6596/1797/1/012017.

[15] N. Godson, H. C. Inyama, C. O. Ohaneme, and F. E. Ozioko, "The impact of using SMS alert-based e-health in increasing outpatients' access to health care resources in Nigeria," *IOSR Journal of Computer Engineering*, vol. 18, no. 6, pp. 23–29, 2023. doi: 10.9790/0661-1806032329.

[16] M. Rubin Julis and S. Alagesan, "Spam detection in SMS using machine learning through text mining," *International Journal of Scientific and Technology Research*, vol. 9, no. 2, pp. 498–503, 2020.

[17] S. Ballı and O. Karasoy, "Development of content-based SMS classification application using Word2Vec-based feature extraction," *IET Software*, vol. 13, no. 4, pp. 295–304, 2019.