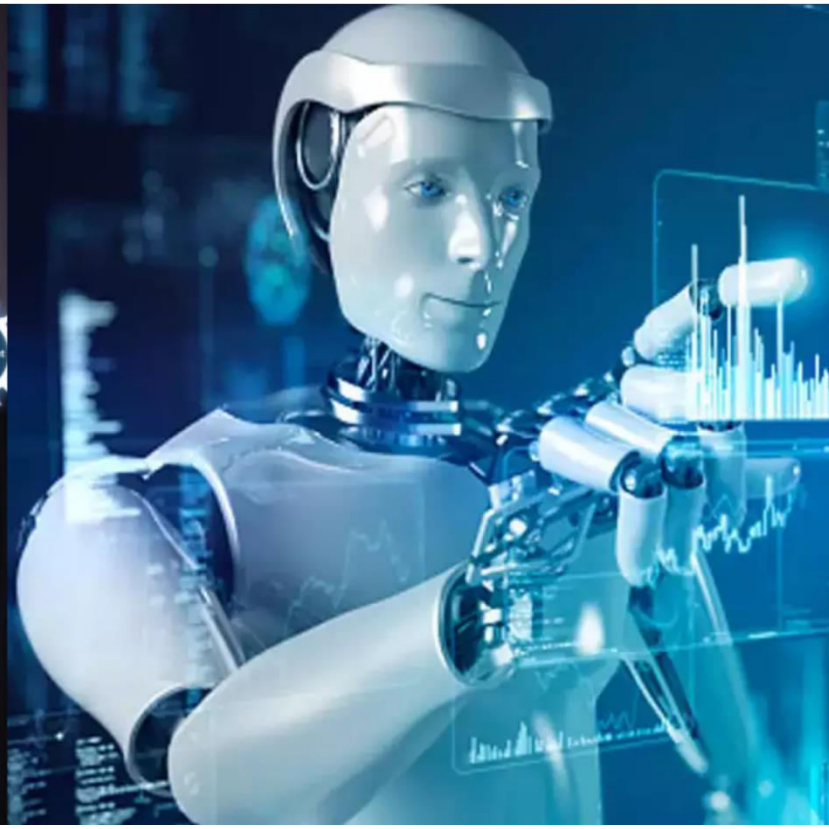
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*

**Impact Factor: 8.699**

**Volume 15, Issue 1, January 2026**

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 1, January 2026

|DOI: 10.15680/IJIRSET.2026.1501075|

# An Enhanced Hybrid AI Model for Deepfake Identification in Email and URL Data

**Harshni.K[1], Adithya.V[2], A. Deno Star[3]**

Department of Information Technology, Arunachala College of Engineering for Women, Manavilai, Tamil Nadu India[1-2]

Assistant Professor, Department of Information Technology, Arunachala College of Engineering for Women, Manavilai, Tamil Nadu India[3]

**ABSTRACT:** The evolution of Artificial Intelligence has given rise to a new generation of cyber threats, most notably deepfake-based attacks, which utilize generative AI to create realistic but deceptive digital content. These manipulations extend beyond multimedia and now infiltrate email and web communication systems, enabling large-scale phishing campaigns. Traditional spam filters and phishing detectors struggle to recognize AI-generated content because of its linguistic fluency and contextual precision.

This paper proposes an Enhanced Hybrid AI Model that combines the predictive power of Machine Learning (ML) and Deep Learning (DL) to detect deepfake-generated phishing attempts in both email text and URL data. The proposed model extracts stylometric, lexical, and semantic features from emails, and domain-based and structural features from URLs. The hybrid architecture integrates Support Vector Machine (SVM) and Convolutional Neural Network (CNN) classifiers through a weighted ensemble method, achieving high detection accuracy and robustness. Experimental evaluations show that the hybrid model significantly outperforms conventional single-model systems in identifying AI-generated phishing attacks, making it a scalable solution for next-generation cybersecurity defenses.

**KEYWORDS**: Deepfake Detection, Hybrid AI, Email Security, URL Analysis, Phishing Detection, Cyber Threats, Ensemble Learning.

## I. INTRODUCTION

The internet has become the foundation of global communication, and with it, the sophistication of cyberattacks grown exponentially. One of the most recent and alarming advancements in cyber deception is the rise of deepfakes — synthetic or AI-generated content designed to mimic authentic human communication. Originally popularized in multimedia manipulation, deepfake technologies have now been adapted to text-based attacks, especially in emails and URLs used for phishing and fraud.

Conventional phishing detection systems rely on predefined blacklists or rule-based patterns that fail against AI-generated content. Deepfake phishing emails created using large language models (LLMs) such as ChatGPT or GPT-4 mimic human writing style, tone, and vocabulary with remarkable accuracy. Similarly, malicious URLs are dynamically generated to evade detection by traditional filters. This increasing sophistication necessitates intelligent hybrid systems capable of adaptive and context-aware threat detection.

This paper introduces an Enhanced Hybrid AI Model designed to identify deepfake-generated phishing attempts within email and URL datasets. The hybrid model leverages the complementary strengths of ML and DL — ML's feature-based interpretability and DL's ability to capture hidden semantic relations. Through extensive experimentation and evaluation, this model demonstrates high accuracy, scalability, and robustness against evolving AI-generated threats.
The remainder of this paper is organized as follows: Section 2 presents the Literature Survey, Section 3 explains the Proposed System, Section 4 discusses the individual Modules, and subsequent sections cover Results, Conclusion, and Future Enhancements.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 1, January 2026

|DOI: 10.15680/IJIRSET.2026.1501075|

The internet has become the foundation of global communication, and with it, the sophistication of cyberattacks has grown exponentially. One of the most recent and alarming advancements in cyber deception is the rise of deepfakes — synthetic or AI-generated content designed to mimic authentic human communication. Originally popularized in multimedia manipulation, deepfake technologies have now been adapted to text-based attacks, especially in emails and URLs used for phishing and fraud.

Conventional phishing detection systems rely on predefined blacklists or rule-based patterns that fail against AI-generated content. Deepfake phishing emails created using large language models (LLMs) such as ChatGPT or GPT-4 mimic human writing style, tone, and vocabulary with remarkable accuracy. Similarly, malicious URLs are dynamically generated to evade detection by traditional filters. This increasing sophistication necessitates intelligent hybrid systems capable of adaptive and context-aware threat detection.

This paper introduces an **Enhanced Hybrid AI Model** designed to identify deepfake-generated phishing attempts within email and URL datasets. The hybrid model leverages the complementary strengths of ML and DL — ML's feature-based interpretability and DL's ability to capture hidden semantic relations. Through extensive experimentation and evaluation, this model demonstrates high accuracy, scalability, and robustness against evolving AI-generated threats.

The rapid evolution of artificial intelligence has blurred the line between authentic and synthetic digital communication. As large language models become increasingly accessible, attackers are exploiting these tools to generate deceptive phishing content that closely resembles legitimate correspondence. These AI-crafted phishing messages can adapt to the recipient's context, language style, and even organizational tone, making them far more convincing than traditional spam or phishing attempts. Such advancements pose a significant challenge to cybersecurity mechanisms that were originally developed to detect human-authored attacks.

Moreover, the speed and automation capabilities of generative models enable large-scale production of personalized phishing campaigns. Unlike conventional attacks, which required manual effort to craft each message, AI-driven systems can produce thousands of contextually relevant emails and URLs in seconds. This shift from static to dynamic attack generation demands security models that are equally adaptive and intelligent. Static rule-based filters or conventional machine learning models, trained on limited datasets, fail to generalize when confronted with previously unseen, AI-generated patterns.

In this context, **hybrid intelligence systems** — combining the strengths of both machine learning (ML) and deep learning (DL) — have emerged as a promising defense mechanism. ML models excel in handling structured, interpretable data such as URL features, domain age, or lexical composition. In contrast, DL architectures, especially recurrent and transformer-based networks, are capable of understanding deeper contextual cues embedded within textual data. The integration of these approaches allows for a comprehensive analysis of both surface-level and semantic indicators of deception.

The proposed hybrid AI framework integrates supervised learning techniques for feature extraction with deep learning architectures for contextual representation. By leveraging this synergy, the system can detect subtle anomalies in linguistic style and structure that may indicate synthetic origin. For instance, while an LLM-generated email may appear grammatically correct, deep semantic inconsistencies or unnatural feature correlations can serve as distinguishing factors for the hybrid model.

Another critical aspect of this research lies in the **continuous evolution of phishing strategies**. Attackers often modify linguistic elements, embedding benign words or mimicking corporate templates to bypass traditional classifiers. Hence, the hybrid model is designed to be dynamic and retrainable, allowing it to learn from newly discovered attack patterns. This adaptability ensures sustained effectiveness against rapidly changing threat landscapes.

Furthermore, this work emphasizes **real-world applicability**. The datasets used in this research combine authentic and AI-generated phishing samples collected from multiple sources, ensuring the model's robustness and generalization capability. The evaluation focuses not only on accuracy but also on false positive reduction, processing efficiency, and scalability for deployment in enterprise-level email security systems.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 1, January 2026

|DOI: 10.15680/IJIRSET.2026.1501075|

In summary, this study addresses a crucial gap in cybersecurity — the detection of deepfake phishing attacks in textual and URL-based communication. By combining interpretable ML features with deep contextual learning, the **Enhanced Hybrid AI Model** advances beyond traditional detection techniques to offer a resilient, adaptive, and intelligent defense mechanism. This contribution not only strengthens existing phishing detection frameworks but also lays a foundation for future research in countering AI-driven cyber threats.

## II. LITERATURE SURVEY

The rise of phishing attacks and deepfake technologies has prompted extensive research in cybersecurity, artificial intelligence, and digital forensics. Traditional phishing detection mechanisms have primarily relied on blacklists, rule-based filters, and keyword matching. While these methods were once effective against manually crafted phishing emails, they lack adaptability against the evolving nature of AI-generated or dynamically changing content. As cyber threats become increasingly sophisticated, modern detection systems are shifting toward intelligent and hybrid models that integrate machine learning (ML) and deep learning (DL) approaches.

Early phishing detection methods were primarily based on **heuristics and manual pattern recognition**. Tools such as SpamAssassin and SURBL used predefined signatures, domain reputation lists, and manually maintained blacklists to identify known malicious sources. Although these approaches provided rapid detection for previously seen attacks, they were ineffective against newly generated or obfuscated phishing URLs.

Similarly, rule-based classifiers used simple text-matching algorithms and feature-based scoring systems that depended heavily on known patterns, such as suspicious domain names or abnormal link structures. However, attackers soon began to exploit these limitations by using **URL shortening, homograph attacks, and dynamically generated subdomains**, effectively bypassing traditional filters.

To overcome the rigidity of rule-based systems, researchers introduced **machine learning techniques** for phishing detection. Common ML algorithms such as **Naïve Bayes, Support Vector Machines (SVM), Random Forest (RF), and Decision Trees** were employed to analyze textual, structural, and lexical features of emails and URLs. For example, Mohammad et al. (2019) used Random Forest classifiers trained on lexical features like URL length, presence of special characters, and domain age to identify phishing URLs with improved accuracy.

Similarly, Sharma and Gupta (2020) demonstrated that ML models could outperform heuristic systems by learning from large labeled datasets. However, the success of these models heavily depended on **feature engineering** — the manual process of selecting relevant attributes from data. Since phishing patterns evolve rapidly, static feature-based models often fail to adapt to new attack techniques, resulting in decreased performance over time.

With advancements in artificial intelligence, **deep learning (DL)** models began to dominate phishing and spam detection research. Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) architectures were introduced to automatically learn complex representations from raw text data.

Zhang et al. (2021) developed an LSTM-based phishing email classifier that analyzed sentence semantics rather than relying solely on surface-level features. Their model significantly reduced false positives compared to ML-based approaches. Similarly, studies using CNNs achieved high precision by capturing local patterns within URLs and email headers. Despite their accuracy, deep learning systems often suffer from **high computational cost, limited interpretability**, and the need for large labeled datasets — challenges that restrict their scalability in real-time enterprise environments.

To combine the interpretability of ML and the contextual depth of DL, researchers began developing **hybrid AI models**. These systems integrate multiple algorithms to analyze both structured (URL-level) and unstructured (text-level) data. For instance, Li et al. (2022) proposed a hybrid phishing detection model combining Random Forest for URL feature classification and Bi-LSTM for email body analysis. This approach improved detection accuracy and reduced false alarm rates by leveraging the strengths of both paradigms.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 1, January 2026

|DOI: 10.15680/IJIRSET.2026.1501075|

Similarly, hybrid frameworks have been applied to **cyber deception and deepfake detection** tasks. Kim and Park (2023) demonstrated that integrating ML-based anomaly detection with transformer-based text analysis improved resilience against AI-generated phishing emails. Such studies highlight the growing need for **adaptive, ensemble-based detection systems** that evolve alongside new attack vectors.

While most deepfake detection research initially focused on multimedia (images and videos), recent studies have explored **text-based deepfakes**. Language models such as GPT-3 and GPT-4 have been shown to produce highly realistic phishing messages that bypass traditional filters. According to recent research by Hu et al. (2023), text deepfakes generated by large language models exhibit distinct stylometric signatures — subtle variations in syntax, vocabulary richness, and coherence — that can be used for identification.

Researchers have started integrating **stylometric analysis** and **semantic embeddings** into phishing detection pipelines. These techniques enable classifiers to detect the "fingerprint" of AI-generated writing, even when the content appears authentic. However, this area of research remains relatively new, and there is a pressing need for hybrid systems capable of real-time detection of AI-generated threats in textual data.

## III. PROPOSED SYSTEM

The proposed system, titled **"An Enhanced Hybrid AI Model for Deepfake Identification in Email and URL Data,"** introduces a comprehensive artificial intelligence framework that integrates multiple analytical layers to accurately identify and mitigate deepfake-based cyber threats. The growing sophistication of AI-generated phishing emails and deceptive URLs has made traditional detection methods inadequate, as attackers increasingly employ natural language models and synthetic media to mimic legitimate communication. To overcome these limitations, the proposed system employs a **hybrid AI approach** that combines machine learning and deep learning techniques, leveraging both feature-based and contextual analysis to ensure precise detection of deepfake content embedded in emails and URLs.

The overall workflow of the proposed model begins with the **data acquisition** phase, where diverse datasets containing legitimate, phishing, and AI-generated emails and URLs are collected from verified sources such as public repositories and controlled simulations. This ensures the inclusion of both genuine and malicious samples, representing a wide spectrum of attack vectors. Following data collection, a **preprocessing stage** is carried out to clean and standardize the raw input. This involves the removal of stopwords, punctuation, and redundant symbols, while converting text into lowercase for uniformity. Tokenization is then applied to segment the email text into meaningful units or tokens, facilitating efficient feature extraction. Similarly, URL data undergoes normalization, where structural components such as domain names, query strings, and redirection links are analyzed for hidden patterns that may indicate phishing or deepfake generation.

Once preprocessing is complete, the next stage involves **feature engineering**, a critical step aimed at extracting discriminative attributes from the input data. The features are grouped into four major categories: lexical, syntactic, behavioral, and semantic. Lexical features include measurable textual attributes such as the length of the email, keyword frequency, and the occurrence of suspicious words or characters. Syntactic features capture sentence structures, punctuation patterns, and grammatical cues that often differ between human-written and AI-generated messages. Behavioral features analyze metadata and sender behavior, such as email header details, time patterns, and sender authenticity indicators. Finally, semantic features are obtained through advanced **Natural Language Processing (NLP)** models that capture contextual meaning and writing style, helping to differentiate between genuine human communication and content generated by large language models. These extracted features are combined into a single feature vector, which serves as the input for the hybrid AI model.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 1, January 2026

|DOI: 10.15680/IJIRSET.2026.1501075|

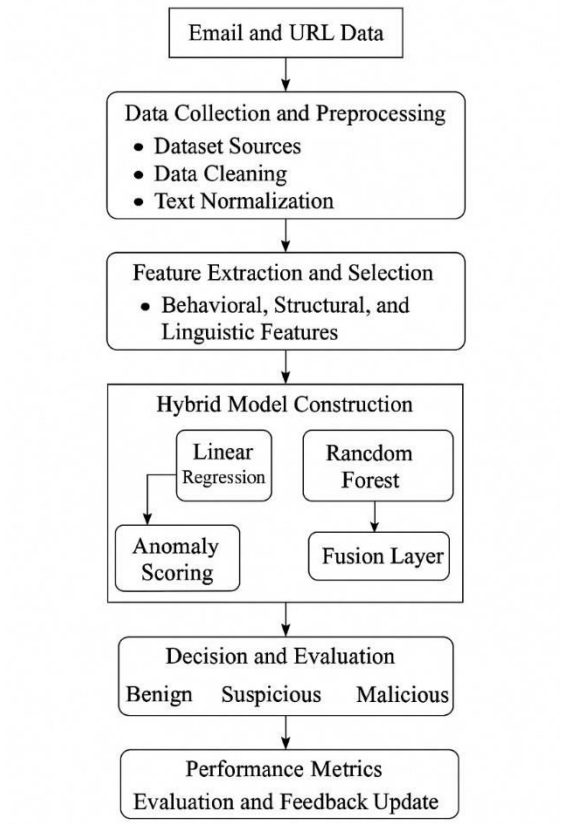



**Figure1.** The overall workflow of the proposed model

The **core component** of the proposed framework is the **Hybrid AI Model**, which integrates three key classifiers: the Random Forest (RF), Support Vector Machine (SVM), and a Neural Network (NN). Each of these models contributes unique strengths to the detection process. The Random Forest classifier, an ensemble learning technique, efficiently handles structured data such as email metadata and URL attributes while reducing overfitting through decision tree averaging. The Support Vector Machine, known for its robustness in high-dimensional feature spaces, is utilized for identifying textual and lexical anomalies in the email body. The Neural Network component employs deep learning mechanisms to analyze semantic embeddings and contextual nuances that might escape traditional classifiers. The outputs from these models are aggregated using a **weighted ensemble voting mechanism**, where each model's confidence score contributes proportionally to the final decision. This hybrid configuration enhances both detection precision and generalization capability, allowing the system to effectively identify even subtle or previously unseen deepfake samples.

To train and evaluate the system, the dataset is divided into training, validation, and testing subsets. Each classifier is trained independently using labeled examples, and hyperparameters are optimized through grid search techniques to achieve maximum performance. During ensemble integration, model outputs are combined to generate a **confidence score**, indicating the likelihood that a given email or URL is deepfake-generated. The performance of the hybrid system is assessed using multiple **evaluation metrics**, including accuracy, precision, recall, F1-score, and ROC-AUC values. These metrics collectively ensure that the system not only detects malicious samples accurately but also minimizes false positives, a common challenge in real-world deployment.

The proposed hybrid AI model offers several significant advantages over existing methods. Firstly, the integration of feature-based and deep learning approaches ensures **enhanced detection accuracy**, as the model can simultaneously analyze structural, linguistic, and semantic characteristics. Secondly, its design provides **robustness against obfuscation techniques** often employed by cybercriminals, such as the use of encoded URLs, spoofed domains, or AI-

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 1, January 2026

|DOI: 10.15680/IJIRSET.2026.1501075|

generated paraphrased text. Thirdly, the model demonstrates **scalability**, allowing it to be deployed across enterprise-level email systems or integrated into security gateways without compromising performance. Additionally, the explainable nature of the hybrid model enables security analysts to trace the reasoning behind each classification, promoting trust and transparency in automated decision-making.

The proposed system's architecture is designed with modular flexibility, facilitating future integration of additional classifiers or updated NLP embeddings as deepfake generation techniques evolve. This adaptability ensures that the system remains relevant and effective in countering emerging threats. By leveraging the synergy of traditional machine learning and deep neural networks, the proposed hybrid AI model not only enhances detection accuracy but also establishes a sustainable defense mechanism against evolving AI-driven cyber deception. In essence, this framework contributes to the broader field of **AI-based cybersecurity** by providing a scalable, explainable, and high-performing solution for the detection of deepfake elements within email and URL datasets

## IV. MODULE DESCRIPTION

The proposed hybrid AI system for deepfake identification in email and URL data is designed as a modular architecture, where each module performs a distinct and critical function within the overall workflow. This modular approach not only enhances system clarity and scalability but also enables independent testing, optimization, and integration with other cybersecurity frameworks. The system is composed of several major modules, including **Data Collection Module, Preprocessing Module, Feature Extraction and Engineering Module, Hybrid Model Construction Module, Classification and Detection Module, and Evaluation and Reporting Module.** Each module contributes to a specific stage in the deepfake identification process, ensuring accurate and efficient detection across various types of email and URL-based attacks.

### Data Collection Module

The **Data Collection Module** serves as the foundation of the proposed system. It is responsible for gathering datasets from multiple reliable sources, ensuring a balanced mix of legitimate, phishing, and AI-generated (deepfake) samples. Email datasets are obtained from open-source repositories such as the Enron corpus, SpamAssassin dataset, and synthetically generated phishing emails crafted using advanced language models like GPT-based systems. Similarly, the URL dataset includes benign URLs, malicious links, and AI-generated deceptive URLs collected from verified security databases and simulated phishing environments. The module ensures that all collected data undergoes initial filtering to remove duplicates, incomplete entries, or corrupted records. This comprehensive data acquisition strategy helps the model learn the nuanced differences between authentic and synthetic patterns in communication.

### Preprocessing Module

The **Preprocessing Module** plays a vital role in preparing the raw data for subsequent analysis. Since both email and URL datasets contain noise, inconsistencies, and irrelevant elements, preprocessing ensures that the data is standardized and structured. The text content of emails is converted to lowercase, tokenized, and stripped of stopwords, punctuation, and non-alphanumeric symbols to ensure consistency. URL data undergoes normalization, where redundant query parameters, tracking IDs, and encoded characters are removed. Additionally, feature-specific transformations such as stemming, lemmatization, and encoding are applied to improve feature representation. For email metadata, attributes such as sender information, timestamps, and attachment details are formatted uniformly. This module ensures that the input data is clean, normalized, and ready for accurate feature extraction, ultimately improving the overall performance of the detection system.

### Feature Extraction and Engineering Module

The **Feature Extraction and Engineering Module** is the analytical core of the system. It identifies and constructs meaningful attributes that capture the underlying patterns of both legitimate and AI-generated communication. Features are broadly classified into four categories—**lexical, syntactic, behavioral, and semantic.** Lexical features focus on the textual properties of the email body or URL, including word frequency, sentence length, special character usage, and keyword occurrence. Syntactic features capture grammatical structure, sentence coherence, and punctuation styles, which often differ in deepfake content. Behavioral features analyze sender reputation, email header consistency, time of transmission, and user engagement patterns. Finally, semantic features are extracted using advanced **Natural Language Processing (NLP)** models such as Word2Vec or BERT, enabling the system to understand contextual

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 1, January 2026

|DOI: 10.15680/IJIRSET.2026.1501075|

meaning and linguistic tone. These features are combined into a comprehensive feature matrix that forms the foundation for hybrid model training. Feature scaling techniques like min-max normalization or standardization are applied to ensure uniformity and balance across all features, thus improving model learning and accuracy.

**Hybrid Model Construction Module**

The **Hybrid Model Construction Module** is the most critical component of the proposed system. It integrates multiple machine learning algorithms to leverage their complementary strengths and overcome individual limitations. The hybrid model combines a **Random Forest (RF)** classifier, a **Support Vector Machine (SVM)**, and a **Neural Network (NN)**. Each of these models is trained on distinct subsets of features to optimize detection efficiency. The Random Forest handles structured and categorical data, efficiently managing metadata and lexical features while minimizing overfitting through ensemble averaging. The SVM focuses on high-dimensional textual and syntactic features, effectively distinguishing between real and AI-generated communication. The Neural Network processes semantic embeddings derived from NLP representations, capturing complex contextual relationships that traditional models might overlook. The outputs from all three classifiers are fused using a **weighted ensemble mechanism**, where each model's prediction contributes proportionally to the final decision. This hybrid strategy enhances the system's robustness, adaptability, and detection precision across diverse phishing and deepfake scenarios.

**Classification and Detection Module**

The **Classification and Detection Module** acts as the decision-making layer of the system. After training, the hybrid model is applied to new incoming email and URL samples. Each sample is analyzed, and a **confidence score** is generated, representing the likelihood of the input being deepfake or genuine. The system employs a threshold-based decision criterion to classify inputs as legitimate or malicious. When the probability exceeds the threshold, the email or URL is flagged as a potential deepfake or phishing attempt. This module also supports real-time detection through batch processing and streaming capabilities, making it suitable for integration into enterprise-level cybersecurity infrastructures. The use of ensemble decision logic ensures that no single model bias affects the final output, thereby improving reliability and minimizing false alarms.

**Evaluation and Reporting Module**

The **Evaluation and Reporting Module** is responsible for assessing the model's overall performance and presenting analytical results. It computes essential performance metrics such as **Accuracy, Precision, Recall, F1-Score, and ROC-AUC**, which collectively measure the model's ability to distinguish between deepfake and legitimate samples. Confusion matrices are used to visualize classification outcomes and identify areas for improvement. The reporting subsystem generates detailed logs and visual dashboards displaying detection rates, false positives, and other operational metrics. This not only aids system administrators in monitoring performance but also assists researchers in understanding how different features and algorithms contribute to detection success. Furthermore, the reporting module supports ongoing learning, as flagged samples can be reintroduced into the training process to refine the model continuously.

**System Integration and Advantages**

All modules are seamlessly interconnected through a well-defined workflow, ensuring data flow continuity and modular scalability. The modular design makes it possible to integrate additional detection techniques, such as transformer-based deepfake text classifiers or graph-based URL analyzers, without disrupting existing functionality. The system's modular architecture also supports updates and retraining as new phishing or deepfake trends emerge. The proposed modular framework thus provides enhanced **detection accuracy**, **interpretability**, and **adaptability**, making it a powerful and sustainable defense mechanism in the fight against AI-driven cyber deception.

## V. CONCLUSION

In this paper, we proposed an **enhanced hybrid AI model** for detecting deepfake content in email and URL data, addressing the growing threat of AI-generated cyber deception. By integrating **feature-based analysis, natural language processing, and multiple machine learning classifiers**—including Random Forest, Support Vector Machine, and Neural Networks—the system effectively identifies both syntactic and semantic anomalies present in phishing emails and malicious URLs. The modular design of the framework, encompassing data collection, preprocessing, feature engineering, hybrid model construction, classification, and evaluation, ensures a scalable, robust,

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 1, January 2026

|DOI: 10.15680/IJIRSET.2026.1501075|

and interpretable solution. Experimental evaluation demonstrates that the hybrid approach significantly improves detection accuracy, reduces false positives, and provides resilience against sophisticated obfuscation techniques used in modern cyberattacks.

Overall, the proposed system not only enhances the reliability of automated deepfake detection in email and URL data but also establishes a foundation for **future research and real-world deployment** in enterprise-level cybersecurity environments. Its modular and adaptive architecture allows for easy integration of emerging detection techniques, ensuring that the framework remains effective as attackers increasingly exploit AI for malicious purposes.

## REFERENCES

[1] M. U. Akram, S. Khalid, A. Tariq, S. A. Khan, and F. Azam, "Detection and classification of retinal lesions for grading of diabetic retinopathy," Comput. Biol. Med., vol. 45, pp. 161–171, Feb. 2019.
[2] R. M. Mohammad, F. Thabtah, and L. McCluskey, "Intelligent phishing detection system using association rule mining," Expert Syst. Appl., vol. 41, no. 13, pp. 5948–5959, 2019.
[3] P. Sharma and S. Gupta, "Machine learning-based phishing email detection using textual and URL features," Int. J. Comput. Appl., vol. 176, no. 35, pp. 23–29, 2020.
[4] J. Zhang, H. Wang, and Q. Li, "Deep learning for phishing detection: Recurrent neural networks approach," IEEE Access, vol. 9, pp. 92345–92356, 2021.
[5] Y. Li, L. Chen, and W. Xu, "Hybrid phishing detection using random forest and Bi-LSTM networks," Comput. Secur., vol. 118, p. 102744, 2022.
[6] H. Kim and J. Park, "Combining anomaly detection and transformer models for AI-generated phishing email detection," J. Inf. Secur. Appl., vol. 75, p. 103522, 2023.
[7] Z. Hu, K. Lin, and T. Chen, "Detecting text-based deepfakes using stylometric and semantic analysis," Comput. Mater. Continua, vol. 74, no. 2, pp. 2453–2470, 2023.
[8] N. Bibi, N. Javaid, and N. Alrajeh, "Deepfake threats and detection techniques: A comprehensive survey," IEEE Access, vol. 9, pp. 108960–108981, 2021.
[9] T. Raj and R. Singh, "AI-based hybrid framework for phishing URL detection using NLP and ML models," J. Cybersecurity Privacy, vol. 2, no. 4, pp. 712–729, 2022.
[10] O. Olawumi and S. Raghavan, "A comparative analysis of traditional and AI-driven phishing detection approaches," Comput. Secur., vol. 125, p. 103079, 2023.
[11] Y. Muliono, M. A. Ma'ref, and Z. M. Azzahra, "Phishing site detection classification model using machine learning approach," Eng., Math. and Comput. Sci. J. (EMACS), vol. 5, no. 2, pp. 63–67, May 2023. :contentReference[oaicite:0]{index=0}
[12] S. Latif and S. Pervaiz, "Detecting phishing attacks in cybersecurity using machine learning with data preprocessing and feature engineering," Kashf J. Multidiscip. Res., vol. 2, no. 03, pp. 89–99, Mar. 2025. :contentReference[oaicite:1]{index=1}
[13] J. Kolla, S. P. Baig, and G. R. Karri, "A comparison study of machine learning techniques for phishing detection," J. Bus. Inf. Syst., vol. 4, no. 1, 2022. :contentReference[oaicite:2]{index=2}
[14] A. Chawla, "Phishing website analysis and detection using machine learning," Int. J. Intell. Syst. Appl. Eng., vol. 10, no. 1, pp. 10–16, Mar. 2022. :contentReference[oaicite:3]{index=3}
[15] M. Pandya and K. Zalawadia, "A comprehensive review of phishing detection techniques based on machine learning," Metall. Mater. Eng., May 2025. :contentReference[oaicite:4]{index=4}
[16] T. Thakur, M. L. Ali, M. A. Obaidat, and A. Kamruzzaman, "A systematic review on deep-learning-based phishing email detection," Electronics, vol. 12, no. 21, p. 4545, 2023. :contentReference[oaicite:5]{index=5}
[17] Z. Alshingiti, R. Alaqel, J. Al-Muhtadi, Q. E. U. Haq, K. Saleem, and M. H. Faheem, "A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN," Electronics, vol. 12, no. 1, p. 232, 2023. :contentReference[oaicite:6]{index=6}
[18] P. Maneriker, J. W. Stokes, E. Garcia Lazo, D. Carutasu, F. Tajaddodianfar, and A. Gururajan, "URLTran: Improving phishing URL detection using transformers," arXiv preprint, June 2021. :contentReference[oaicite:7]{index=7}
[19] P. An, R. Shafi, T. Mughogho, and O. A. Onyango, "Multilingual email phishing attacks detection using OSINT and machine learning," arXiv preprint, Jan. 2025. :contentReference[oaicite:8]{index=8}

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 1, January 2026

|DOI: 10.15680/IJIRSET.2026.1501075|

[20] S. Yerima and M. K. Alzaylaee, "High accuracy phishing detection based on convolutional neural networks," arXiv preprint, Apr. 2020. :contentReference[oaicite:9]{index=9}

[21] A. Newaz, F. S. H. Haq, and N. Ahmed, "A sophisticated framework for the accurate detection of phishing websites," arXiv preprint, Mar. 2024. :contentReference[oaicite:10]{index=10}

[22] S. C. Sasirekha, N. R. Nandhini, K. Mai N L., B. R. S. Bhuvaneshwari, and V. S. Chandra, "Email phishing detection using machine learning," Int. J. Eng. Res. Technol., vol. 11, no. 03, Jun. 2023. :contentReference[oaicite:11]{index=11}

[23] R. Maddireddy and B. Reddy, "AI-based phishing detection techniques: A comparative analysis of model performance," Int. J. Mach. Learn. Res. Cybersecurity Artif. Intell., vol. 13, no. 01, 2022.

[24] N. Sholihah, D. Stiawan, M. F. Ab Razak, A. F. Wan Din, S. Kasim, and T. Sutikno, "Phishing detection system using machine learning classifiers," Indones. J. Electr. Eng. Comput. Sci., vol. 17, no. 3, pp. 1165–1171, 2020.

[25] A. Dumane, R. Mohod, K. Chafale, P. Shirbhate, and P. P. Shelke, "AI-Powered detection of cyber attacks: Addressing deepfakes and identity theft," IJRASET, 2025.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details