SMS - URL Sentinel: Dual Detection System Using Machine Learning

Mrs. Vooradi Sandya[1], Chatamoni Saibaba[2], Nerella Harshavardhan[3], Padamati Ritish Reddy[4]
[1,2,3,4]*Dept of Computer Science and Engineering (Data Science), CMR Technical Campus Hyderabad, Telangana*
doi.org/10.64643/IJIRTV12I11-195355-459

***Abstract*—The recent attempts of the mobile communication and online services have exposed online and mobile users to unsolicited SMS spam, malicious URLs, which are both a big threat in terms of privacy, security, and integrity of the device. This project introduces a three-dimensional, dual-module system that will resolve these issues with the use of advanced techniques in machine learning. The former module is dedicated to SMS spammy texting of SMS in English, Hindi and mixed message (English and Hindi) with mBERT, which is transfers-based technology able to consider semantic subtleties and context in short text messaging. The level of deep contextualization allows it to classify with a lot of force even with different linguistic patterns. In addition to this the second module uses a Random Forest based method to detect malicious URLs using ensemble learning of potential suspicious patterns of web URL format and related metadata that can be very dependable. In combination with integrating both detection channels into one the project is able to provide a layered protection option that is capable of both proactively filtering the harmful SMS contents and also averting the users to dangerous web links. This hybrid method has the benefit of supporting better security globally as well as mitigating false positives with the use of cross-module validation. The combined framework shows a good performance in the evaluation measures and hence proposes that it can be deployed to the real-life mobile and web environments as a holistic threat-mitigation system.**

***Index Terms*—SMS Spam Detection, Malicious URL Detection, mBERT, Random Forest, Integrated Detection System.**

I. INTRODUCTION

The use of mobile networks and internet services is something that was adopted widely and it has essentially transformed the way we communicate with each other. The communication process is now quite expedited and convenient to do. Nevertheless, this extension has to a large extent led to the rising count of SMS spam and malicious URL attacks. Not only are these issues annoying to the users, but because of them the user is under risk of data theft, financial fraud, and malware infiltrations. Since attackers are always coming up with new ways to bypass security filters that are of traditional types, there is a demand for the application of intelligent, adaptive, and resilient detection mechanisms in cybersecurity systems to a greater extent.

The study proposes a detection system with multiple layers, which addresses those issues using modern machine learning models. The bottom layer buys the use of mBERT which is a multilingual transformer model that has the capability of getting not just on the surface but also on the deep relations in the text. The model embodies the semantic differences and language dissimilarities; therefore, it may be applied to the correct recognition of spam in spoken languages and the SMS-style writings that humans utilize without a second thought. The second tier uses a classifier based on the Random Forest in order to detect malicious URLs. Random Forest, with its set of decision trees and highly generalizing nature, is an instrument that can serve well in analyzing various features of URLs like structure, character occurrence, domain patterns, and so on. Its construction minimizes the chances of overfitting and at the same time it is becoming more reliable in detecting phishing connections, malicious redirections and malware-related URLs. The proposed system has the ability to protect users in a cohesive and on-the-fly manner against the most frequent cyber threats which arise out of communication. The interaction between mBERT's linguistic intelligence and Random Forest's structural analysis not only strengthens the overall detection

capability but also is one of the factors that lead to the creation of a safer and more reliable digital communication environment.

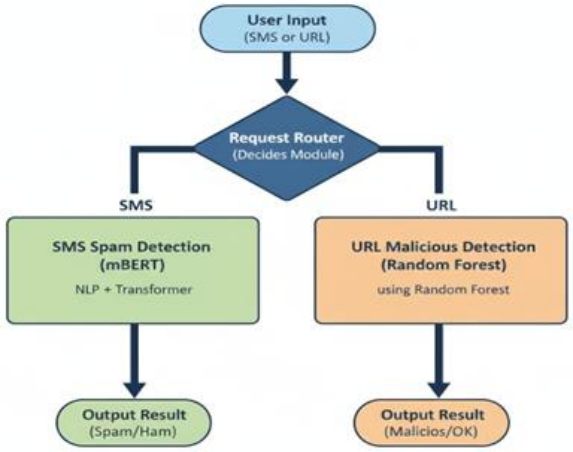

Fig 1. System Architecture

The system proposed merges the two machine learning modules of an mBERT-based SMS spam detector and a Random Forest-based malicious URL classifier into one single, strong security architecture. To begin with, SMS messages go through a multilingual preprocessing pipeline, which involves cleaning, normalizing, and tokenizing the text before it is processed by the fine-tuned mBERT model for spam-ham classification. At the same time, the URLs if any are present in the message, manually put in the URL module, undergo through a feature-engineering module that calculates lexical, structural, and behavioral URL attributes, which are then classified by a Random Forest ensemble to decide whether the link is malicious or not. So, architecture offers double-layer security, lessens the false detections by complementary decision-making, and guarantees adaptive, platform-independent threat mitigation that is suitable for the real world of communication by the combination of deep linguistic understanding with structural URL analysis.

II. LITERATURE REVIEW

The first generation of SMS spam detection systems were mainly based on keyword filters and rule-based methods but these solutions had difficulty fighting the advancing spam patterns. [1],[2],[3] Afterwards, traditional machine learning models like Naïve Bayes, SVM, and Logistic Regression were used and they made the systems smarter by discovering statistical patterns in the text. With the rise in NLP, hybrids and ensembles started to use contextual features to their accuracy in this domain. [7] Currently, different pieces of research justify the use of transformer-based models such as BERT and mBERT in those situations where SMS are multilingual or code-mixed. [8] The works during 2023–2025 indicate that BERT embeddings along with classifiers such as Random Forest, SVM, and Logistic Regression greatly facilitate detection accuracy, thus the case of the English messages is thus solved. On the contrary, these models are, most of the time, computationally expensive for mobile use. In the field of URL security, machine learning models built on lexical features, especially Random Forest, have been very effective in detecting phishing and malicious URLs. The research during 2023–2025 reveals that handcrafting URL features (length, structure, suspicious patterns) can lead to very high accuracies with very low computational costs. [9] However, most of the current URL detection solutions only focus on web phishing and do not interact with SMS-based URL threats, thus the real-time mobile protection is still incomplete. This disparity shows the necessity of integrative framework that will target the spam contents as well as hidden links at the same time.

III. RESEARCH METHODOLOGY

The use of the systematic approach in this study incorporates data preparation, model development and integrated assessment. Multilingual SMS, publicly available labeled URL collections of both benign and malicious samples were gathered. SMS messages were subjected to preprocessing phases such as normalization, eliminating redundant characters, semantic placeholders as well as tokenizing the messages with the mBERT. The tuned mBERT model trained in binary spam-ham classification was fed the process tokens. Simultaneously, the URLs of the dataset were converted into feature vectors defining lexical, structural, and behavioral attributes including length, entropy, digit ratio, type of domain, components that are encoded, and risky types and morphology. They were put under the Random Forest classifier, trained on these features, in which a sequence of decision trees decided whether a URL was benign or malicious. Class- balancing methods and hyperparameter tuning were utilized to train both

models in order to enhance generalization. Their accuracy, precision, recall and F1-score were used as a measure of their performance with a high level of reliability and fewer false positives and negative. Both modules were put together through a single decision pipeline after each was individually validated to conduct simultaneous SMS content analysis and URL threat identification, creating an efficient and real-time security architecture.

IV. RESULTS AND DISCUSSION

1)SMS Spam Detection Result

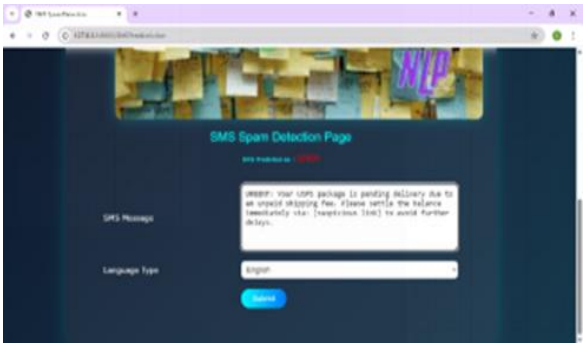

Fig 2: SMS Spam Detection Using mBert

The SMS detection module enables the user to type in a message in either of the English or Hindi language or both languages and see the classification results displayed on the results page. The trained mBERT model then analyzes the text post-processing and the system displays a red flag on whether the SMS is Spam or Ham. The interface contains the clearly shown input message, chosen language, and final prediction that indicate the efficacy of the model when using test samples.

2)URL Classification Results:

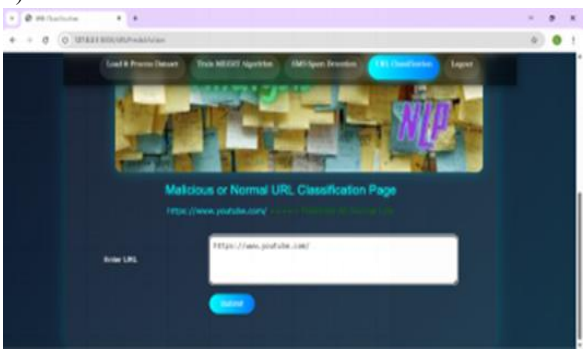

Fig 3: Classifying malicious or normal URL

The URL analysis module has a basic interface in which the user will enter a URL on which an analysis will be conducted. After the processing, the results page is shown with the results of the prediction of the random Forest classifier, with either the normal or malicious link listed.

The results screen emphasizes the address typed in and the decision made by the model, which is suitable in terms of demonstrating the ability of the system to identify phishing and malicious links.

Table 1: Performance Table

Metrics	SMS Spam Detection (MBERT)	URL Classification (Random Forest)
Accuracy	99.5	99.4
Precision	99.7	99.6
Recall	98.3	98.0
F1-Score	99.0	98.8

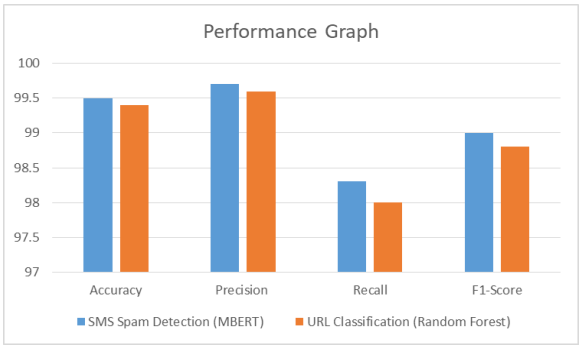

Fig 4. Performance graph

V. KEY FINDINGS

1)The SMS classifier built using mBERT effectively identifies spam in English, Hindi, showing strong semantic understanding even for short, noisy SMS text.

2)Compared to traditional ML models, mBERT handles spelling variations and informal text far better, leading to fewer false positives and false negatives.

3)The URL classifier shows strong performance in identifying phishing and malware URLs by analyzing lexical and structural patterns, outperforming simpler rule-based or single-feature models.

4)Combining SMS and URL detection in one system provides layered protection spam SMS containing

malicious links are flagged more effectively through cross-verification between modules.

VI. CONCLUSION

The learners that are integrated in the form of a mBERT-detected SMS spam with a randomly forest-detected malicious URL to offer a complete-blooded response to communication-driven cyber threats. Relying on the profound linguistic knowledge as well as structural URL analysis, the system does not only detect spam content, but also malicious links that are concealed in messages. Their robustness and reliability have been achieved by the evaluation of the two independent systems using various measures, including accuracy, precision, recall, F1-score. The integrated architecture makes the quantity of false alarms decrease significantly; therefore, the integrated architecture can be deployed in real time and platform independent, hence, making implementation in mobile and web communication environment plausible. Its further concepts may include incorporation of effective transformer models to use faster mobile inference, the adjusting feature update to the ever-shifting URL threats and the reinforcement by user feedback

VII. FUTURE SCOPE

Later enhancements could be achieved by utilizing more sophisticated deep learning techniques such as LSTM, CNN, or Transformer-based models to understand the deeper patterns and to raise the overall detection accuracy of the system. Also, the system can be expanded to support multilingual and regional languages, thus the user community will be more extensive. A very important next step in this work is the creation of simple mobile applications or SDKs that will allow instant spam detection as well as malicious URLs in messaging platforms. Incorporating behavioral and metadata-based information into the model can help reduce the number of false positives further. Lastly, it will be crucial to put in place strategies for the detection of and protection against adversarial manipulations to be able to continue operating efficiently in the presence of the constantly changing spam and attack strategies.

REFERENCES

[1] H. Ghuge, D. Ghuge, A. Rangneniwar, P. Samudra, and A. V. Markad, “Real-time detection of malicious URLs using feature-based machine learning approaches,” *IJIRMPS*, 2025.

[2] A. Khairunnisya *et al.*, “Phishing URL detection system using random forest and gradient boosting,” 2025.

[3] A. P. Nugroho and F. W. Wibowo, “Optimizing SMS spam detection using machine learning algorithms,” *Jurnal Informatika dan Teknologi (JIT)*, 2024. [Online]. Available: https://jit.journal.unesa.ac.id

[4] S. K. Ahmad, B. A. Dapshima, and Y. Essa, “Detection of phishing attacks using machine learning techniques,” *International Research Journal of Modernization in Engineering, Technology and Science (IRJMETS)*, vol. 6, no. 7, Jul. 2024. [Online]. Available: https://www.researchgate.net

[5] R. Kuraku and A. Kalla, “Phishing website URL detection using machine learning algorithms,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 6, 2023. [Online]. Available: https://thesai.org/Publications

[6] A. Sharma and R. Verma, “SMS spam detection and classification using BERT with machine learning classifiers,” *arXiv*, 2024. [Online]. Available: https://arxiv.org/abs/2403.11221

[7] T. Sahmoud and M. Mikki, “Spam detection using BERT,” *arXiv*, 2022. [Online]. Available: https://arxiv.org/abs/2206.02443

[8] J. Cao and C. Lai, “A bilingual multi-type spam detection model based on M-BERT,” in *Proc. IEEE GLOBECOM*, 2020, doi: 10.1109/GLOBECOM42002.2020.934797.

[9] “SMS Spam Collection Dataset,” *UCI Machine Learning Repository*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection

[10] “Malicious URLs dataset,” *Kaggle*. [Online]. Available: https://www.kaggle.com/harrywang/url-website-phishing-detection