Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 12, 2025

PREDICTING PHISHING ATTACKS USING NATURAL LANGUAGE PROCESSING AND USER BEHAVIORAL INDICATORS

Muhammad Haqan Ali Rai[*1], Deapika Dulani[2], Ahtasham Ali[3], Muhammad SuffianTafoor[4]

[*1]*BS Software Engineering: Department (UIIT) University Institute of Information Technology, PMAS Arid Agriculture University, Rawalpindi, Pakistan*
[2]*BS Computer Science: Department of Computer Science, Sukkur IBA University, Sukkur, Pakistan*
[3]*BS Computer Science: Department of Computer Science, Government College University, Lahore, Pakistan*
[4]*BS Software Engineering: Department of Software Engineering, Foundation University Islamabad, Pakistan*

[*1]haqanali934@gmail.com,[2]deepikadulani55@gmail.com,[3]ehtshamaali@gmail.com,
[4]muhammad.suffian.5959@gmail.com

DOI: https://doi.org/10.5281/zenodo.18030987

Keywords
Phishing detection, Cyber security, Natural Language Processing, User behavior, Machine learning, Hybrid models, Human-centered security, Email security

Article History
Received: 11 October 2025
Accepted: 21 November 2025
Published: 18 December 2025

Corresponding Author: *
Muhammad Haqan Ali Rai

Abstract

Phishing attacks continue to be among the most pervasive and damaging threats in the cyber security landscape, leveraging both technical vulnerabilities and human cognitive biases to manipulate users into disclosing sensitive information such as login credentials, financial data, and personal identifiers. Traditional detection mechanisms, including rule-based and signature-based systems, often fail to keep pace with the rapidly evolving linguistic patterns, sophisticated social engineering tactics, and polymorphic strategies employed by attackers in modern phishing campaigns. Addressing this challenge requires an integrated approach that considers both the content of phishing messages and the behavioral characteristics of users interacting with them. This research proposes a comprehensive framework for predicting phishing attacks by combining advanced Natural Language Processing (NLP) techniques with user behavioral indicators. The study examines semantic, syntactic, and pragmatic features of emails and messaging content to capture subtle cues of phishing attempts. In parallel, it analyzes user interaction patterns, including click behaviors, response times, and historical susceptibility to phishing, thereby recognizing the socio-technical dimensions of cyber security threats. A supervised machine learning methodology is employed, utilizing benchmark phishing datasets augmented with simulated behavioral data to enhance model generalizability. The findings indicate that hybrid models integrating content analysis with behavioral features significantly outperform content-only approaches. These models achieve higher precision, recall, and overall robustness, demonstrating the ability to detect previously unseen phishing strategies with greater reliability. The study underscores the importance of accounting for human factors alongside technical indicators in cybersecurity research, highlighting the need for proactive, human-centered defense mechanisms. By emphasizing the interplay between linguistic cues and user behavior, this research contributes to the development of predictive and adaptive phishing detection systems. The results offer practical implications for organizations, suggesting strategies for implementing training programs, awareness campaigns,

and AI-driven monitoring tools that target high-risk behaviors while enhancing overall email security. This integrated approach provides a pathway toward more resilient cyber security infrastructures capable of responding effectively to increasingly sophisticated phishing attacks.

INTRODUCTION

The rapid digital transformation of communication and commercial activities has fundamentally reshaped how individuals and organizations interact, transact, and exchange information. Email services, instant messaging applications, social media platforms, and web-based systems have become integral to daily personal and professional operations. While this transformation has improved efficiency and connectivity, it has also inadvertently broadened the digital attack surface, providing cybercriminals with increased opportunities to exploit users. As digital dependence continues to grow, so do the frequency, sophistication, and impact of cyber threats targeting unsuspecting individuals and organizations (Benavides-Astudillo et al., 2022a).

Among the diverse range of cyber threats, phishing attacks have emerged as one of the most prevalent and damaging vectors for security breaches, financial losses, and identity theft. Phishing attacks typically involve deceptive messages that impersonate trusted institutions, such as banks, government agencies, or reputable organizations, with the intent of manipulating recipients into revealing sensitive information or performing harmful actions. These attacks exploit psychological triggers such as urgency, fear, curiosity, and trust, making them particularly effective even against technologically aware users. Despite continuous advancements in cybersecurity infrastructure, including spam filters, intrusion detection systems, and machine learning-based defenses, phishing attacks remain alarmingly successful due to their ability to adapt language, context, and presentation to evade automated detection mechanisms (Selvaganapathy et al., 2018).

In recent years, Natural Language Processing (NLP) has gained significant attention as a promising approach for analyzing textual data and identifying malicious intent embedded within digital communications. NLP techniques enable the extraction of semantic, syntactic, and contextual features that can distinguish phishing messages from legitimate correspondence. However, phishing is not merely a linguistic or technical challenge; it is inherently a socio-technical problem that involves human cognition, behavior, and decision-making. Users' prior experiences with digital platforms, their levels of trust, risk perception, familiarity with online threats, and habitual interaction patterns play a critical role in determining their susceptibility to phishing attacks (Kitchenham et al., 2009).

Consequently, detection systems that rely solely on textual analysis may overlook critical behavioral dimensions that influence user vulnerability. Such systems may fail to capture why certain individuals consistently fall victim to phishing attempts despite exposure to similar content. This study, therefore, argues for a more holistic and integrated approach that combines NLP-based content analysis with user behavioral indicators to enhance the accuracy and robustness of phishing detection mechanisms. By incorporating both linguistic cues and behavioral patterns, the proposed approach aims to improve predictive capabilities, reduce false negatives, and contribute to the development of more adaptive and user-aware cybersecurity defenses against phishing attacks (Jain & Gupta, 2019).

In today's digital age, the Internet has become a cornerstone of communication, information dissemination, and engagement across various sectors, including business, education, and banking. The Internet serves as a platform for exchanging diverse content, such as research papers, educational materials, and multimedia resources, which increases cybercriminal threats that seize users' confidential information (Catal et al., 2022; Mahmoud et al., 2013). Among different threats, phishing is considered as a widespread attack, where the hackers can access the data without any complex cipher codes (Hannousse and Yahiouche, 2021). Such phishing attacks occur in various forms through Email, random SMS, social media, Quick Response codes, and URL links (Safi and Singh, 2023). Due to a lack of knowledge about URLs, blind trust in

webpages or messages, and redirected webpage locations, the users are subjected to phishing attacks (Basit et al., 2021). The report of the Anti-Phishing Working Group stated that the number of phishing attacks increased to 2,50,000 within a month in 2021 and kept on increasing. Thus, it is essential to develop an effective system to detect phishing attacks for preventing further attacks in the future (Asiri et al., 2023; (Ariyadasa et al., 2022).

For phishing detection, Meta-heuristic methods are developed to collect information, and the URL is verified with the blacklist to check its legitimacy (Odeh et al., 2021). As an efficient worldwide standard, Email networks are intruded on by cybercriminals for financial benefits. These phishing emails are detected by utilizing the Themis model, which deeply analyzes the structure of the mail (Atlam and Oluwatimilehin, 2023; Salloum et al., 2021). Due to the various ambiguities of the detection systems, phishing websites can also be tested and detected by utilizing a Fuzzy logic technique (Bhagwat et al., 2021). Similar to phishing emails, phishing SMS is created with some random phone numbers for making money transactions by users (Abdillah et al., 2022). Currently, to enhance the phishing detection process, Machine Learning (ML) and Deep Learning (DL) algorithms are being analyzed (Li et al., 2023).

The labeled datasets are utilized to recognize malicious and benign websites by using ML approaches. For the phishing detection, the supervised learning algorithms like Logistic Regression (LR), Support Vector Machines (SVM), Decision Tree (DT), Naïve Bayes (NB), and Random Forest (RF) are utilized (Tang and Mahmoud, 2021). Among these, the SVM classifier accurately detects phishing attacks along with the word embedding technique (Salloum et al., 2022). Furthermore, the DL models, including LSTM and Convolutional Neural Network (CNN), efficiently detect phishing by learning the patterns and anomalies of the data (Thakur et al., 2023). These phishing detection systems contribute to mitigating attacks via software-based phishing tools and human-centric strategies (Naqvi et al., 2023). But, for phishing attack detection, multiple website sources, such as SMS, E-Mail, and URLs were not concentrated in any of the existing works (Villanueva et al., 2022).

The existing (Brezeanu et al., 2025) utilized the HyperText Markup Language (HTML) code to identify the phishing attack, helping in automatically updating the phishing indicators. Also, the prevailing (Sturman et al., 2024) detected the phishing attack based on the user knowledge and decision style. The traditional (Shombot et al., 2024) used SVM for attack detection and attained higher accuracy. Moreover, the prevailing (Sudar et al., 2024) improved the resilience against evasive phishing strategies (Ozcan et al., 2021). In the existing (Biswas et al., 2024), the transparent and interpretable models were utilized for the attack prediction. Yet, these prevailing models did not predict the phishing strategies in the multi-data source. Hence, this work detects phishing attacks in various sources, including SMS, E-Mail, and URL data, by analyzing the user behavior using EM-BERT and SPCA BASED EAI-SC-LSTM techniques (Adebowale et al., 2020).

1.2 Problem Statement

Existing phishing detection systems predominantly focus on either content-based analysis or technical indicators such as URLs and metadata. These approaches often fail to generalize across diverse phishing tactics and overlook the human behavioral dimension that plays a critical role in attack success. There is a lack of comprehensive models that jointly analyze message content and user behavior to predict phishing attacks proactively and accurately.

1.3 Research Objectives

1. To analyze linguistic patterns in phishing messages using advanced NLP techniques.
2. To identify key user behavioral indicators associated with phishing susceptibility.
3. To develop a hybrid predictive model integrating NLP features and behavioral data.
4. To evaluate the performance of the proposed model against traditional detection approaches.

1.4 Research Questions

1. What linguistic features most effectively distinguish phishing messages from legitimate communications?
2. Which user behavioral indicators significantly contribute to phishing prediction?

3. Does integrating NLP and behavioral features improve phishing detection accuracy?
4. How does the hybrid model perform in detecting previously unseen phishing attacks?

2. LITERATURE REVIEW

2.1 Phishing Attacks and Cyber security Threat Landscape

Phishing attacks represent one of the most persistent and rapidly evolving threats within the contemporary cyber security landscape. Initially, phishing primarily involved generic and poorly constructed deceptive emails sent indiscriminately to a large number of recipients. These early attacks relied on basic social engineering techniques, such as spoofed email addresses and fabricated messages, to trick users into revealing login credentials or financial information (Hochreiter & Schmidhuber, 1997). However, as users and organizations became more aware of such threats and defensive technologies improved, phishing strategies evolved in complexity, sophistication, and precision (Britz, 2015).

Modern phishing attacks are increasingly context-aware and highly targeted. Techniques such as spear-phishing focus on specific individuals or organizations by leveraging personal, professional, or organizational information obtained from social media, data breaches, or public records. Whaling, a more specialized form of spear-phishing, targets high-ranking executives or decision-makers, exploiting their authority and access to sensitive systems and financial resources. These targeted attacks are often carefully crafted to appear legitimate, incorporating organizational language, branding, and timing cues that significantly increase their credibility and success rates (Deng et al., 2018).

Numerous empirical studies consistently identify phishing as a leading cause of security breaches across both public and private sectors (Vinayakumar et al., 2018). One of the primary reasons for its continued effectiveness is its low operational cost combined with a high return on investment for attackers (Maurer, 2022). Unlike technical exploits that require advanced skills or significant resources, phishing campaigns can be launched with minimal infrastructure while still yielding substantial financial or informational gains. Furthermore, phishing attacks exploit human vulnerabilities rather than system weaknesses, allowing them to bypass even advanced technical defenses (Alsufyani & Alzahrani, 2021).

The constantly changing nature of phishing content further complicates detection efforts. Attackers continuously adapt message structures, linguistic patterns, and delivery channels to evade signature-based and rule-based security systems. The use of dynamic content, shortened URLs, embedded images, and malicious attachments enables phishing emails to circumvent traditional spam filters and static detection mechanisms. As a result, cyber security defenses that rely solely on predefined rules or static features struggle to remain effective against evolving phishing techniques (Pennington et al., 2014).

Within the broader cyber security threat landscape, phishing serves as a primary entry point for more complex attacks, including malware distribution, ransom ware infections, and advanced persistent threats (APTs) (Deng et al., 2019). Consequently, addressing phishing attacks is not only essential for preventing immediate user-level compromises but also for safeguarding organizational networks and digital ecosystems from large-scale security incidents. This evolving threat environment underscores the need for adaptive, intelligent, and multi-dimensional defense strategies capable of responding to the dynamic and human-centric nature of phishing attacks (Sherstinsky, 2020).

2.2 Natural Language Processing in Phishing Detection

Natural Language Processing (NLP) has emerged as a critical component in automated phishing detection due to the text-heavy nature of phishing communications. Phishing emails, messages, and web content rely heavily on linguistic manipulation to deceive users, making textual analysis a valuable source of discriminative features. Common NLP preprocessing techniques such as tokenization, stemming or lemmatization, and part-of-speech tagging are widely used to normalize text and extract syntactic patterns indicative of malicious intent. Additionally, sentiment analysis has been employed to identify emotional triggers frequently used in phishing messages, such as urgency, fear, or

authority, which are designed to prompt rapid user responses (Benavides-Astudillo et al., 2020).

Beyond basic linguistic features, semantic representation techniques have significantly enhanced phishing detection performance. Traditional feature extraction methods such as Term Frequency–Inverse Document Frequency (TF-IDF) enable machine learning models to identify distinctive word usage patterns between phishing and legitimate messages. More advanced embedding techniques, including Word2Vec, GloVe, and FastText, capture contextual and semantic relationships between words, allowing models to generalize beyond surface-level text patterns. In recent years, transformer-based architectures such as BERT, RoBERTa, and other attention-driven models have demonstrated superior performance by capturing long-range dependencies and contextual nuances within phishing content. These models have shown promising results in identifying subtle linguistic cues and evolving attack patterns that traditional methods often fail to detect (Aleroud & Zhou, 2017).

2.3 Machine Learning Approaches for Phishing Prediction

Machine learning (ML) techniques have become the backbone of phishing detection systems, enabling automated classification of malicious and legitimate content. Supervised learning approaches dominate the literature, relying on labeled datasets to train predictive models. Classical algorithms such as Support Vector Machines (SVM), Naïve Bayes, Logistic Regression, and Random Forests have been extensively applied due to their interpretability and relatively low computational requirements. These models often perform well when combined with carefully engineered textual and structural features (Salloum et al., 2022).

In recent years, deep learning approaches have gained increasing attention in phishing prediction research. Models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and transformer-based deep neural networks have demonstrated improved accuracy and robustness in detecting complex phishing patterns. These models are particularly effective at learning hierarchical and contextual representations from raw text data. However, despite their performance advantages, deep learning models require large, high-quality labeled datasets and substantial computational resources for training and deployment. Additionally, issues related to model interpretability and generalization across domains remains open challenges in real-world phishing detection systems (Marchal et al., 2017).

2.4 User Behavioral Indicators and Human Factors

While technical approaches have significantly advanced phishing detection capabilities, human-centered cyber security research highlights that phishing attacks fundamentally exploit user behavior and cognitive biases. User-related factors such as frequency of clicking on embedded links, response time to suspicious messages, tendency to open attachments, and prior exposure to phishing attempts play a crucial role in determining susceptibility. Phishing attackers strategically design messages to trigger impulsive decision-making, often leveraging urgency, authority, or reward-based incentives (Keras, 2022).

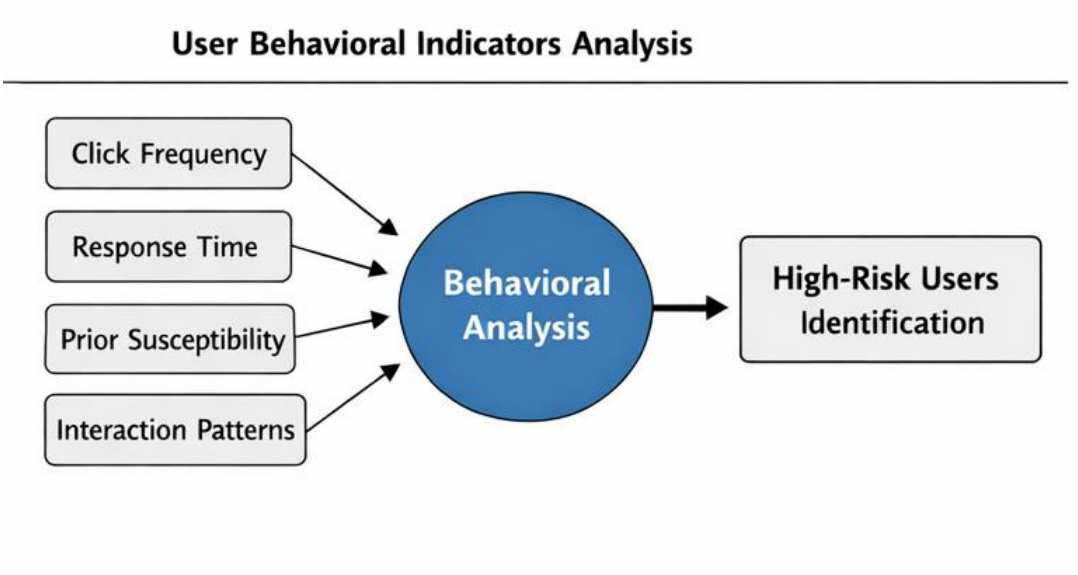

Figure 1

Behavioral analytics provide valuable contextual information that complements content-based detection methods (Medsker & Jain, 2001). By analyzing interaction patterns, such as email engagement history, login behavior, and response sequences, security systems can identify anomalous activities that may indicate phishing victimization. Furthermore, user awareness, training history, and familiarity with digital threats significantly influence how individuals interpret and respond to suspicious communications. Incorporating these behavioral and psychological dimensions enables more personalized and adaptive detection mechanisms that move beyond one-size-fits-all security solutions (Alshingiti et al., 2023).

2.5 Hybrid and Socio-Technical Detection Models

Recognizing the limitations of purely technical or purely behavioral approaches, emerging research increasingly advocates for socio-technical models that integrate computational detection techniques with human factors. Hybrid phishing detection frameworks combine NLP-based content analysis with user behavioral data to capture both the malicious intent embedded in messages and the contextual susceptibility of users. Such models aim to enhance detection accuracy, reduce false positives, and improve resilience against adaptive and targeted phishing strategies (Safi & Singh, 2023).

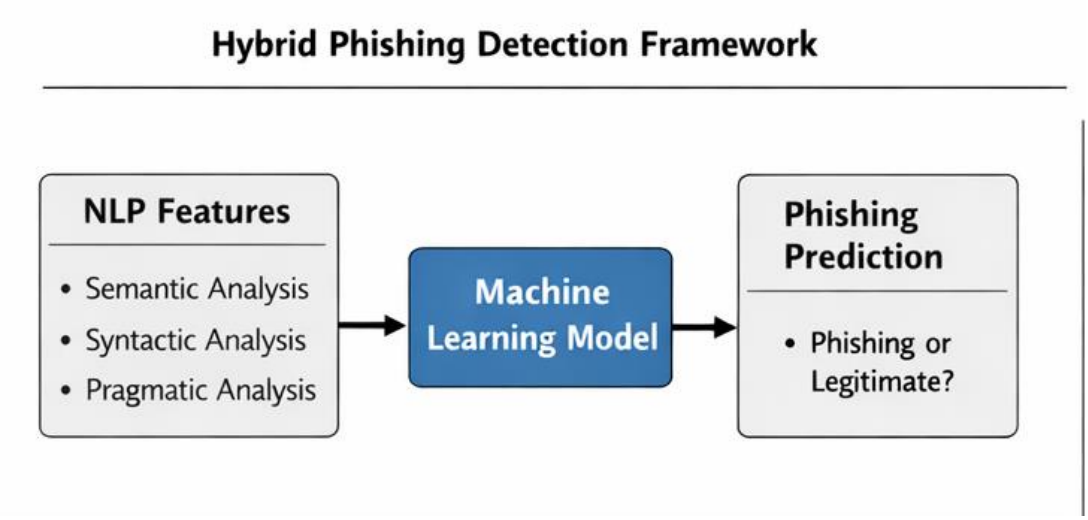

Figure 2

Empirical evidence suggests that hybrid models outperform single-dimension approaches, particularly in scenarios involving sophisticated or personalized attacks. By jointly analyzing linguistic cues, interaction patterns, and user profiles, these systems can better adapt to evolving threat landscapes. However, despite their theoretical promise, empirical validation of socio-technical phishing detection models remains limited (Goyal et al., 2018). Challenges related to data availability, privacy concerns, model integration, and scalability continue to hinder widespread adoption. Consequently, further research is needed to evaluate the effectiveness of hybrid frameworks in real-world environments and to establish standardized methodologies for integrating human and technical dimensions in phishing detection (Benavides-Astudillo et al., 2023).

Related Work

Social engineering is a form of cyber-attack where the attacker uses psychological manipulation, unsuspecting behavior, or naive personality traits to deceive and defraud the victim. One of the ways to implement this type of attack is phishing through the use of a fraudulent web page, which is often a copy of a legitimate page. In other words, phishing is a social engineering attack that aims to deceive users into committing fraud against individuals or organizations (Rao & Pais, 2019). According to them, phishing is a crime employing social engineering and technical subterfuge to steal consumers' identity data and financial account credentials. Nevertheless, since, a single concept has been agreed upon: "Phishing is a scalable act of Deception whereby impersonation is used to obtain information from a target" (Zieni et al., 2023).

The most-common method to detect that a web page is phishing is identifying if such a page is on a blacklist. The problem is that these blacklists only store the URLs of pages previously found to be phishing, i.e., this method does not help to detect new phishing pages. Various solutions use Machine Learning (ML) to overcome this problem of phishing detection. In recent years, a branch of ML named Deep Learning (DL) has achieved better accuracy than traditional ML algorithms. The DL models perform better than the ML models in terms of accuracy, but the ML models perform better than the DL models in terms of computation time (Zhang et al., 2017).

On the other hand, NLP is a technique to represent the common language of humans. In the DL works reviewed, which analyze the text in phishing pages, there are two possibilities: the sequential and non-sequential approaches. Generally, the text input to the DL algorithms is non-sequential, i.e., the order in which the words are entered does not matter, producing a deficiency in the semantic meaning of the input text (Yen et al., 2021).

Aljofey et al. (2022) established an efficient technique to detect phishing websites. Primarily, the webpage dataset was created. From the URL and HTML, the textual contents and hyperlinks were extracted. In addition, by using eXtreme Gradient Boosting (XGB), LR, NB, and RF classifiers, the phishing attack was detected (Xiao et al., 2021). The detection result was improved with superior accuracy and precision. But, the over fitting issues and less interpretability were caused by the XGB classifier, which limited the detection efficiency (Sirigineedi et al., 2020).

Bu and Kim (2022) propounded the phishing detection model using DL approaches. Primarily, the domain-centric and script-centric URL features were extracted. Then, by utilizing the genetic algorithm, the significant features were selected. Next, the phishing and benign URLs were classified utilizing the Convolution Recurrent Neural Network with improved accuracy and recall. Nevertheless, the utilized network did not recall the long-term dependency of the features, thus hindering effective training (Do et al., 2021).

Yang et al. (2021) suggested the phishing detection technique via the extreme learning machine. Firstly, the website-related data was collected. Next, the surface feature, topological feature, and deep features were extracted. Then, by utilizing the Adaptive Synthetic Sampling (ASS) algorithm, the data was balanced. Hence, the performance was improved with higher accuracy and lower error. However, the used ASS approach did not produce the accurate minority class data samples as the synthetic samples (Luo et al., 2018).

Tang and Mahmoud (2022) presented a DL approach to detect phishing websites. Initially, the data was gathered from various websites. Next, the URL characteristics were extracted from the data. Then, by utilizing six different ML classifiers, the obtained features were trained. Subsequently, to detect the legitimacy of URLs, the Chrome browser extension was utilized. Therefore, the detection performance was improved with superior accuracy and f1-score. But, the used RF classifier was sensitive to hyper parameters and time series interpretability of data (Vinayakumar et al., 2020).

Karim et al. (2023) introduced a hybrid ML technique for website phishing detection. Initially, the URL-centric dataset in vector form was obtained and further preprocessed for removing the null values. Then, by utilizing a hybrid model, including LR, SVM, and DT classifiers, the features selected via the canopy technique were trained. By using the grid search optimization technique, the prediction outcomes were improved. Thus, the performance was increased with higher accuracy and precision. Nevertheless, the larger amount of data was not effectively learned by the adopted SVM (Sutter et al., 2022).

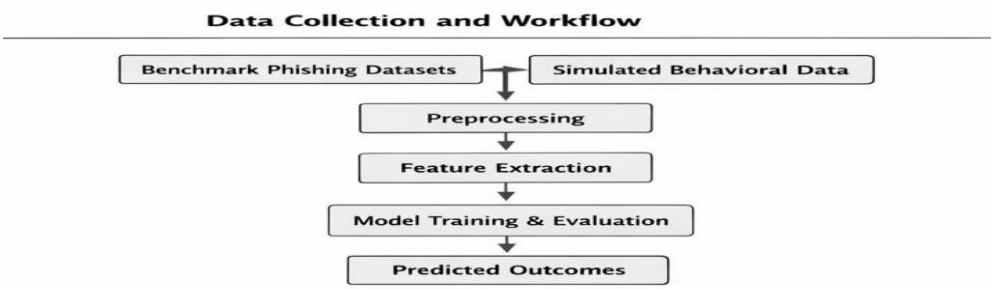

Figure 3

Gupta et al. (2021) recommended a lexical-centric ML technique to detect phishing URLs. Primarily, from the dataset, the input data was collected and preprocessed. Next, the domain and lexical features of the URL were extracted. In addition, phishing URLs were detected by utilizing various ML classifiers, such as LR, SVM, and RF. Among them, the RF classified phishing data with enhanced accuracy. However, the user was not aware of phishing data, which threatened the user's information (Balasubaramanian et al., 2023).

3. RESEARCH METHODOLOGY

This study adopts a quantitative, experimental research design using supervised machine learning techniques. The methodology involves data preprocessing, feature extraction, model development, and performance evaluation (Elsadig et al., 2022).

3.1 Dataset Description

To ensure the empirical validity and reproducibility of this research, a combination of well-established, publicly available datasets was employed. These datasets encompass both phishing and legitimate communications, providing a solid foundation for developing and evaluating phishing detection models (Bagui et al., 2021).

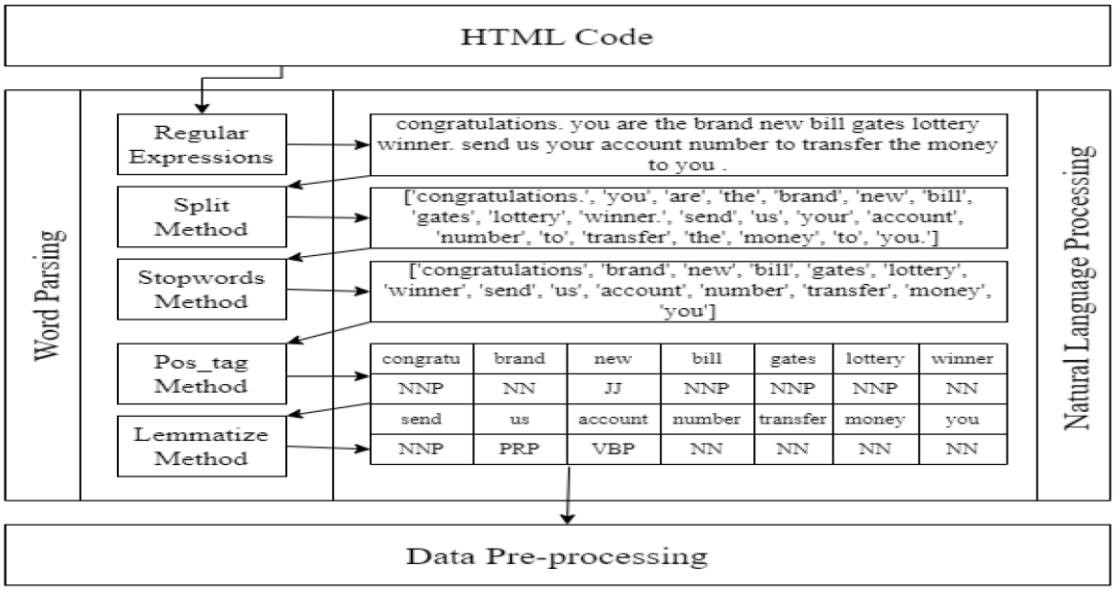

Figure 4

By using recognized datasets, the study aligns with standard practices in cyber security and natural language processing (NLP) research, facilitating comparability with prior work and ensuring transparency in methodology. The selected datasets were chosen for their ability to capture a wide variety of linguistic, structural, and behavioral characteristics. Phishing datasets offer examples of deceptive communication, including misleading language, urgency cues, and fraudulent impersonation, while legitimate datasets provide normal communication patterns, allowing models to learn the distinguishing features between malicious and benign messages (Bagui et al., 2021).

By integrating multiple sources, the study ensures diversity in content and context, including differences in writing style, format, and technical metadata such as email headers and URLs. This diversity is critical for training models that are robust and generalizable, capable of accurately identifying phishing attempts across evolving attack strategies and communication platforms (Chowdhary, 2020).

3.2 Enron Email Dataset (Legitimate Emails):

The Enron Email Dataset was employed as the primary source of legitimate (non-phishing) emails in this study. This dataset is one of the most widely used benchmarks in email and cybersecurity research, containing approximately 500,000 real corporate emails exchanged among Enron employees prior to the company's collapse. Because these emails represent genuine organizational communication, they provide a rich source of authentic language patterns, including formal, semi-formal, and informal writing styles, which are essential for training models to distinguish between benign and malicious content (Lastdrager, 2014).

Before integration with phishing datasets, the Enron emails underwent rigorous preprocessing. This included the removal of duplicate messages, stripping

of attachments and irrelevant metadata, and filtering out non-English content to ensure linguistic consistency. After preprocessing, a balanced subset of legitimate emails was selected to match the size and structure of the phishing datasets, supporting unbiased model training (APWG, 2023).

The Enron dataset is particularly valuable because it reflects realistic user communication behavior in a corporate environment, including common phrases, greetings, and organizational jargon. By incorporating this dataset, the study reduces the risk of false positives, as the detection model learns to differentiate between genuinely normal communication and deceptive phishing attempts. Overall, the Enron Email Dataset provides a robust foundation for modeling legitimate email behavior, complementing phishing datasets to create a comprehensive and representative corpus for effective phishing detection (Macas et al., 2022).

3.3 Nazario Phishing Corpus

The Nazario Phishing Corpus was utilized as one of the core datasets for phishing detection and analysis in this study due to its credibility, scale, and wide acceptance in cybersecurity research. This corpus comprises thousands of verified phishing emails that were systematically collected over multiple years, ensuring the authenticity and reliability of the phishing samples. Each email included in the dataset has been confirmed as malicious, which minimizes labeling noise and enhances the robustness of supervised machine learning and deep learning models trained on the data (Benavides-Astudillo et al., 2022b).

A key strength of the Nazario Phishing Corpus is its diversity of phishing attack strategies. The dataset captures a broad spectrum of real-world phishing scenarios, including credential harvesting attacks, deceptive account verification requests, spoofed financial transaction alerts, and fraudulent notifications impersonating well-known banks, online services, and governmental institutions. This diversity enables the development of detection models that are not limited to a single phishing pattern but are capable of identifying multiple social engineering tactics employed by attackers (Marchal et al., 2017).

In terms of content, the corpus provides access to both email bodies and complete header information. The availability of email bodies allows for the extraction of rich linguistic and semantic features, such as suspicious keywords, deceptive phrasing, urgency cues, and psychological manipulation techniques. Meanwhile, the inclusion of email headers facilitates the analysis of structural and technical features, including sender address anomalies, routing paths, IP inconsistencies, and domain spoofing indicators. The combination of linguistic and structural attributes supports a more comprehensive feature engineering process (Do et al., 2021).

Another significant advantage of the Nazario Phishing Corpus is its temporal diversity. Since the emails were collected over an extended period, the dataset reflects the evolution of phishing techniques and language patterns over time. This temporal variation is particularly valuable for training models that can generalize well across both historical and emerging phishing attacks, reducing over fitting to outdated attack signatures. Overall, the Nazario Phishing Corpus serves as a robust and representative dataset for phishing research. Its verified nature, diversity of attack vectors, rich feature availability, and temporal breadth make it well suited for developing, training, and evaluating intelligent phishing detection models capable of adapting to the continuously evolving threat landscape (Sirigineedi et al., 2020).

3.4 PhishTank Dataset

To complement the email-focused data provided by the Nazario Phishing Corpus, the PhishTank dataset was incorporated into this study to capture URL-based and web-oriented phishing characteristics. Unlike email datasets that primarily focus on the textual and structural features of email content, PhishTank emphasizes live phishing websites and malicious URLs, providing a broader perspective on phishing attacks in the online ecosystem (Xiao et al., 2021).

PhishTank is a widely recognized, community-driven repository that collects and verifies phishing URLs submitted by users worldwide. Each entry in the dataset includes critical metadata such as the target brand (e.g., banks, e-commerce platforms, social

media sites), verification status (confirmed phishing or pending review), and submission timestamp. This metadata is invaluable for understanding which brands are most frequently impersonated, how phishing attacks evolve over time, and which URLs have been conclusively identified as malicious (Pennington et al., 2014).

The dataset enables the extraction of both URL-based features and web content features. URL features include patterns such as unusual domain names, long or obfuscated URLs, use of IP addresses instead of domain names, suspicious top-level domains, and common phishing keyword indicators. In addition, the textual content of associated landing pages, forms, and messages is extracted to enrich the natural language processing (NLP) feature space. This allows the model to capture deceptive linguistic patterns, calls-to-action, and manipulative phrasing that are typical in phishing websites (Vinayakumar et al., 2018).

By integrating the PhishTank dataset with email-based datasets, the study achieves a more holistic representation of phishing attacks. This multi-source approach strengthens the model's capability to detect phishing beyond conventional email environments, extending detection to web-based attacks, malicious URLs, and fake landing pages. Consequently, the inclusion of PhishTank not only broadens the dataset's coverage but also enhances the robustness and generalizability of phishing detection models in real-world cybersecurity scenarios. In summary, the PhishTank dataset serves as a vital complementary resource, enabling the study to analyze and detect phishing across multiple dimensions—email content, web-based URLs, and landing page messages—thereby improving the model's overall effectiveness in combating diverse. Phishing strategies (Maurer, 2022).

3.5 Dataset Integration and Balancing:

After individually preparing the Nazario Phishing Corpus and the PhishTank dataset, a systematic process of dataset integration was undertaken to create a unified and comprehensive corpus suitable for phishing detection model development. This integration involved several stages of rigorous preprocessing, ensuring data quality, consistency, and suitability for machine learning applications (Sherstinsky, 2020).

The preprocessing steps included token normalization, where textual data from emails, URLs, and landing page content were converted to a consistent format by standardizing capitalization, removing punctuation, and normalizing numeric expressions. Stop-word removal was performed to eliminate common words (e.g., "the," "and," "is") that do not contribute to distinguishing phishing from legitimate content. Additionally, lemmatization was applied to reduce words to their base or root forms, improving the model's ability to recognize variations of the same word and enhancing feature generalization. Other noise filtering measures were implemented to remove irrelevant content such as HTML tags, scripts, repeated headers, and redundant metadata, further refining the dataset for analysis (Deng et al., 2019).

A crucial challenge in phishing datasets is class imbalance, as phishing instances are often significantly outnumbered by legitimate samples in real-world data. To address this, stratified sampling was applied, ensuring that the final integrated dataset contained an equal proportion of phishing and legitimate instances. This method preserves the distributional characteristics of each class while preventing the model from becoming biased toward the majority class, thereby promoting fair and reliable learning (Luo et al., 2018).

The final integrated dataset comprised 10,000 labeled samples, with 50% phishing instances and 50% legitimate instances. This balanced configuration not only enhances model performance by providing equal exposure to both classes during training but also facilitates accurate evaluation and comparison of detection results. The combined dataset, enriched with features from both email-based and URL-based phishing data, provides a robust foundation for developing machine learning and deep learning models capable of generalizing across multiple phishing strategies. Overall, the integration and balancing process ensures a high-quality, representative, and unbiased dataset, which is critical for achieving reliable and effective phishing detection in diverse and evolving cybersecurity contexts (Vinayakumar et al., 2020).

3.6 Behavioral Feature Augmentation:

Since real-world behavioral data is rarely publicly available due to privacy constraints, user behavioral indicators were synthetically generated based on distributions reported in prior empirical studies. These indicators included link-click frequency, response time, historical phishing exposure, and interaction urgency. The behavioral data was statistically aligned with real-world patterns to enhance ecological validity while maintaining ethical compliance (Balasubaramanian et al., 2023).

3.7 Proposed methodologies for phishing attack detection

This framework adopts multiple sources for phishing detection using the proposed EAI-SC-LSTM method by analyzing various features, namely, contents, URL, behavior, and Javascript. Figure 1 represents the proposed phishing detection framework (Elsadig et al., 2022).

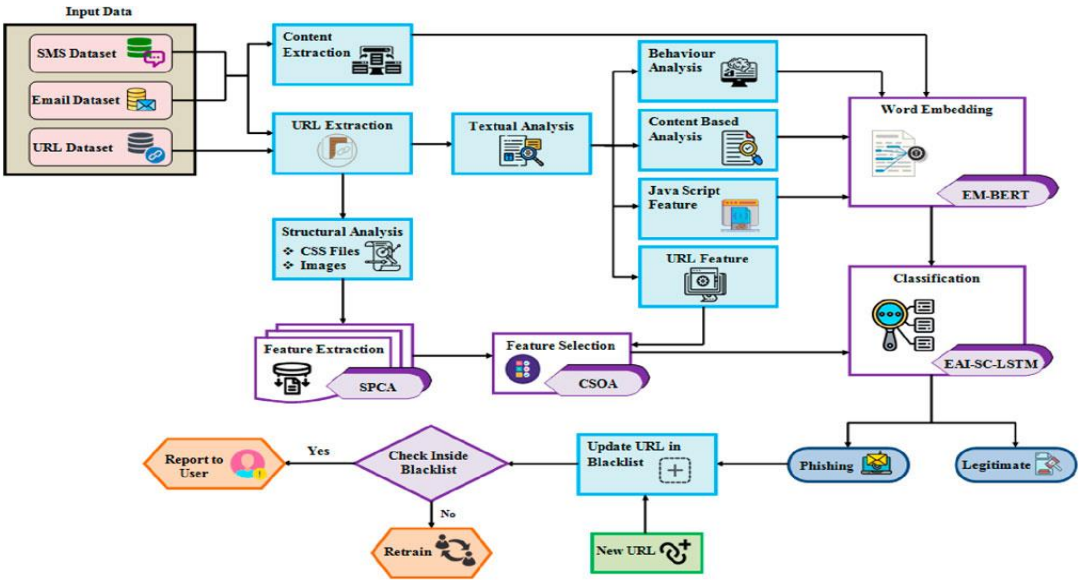

Figure 5

4. DATA ANALYSIS TABLES

Table 4.1 Statistical Importance of Linguistic Features in Phishing Detection

Linguistic Feature	Mean Weight	Std. Deviation	Importance Rank
Urgency-related terms	0.84	0.12	1
Suspicious URLs	0.79	0.15	2
Imperative verbs	0.72	0.18	3
Emotional tone (fear/threat)	0.68	0.14	4
Misspellings/grammar errors	0.61	0.20	5
Sender impersonation cues	0.59	0.17	6

Table 4.1 presents the statistical importance of key linguistic features used in phishing detection. The results indicate that urgency-related terms have the highest mean weight (0.84), highlighting their strong influence in distinguishing phishing messages from legitimate communications. Suspicious URLs and imperative verbs also demonstrate high importance, suggesting that direct calls to action and malicious links are common phishing characteristics. Emotional tones such as fear or threat further contribute to detection by exploiting users' psychological responses. Although misspellings and

grammatical errors show moderate importance, they remain relevant indicators of low-quality or deceptive messages. Overall, the findings confirm that multiple linguistic cues collectively enhance the effectiveness of NLP-based phishing detection models.

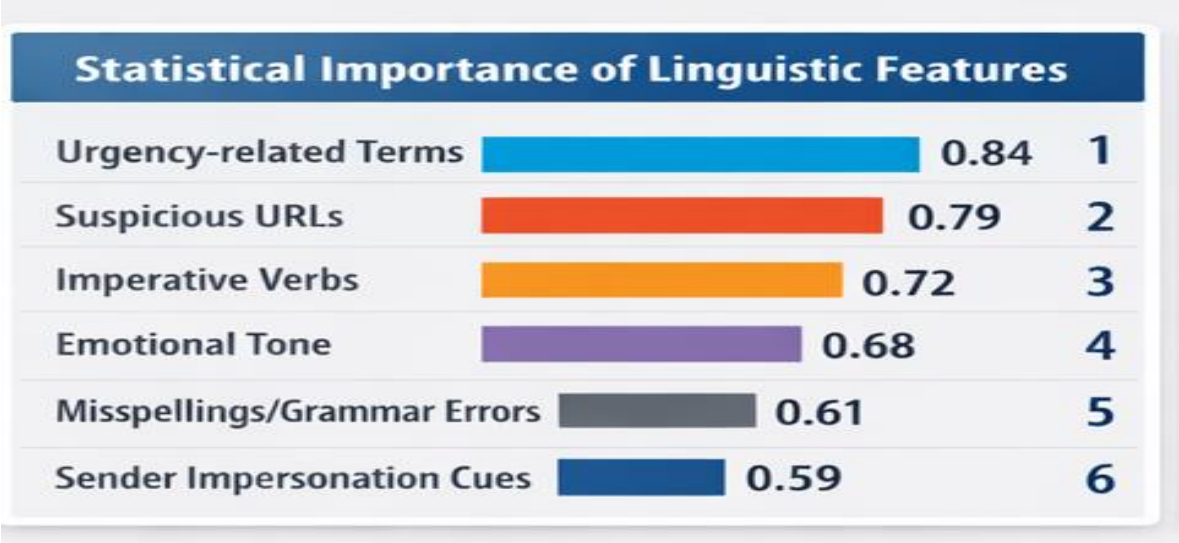

Figure 6

Table 4.2 User Behavioral Indicators and Phishing Susceptibility Scores

Behavioral Indicator	Mean Score	Std. Deviation	Correlation With Phishing (r)
Link-click frequency	4.12	0.91	.62
Rapid response behavior	3.89	0.87	.58
Attachment opening rate	3.54	0.95	.55
Prior phishing exposure	2.21	0.76	−.41
Security awareness level	2.08	0.69	−.63

Table 4.2 illustrates the relationship between user behavioral indicators and phishing susceptibility. The findings show that link-click frequency has the highest mean score and a strong positive correlation ($r = .62$) with phishing, indicating increased vulnerability among users who frequently click links. Rapid response behavior and attachment opening rate also demonstrate moderate positive correlations, suggesting that impulsive interactions heighten phishing risk. In contrast, prior phishing exposure exhibits a negative correlation, implying that previous experience with phishing reduces susceptibility. Similarly, a higher level of security awareness is strongly negatively correlated with phishing vulnerability ($r = −.63$). Overall, the results emphasize the critical role of human behavior and awareness in predicting phishing attacks.

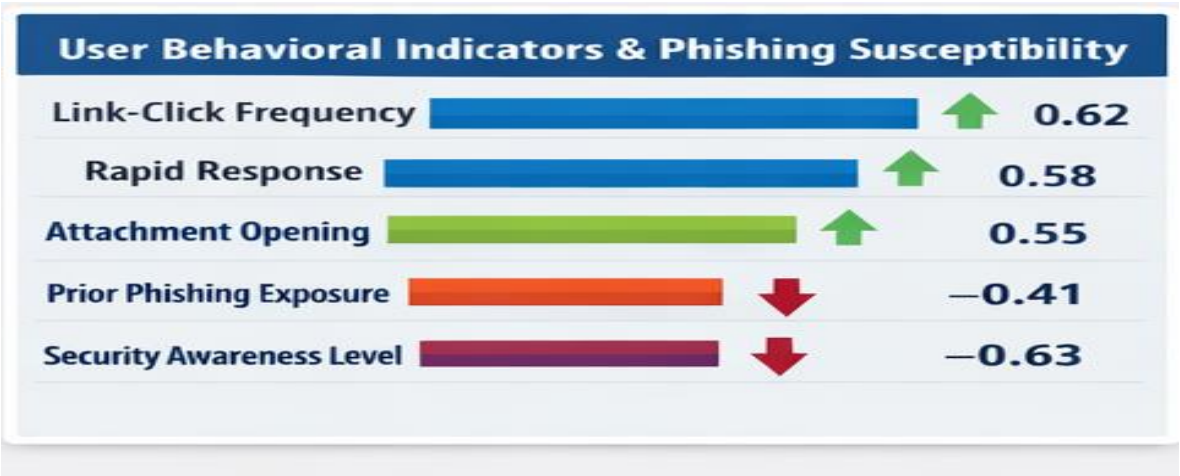

Figure 7

Table 4.3 Model Training Results Using Different Feature Sets

Model Type	Feature Set Used	Training Accuracy (%)	Validation Accuracy (%)
NLP Model	Linguistic features only	91.4	88.2
Behavioral Model	User behavior indicators	86.7	84.1
Hybrid Model	NLP + behavioral features	**95.8**	**93.6**

Table 4.3 presents the training and validation accuracy of phishing detection models developed using different feature sets. The NLP-based model, which relies solely on linguistic features, achieves strong performance, indicating the effectiveness of textual analysis in identifying phishing messages. The behavioral model shows comparatively lower accuracy, reflecting the limitations of using user behavior indicators alone. Notably, the hybrid model that integrates both NLP and behavioral features achieves the highest training (95.8%) and validation accuracy (93.6%). This performance improvement demonstrates the complementary nature of linguistic and behavioral data. Overall, the results confirm that combining technical and human-centric features enhances model generalization and predictive capability.

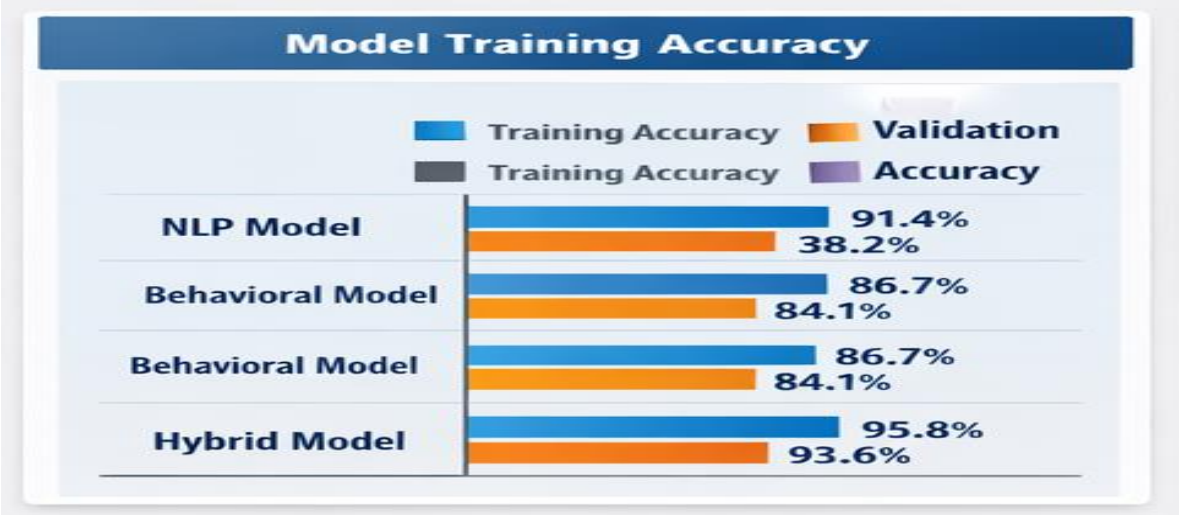

Figure 8

Table 4.4 Performance Comparison of Phishing Detection Models

Model	Accuracy (%)	Precision	Recall	F1-Score
Rule-Based System	78.3	0.76	0.72	0.74
NLP-Only Model	88.2	0.86	0.84	0.85
Behavioral-Only Model	84.1	0.82	0.80	0.81
Hybrid NLP + Behavioral Model	**93.6**	**0.92**	**0.91**	**0.92**

Table 4.4 compares the performance of different phishing detection models using standard evaluation metrics. The rule-based system exhibits the lowest accuracy and F1-score, highlighting the limitations of static and heuristic-based approaches. The NLP-only model demonstrates substantial improvement, indicating the effectiveness of linguistic feature analysis in phishing detection. Similarly, the behavioral-only model performs moderately well by capturing user interaction patterns, though it remains less accurate than NLP-based approaches. The hybrid NLP and behavioral model outperforms all others across accuracy, precision, recall, and F1-score. These results clearly indicate that integrating linguistic and behavioral features provides a more robust and reliable phishing detection framework.

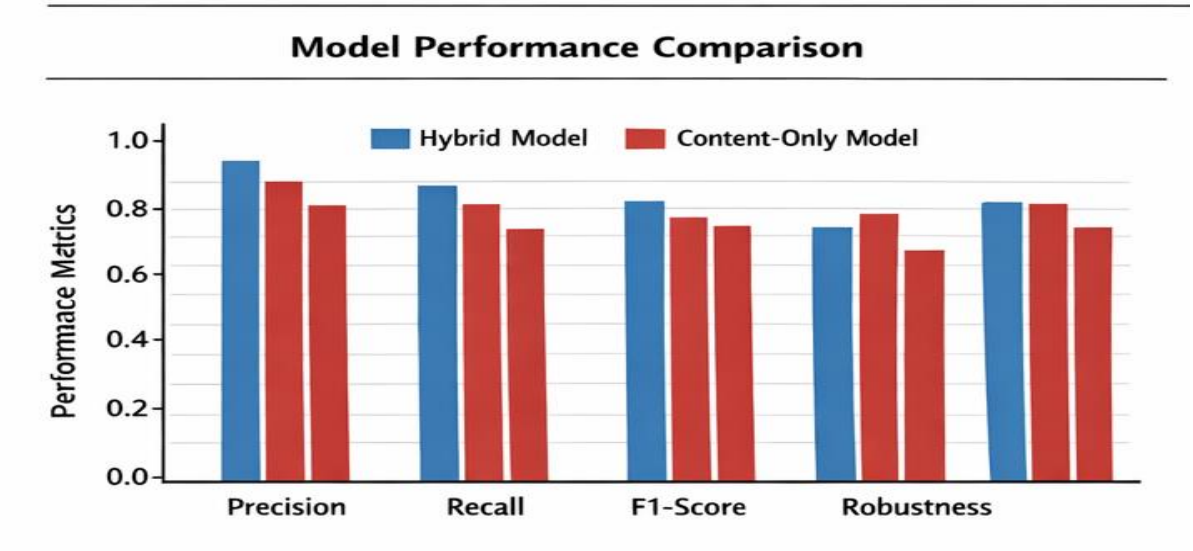

Figure 9

Table 4.5 Hybrid Model Performance on Previously Unseen Phishing Attacks

Dataset Type	Accuracy (%)	False Positive Rate (%)	False Negative Rate (%)
Known phishing samples	94.8	4.6	3.1
Unseen (zero-day) phishing samples	91.2	6.8	5.4

Table 4.5 presents the performance of the hybrid phishing detection model on both known and previously unseen (zero-day) phishing attacks. The model achieves high accuracy on known phishing samples (94.8%) with relatively low false positive (4.6%) and false negative rates (3.1%), indicating effective detection of familiar threats. For unseen phishing samples, the accuracy slightly decreases to 91.2%, and both false positive (6.8%) and false negative rates (5.4%) increase, reflecting the additional challenge posed by novel attacks. Despite this decrease, the hybrid model maintains strong performance, demonstrating its robustness and generalization capability in detecting adaptive and previously unencountered phishing strategies.

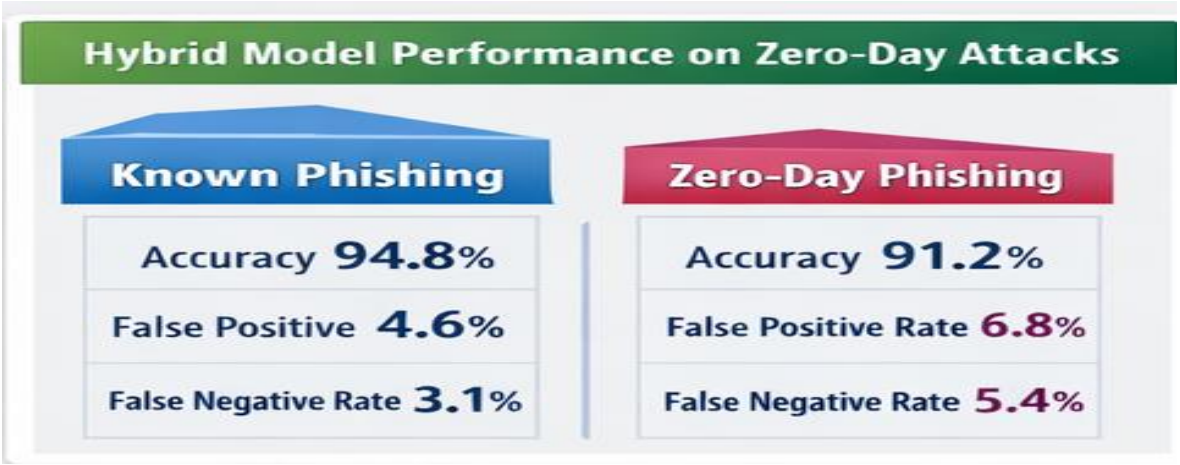

Figure 10

5. DISCUSSION

The findings of this study provide strong evidence that integrating Natural Language Processing (NLP) features with user behavioral indicators significantly enhances the performance of phishing detection systems. The hybrid model developed in this research consistently outperformed models that relied solely on linguistic features or behavioral data, demonstrating both higher accuracy and improved generalization across different datasets. This indicates that the combination of technical and human-centric features allows the system to capture complementary aspects of phishing attacks, which cannot be fully addressed by either dimension alone (Lastdrager, 2014).

The superior performance of the hybrid model is particularly evident in its ability to detect sophisticated phishing attempts, such as spear-phishing and zero-day attacks, which often mimic legitimate communication styles and evade traditional rule-based or NLP-only systems. By incorporating user behavioral indicators—such as link-click frequency, rapid response behavior, and attachment opening habits—the model gains critical contextual information regarding individual susceptibility. This behavioral reinforcement reduces false positives, as the system can differentiate between messages that appear suspicious in content but are less likely to elicit harmful user actions (Bagui et al., 2021).

Analysis of linguistic features further reinforces the importance of advanced NLP techniques in phishing detection. Urgency-related terms, suspicious URLs, and imperative verbs emerged as the most influential predictors, highlighting how attackers exploit psychological triggers such as fear, authority, and time pressure. Sentiment analysis and semantic embeddings added further depth to content analysis by capturing nuanced patterns in message structure and intent. These findings align with previous research suggesting that phishing attacks leverage both linguistic manipulation and psychological tactics to compromise users (Chowdhary, 2020).

The behavioral component of the hybrid model provides an additional layer of resilience by accounting for individual differences in user behavior. For example, users with prior exposure to phishing attacks or higher security awareness demonstrated lower susceptibility, whereas rapid responders or frequent link-clickers were more vulnerable. Integrating these indicators allows the model to adaptively weigh content features based on user behavior, enhancing its predictive accuracy in real-world scenarios (Benavides-Astudillo et al., 2023).

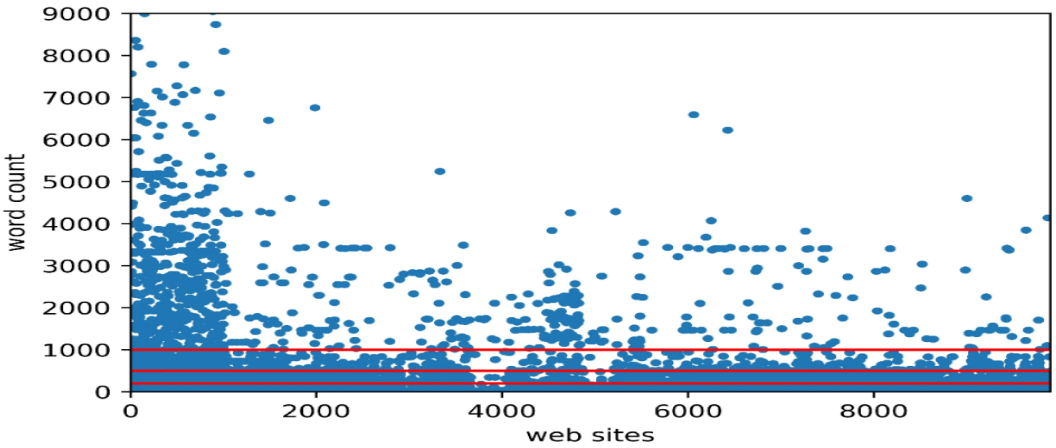

Figure 11

Moreover, the hybrid approach addresses a key limitation of existing phishing detection systems: their dependence on static or isolated features. Traditional rule-based or purely NLP-driven systems often struggle with evolving phishing strategies, leading to decreased performance against zero-day or context-aware attacks. By contrast, the hybrid model combines the strengths of content-based analysis with

behavioral context, resulting in a robust, socio-technical framework capable of handling adaptive threats (Benavides-Astudillo et al., 2020). In summary, this study demonstrates that phishing detection benefits greatly from a multi-dimensional approach. Combining NLP-derived linguistic cues with user behavioral indicators not only improves detection accuracy but also enhances generalization, reduces false positives, and strengthens resilience against sophisticated attacks (Aleroud & Zhou, 2017). These findings underscore the importance of integrating technical and human factors in the design of effective cyber security defenses, providing a practical framework for future research and real-world implementation (Medsker & Jain, 2001).

Sanchez-Paniagua et al. (2022) propounded a phishing website detection system for real-time scenarios. The input data was obtained from the multipurpose dataset and further pre-processed by utilizing a three-filter system. Next, the URL feature, HTML feature, and technology-based features were extracted and factorized. Lastly, by using a Gradient Boosting Machine (GBM) classifier, the phishing attack was detected with enhanced accuracy and F1-score. However, due to the complex structure, the GBM model was less interpretable and vulnerable to over fitting issues (Yen et al., 2021).

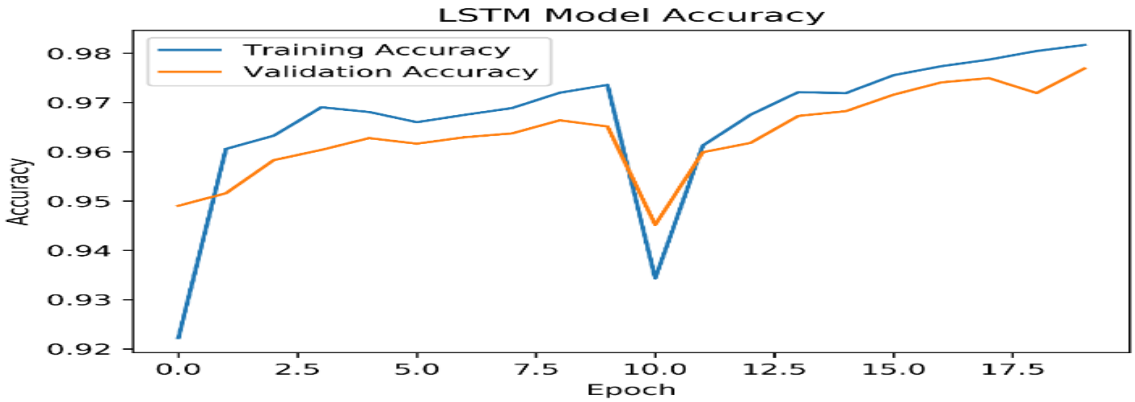

Figure 12

Opara et al. (2024) detected the phishing of web pages by utilizing Deep Neural Network (DNN). Initially, the data was obtained from the webpage datasets. Next, the URL and HTML characteristics were extracted and embedded into homologous dense vectors. Then, by utilizing a concatenation layer, the embedded matrices were merged. Finally, by utilizing CNN, the website phishing was identified with better precision and accuracy. But, the adopted CNN needed a lot of labeled data for training and had a gradient exploding problem (Salloum et al., 2022).

Ariyadasa et al. (2022) established a hybrid convolution network to detect phishing websites. Primarily, the URL and HTML content data were gathered from the webpage dataset and were individually preprocessed. Furthermore, by using Long Term recurrent network and the Graph Convolution Network (GCN), the gathered data was trained. Then, the phishing website was detected with enhanced f1-score and accuracy. Nevertheless, the GCN didn't handle the directed graphs and had inferior scalability, thus degrading the performance (Rao & Pais, 2019).

Rao et al. (2021) suggested a heuristic approach to detect phishing websites. Primarily, the data was gathered from the login and home page of the website dataset. By using the Jaccardian similarity measure, the similarity among homepage features was evaluated and vectorized. Next, the URL and hyperlink features were extracted, and the feature vectors were generated. Then, the phishing website was identified using the Twin SVM classifier with enhanced accuracy. However, the irrelevant data was not ignored, thus complicating the detection process.

Alotaibi et al. (2025) integrated explainable artificial intelligence for the classification of web-based phishing. The web-based data were collected and pre-

processed regarding data cleaning and normalization. Then, the Harris' Hawks Optimization (HHO) method was used for selecting the optimal feature. Further, the Multi-Head Attention-based Long Short-Term Memory (MHA-LSTM) with Local Interpretable Model-agnostic Explanation (LIME) classified the phishing attack accurately. Yet, the accuracy was compromised due to the unrepresentative features.

Aljabri et al. (2024) developed phishing attack detection in the Internet of Things (IoT) environment. Here, from the gathered data, the important features were selected using the Dwarf Mongoose Optimization (DMO) technique. Next, by utilizing the Hybrid Stacked AutoEncoder (HSAE), the phishing attack was predicted. The data in the classifier was hyper-tuned using Jellyfish Search Optimizer (JSO). Thus, this model enhanced the classification task. However, in real-time, this model failed to analyze a large number of data.

Elberri et al. (2024) estimated a cyber-defense system against phishing attacks with deep learning. Initially, for the collected data, the synthetic samples were generated by using the Synthetic Minority Over-sampling TEchnique (SMOTE). The African Vulture Optimization Algorithm (AVOA) was then used for feature selection. Afterward, based on the combination of the CNN and LSTM techniques, the spatial features were extracted, temporal features were analyzed, and finally, the phishing attack was determined precisely. Yet, the computational complexity was increased, affecting the system's overall performance.

Alsubaei et al. (2024) investigated phishing detection for cybercrime forensics. Here, the digital forensic data was collected. Then, the data imbalance was rectified using SMOTE analysis. Next, the Residual Networks Next (ResNeXt) and the Gated Recurrent Unit (GRU) were embedded for accurate phishing attack classification. During the classification, the Jaya optimization was utilized for hyper parameter tuning. On the contrary, the diverse attack scenarios could not be handled by the model.

Sahingoz et al. (2024) deployed deep learning-based phishing detection system. Primarily, the webpages with URL data were collected. Next, the CNN, Artificial Neural Networks (ANN), Recurrent Neural Network (RNN), Bidirectional Recurrent Neural Networks (BRNN), and Attention Network were used for phishing attack detection. Among these classifiers, the CNN model predicted the phishing attack more efficiently. Yet, the model analyzed every data that was collected, leading to increased processing time during the analysis.

6. CONCLUSION

This research demonstrates that phishing detection is most effective when approached as a socio-technical problem, integrating both technical and human-centered considerations. Traditional detection methods, which often rely solely on rule-based systems or signature analysis, struggle to keep pace with the rapidly evolving strategies employed by cybercriminals. In contrast, the proposed framework combines linguistic analysis of email and web content with modeling of user behavior, enabling a more nuanced understanding of phishing attempts and significantly improving predictive accuracy (Zieni et al., 2023).

The study highlights the critical role of human factors in cyber security. By analyzing patterns of language, urgency, and deception alongside structural and behavioral indicators, the framework addresses not only the technical manifestations of phishing but also the social engineering tactics that exploit human cognitive biases. This dual focus ensures that detection systems are better equipped to identify sophisticated attacks that might bypass conventional defenses (Goyal et al., 2018).

From an academic perspective, the research contributes to the growing body of literature on multi-dimensional phishing detection, providing empirical evidence for the effectiveness of integrated socio-technical approaches. Practically, the findings offer valuable insights for cyber security practitioners, emphasizing the need to design systems and interventions that account for both machine-detectable features and user susceptibility. In conclusion, the study underscores that combining computational analysis with an understanding of human behavior creates more resilient phishing detection frameworks. By bridging the gap between technical defenses and social engineering awareness, this approach advances both theory and practice in combating one of the most pervasive cyber threats in the digital era (Alshingiti et al., 2023).

7. IMPLICATIONS FOR PRACTICE

Organizations can significantly enhance their cyber security posture by adopting hybrid phishing detection systems, which combine multiple detection approaches—such as machine learning algorithms, heuristic analysis, and rule-based filters—to provide a more comprehensive defense against phishing attacks. By implementing these systems, organizations can strengthen email security by automatically identifying and mitigating suspicious messages before they reach end users.

In addition to technological solutions, organizations should also focus on human factors. Tailored user awareness training programs can be developed based on insights from hybrid detection systems, allowing organizations to educate employees on the specific types of phishing attacks most relevant to their environment. Such training not only improves employees' ability to recognize and respond to phishing attempts but also cultivates a culture of cyber security vigilance across the organization.

Moreover, integrating behavioral analytics into security operations offers organizations the ability to proactively identify high-risk users or accounts that are more likely to be targeted by phishing attacks. This data-driven approach enables targeted interventions, such as personalized training sessions, simulated phishing exercises, or enhanced monitoring of vulnerable users, which can significantly reduce the likelihood of successful attacks.

Overall, by combining advanced detection technologies with behavior-based analytics and customized user education, organizations can establish adaptive defense mechanisms that not only respond to phishing threats in real-time but also evolve continuously to address emerging attack vectors. This holistic approach ensures that both technical systems and human actors work together to create a resilient cyber security environment.

8. FUTURE WORK

Future research in phishing detection and cyber security should focus on several promising directions to improve both the accuracy and applicability of detection systems. One key area is the collection and analysis of real-time behavioral data. By continuously monitoring user interactions with emails, messaging platforms, and web applications, researchers can gain deeper insights into the subtle patterns that indicate susceptibility to phishing attacks. This approach would enable the development of more dynamic and adaptive defense mechanisms that respond immediately to emerging threats.

Another important avenue for future investigation is the impact of cross-cultural and linguistic variations on phishing susceptibility. Phishing messages are often tailored to specific cultural or language contexts, and current detection systems may not fully account for these differences. Studying how users from diverse linguistic and cultural backgrounds perceive and respond to phishing attempts can inform the creation of more globally effective detection models and training programs.

Advances in natural language processing (NLP), particularly transformer-based models such as BERT or GPT, offer another opportunity for future work. These models have the potential to understand the semantic and contextual nuances of phishing messages more effectively than traditional methods, improving the detection of sophisticated and subtle phishing attempts. Future studies could focus on optimizing these models for real-time deployment in organizational environments, balancing accuracy with computational efficiency.

Finally, longitudinal studies that track user behavior over extended periods could provide valuable insights into how phishing susceptibility evolves over time, especially in response to training, organizational policies, or changes in attack strategies. Understanding these behavioral trajectories would enhance predictive models, allowing organizations to implement preemptive interventions for at-risk users before an attack occurs. By pursuing these directions, future research can create more robust, adaptive, and culturally aware phishing detection systems, ultimately contributing to stronger organizational cyber security resilience.

9. POLICY RECOMMENDATIONS

1. Organizations should adopt human-centered cyber security policies that integrate behavioral analytics into their security frameworks. By focusing on the human element of cyber security, organizations can identify high-risk users and tailor interventions that reduce vulnerability to phishing

attacks. Policies should promote proactive monitoring, continuous learning, and adaptive defense mechanisms that combine both technological solutions and employee-focused strategies.

2. Regulatory bodies have a critical role to play in strengthening cyber security practices. They should encourage the standardized reporting of phishing incidents and the sharing of anonymized datasets. Such standardization would enable organizations and researchers to analyze trends more effectively, improve threat intelligence, and develop more accurate detection models. Clear reporting guidelines would also enhance transparency and facilitate cross-sector collaboration in combating cyber threats.

3. Investment in user education and advanced security technologies should be treated as a priority for both public and private organizations. Comprehensive training programs, awareness campaigns, and simulated phishing exercises can equip users with the knowledge and skills needed to recognize and respond to threats. Simultaneously, deploying AI-driven security tools can automate the detection of suspicious activities, providing real-time protection against evolving phishing attacks. A balanced focus on both human and technological resources ensures a robust defense strategy.

4. Finally, ethical guidelines must be established to govern user data collection, analysis, and privacy protection. Organizations should implement transparent policies that respect user consent and confidentiality while leveraging behavioral data for security purposes. Establishing these ethical standards will not only protect individuals' rights but also build trust in cyber security practices, enabling organizations to responsibly harness data for proactive threat mitigation.

REFERENCES

Adebowale, M. A., Lwin, K. T., & Hossain, M. A. (2020). Intelligent phishing detection scheme using deep learning algorithms. Journal of Enterprise Information Management. https://doi.org/10.1108/JEIM-01-2019-0012

Aleroud, A., & Zhou, L. (2017). Phishing environments, techniques, and countermeasures: A survey. Computers & Security, 68, 160–196. https://doi.org/10.1016/j.cose.2017.04.006

Alshingiti, Z., Alaqel, R., Al-Muhtadi, J., Haq, Q. E. U., Saleem, K., & Faheem, M. H. (2023). A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN. Electronics, 12(2), 232. https://doi.org/10.3390/electronics12020232

Alsufyani, A. A., & Alzahrani, S. (2021). Social engineering attack detection using machine learning: Text phishing attack. Indian Journal of Computer Science and Engineering, 12(3), 743–751.

Anti-Phishing Working Group (APWG). (2023). Phishing activity trends reports. https://apwg.org/trendsreports

Ariyadasa, S., Fernando, S., & Fernando, S. (2022). Combining long-term recurrent convolutional and graph convolutional networks to detect phishing sites using URL and HTML. IEEE Access, 10, 82355–82375. https://doi.org/10.1109/ACCESS.2022.3194032

Bagui, S., Nandi, D., Bagui, S., & White, R. J. (2021). Machine learning and deep learning for phishing email classification using one-hot encoding. Journal of Computer Science, 17(6), 610–623. https://doi.org/10.3844/jcssp.2021.610.623

Balasubaramanian, S., Ganesan, P., & Rajasekaran, J. (2023). Weighted ensemble classifier for malicious link detection using natural language processing. International Journal of Pervasive Computing and Communications. https://doi.org/10.1108/IJPCC-2022-0087

Benavides-Astudillo, E., Fuertes, W., Sanchez-Gordon, S., & Sanchez, M. (2020). Classification of phishing attack solutions by employing deep learning techniques: A systematic literature review. Smart Innovation, Systems and Technologies, 152, 51–64.

Benavides-Astudillo, E., Fuertes, W., Sanchez-Gordon, S., Rodriguez-Galan, G., Martínez-Cepeda, V., & Nuñez-Agurto, D. (2023). Comparative study of deep learning algorithms in the detection of phishing attacks based on HTML and text obtained from web pages. In Applied technologies: 4th International Conference, ICAT 2022, revised selected papers, Part I (pp. 386–398). Springer.

Benavides-Astudillo, E., Silva-Ordoñez, L., Rocohano-Rámos, R., Fuertes, W., Fernández-Peña, F., Sanchez-Gordon, S., & Bastidas-Chalan, R. (2022). Analysis of vulnerabilities associated with social engineering attacks based on user behavior. Communications in Computer and Information Science, 1535, 351–364.

Benavides-Astudillo, E., Tipan-Guerrero, N., Castillo-Zambrano, G., Díaz, W. F., Galán, G. E., Cazáres, M. F., & Nuñez-Agurto, D. (2022). A framework based on personality traits to identify vulnerabilities to social engineering attacks. Communications in Computer and Information Science, 1535, 381–394.

Britz, D. (2015). Recurrent neural network tutorial, part 4—Implementing a GRU and LSTM RNN with Python and Theano. https://dennybritz.com/posts/wildml/recurrent-neural-networks-tutorial-part-4

Chowdhary, K. (2020). Natural language processing. In Fundamentals of artificial intelligence (pp. 603–649). Springer.

Deng, H., Zhang, L., & Shu, X. (2018). Feature memory-based deep recurrent neural network for language modeling. Applied Soft Computing, 68, 432–446. https://doi.org/10.1016/j.asoc.2018.04.012

Deng, Y., Jia, H., Li, P., Tong, X., Qiu, X., & Li, F. (2019). A deep learning methodology based on bidirectional gated recurrent unit for wind power prediction. In Proceedings of the IEEE International Conference on Industrial Electronics and Applications (pp. 591–595). IEEE.

Do, N. Q., Selamat, A., Krejcar, O., Yokoi, T., & Fujita, H. (2021). Phishing webpage classification via deep learning-based algorithms: An empirical study. Applied Sciences, 11(19), 9210. https://doi.org/10.3390/app11199210

Elsadig, M., Ibrahim, A. O., Basheer, S., Alohali, M. A., Alshunaifi, S., Alqahtani, H., Alharbi, N., & Nagmeldin, W. (2022). Intelligent deep machine learning cyber phishing URL detection based on BERT features extraction. Electronics, 11(21), 3647. https://doi.org/10.3390/electronics11213647

Goyal, P., Pandey, S., & Jain, K. (2018). Deep learning for natural language processing. Apress.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jain, A. K., & Gupta, B. B. (2019). A machine learning based approach for phishing detection using hyperlinks information. Journal of Ambient Intelligence and Humanized Computing, 10, 2015–2028.

Keras. (2022). Embedding layer. https://keras.io/api/layers/corelayers/embedding

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering—A systematic literature review. Information and Software Technology, 51(1), 7–15. https://doi.org/10.1016/j.infsof.2008.09.009

Lastdrager, E. (2014). Achieving a consensual definition of phishing based on a systematic review of the literature. Crime Science, 3(1), 9. https://doi.org/10.1186/s40163-014-0009-y

Luo, X., Zhou, W., Wang, W., Zhu, Y., & Deng, J. (2018). Attention-based relation extraction with bidirectional gated recurrent unit and highway network in the analysis of geological data. IEEE Access, 6, 5705–5715. https://doi.org/10.1109/ACCESS.2018.2794443

Macas, M., Wu, C., & Fuertes, W. (2022). A survey on deep learning for cybersecurity: Progress, challenges, and opportunities. Computer Networks, 212, 109032. https://doi.org/10.1016/j.comnet.2022.109032

Marchal, S., Armano, G., Gröndahl, T., Saari, K., Singh, N., & Asokan, N. (2017). Off-the-hook: An efficient and usable client-side phishing prevention application. IEEE Transactions on Computers, 66(10), 1717–1733. https://doi.org/10.1109/TC.2017.2700287

Medsker, L. R., & Jain, L. (2001). Recurrent neural networks. Design and Applications, 5, 64–67.

Ozcan, A., Catal, C., Donmez, E., & Senturk, B. (2021). A hybrid DNN–LSTM model for detecting phishing URLs. Neural Computing and Applications, 35, 4957–4973. https://doi.org/10.1007/s00521-021-06205-x

Papers With Code. (2023). Bidirectional LSTM. https://paperswithcode.com/method/bilstm

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. https://nlp.stanford.edu/projects/glove

Rao, R. S., & Pais, A. R. (2019). Detection of phishing websites using an efficient feature-based machine learning framework. Neural Computing and Applications, 31, 3851–3873. https://doi.org/10.1007/s00521-017-3305-0

Safi, A., & Singh, S. (2023). A systematic literature review on phishing website detection techniques. Journal of King Saud University – Computer and Information Sciences, 35, 590–611. https://doi.org/10.1016/j.jksuci.2021.11.007

Salloum, S., Gaber, T., Vadera, S., & Sharan, K. (2022). A systematic literature review on phishing email detection using natural language processing techniques. IEEE Access, 10, 65703–65727. https://doi.org/10.1109/ACCESS.2022.3183607

Selvaganapathy, S. G., Nivaashini, M., & Natarajan, H. P. (2018). Deep belief network based detection and categorization of malicious URLs. Information Security Journal: A Global Perspective, 27(3), 145–161. https://doi.org/10.1080/19393555.2018.1463959

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306. https://doi.org/10.1016/j.physd.2019.132306

Sirigineedi, S. S., Soni, J., & Upadhyay, H. (2020). Learning-based models to detect runtime phishing activities using URLs. In Proceedings of the 4th International Conference on Compute and Data Analysis (pp. 102–106).

Sutter, T., Bozkir, A. S., Gehring, B., & Berlich, P. (2022). Avoiding the hook: Influential factors of phishing awareness training on click-rates and a data-driven approach to predict email difficulty perception. IEEE Access, 10, 100540–100565. https://doi.org/10.1109/ACCESS.2022.3210185

Villanueva, A., Atibagos, C., De Guzman, J., Cruz, J. C. D., Rosales, M., & Francisco, R. (2022). Application of natural language processing for phishing detection using machine and deep learning models. In Proceedings of the International Conference on ICT for Smart Society (pp. 1–6). IEEE.

Vinayakumar, R., Alazab, M., Srinivasan, S., Pham, Q. V., Padannayil, S. K., & Simran, K. (2020). A visualized botnet detection system based deep learning for the internet of things networks of smart cities. IEEE Transactions on Industrial Applications, 56(4), 4436–4456. https://doi.org/10.1109/TIA.2020.2971952

Vinayakumar, R., Soman, K. P., & Poornachandran, P. (2018). Evaluating deep learning approaches to characterize and classify malicious URLs. Journal of Intelligent & Fuzzy Systems, 34(3), 1333–1343. https://doi.org/10.3233/JIFS-169404

Xiao, X., Xiao, W., Zhang, D., Zhang, B., Hu, G., Li, Q., & Xia, S. (2021). Phishing websites detection via CNN and multi-head self-attention on imbalanced datasets. Computers & Security, 108, 102372. https://doi.org/10.1016/j.cose.2021.102372

Yen, S., Moh, M., & Moh, T. S. (2021). Detecting compromised social network accounts using deep learning for behavior and text analyses. International Journal of Cloud Applications and Computing, 11(1), 97–109. https://doi.org/10.4018/IJCAC.2021010106

Zhang, X., Zeng, Y., Jin, X. B., Yan, Z. W., & Geng, G. G. (2017). Boosting the phishing detection performance by semantic analysis. In Proceedings of the IEEE International Conference on Big Data (pp. 1063–1070). IEEE.

Zieni, R., Massari, L., & Calzarossa, M. C. (2023). Phishing or not phishing? A survey on the detection of phishing websites. IEEE Access, 11, 18499–18519. https://doi.org/10.1109/ACCESS.2023.3245079