SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I7P114

Volume 11 Issue 7, 141-149, July 2024

Original Article

Unmasking Deception: Artificial Neural Networks in Smishing Detection for Cyber Security Fortification

T.V. Mini[1], Jomy John[2], P.D. Siji[3]

[1]*Department of Computer Science, Sacred Heart College, Kerala, India.*
[2]*Department of Computer Science, P M Government College, Kerala, India.*
[3]*Department of Computer Science, St. Joseph's College, Kerala, India.*

[1]*Corresponding Author : tv.mini.12@outlook.com*

Received: 10 May 2024 Revised: 09 June 2024 Accepted: 09 July 2024 Published: 27 July 2024

Abstract - *Smishing is a form of social engineering that employs text messaging to mislead people into revealing confidential information or performing malicious actions. It has become a prevalent and sophisticated cyber threat. The rise of smishing (SMS phishing) poses a significant threat to cyber security, demanding advanced detection and classification methods to safeguard users and organizations. As a deceptive technique targeting mobile users through text messages, smishing requires proactive defense mechanisms. This paper proposes an Artificial Neural Networks (ANNs) model for smishing detection and classification to bolster cyber security. This paper outlines a proposed framework comprising four distinct modules designed for smishing detection and classification. Through simulation, the framework demonstrated exceptional performance, attaining an accuracy rate of 97.66%. Comparative analysis against various machine learning models underscored the superiority of the proposed approach. As mobile devices continue to be integral to daily communication, the implementation of ANN-based solutions serves as a vital component in fortifying cyber defenses, ensuring the security and privacy of individuals and organizations in an increasingly interconnected digital landscape.*

Keywords - *Cyber threats, Text messages, Spam SMS, Phishing, Smishing, Cyber Security, Neural Network, Back propagation algorithm.*

1. Introduction

In the contemporary era, the digital landscape plays a pivotal role in our day-to-day existence. The use of computers and the internet extends far beyond simple access to personal and professional information; it encompasses approaches such as the Internet of Things (IoT) and cryptocurrency. As the world increasingly depends on digital technology, the profound integration of computers and various technological facets introduces novel challenges. Cyber security has emerged as a pressing concern, with individuals and organizations confronting a diverse array of cyber threats across internet infrastructures, spanning from e-banking to e-commerce [1, 2]. Common risks encompass identity theft, phishing, man-in-the-middle attacks, SQL injections, DDoS attacks, and malware. Notably, a significant portion of contemporary internet attacks originate from the direct or indirect application of phishing techniques.

Phishing scams represent a cunning online strategy wherein malicious actors exploit electronic communication channels, such as social media, to manipulate users into taking actions that ultimately benefit the scammer [3]. Typically, these fraudulent activities involve the use of deceptive emails that pose as authentic messages from reputable organizations. Scammers target smartphone users by inundating them with a stream of bogus emails containing malicious links. These emails prompt unsuspecting users to update their information by clicking on a provided link, redirecting them to a deceitful website designed to obtain sensitive data or install harmful software illicitly. To execute a phishing attack, the perpetrator meticulously replicates the official website, creating an eerily accurate mirror image. Given the widespread use of smartphones in India, where text messaging serves as a cost-effective and popular means of communication, scammers exploit this medium to disseminate enticing shopping offers and advertisements [4].

SMS phishing, also known as smishing, is a deceptive practice wherein cybercriminals send fraudulent text messages to smartphone users with the aim of illicitly obtaining sensitive information. Smishing attacks involve the transmission of misleading messages designed to prompt the recipient into taking actions that could lead to financial losses or the unauthorized use of personal information. Perpetrators meticulously craft these deceptive messages to create an illusion of authenticity, intending to deceive the recipient into believing that the communication originates from a legitimate individual or organization. Figure 1 illustrates an example of a smishing SMS, highlighting the carefully orchestrated nature of these malicious attempts.

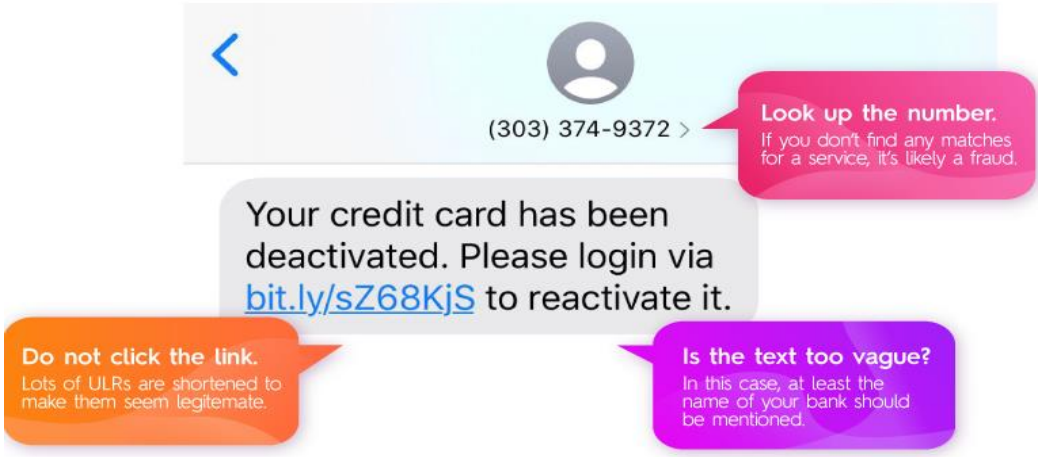

Fig. 1 An example of smishing SMS

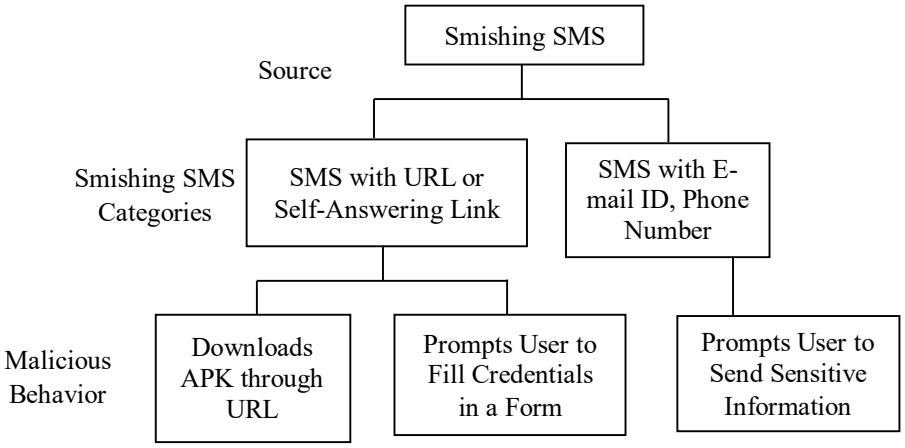

Fig. 2 Malicious activities of a smishing SMS

Smishing attacks represent a deceptive form of communication designed to deceive recipients into taking actions that may lead to financial losses or the compromise of sensitive information. Perpetrators carefully craft these misleading messages to create an illusion of legitimacy, tricking victims into believing they are genuine communications from trusted individuals or organizations. The sophistication of these messages often convinces individuals to divulge crucial information such as credit card numbers, CVVs, professional credentials like usernames and passwords, or personal details, including social security numbers. The tactics employed in smishing SMS messages vary, encompassing offers of enticing discount coupons, gifts, or prizes. Additionally, attackers may pose as legitimate entities, requesting the renewal of Know Your Customer (KYC) details or soliciting user-sensitive information purportedly for activating a bank's debit or credit card. To achieve their objectives, these deceptive messages guide users towards interacting with fraudulent applications, malicious links, or fake websites, with the ultimate goal of illicitly obtaining sensitive user credentials like usernames, passwords, and bank account details. Figure 2 classifies the malicious activities associated with smishing SMS [5].

Detecting and classifying Smishing attacks is a critical aspect of fortifying cyber security defenses against the constantly evolving landscape of cyber threats. In today's expansive digital environment, cyber attackers continually find new avenues to exploit. Smishing, a deceptive practice involving phishing attempts through SMS, has become a formidable weapon in their arsenal. The identification and categorization of Smishing attacks not only protect individuals and organizations from potential financial losses and data breaches but also contribute to enhancing overall cyber security resilience.

The application of advanced technologies, such as machine learning and pattern recognition, plays a pivotal role in this process. By leveraging these tools, security systems can swiftly identify suspicious messages, analyze their content, and categorize them effectively. This proactive approach allows for the timely implementation of countermeasures, providing users with the confidence to navigate the digital realm while bolstering their resilience against the persistent threat of cyber-attacks. The major contribution of this paper includes:

- An intelligent framework for Smishing detection and classification using ANN.
- An effective feature extraction method using Neural Network.
- Evaluating the effectiveness of the suggested method in comparison to various machine learning approaches.

The subsequent section of the paper is organized as follows: Section two presents a review of the existing literature, highlighting areas necessitating further research. Section three elucidates the methodology in detail. Section four delves into a comprehensive discussion of the outcomes resulting from the proposed approach. Finally, in Section 5, the paper concludes by summarizing the findings.

2. Related Works

Jae Woong Joo et al. [6] introduced an enhanced security model for smishing attack detection, employing a Naive Bayes classifier to differentiate between normal and smishing SMS. The proposed model demonstrated effectiveness in analyzing and detecting SMS phishing. Ankit Kumar Jain et al. [7] presented a two-phase framework to rectify smishing SMS from spam SMS. The initial phase rectified spam and ham SMS, while the next phase specifically targets smishing SMS. Simulation results indicated high accuracy in detecting spam messages (94.9%) and filtering smishing messages (96%). Caner Balim and Efnan Sora Gunal [8] proposed machine learning models for smishing detection, conducting experiments on a substantial dataset containing both legitimate and smishing messages.

The proposed model exhibited promising detection performance. Oluwatobi Noah Akande et al. [9] implemented rule-based classifiers for spam and ham SMS separation, utilizing a mobile application for smishing detection. The developed application provided the analysis outcomes to consumers, contributing to effective spam and smishing differentiation.

Heider A. M. Wahsheh and Mohammed S. Al-Zahrani [10] developed two automated approaches for smishing detection. The initial model employed a lightweight cryptographic SMS approach to secure messages and identify potential phishing content. Experimentation with ensemble learning classifiers demonstrated that RF achieved superior results, making it the suggested framework for anti-smishing efforts.

Gunikhan Sonowal and K. S. Kuppusamy [11] introduced an innovative anti-smishing framework named SmiDCA, which scrutinized smishing messages by extracting 39 features for detection. Four established machine-learning algorithms were employed to differentiate between smishing and legitimate messages. The experimental outcomes revealed an impressive accuracy of 96.40%, particularly with the assistance of the RF classifier. Diksha Goel and Ankit Kumar Jain [12] proposed a unique approach for detecting smishing attacks utilizing an NB classifier to filter text messages. This approach analyzed message content, extracting commonly used words in smishing messages.

Neda Abdelhamid et al. [13] developed an intelligent framework for phishing detection, exploring various machine-learning techniques to identify optimal anti-phishing tools. Comparative experiments on real phishing datasets demonstrated that covering approach models, particularly for novice users, were more suitable as anti-phishing solutions.

Ping Yi et al. [14] focused on analyzing phishing website features, presenting two types of features for web phishing detection. They introduced a deep belief network to detect phishing websites and attained a notable TPR of 90%.

Jian Mao et al. [15] proposed a robust phishing detection approach, Phishing-Alarm, leveraging the CSS features of web pages. The proposed methodology involved identifying effective CSS features and developing algorithms for efficient page similarity evaluation. The effectiveness of the phishing alarm was demonstrated through a prototype integrated into the Google Chrome browser, showcasing its prowess in real-world phishing sample evaluations.

Although a large portion of recent research has been successful in classifying messages as spam or authentic, it is still not possible to classify messages as malicious—that is, as smishing. Many of the current techniques rely on feature-based algorithms that are applied to SMS data in order to extract particular features from the message content.

Nevertheless, these techniques frequently ignore these traits' behavioral elements, which are essential for accurately identifying SMS-based phishing assaults. The focus on feature behavior instead of just their presence is shown to be a crucial component missing from modern methods for addressing the particular difficulties presented by smishing detection.

3. Materials and Methods

Smishing involves malicious texts leading users to harmful websites or applications, and it is a subset of spam SMS. This paper develops a smishing detection and classification model employing an ANN to differentiate between smishing and legitimate messages.

The proposed approach aims to categorize messages into either smishing or HAM (legitimate), where smishing includes malicious messages, and HAM encompasses both legitimate and spam messages sent for advertising or commercial purposes. The detailed block schematics of the suggested approach are illustrated in Figure 3.

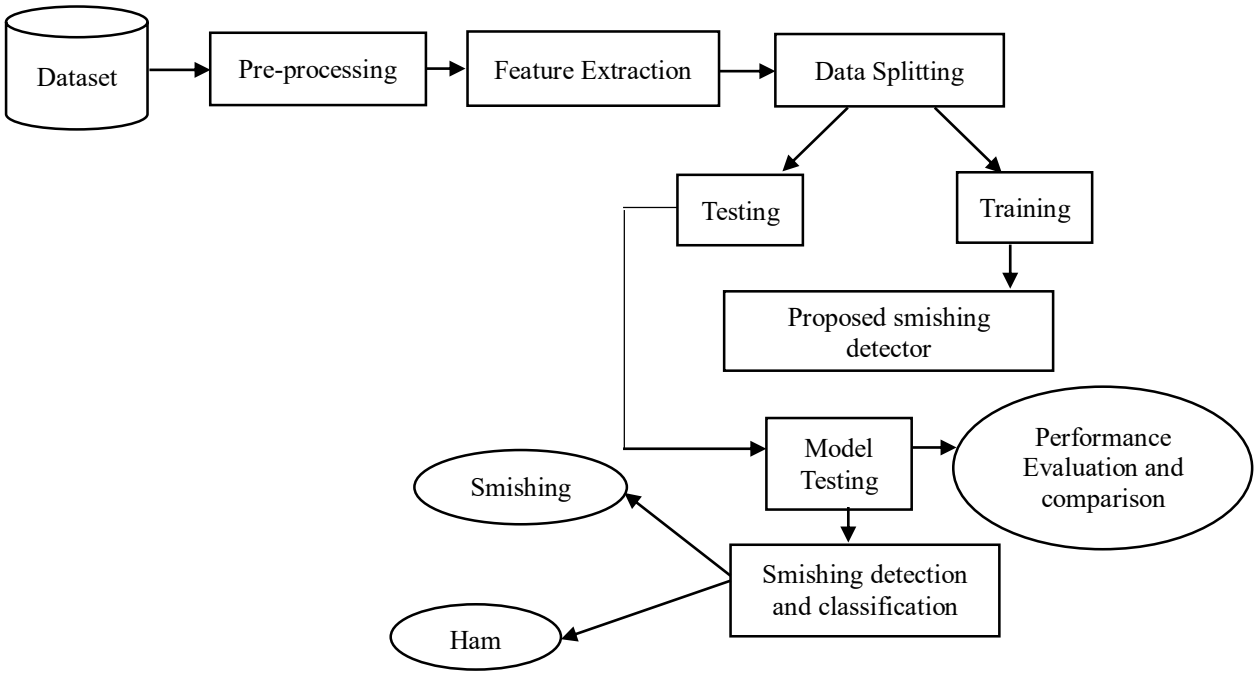

Fig. 3 Block diagram of the proposed methodology

3.1. Dataset Description

The dataset gathered from Almeida's research [16] comprises 5574 text messages, with 4827 categorized as ham and 747 as spam. Among the spam messages, 254 are identified as smishing, aiming to obtain sensitive user information illicitly. These smishing messages were manually extracted. Additionally, real-time smishing messages were collected for evaluating the proposed model. The dataset includes both ham and smishing messages and smishing messages originated from images and were converted into text.

3.2. Pre-processing

Before moving on to feature extraction and model evaluation, the text messages undergo pre-processing. This step involves preparing the text data for analysis and machine learning algorithms. As a crucial preparatory phase, text pre-processing encompasses actions such as converting all words to lowercase, eliminating punctuation marks and special characters present in text messages, and applying stemming [17].

3.2.1. Lower Casing

Converting each word to lowercase is an essential step in text pre-processing. It helps in reducing the dimensionality of the data, improving the accuracy of text analysis, and enhancing the performance of machine learning models. By treating all words as lowercase, we eliminate the distinction between different capitalizations of the same word, effectively reducing the number of unique words in the corpus.

3.2.2. Stemming

Stemming is a text pre-processing technique that reduces words to their root or base form using heuristics. This involves removing prefixes and suffixes, such as "-ing," "-ed," "-es," and "-s," to group related words together and reduce the dimensionality of the text data.

3.2.3. Removal of Punctuation and Special Strings

Removal of punctuation and special strings is a crucial step in text pre-processing, as it helps to clean and normalize the text data, making it more suitable for analysis and learning algorithms.

3.3. Feature Extraction

The key features of the dataset are identified through the use of ANN. Within ANN, feature extraction involves monitoring the minimum error value associated with the chosen feature [18]. These features are derived by emphasizing the network's minimal error and maximal accuracy. The most prominent features in the dataset are listed in Table 1.

3.4. Proposed Smishing Detection and Classification Model

The proposed model for detecting and classifying smishing utilizes a back propagation neural network [19]. NN, inspired by the structure of the human brain, are a form of supervised machine learning that tackles problems by constructing interconnected nodes resembling neurons. These networks learn from datasets, developing the ability to make informed decisions autonomously.

In the context of NNs, a target output is essential, paired with a set of inputs, allowing the network to compare its predictions with the actual output. An activation function is employed to generate a predicted or actual output, and a loss function quantifies the disparity between the predicted and target outputs. To minimize this error, weights and biases associated with the network are updated, facilitated by an optimization function. This function, in turn, employs gradient descent a process where the partial derivative of the loss function is calculated to determine the slope of the error function. Feedforward involves predicting the output, while backpropagation [20] focuses on computing and minimizing the error. The architecture of the NN is visually represented in layers, as depicted in Figure 4.

The input layer, serving as the initial layer, connects to multiple hidden layers, ultimately leading to the output layer, which functions as the final layer of the system. Each neuron in the input layer receives inputs, and Equation (1) is employed to compute the weighted summation of these inputs. The weighted sum is then fed into an activation function, expressed by Equation (2).

$$S_j = b_j + \sum_{i=1}^{I} W_{ij}\, x_i \quad (1)$$

$$a_j = \frac{1}{(1+e^{-s_j})} \quad (2)$$

The predicted output of the hidden layer is represented in Equation (3).

$$S_{k=} b_k + \sum_{j=1}^{J} W_{jk} a_j \quad (3)$$

Table 1. Feature description

Features	Description
url	A link present in the smishing message redirects the user to fraudulent websites.
email_id	An email id of the attacker is provided in the SMS.
phone_no	The phone number of the attacker is included in the text message.
leet_words	Words in which lookalike symbols and numerals are used in place of alphabets to give a similar look.
smishing_keywords	Words included in the SMS prompt the user to contact the attacker.
Symbols	Symbols like %, /etc are included in the SMS to delude the user.
special_characters	Special characters like $! etc., are included in the SMS to attract the user.

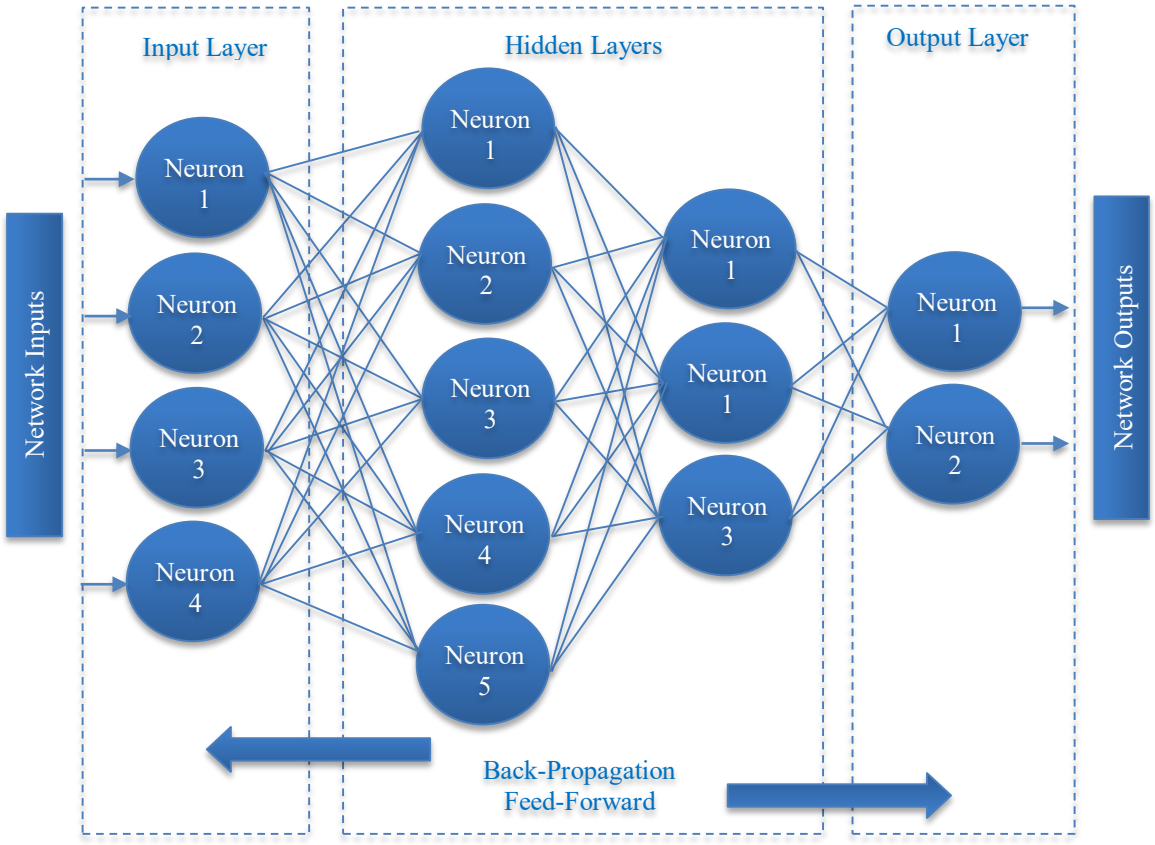

Fig. 4 Backpropagation in ANN

The final output, a_k, which is the actual predicted output of the model attained by examining each layer within the network, can be represented as in Equation (4).

$$a_k = \frac{1}{(1+e^{-s_k})} \tag{4}$$

After the completion of the feed-forward phase, it is necessary to engage in backward propagation to calculate the error.

$$E = \frac{1}{2}(t_k - a_k)^2 \tag{5}$$

Minimizing the error involves computing the partial derivative of the error function, known as gradient descent, to determine the slope of the error function. A positive slope signifies the necessity to reduce weights, whereas a negative slope implies the need to augment weights. Equation (6) is employed to compute the gradient of the error.

$$\frac{\partial E}{\partial a_k} = \frac{\partial[\frac{1}{2}(t_k - a_k)^2]}{\partial a_k} \tag{6}$$

The new weights are calculated by using Equation (7).

$$W_{kj}^{new} = W_{kj}^{old} + \eta \frac{\partial E}{\partial W_{kj}} + \alpha \Delta W_{kj}^{old} \tag{7}$$

The smishing detection and classification model is composed of four phases. The block schematics of the suggested smishing detection and classification approach are illustrated in Figure 5. The detailed algorithm is also explained below

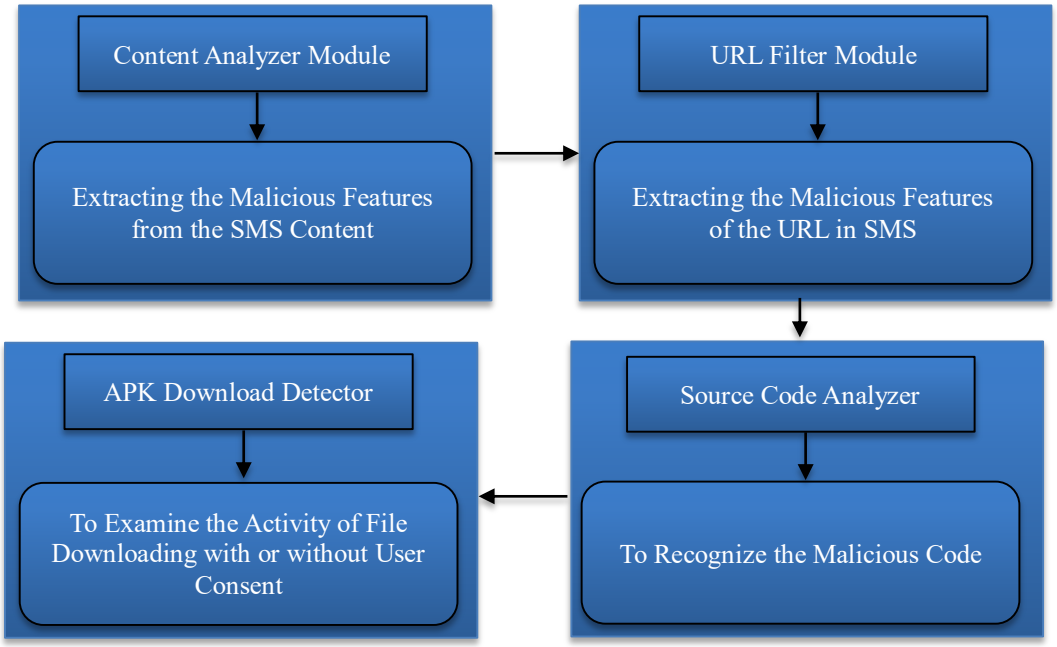

Fig. 5 Proposed smishing detection and classification model

Algorithm 1. Smishing detection and classification

Input : Incoming text messages (currentsms), current_URL (currentsms.URL)
Output : Smishing Message, Ham Message

Step 1 : Check for patterns in the SMS body.
Step 2 : If the SMS body matches a blacklist, classify it as smishing.
Step 3 : If the SMS body contains a URL, check the URL for various features.
Step 4 : If the URL matches a blacklist or has three or more of the checked features, classify it as smishing.
Step 5 : If the URL is legitimate, retrieve the HTML source code of the webpage and check for a form tag.
Step 6 : If the source code contains a form tag and the domain name matches the extracted domain name, classify the SMS as legitimate.
Step 7 : If the source code does not contain a form tag or the domain name does not match, use an APK download detector algorithm for further classification.
Step 8 : Check the target URL of the extracted URL for .apk in the basename.
Step 9 : If user consent is taken, classify the SMS as Ham.
Step 10 : If user consent is not taken, classify the SMS as Smishing.
Step 11 : If the basename of the target URL does not contain .apk, classify the SMS as Ham.

Table 2. Hyperparameters

Hyperparameters	Values
Number of Neurons	12
Activation function	sigmoid
Momentum	0.8
Learning Rate	0.01
Loss Function	Binary cross entropy
Epochs	50
Optimizer	Stochastic Gradient Descent

4. Results and Discussion

4.1. Hardware and Software Setup

The proposed model was executed after the dataset was collected. The entire procedure was carried out in Python and TensorFlow on Google Collaboratory, where the model was built and trained.

The entire dataset was split into training, testing, and validation sets. Various hyper-parameter values are experimented with in the network to achieve optimal performance. The value of hyper-parameters that yield the highest accuracy on the dataset is noted, which is shown in Table 2.

4.2. Performance Evaluation

Accuracy and loss plots are crucial for evaluating the performance of a model during training. The accuracy plot shows the effectiveness of the model in making correct predictions, usually improving over time. Meanwhile, the loss plot reflects the decreasing error between predicted and actual values, signaling the convergence of the model to an optimal state. The accuracy plot and loss plot of the suggested model are visualized in Figure 6. The classification report of the suggested model is tabulated in Table 3.

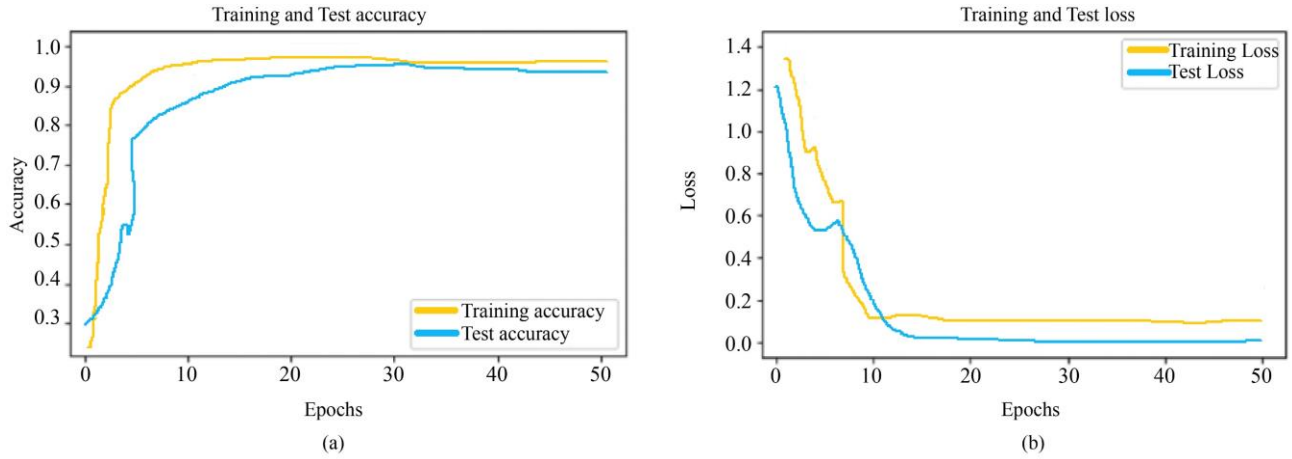

Fig. 6(a) Accuracy plot, and (b) Loss plot.

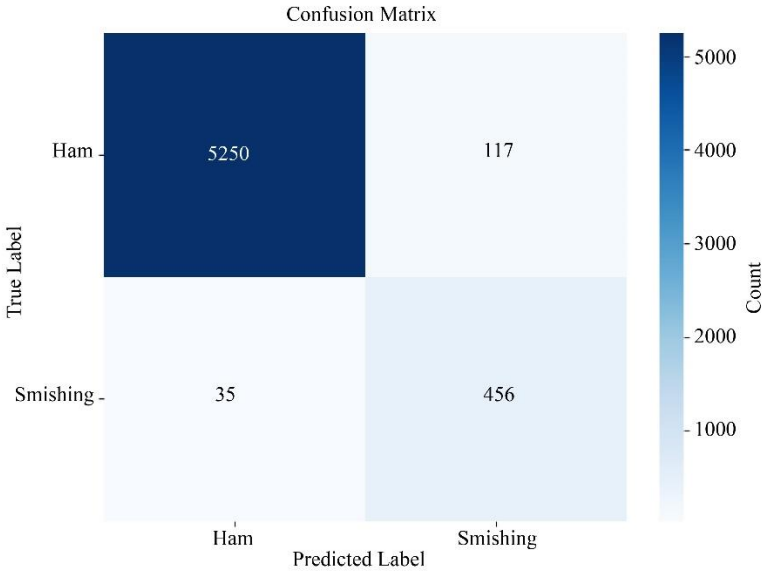

Fig. 7 Confusion matrix

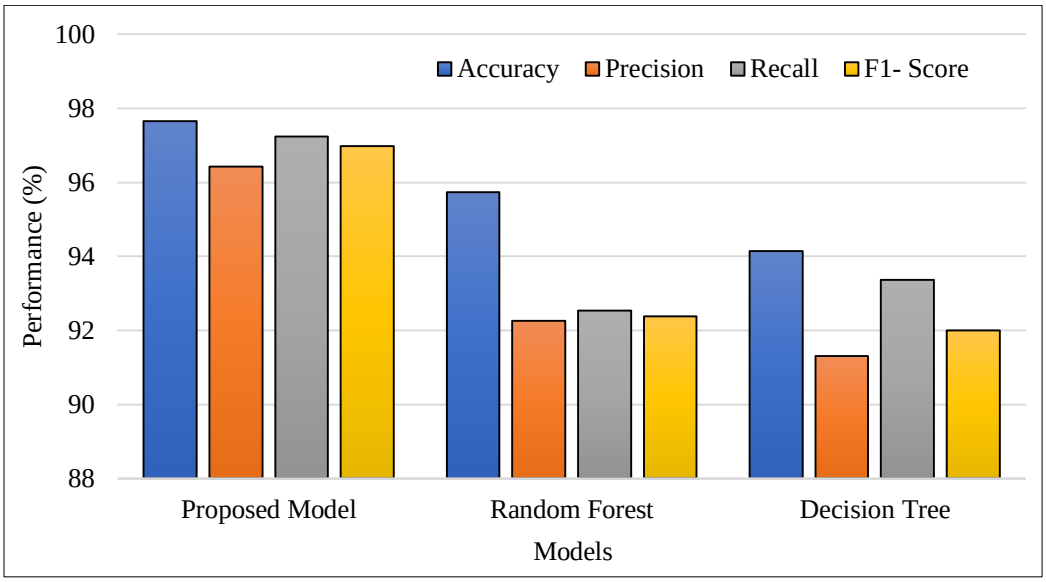

Fig. 8 Performance comparison

Table 3. Classification report

Performance Metrics	Result Obtained (%)
Accuracy	97.66
Precision	96.42
Recall	97.23
F1-score	96.98

The confusion matrix plays a crucial role in evaluating the performance of a model, offering a comprehensive breakdown of predicted and actual class assignments [21]. In this case, the confusion matrix contains two classes: smishing and ham. The confusion matrix of the proposed model is illustrated in Figure 7.

Figure 8 depicts the performance comparison of the suggested model with various ML models. The proposed method exhibits superior overall performance, achieving an impressive accuracy of 97.66%, surpassing both Random Forest and Decision Tree. These results emphasize the efficacy of the proposed method in smishing detection, suggesting its potential for robust and accurate classification in a cyber-security context.

5. Conclusion

With the increasing prevalence of mobile devices and the growing reliance on text messages for communication, smishing has emerged as a potent threat vector for cyber-attacks. Detecting and classifying smishing attempts is crucial for safeguarding individuals and organizations against phishing scams that aim to compromise sensitive information, such as login credentials and financial details. The ability to accurately identify and thwart smishing attacks enhances overall cyber security by preventing unauthorized access, financial losses, and the potential compromise of personal or corporate data.

This paper proposes an effective smishing detection and classification model using ANN. The suggested framework primarily comprises four distinct modules designed for the detection and classification of smishing. The experimental findings revealed that the proposed approach achieved outstanding performance with an accuracy of 97.66% and outperformed various machine learning models.

The integration of ANN technology in smishing detection is pivotal for staying ahead of cyber adversaries, contributing significantly to the ongoing efforts to safeguard sensitive information and ensure the integrity of digital communications.

Acknowledgements

The author expresses profound appreciation to the supervisor for providing guidance and unwavering support throughout the course of this study.

References

[1] Ross Anderson et al., *Measuring the Cost of Cybercrime*, The Economics of Information Security and Privacy, Springer, Berlin, Heidelberg, pp. 265-300, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[2] What Are the Most Common Cyber Attacks?, 2022. [Online]. Available: https://www.cisco.com/c/en_in/products/security/common-cyberattacks.html#~typesof-cyber-attacks

[3] Mahmoud Khonji, Youssef Iraqi, and Andrew Jones, "Phishing Detection: A Literature Survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 2091-2121, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[4] Worldwide Texting Statistics, 2018. [Online]. Available: https://shso.vermont.gov/content/worldwide-texting-statistics

[5] Sandhya Mishra, and Devpriya Soni, "Smishing Detector: A Security Model to Detect Smishing through SMS Content Analysis and URL Behavior Analysis," *Future Generation Computer Systems*, vol. 108, pp. 803-815, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jae Woong Joo et al., "S-Detector: An Enhanced Security Model for Detecting Smishing Attack for Mobile Computing," *Telecommunication Systems*, vol. 66, pp. 29-38, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[7] Ankit Kumar Jain, Sumit Kumar Yadav, and Neelam Choudhary, "A Novel Approach to Detect Spam and Smishing SMS Using Machine Learning Techniques," *International Journal of E-Services and Mobile Applications*, vol. 12, no. 1, pp. 21-38, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Caner Balim, and Efnan Sora Gunal, "Automatic Detection of Smishing Attacks by Machine Learning Methods," *1st International Informatics and Software Engineering Conference*, Ankara, Turkey, pp. 1-3, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[9] Oluwatobi Noah Akande et al., "Development of a Real Time Smishing Detection Mobile Application Using Rule Based Techniques," *Procedia Computer Science*, vol. 199, pp. 95-102, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Heider A.M. Wahsheh, and Mohammed S. Al-Zahrani, "Lightweight Cryptographic and Artificial Intelligence Models for Anti-Smishing," *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, pp. 483-496, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Gunikhan Sonowal, and K.S. Kuppusamy, "SmiDCA: An Anti-Smishing Model with Machine Learning Approach," *The Computer Journal*, vol. 61, no. 8, pp. 1143-1157, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[12] Diksha Goel, and Ankit Kumar Jain, "Smishing-Classifier: A Novel Framework for Detection of Smishing Attack in Mobile Environment," *Smart and Innovative Trends in Next Generation Computing: Third International Conference*, Dehradun, India, pp. 502-512, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[13] Neda Abdelhamid, Fadi Thabtah, and Hussein Abdel-Jaber, "Phishing Detection: A Recent Intelligent Machine Learning Comparison Based on Models' Content and Features," *2017 IEEE International Conference on Intelligence and Security Informatics*, Beijing, China, pp. 72-77, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[14] Ping Yi et al., "Web Phishing Detection Using a Deep Learning Framework," *Wireless Communications and Mobile Computing*, vol. 2018, no. 1, pp. 1-9, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[15] Jian Mao et al., "Phishing-Alarm: Robust and Efficient Phishing Detection via Page Component Similarity," *IEEE Access*, vol. 5, pp. 17020-17030, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[16] Tiago A. Almeida, Jose Maria G. Hidalgo, and Akebo Yamakami, "Contributions to the Study of SMS Spam Filtering: New Collection and Results," *Proceedings of the 11th ACM Symposium on Document Engineering*, Mountain View California USA, pp. 259-262, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[17] S. Vijayarani, M.J. Ilamathi, and M. Nithya, "Preprocessing Techniques for Text Mining - An Overview," *International Journal of Computer Science & Communication Networks*, vol. 5, no. 1, pp. 7-16, 2015. [Google Scholar] [Publisher Link]

[18] M. Soranamageswari, and C. Meena, "Statistical Feature Extraction for Classification of Image Spam Using Artificial Neural Networks," *2010 Second International Conference on Machine Learning and Computing*, Bangalore, India, pp. 101-105, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[19] Jing Li et al., "Brief Introduction of Back Propagation (BP) Neural Network Algorithm and Its Improvement," *Advances in Computer Science and Information Engineering*, vol. 2, pp. 553-558, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[20] Back Propagation Process in Deep Neural Network, 2022. [Online]. Available: https://www.javatpoint.com/pytorch-backpropagation-process-in-deep-neural-network

[21] Robert Susmaga, "Confusion Matrix Visualization," *Proceedings of the International Intelligent Information Processing and Web Mining*, Zakopane, Poland, pp. 107-116, 2004. [CrossRef] [Google Scholar] [Publisher Link]