ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun010

An Intelligent Web-Based System for Detecting Smishing Message Using Hybrid Machine Learning Technique

Muzammil Sunusi Umar[1*]; Dr. Sandeep Kumar[2]; Muhammad Ibrahim Isah[3]; Bello Bello Musa[4]; Usman Ibrahim Usman[5]; Abdurrazaq Jibril Baba[6]

[1,2,3,4,5,6]Shobhit Institute of Engineering and Technology, Department of Computer Science and Engineering

Corresponding Author: Muzammil Sunusi Umar[1*]

Publication Date: 2026/06/19

Abstract: The increasing dependency on mobile connectivity has turned smishing attacks into a significant cybersecurity threat which leads the stealing of sensitives financial and credential information by using some modern technique and tools likes phishing through content injection and social engineering. This research employs a hybrid machine learning technique to create an intelligent web-based system for classifying smishing, spam, and legitimate SMS messages. To enhance detection capabilities, the study implements a supervised learning approach with various classifiers, including Naïve Bayes, Support Vector Machine, Random Forest, and Logistic Regression, all of which are combined through an hybrid machine learning strategy. The preprocessing and transforming of SMS data are publicly available. Also, the research used TF- IDF and N-gram feature extraction methods. The proposed method was evaluated using multi- class classification metrics and achieved accuracy of 97.58% and F1-score of 97.58%. Experimental results justify that the hybrid ensemble model smashed specific classifiers, getting consistent performance and high accuracy over all three categories. The system best at identifying deceptive SMS messages by its notable recall rate for smishing content. To confirm the whole user experience and allow for automated SMS detection, a classification model was integrated into a web-based platform. The hybrid machine learning technique gives a dependable approach for identifying the SMS threats that confirm by the results. The research also plays a vital role in the field of cybersecurity and forensics by presenting an intelligent smishing and phishing detection.

Keywords: *Cybersecurity, Smishing, Phishing, Hybrid Machine Learning, SMS Security, Text Classification.*

How to Cite: Muzammil Sunusi Umar; Dr. Sandeep Kumar; Muhammad Ibrahim Isah; Bello Bello Musa;Usman Ibrahim Usman; Abdurrazaq Jibril Baba. (2026) An Intelligent Web-Based System for Detecting Smishing Message Using Hybrid Machine Learning Technique. *International Journal of Innovative Science and Research Technology*, 11(6), 592-598. https://doi.org/10.38124/ijisrt/26jun010

I. INTRODUCTION

In today's globe, the uses of emerging technology such as mobile device and computer system become a commonly way of messaging communication, which enable scammers to compromise the integrity and confidentiality of the user's personal data(Xu et al., 2025). The cybersecurity expert and other related security expert contribute to low the risk of using the technology by deploying many modern machine learning approaches such as the proposed intelligent system which enable the detection of smishing messages. Smishing is a type of social engineering attack targeting the mobile short message service (SMS). This techniques involve the tricking of users by sending malicious messages, harmful link, phone numbers in order to reveal financial details such as credit card details, bank account details, and password with intension of stealing(Chichwadia and Mpekoa, 2024). Traditional way of detecting smishing and phishing mostly based on heuristic or signature-based approaches, these techniques active against common attack patterns and it is very difficult for the analysis of sophisticated emerging technology especially one with encrypted massage. This reason change the mindset of researchers switch into machine learning (ML) approaches for better security detection(Altan et al., 2025). In phishing detection system, common techniques include Naïve Bayes, Random Forests, Support Vector Machines(SVM) and Logistic Regression can be used to analyze the textual features extraction by the machine learning to have the access of smishing and legitimate message(Kumar et al., 2025). The existing detection system are mostly used to detect only spam message or smishing with low accuracy in some sophisticated messages, no any existing

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun010

user-friendly interface such as the proposed web-based system that can simply classify the both spam, smishing and ham messages. The primary objective of this research is to design and implement an intelligent web-based system for detecting smishing message using hybrid machine learning techniques. The performance of the proposed model will be examined using evaluation metrics such as, accuracy, precision, recall and F1-score. The main contribution of this research is to implement a multi-class classification model capable of classifying legitimate, spam and smishing with comprehensive experimental evaluation to improve detection performance compared to traditional single-model approaches. This paper was organized as: Section II reviews related work on phishing and smishing detection techniques. Section III presents the proposed intelligent web-based detection system and methodology. Section IV describes the experimental setup, datasets, and evaluation metrics. Section V discusses the experimental results and performance analysis. Finally, Section VI concludes the paper and outlines directions for future research.

II. RELATED WORKS

➢ *Overview of Smishing and Phishing Attacks*

The term phishing alternatively called as "smishing" when occurred via SMS, it also a form of social engineering whereby the criminals use a tactics of impersonating the legitimate entity to scatter and steal financial data by sending SMS message.(Rajput and Mishra, 2025). The security researchers were emerged during the early 2000s to start discovering the vulnerabilities in the SMS service on different device including Android and other platform like iOS. As the smartphone communication develops very fast with vulnerabilities exploitation, initiating smishing set as significant cybersecurity attacks(Goel et al., 2024). These cyber threats have been managed by using both traditional and modern tactics to adapt the situation to low the risk of using the technology. Many researchers were conducted research on the email phishing which leads the awareness of cyber attacks and using modern technology with more percussion but few of researches where focus on SMS phishing which has been on the ground for many eras. There also an existing methods of detecting such attack (smishing) which currently researches are ongoing.

➢ *Traditional Smishing and Phishing Detection Techniques*

Techniques such as keyword matching, blacklist filtering, and signature-based detection were commonly employed as rule-based and heuristic methods in the early days of smishing and phishing detection. These approaches were computationally efficient, they lacked flexibility and struggled to adapt the new or disguised attack patterns (Cagatay Catal et al., n.d.). Jain and Gupta proposed rule-based data mining approach for detecting smishing message by using nine effective rules, the study used text normalization techniques to convert SMS message into standard text before training the rule-based classification, the key finding of the research was the ability of the algorithms to classify smishing message with legitimate, it also limited by it is dependence on predefined rules and it is only focus on SMS-based(Jain and Gupta, 2018). Many researchers were switched to machine learning techniques just to strengthen the security and fill the gaps of the traditional method to align the modern technologies.

➢ *Machine Learning-Based Smishing and Phishing Detection*

The popularity rising of machine learning algorithms reduces from their ability to detect patterns in data and adapt to evolving risks and common classifiers utilized for detecting SMS phishing include Naïve Bayes, Support Vector Machines (SVM), Random Forests, Logistic Regression, and k-Nearest Neighbors (Kumar et al., 2025). In some study has demonstrated that machine learning classifiers, when trained on labeled SMS datasets, significantly outperform traditional methods. Some aspects like feature selection, dataset integrity, and model architecture play crucial roles in determining effectiveness (Aparna et al., 2025). The recent study proposed a machine learning based method that utilizer SVM and Random classifier with limited ability to the normalization process to expand message words based on their contextual relevance from related concepts (Goel et al., 2024).

➢ *Feature Extraction and Text Representation Techniques*

Feature extraction is an important aspect of systems that use machine learning for detecting smishing. When dealing with SMS messages, different text processing techniques such as stemming, normalization, eliminating stop words, and tokenization are commonly employed (Saidat et al., 2024).

➢ *The Common Feature Representation Techniques Include:*

- Term frequency-Inverse Document Frequency (TF-IDF)
- Bag-of-Words (BoW)
- N-Grams

Utilizing Term frequency-inverse document frequency (TF-IDF) combined with supervised machine learning classifiers provides a solid foundational performance for classifying phishing text(Tamal et al., 2024).

➢ *Hybrid and Ensemble Machine Learning Approaches*

To improve the precision and reliability of detections, hybrid machine learning systems use different classifiers and techniques for feature extraction. These approaches aim to reduce the limitations of each algorithm while amplifying their advantages (Elbehiery, 2025). For instance, voting classifiers and ensembles like Random Forest have shown superior performance compared to single classifiers, particularly in terms of accuracy and recall(Xu et al., 2025). The research by Routhu Srinivasa Rao focused on extracting diverse features from URLs to detect phishing attacks effectively based on hybrid machine learning. the ML algorithms achieved an accuracy exceeding 90%, reflecting the effectiveness of selected features(Rao et al., 2025).

➢ *Deep Learning and NLP-Based Detection Techniques*

The use of deep learning and Natural Language Processing (NLP) techniques for detecting smishing attacks by recent research. Various models, such as Long Short-Term Memory networks (LSTMs), Gated Recurrent Units (GRUs),

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun010

Convolutional Neural Networks (CNNs), and those based on transformer architectures, have shown encouraging results(Mahmud et al., 2024).

The article titled SecureNet: A Comparative Study of DeBERTa & LLMs for Phishing Detection (2024) highlights that contextual embeddings significantly improve the performance of classification tasks when evaluating transformer-based models(Mahendru and Pandit, 2024). The implementation of deep learning models in web applications or lighter systems might face challenges due to their substantial need for computational power and the requirement for extensive datasets(Cagatay Catal et al., n.d.).

➢ *Web-Based Intelligent Detection Systems*

The introduction of visual interfaces for engaging with machine learning models significantly boosts the ease of access and user-friendliness of web-based detection systems. Also integrating Machine learning models into web environments increases user involvement and enables immediate detection capabilities(Ishaq et al., 2025). However, the challenges of deployment in much of the existing research tends to overlook and focusing primarily on enhancing model performance instead, That has led to a disconnect between scholarly studies and real-world(Vennela et al., 2026).

Table 1 Summary of Hybrid and Ensemble ML Approaches

Study	Key Findings	Methods
(Jain and Gupta, 2018)	The proposed model achieved a true negative	Rule-based
(Rao et al., 2025)	extracting diverse features from URLs to detect phishing attacks with accuracy exceeding 90%.	Hybrid machine learning
SmishNet (2025)	Superior detection accuracy	CNN + LSTM + attention
Elkholy et al. (2025)	Improved accuracy and reduced false positives	Hybrid ML ensemble
Malicious SMS Detection Using Ensemble Learning (2025)	Enhanced recall on imbalanced datasets	Ensemble + SMOTE
Hybrid BERT + CNN Model (2025)	Effective multilingual smishing detection	Deep hybrid NLP

The Research Gaps Identified based on the reviewed literatures the followings gaps were identified. First, Limited focus on SMS phishing (smishing) only, compared to email phishing. Second, Inadequate integration of hybrid ML model into deployable web-based systems. Third, Overdependent on single-classifier models. And fourth Lack of balance on system high accuracy and practicality

III. METHODOLOGY

This study uses hands-on and guided approach to machine learning, which is well-suited for challenges related to categorizing labeled data. Then, through supervised learning, models can identify patterns from data that has already been labeled and apply this knowledge to new, unlabeled messages. This capability is crucial for the objective of research for distinguishing between harmful and legitimate SMS messages.

➢ *System Architecture*

To certify the user-friendly and functional, the proposed detection system was integrated into a web-based platform. Also, the system was designed with a modular framework, allowing for easy upgrades, straightforward maintenance, and expansion in the future.

➢ *The Following are Five Main Components Contains in the System:*

- SMS user-interface
- Text preprocessing module
- Feature extraction module
- Hybrid classification engine
- Result representation interface

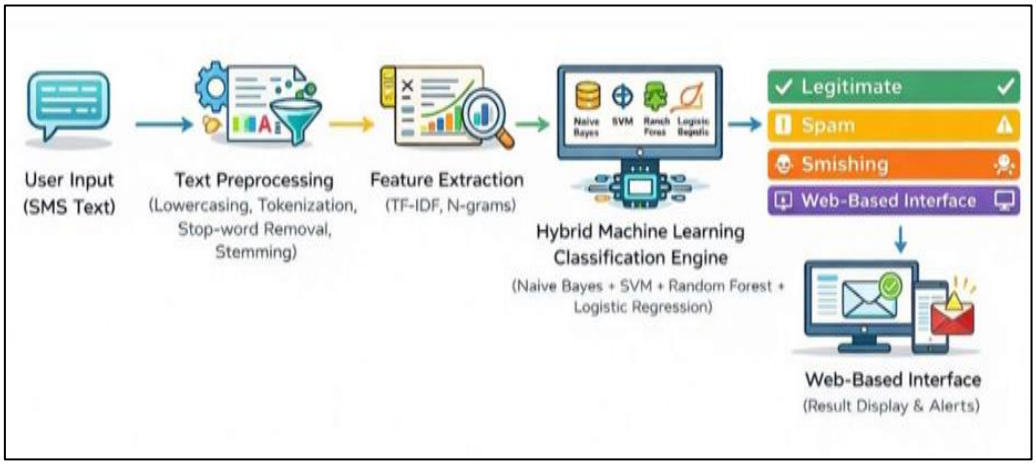

Fig 1 Proposal System Architecture

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun010

➢ *Dataset*

The datasets used in this research are obtained from publicly available SMS phishing and spam datasets which are widely used in prior studies to ensure reproducibility and reliability (Munoz and Islam, 2025), which aligns with modern studies and allows for a fair comparison with past research (Aparna et al., 2025). The datasets consist of 10,191 SMS messages that are categorized as ham or smishing/spam. To maintain consistency during the text preprocessing and feature extraction stages, only messages in English were considered.

➢ *Data Exploration*

The dataset has been explored to examine and prepare collected data before model development, to check and understand the class distribution, message length by class and URL/Email/Phone of the message, as the result illustrate in the following figure:

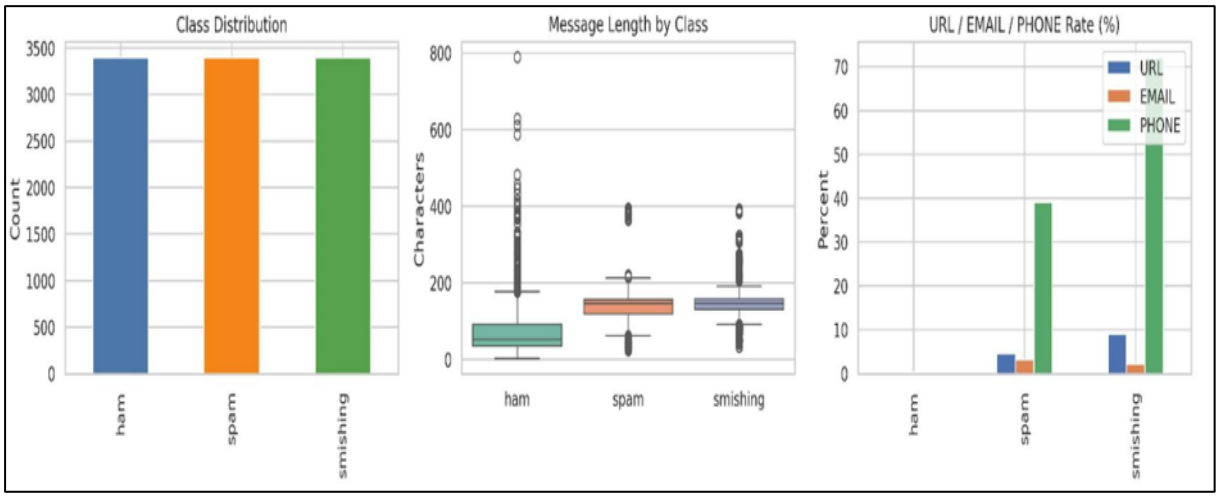

Fig 2 Dataset Class Distribution, Message Length by Class and URL/EMAIL/PHONE Rate

➢ *Data Preprocessing*

The raw messages often contain different types of noise that include Abbreviation, punctuation, URLs, numbers, and inconsistent formatting. If this noise is not handled it, then it can hinder the effectiveness of the machine learning models. To make sure that the SMS text is tidy and standardized before extracting features, a process of data-preprocessing was implemented.

➢ *Below are the Preprocessing Steps:*

- Conversion of all text to lowercase
- Removal of noise; punctuation, numbers and special characters.
- Tokenization of SMS text to individual words.
- Removal of known used stop words
- Application of lemmatization and stemming

All these measures help in lowering the number of dimensions, enhancing the learning effectiveness of classification models and removing unnecessary data (Tamal et al., 2024).

0	ham	Hi! You just spoke to DEEPAK. We'd like to kno...	hi spoke deepak like know satisfi experi repli...
1	ham	Yay can't wait to party together!	yay wait parti togeth
2	ham	At what time are you coming.	time come
3	smishing	Dear customer your PAY2TMKYC has been expired,...	dear custom pay tmkyc expir bl cked within hou...
4	ham	Yo you around? A friend of mine's lookin to pi...	yo around friend mine lookin pick later tonight
...	...	...	...
10185	spam	pdate_now - double mins and 1000 txts on orang...	pdate doubl min txt orang tariff latest motoro...
10186	spam	mobile club: choose any of the top quality ite...	mobil club choos top qualiti item mobil cfca
10187	spam	thanks for your ringtone order, reference numb...	thank rington order refer number mobil charg t...
10188	spam	u can win £100 of music gift vouchers every we...	win music gift voucher everi week initi txt wo...
10189	spam	convey the official england poly ringtone or c...	convey offici england poli rington colour flag...

Fig 3 Dataset After Text Preprocessing and Cleaning

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun010

➢ *Feature Extraction Techniques*

The refined SMS texts were transformed into numerical formats after the initial cleaning process, making them suitable for processing by machine learning models. Features such as N-gram and Term Frequency-Inverse Document Frequency (TF-IDF) has been employed in this research, which are widely recognized for their effectiveness and simplicity in text classification tasks(Elbehiery, 2025). Also, TF-IDF assigns a greater importance to terms that are more relevant within a specific message but appear less frequently across the entire dataset, it enables algorithms to detect patterns that distinguish legitimate messages from smishing attempts.

➢ *Classification Model*

Taking a consideration on their show ability in smishing and phishing detections study, a number of common machine learning classifier were selected to provide high baseline performance. Which include:

- Support Vector Machine (SVM)
- Naïve Bayes
- Random Forest
- Logistic Regression

A recent study also noted that, every classifier offers distinct benefits such as the use of ensemble methods, reasoning grounded in probabilities, classification based on margins, and linear boundaries for decision-making(Kumar et al., 2025).

➢ *Hybrid Classification Technique*

In hybrid classification techniques, no individual model reliably reaches top performance across every dataset, even though certain classifiers may excel in their respective tasks. hybrid machine learning approach was employed to overcome this limitation. This hybrid model utilizes an ensemble voting technique to merge predictions from various classifiers. Also, this strategy enhances robustness and reduces the likelihood of misclassification, by leveraging the unique decision-making patterns of different classifiers (Xu et al., 2025).

➢ *Training Configuration*

The dataset was classified into two parts: one for training and the other for testing, following a division of 70% for training and 30% for testing. The testing portion was kept aside to assess how well the model perform; the training portion was utilized to create the classification models. This process of dividing the data is used in studies that focused on detecting phishing to guarantee a trustworthy evaluation of how well the models can generalize to new situations.

IV. RESULT AND DISCUSSION

➢ *Evaluation Metrics*

To evaluate the performance of the proposed system, the Accuracy, Precision, Recall, and F1-score metrics were used, in consider with the previous research that ensure an equitable evaluation of every category, the overall accuracy, recall, and F1-score were computed for the multi-class classification challenge

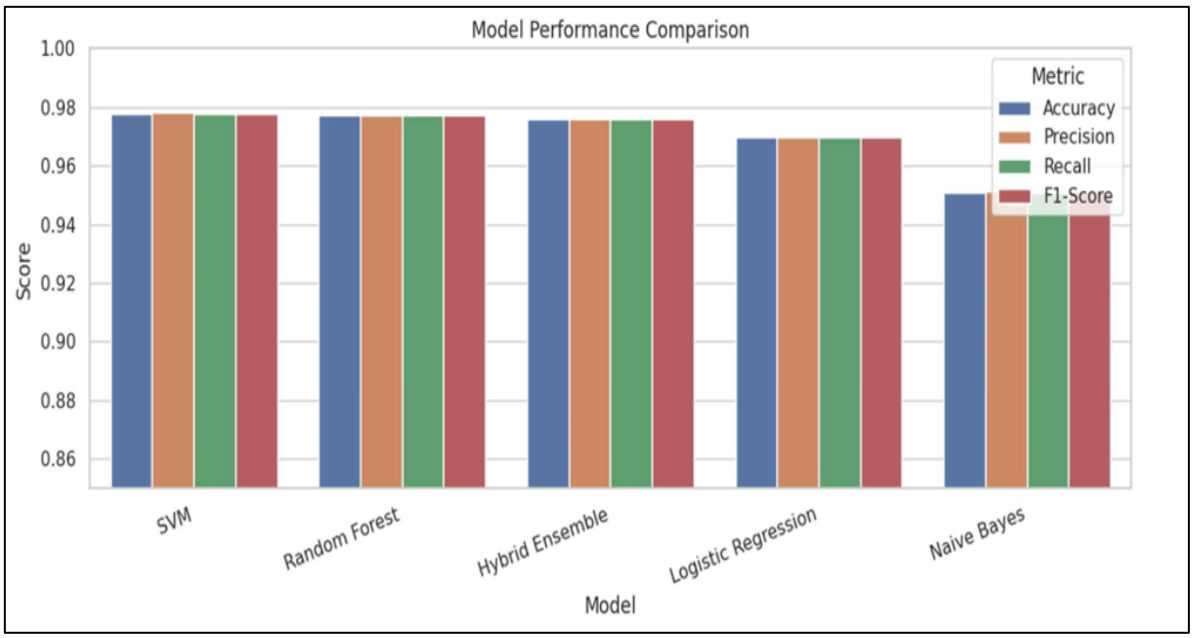

Fig 4 Model Performance Comparison

➢ *Experimental Result*

For the result, forecasts from several classifiers through a collective voting system was integrates by the suggested hybrid framework. Also, this strategy seeks to minimize the limitations of each model and enhance the overall effectiveness of detection.

Table 2 Performance of the Proposed Hybrid Model

Model	Recall (%)	Precision (%)	Accuracy (%)	F1-Score (%)
Hybrid Ensemble Model	97.58	97.58	97.58	97.58

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun010

The hybrid approach appears to indicate the most effective in all evaluation metrics. This improvement may suggest that combining classifiers enhances reliability and reduces errors in categorization particularly in distinguishing smishing from spam texts.

➢ *Performance Comparison*

In this part, we will showcase the outcomes achieved by each machine learning classifier utilized in the research. The results may suggest that models based on probability, such as Naïve Bayes, are surpassed by more advanced classifiers that utilize ensemble methods and margin techniques, including Support Vector Machine (SVM) and Random Forest. To allows the results align with recent studies on identifying SMS phishing threats (Aparna et al., 2025).

Table 3 Performance of Individual Classifiers

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Naïve Bayes	92.1	91.4	90.8	91.1
Logistic Regression	94.3	93.8	93.1	93.4
Support Vector Machine	95.2	94.9	94.3	94.6
Random Forest	96.1	95.7	95.2	95.4

Table 4 Comparison with Existing Studies

Study	Approach	Accuracy
Kumar et al. (2025)	SVM-based model	95.0
Elkholy et al. (2025)	TF-IDF + ML	95.6
Abdullahi et al. (2025)	ML ensemble	96.2
Proposed system	Hybrid ML + Web-based	97.58

➢ *Key Finding*

The findings from this research indicate that hybrid machine learning techniques significantly enhances the detection of smishing, spam, and legitimate SMS messages. By merging multiple classifiers, the system is able to recognize a wide range of textual patterns while also reducing the limitations associated with single models. Also, the proposed system web-based design suggests that, it could potentially be applied in real-world scenarios and integrated with mobile and corporate security frameworks in the future.

V. CONCLUSION

Developed and evaluated an intelligent web-based system for detecting SMS messages as legitimate, spam, and smishing categories using hybrid machine learning techniques has been achieve by the study. The system was designed to provide accurate and practical detection of SMS-based threats, with particular emphasis on smishing attacks. To achieve higher overall accuracy and more balanced performance across all three classes, the hybrid ensemble model consistently outperformed individual machine learning classifiers, hybrid model is likely associated with strong recall for smishing messages and indicating its effectiveness in identifying malicious SMS content. For The results, it improves that combining multiple classifiers improves robustness and minimizes misclassification, especially in cases where spam and smishing messages share similar characteristics. These findings may suggest that hybrid machine learning techniques are well suited for intelligent SMS threat detection. The successful integration of the classification model into a web-based interface also suggests the practical applicability of the proposed system for real-world deployment.

It may suggest that intelligent hybrid machine learning systems can play a crucial role in mitigating these threats in the research. By combining accuracy and practical implementation, the proposed system provides a solid foundation for future advancements in SMS-based cybersecurity solutions.

REFERENCES

[1]. Altan, I., Bachir, A., Parbhulkar, Y., Rizvi, A.M., Farazi, M., 2025. Dual-Path Phishing Detection: Integrating Transformer-Based NLP with Structural URL Analysis. https://doi.org/10.48550/arXiv.2509.20972

[2]. Aparna, D.G., Krishna, B.V., Reddy, C.K., Latha, K., Akshitha, M., 2025. SMS PHISHING DETECTION USING MACHINE LEARNING TECHNIQUES 10.

[3]. Cagatay Catal, Gorkem Giray, Bedir Tekinerdogan, Sandeep Kumar, Suyash Shukla, n.d. Applications of deep learning for phishing detection: a systematic literature review | Knowledge and Information Systems | Springer Nature Link [WWW Document]. URL https://link.springer.com/article/10.1007/s10115-022-01672-x (accessed 2.9.26).

[4]. Chichwadia, A.E., Mpekoa, N., 2024. Detecting Smishing and Vishing Attacks using Machine Learning. Int. J. Intell. Comput. Res. 15, 1234–1241. https://doi.org/10.20533/ijicr.2042.4655.2024.0151

[5]. Elbehiery, H., 2025. An Efficient Phishing Detection Framework Based on Hybrid Machine Learning Models. Sustain. Mach. Intell. J. 11. https://doi.org/10.61356/SMIJ.2025.11525

[6]. Goel, D., Ahmad, H., Jain, A.K., Goel, N.K., 2024. Machine Learning Driven Smishing Detection Framework for Mobile Security [WWW Document]. arXiv.org. URL https://arxiv.org/abs/2412.09641v1 (accessed 2.9.26).

[7]. Ishaq, A., Iro, Z., Musa, A., Ayuba, A., Maijamaa, B.,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun010

Miyim, A., 2025. An Enhanced Hybrid CNN-LSTM with Attention Mechanism for SMS Phishing Detection.

[8]. Jain, A.K., Gupta, B.B., 2018. Rule-based framework for detection of smishing messages in mobile environment. Procedia Comput. Sci. 125, 617–623.

[9]. Kumar, V., Parmar, P., Singh, V., Kumar, S., Pawar, P., 2025. Phishing URL Detection Using Machine Learning: Harnessing Data Analysis to Strengthen Cyber Security. pp. 361–378. https://doi.org/10.1007/978-981-96-6715-4_26

[10]. Mahendru, S., Pandit, T., 2024. SecureNet: A Comparative Study of DeBERTa and Large Language Models for Phishing Detection, in: 2024 IEEE 7th International Conference on Big Data and Artificial Intelligence (BDAI). pp. 160–169. https://doi.org/10.1109/BDAI62182.2024.10692765

[11]. Mahmud, T., Prince, M.A.H., Ali, M.H., Hossain, M.S., Andersson, K., 2024.

[12]. Enhancing Cybersecurity: Hybrid Deep Learning Approaches to Smishing Attack Detection. Systems 12, 490. https://doi.org/10.3390/systems12110490

[13]. Munoz, M., Islam, M., 2025. A Balanced Dataset for Spam and Smishing Detection using Large Language Models (LLMs) 1. https://doi.org/10.17632/vmg875v4xs.1

[14]. Rajput, Y., Mishra, K., 2025. The Evolution of SMS Phishing (Smishing) Detection: A Comprehensive Review of Heuristic, Machine Learning and Natural Language Processing Techniques. Appl. Sci.

[15]. Rao, R.S., Kondaiah, C., Pais, A.R., Lee, B., 2025. A hybrid super learner ensemble for phishing detection on mobile devices. Sci. Rep. 15, 16839. https://doi.org/10.1038/s41598-025-02009-8

[16]. Saidat, M.R.A., Yerima, S.Y., Shaalan, K., 2024. Advancements of SMS Spam Detection: A Comprehensive Survey of NLP and ML Techniques. Procedia Comput. Sci., 6th International Conference on AI in Computational Linguistics 244, 248–259. https://doi.org/10.1016/j.procs.2024.10.198

[17]. Tamal, M.A., Islam, M.K., Bhuiyan, T., Sattar, A., Prince, N.U., 2024. Unveiling suspicious phishing attacks: enhancing detection with an optimal feature vectorization algorithm and supervised machine learning. Front. Comput. Sci. 6. https://doi.org/10.3389/fcomp.2024.1428013

[18]. Vennela, A., Akarapu, R.B., Rakshith, B.L., Asirvatham, L.G., Sunil, G., 2026.

[19]. Intelligent cybersecurity systems for phishing attack detection - An overview. Comput. Electr. Eng. 130, 110829. https://doi.org/10.1016/j.compeleceng.2025.110829

[20]. Xu, H., Qadir, A., Sadiq, S., 2025. Malicious SMS detection using ensemble learning and SMOTE to improve mobile cybersecurity. Comput. Secur. 154, 104443. https://doi.org/10.1016/j.cose.2025