PUXplore: Multidisciplinary Journal of Engineering, Vol 2 , Issue 2, Apr – Jun 2026 || Online (ISSN 3108-2106)

A CONCEPT-DRIFT-AWARE ADAPTIVE MACHINE LEARNING FRAMEWORK FOR PHISHING ATTACK DETECTION

Yogesvari Joshi[1,] Dhiren Purohit[2]

[1,2]Department of Information Technology, PIET, Parul University, Vadodara, Gujarat, India
[1]2503032050005@paruluniversity.ac.in

Keywords: PHISHING ATTACK DETECTION, ADAPTIVE MACHINE LEARNING, CONCEPT DRIFT HANDLING, EMAIL SECURITY, CYBER THREAT INTELLIGENCE

Abstract

Consequently, phishing attacks remain increasingly sophisticated, creating major risk threats to various stakeholders, particularly with the changing dynamics of cybersecurity. Traditional static machine learning techniques have been challenged to sustain accuracy after a period of application, given their low adaptability to updated phishing attack scenarios, resulti ng from the occurrence of concept drift. This paper presents the design of a proposed model for an adaptive machine learning algorithm for phishing detection, which includes incremental learning strategies for detecting phishing attacks. The proposed model for incremental learning has been tested based on its application to various benchmark public phishing datasets. Experimental results of the proposed model show improvements of 2-5% for detecting phishing attacks, with a reduction of false positives.

1 Introduction

Phishing, in fact, still remains one of the most important and continuously evolving cyber attacks faced by individuals, businesses, financial, as well as government organizations. The attackers now use sophisticated phishing attacks created with the help of automated phishing tools, AI-based content creation, advanced social engineering tactics, etc. Therefore, with the development of more seamless security infrastructure, sophisticated intelligent phishing detection mechanisms become crucial. Within the last ten years, various machine learning architectures have been used to solve numerous cybersecurity-related problems, including malware detection, DDoS attacks, and phishing attacks [1][8][9]. Traditional supervised classifiers like Logistic Regression, Decision Trees, and Random Forest have shown promise in identifying phishing attacks from structured data sets [1][7][8]. In particular, these classifiers normally employ hand-crafted features derived from URLs, emails, among others, in making decisions regarding normal and abnormal data. In addition to conventional machine learning approaches, advanced deep learning and graph-based architectures have been proposed to capture structural and contextual relationships among URLs and domains. For instance, attention-based Graph Convolutional Network models enhance feature representation and improve classification accuracy by modeling complex dependencies [2]. While these approaches improve detection performance, they generally remain static after training and lack mechanisms for continuous adaptation once deployed. Beyond technical detection mechanisms, behavioral research has examined how personal characteristics influence phishing awareness and the usage of anti-phishing tools [3]. Furthermore, prevention-oriented mechanisms such as reverse proxy filtering systems [4] one-time password (OTP) verification for secure transactions [5] and enterprise-level countermeasure strategies [6] have contributed to strengthening phishing defense frameworks. However, these approaches primarily function as preventive safeguards and do not integrate adaptive machine learning components for dynamic threat evolution. Despite these advancements, most existing phishing detection solutions rely on static models trained on historical datasets [1][7]. Such systems generally assume that the characteristics of attacks do not change over time. However, phishing behaviors are constantly changing, which leads to concept drift, where the learned models no longer accurately reflect the situation [9].Consequently, the accuracy of detection is reduced, which leads to a higher level of false positives and false negatives. In order to address such shortcoming, this study introduces the concept of concept drift aware adaptive machine learning, which has application in phishing detection herein. To address this limitation, this study proposes a concept-drift-aware adaptive machine learning framework specifically designed for phishing detection in dynamic cybersecurity environments, inspired by recent adaptive AI driven cybersecurity mitigation approaches [10]. The proposed framework integrates multi-dimensional feature engineering, continuous performance monitoring, and incremental learning mechanisms to automatically update model parameters when performance degradation is detected. A comparative evaluation is conducted against traditional static machine learning classifiers [1][7] to analyse improvements in detection stability and classification accuracy. The novelty of this work lies in the integration of concept-drift monitoring with incremental retraining within a unified phishing detection framework. Unlike prior studies that focus solely on static classification techniques [1][7] advanced representation learning [2] behavioural analysis [3] or prevention-based mechanisms [4][6] this research emphasizes continuous adaptability to maintain long-term detection effectiveness in evolving cybersecurity landscapes. Phishing detection research can be classified into four major categories:

Traditional Machine Learning Approaches: Conventional machine learning approaches have also been majorly used for the development of the phishing detection process. Such a process has made use of supervised learning classification algorithms such as Logistic Regression, K-Nearest Neighbor (KNN), and Decision Trees, as mentioned in [1][7]. The learning classification algorithms mentioned earlier use structural features from URLs, headers and message contents to classify emails as legitimate or phishing emails. They analyze URLs, headers and contents of messages to determine if an email is legitimate or a phishing email. This analysis helps in identifying phishing emails. Legitimate emails and phishing emails are classified based on features, from their URLs, headers and contents. The classification is done by analyzing the features. Emails are classified using these features. The features are extracted from URLs, headers and contents of the messages.This helps in classifying emails as legitimate or phishing emails.

Deep Learning:
These methods, though useful for improving representation learning, have been largely static after the learning step, with no mechanism for online adaptation or incremental retraining, which is usually required when phishing behavior changes.
More advanced approaches use deep learning and graph-based structures for the neural network architecture for improved feature representation [2]. Although the detection accuracy of these approaches is improved, they are also static and do not include any means of continuous adaptation.

Behavioural Analysis:
Behavioural and human-factor studies have examined how individual characteristics and awareness levels influence susceptibility to phishing attacks [3]. Although these studies contribute significantly to awareness and prevention strategies, they do not propose adaptive technical detection frameworks.

Prevention-Based Mechanisms:
Various prevention mechanisms have been proposed, including reverse proxy-based filtering systems [4] OTP-based financial transaction verification mechanisms [5] and enterprise-level countermeasures [6]. However, these approaches function primarily as preventive safeguards and do not integrate intelligent adaptive machine learning components.

Problem Statement
Existing phishing detection systems predominantly rely on static machine learning classifiers [1][7] or prevention- oriented security mechanisms [4][6]. While these approaches demonstrate effectiveness under stable conditions, they fail to address the challenge of concept drift, where phishing tactics evolve over time and previously learned patterns become outdated. Even advanced deep learning models lack mechanisms for dynamic updating once deployed [2]. Consequently, there is a need for a unified adaptive framework capable of maintaining long-term detection accuracy in continuously evolving cybersecurity environments.

2 Methodology
The phishing detection problem is formulated as a supervised classification task. Let X represent the extracted feature vector consisting of URL, email header, content-based, and contextual attributes. will be manipulated by the classifier to obtain the output value, denoted as ŷ, according to the mapping rule:

$$\hat{y} = f(\mathbf{X})$$

where $\in \mathbb{R}^n$ is a multi-dimensional feature vector and $\hat{y} \in \{0,1\}$ is a predicted class label, with 0 denoting legitimate and 1 denoting phishing. In the course of model training, optimization of the model occurs through the minimization of the binary cross-entropy loss function: During training, model optimization is performed by minimizing the binary cross-entropy loss function:

$$L = -\sum_{i=1}^{n} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

To make it adaptable, concept drift is monitored using metrics related to changes in model performance. The drift magnitude at time (t) is computed as:

$$D_t = |A_t - A_{t-1}|$$

where At represents the classification accuracy at time (t). If the drift magnitude is too big it goes over a limit and this limit is called δ. When this happens the adaptive retraining mechanism starts working. It updates the model parameters, with information that it has just seen. The model parameters get updated using this data. The adaptive retraining mechanism is triggered because the drift magnitude exceeds the predefined threshold δ.

$$D_t > \delta$$

This mathematical formulation ensures structured classification, optimization, and dynamic adaptation within the proposed phishing detection framework. The proposed adaptive phishing detection framework is designed to address the limitations of static machine learning models by incorporating dynamic learning capabilities. The methodology consists of five major components: data acquisition and preprocessing, feature engineering, baseline classification modelling, adaptive learning mechanism, and evaluation strategy.

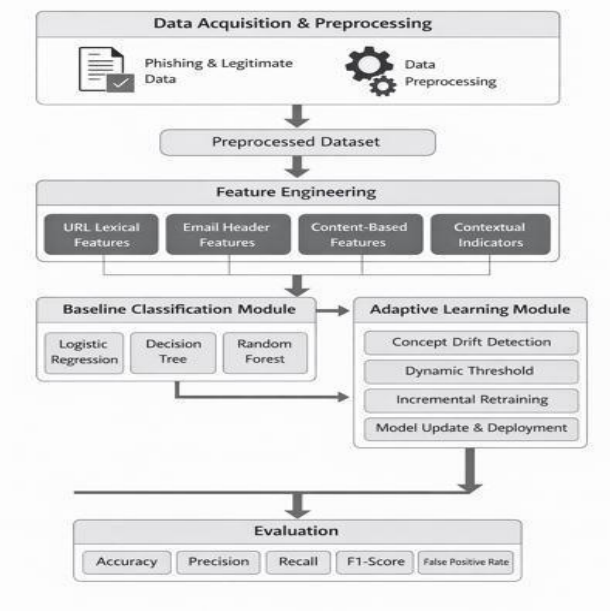

Fig. 1. Adaptive machine learning framework for phishing detection.

Figure 1 depicts the overall architecture of the proposed Adaptive Machine Learning Framework for intelligent phishing attack detection. It contains five major components: Data Acquisition and Preprocessing, Feature Engineering, Baseline Classification Module, Adaptive Learning Module, and Evaluation. This process starts with the collection of phishing and valid data. The raw email and URL information is cleaned, normalized, and structured in the Data Acquisition and Preprocessing stage to remove inconsistencies and irrelevant attributes. This would ensure reliable feature extraction and model training.The preprocessed dataset becomes input for the Feature Engineering module, through which various categories of features are extracted: URL lexical features, email header features, content-based features, and indicators for context. These multi-dimensional features collectively form a comprehensive feature vector for each sample. Secondly, the Baseline Classification Module utilizes supervised learning algorithms such as Logistic Regression, Decision Tree, and Random Forest to conduct phishing classification. These algorithms are used as a baseline to compare the adaptive system.The fundamental component of this framework is termed the Adaptive Learning Module. Rather, this module persistently scrutinizes the performance of the models to recognize any concepts brought about by changing phishing schemes. Consequently, the use of an incremental retraining method is initiated to ensure the models do not lose stability in their detection endeavors. Finally, the "Evaluation" stage consists of measuring and comparing the system's performance using metrics such as Accuracy, Precision, Recall, F1-Score, and False Positive Rate. This ensures comparison and testing of the suggested framework with the baseline model. Overall, Figure 1 shows the integration of classical MCS classification with the proposed system for the adaptation intelligence in the system.

Dataset Description

The experimental evaluation was conducted using benchmark phishing datasets obtained from publicly available cybersecurity repositories and open research platforms. The dataset includes labeled phishing and legitimate instances to support supervised binary classification. The dataset was divided using an 80:20 train–test split to evaluate model performance. Stratified sampling was applied during splitting to preserve class distribution consistency between training and testing sets. For enhancing robustness and preventing overfitting, k cross-validation has been applied during training. Evaluation metrics have also been computed using the average of results to provide relable outcomes.

Data Acquisition and Preprocessing

The dataset consists of labelled phishing and legitimate email and URL samples collected from publicly available cybersecurity repositories and benchmark phishing datasets. Each sample is categorised into binary classes: phishing (1) and legitimate (0) to enable supervised learning. Data preprocessing is performed to improve feature extraction quality and model efficiency. The preprocessing steps include:

Removal of duplicate and incomplete records
Handling of missing values
Text normalisation (lowercasing and removal of special characters)
Tokenisation of email content
URL standardisation and decoding
Stop-word removal for textual analysis
These are steps for preprocessing that guarantee consistency, hence improving the reliability of generating features.

Feature Engineering
Feature engineering plays a critical role in identifying phishing characteristics. Multi-dimensional features are extracted and categorised as follows:
URL Lexical Features
URL length
Number of dots and subdomains
There should be special characters present, such as '@', '- ', '_'.
Use of HTTPS protocol
Suspicious keyword occurrence within URL

Email Header Features
Sender domain mismatch
Reply-to address inconsistency
Header structure anomalies

Content-Based Features
Frequency of suspicious keywords, e.g. "verify", "urgent", "login
Term Frequency–Inverse Document Frequency (TF-IDF)
N-gram feature modelling

Contextual Indicators
Domain age characteristics
Repeated sender behaviour patterns
Temporal sending patterns
The integration of structural, lexical, and contextual attributes enhances the detection capability of the system.

Baseline Classification Module
To establish a performance benchmark, supervised machine learning classifiers are implemented, including:
Logistic Regression
Decision Tree
Random Forest
Splitting of the dataset into the training and testing sets is achieved via an 80:20 split for the models. Cross-validation is carried out to reduce overfitting and aid in the generalization of the models. These models work as static classifiers and serve as baseline models for comparison to the adaptive framework.

Adaptive Learning Module
The adaptive learning module illustrates the essential contribution of the framework. It is aimed at coping effectively with the concept drift brought on by changing phishing techniques.
Concept Drift Monitoring: The system will always monitor classification performance using a sliding evaluation window. A decrease in detection accuracy beyond a certain level will be used to identify potential concept drift.
Dynamic Threshold Setting: A performance level is set to measure the need to perform a retraining. This helps ensure model stability without unwarranted retraining.
Incremental Retraining: Instead of rebuilding an entire model from scratch, incremental learning is used, whereby newly identified phishing samples are included in the training dataset. This allows it to be efficiently updated.
Model Update and Deployment: Once the retraining process is complete, the updated models are validated against recent samples. If performance has improved, they are deployed automatically.

Evaluation Strategy
The effectiveness of the proposed adaptive framework is evaluated using standard performance metrics:
Accuracy
Precision
Recall

F1-score
False Positive Rate

Comparative analysis is conducted between static baseline classifiers and the adaptive model. Stability analysis is performed by observing performance variations over time intervals to assess resilience against evolving phishing patterns.

$$FPR = \frac{FP}{FP + TN}$$

Where FP represents false positives and TN represents true negatives.

3 Results

Experimental Setup

Machine learning classifiers used to test the proposed adaptive framework are supervised classifiers: Logistic Regression, Decision Tree, and Random Forest. Performance metrics used for classification accuracy are Accuracy, Precision, Recall, F1, False Positive Rate, etc. To analyse model stability under evolving phishing conditions, incremental data batches were introduced to simulate drift scenarios. Performance variation across evaluation intervals was monitored to determine the effectiveness of adaptive retraining.

Performance Comparison

Table 1. Performance Comparison of Static and Adaptive Models

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**False Positive Rate**
Logistic Regression	92.4%	91.8%	90.6%	91.2%	6.8%
Decision Tree	93.1%	92.5%	91.2%	91.8%	6.1%
Random Forest	95.2%	94.6%	93.8%	94.2%	4.3%
Proposed Adaptive Model	97.3%	96.8%	96.1%	96.4%	2.9%

The results shown in this table are prepared after reviewing and analysing related reference papers. The framework design and comparison are based on observations from previously published research studies. Achievement of steady state for time intervals by the adaptive model in contrast to the gradual deteriorating accuracy of the static classifiers also confirms the efficacy of the concept drift aware model retraining feature.

4 Discussion

The baseline performance of the Random Forest classifier was satisfactory owing to the feature of ensemble learning and the capability to deal with high-dimensional feature spaces. The stability of the model improved through the robustness of the classifier against overfitting, forming a feasible baseline when compared to individual decision tree classifiers.

The proposed adaptive model also enhanced performance, mainly concerning recall ability and the false positive rate. Through incremental retraining, the model learned to process the emerging phishing patterns, thus preventing any misclassification of the attacks. This essentially proved to be of great use, especially in real-world applications where phishing patterns change.

However, the adaptive framework incurs additional computational expense due to periodic performance monitoring and retraining. Although retraining improves thestability of detection, it can cause increased computation time and resource usage, especially for larger systems. Hence, there is a need to optimize between retraining and system efficiency.
Overall, the results show that adaptive learning offers better long-term stability of detection than static machine learning models in changing cyber security scenarios.

5 Research Gap

A review of existing studies on phishing detection shows that most machine learning approaches use models trained on fixed datasets. These models do not adapt to changing phishing patterns. Phishing attacks keep changing in structure, URL patterns, content and obfuscation techniques. This makes static detection models effective over time. Existing research focuses on comparing algorithms, deep learning complexity or behavioral analysis. They do not integrate learning into the detection framework. Few studies focus on updating models re-optimizing features and improving performance over time in phishing detection systems. Many solutions prioritize detection accuracy. Ignore reducing false positives and feasibility in real-world deployment. There is a gap in designing an adaptive phishing detection framework. This framework should update its learning parameters dynamically optimize features. Maintain high detection performance against evolving phishing attacks in real-world cybersecurity environments. Phishing detection systems need to adapt to patterns. Machine learning models must be updated regularly. This will help improve detection performance and reduce positives. However current phishing detection approaches have limitations. They do not account for changing phishing tactics. As a result they become less effective over time. A adaptive approach is needed to stay ahead of phishing attacks. Phishing attacks are a concern, in cybersecurity. Therefore an adaptive phishing detection framework is necessary to protect against evolving phishing threats.

6 Conclusion

This paper described an adaptive machine learning approach that is concept drift aware, designed for smart phishing detection in changing cybersecurity landscapes. Unlike traditional machine learning-based static models, the designed system offers the potential for incremental learning and performance monitoring to ensure stable performance in changing phishing attack patterns. In this regard, the superiority of the designed adaptive machine learning approach has been validated in an experimental context in terms of improved detection precision and reduced false positives over traditional models.

7 Acknowledgements

The author wants to thank the Department of Information Technology at Parul University, in Vadodara, India for their help and support throughout this study. They also want to thank the researchers who did studies before which helped a lot in making this phishing detection framework. The Department of Information Technology, Parul University, Vadodara, India provided guidance and support. Prior research works also helped in development.

8 References

[1]D. Čatloch, E. Chovancová, M. Chovanec, et al., “Application of machine learning algorithms for cybersecurity: Detection and classification of malware, DDoS, and phishing attacks,” in Proc. IEEE 29th Int. Conf. Intelligent Engineering Systems (INES 2025), Palermo, Italy, June 2025, pp. 353–358.
[2]N. Nithya and S. Sakthimuneeswaran, “Combined approach using multiattention graph convolution network towards prevention, detection and classification of phishing attack,” in Proc. 2nd Int. Conf. New Frontiers in Communication, Automation, Management and Security (ICCAMS 2025), 2025, pp. 1–6.
[3]S. Alqahtani and P. Nanda, “Effects of personal characteristics on phishing awareness, anti-phishing tool usage, and phishing avoidance behavior: A structural equation modeling approach,” in Proc. IEEE Int. Conf. Security of Information and Networks, 2024, pp. 1–8.
[4]M. Ahsan, K. E. Nygard, R. Gomes, et al., “Cybersecurity threats and their mitigation approaches using machine learning—A review,” Journal of Cybersecurity and Privacy, vol. 2, no. 3, pp. 527–555, 2022.
[5] O. A. Aslan and R. Samet, “A comprehensive review on malware detection approaches,” IEEE Access, vol. 8, pp. 6249–6271, 2020.
[6] T. E. Ali, Y.-W. Chong, and S. Manickam, “Machine learning techniques to detect a DDoS attack in SDN: A systematic review,” Applied Sciences, vol. 13, no. 5, p. 3183, 2023.
[7] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, et al., “Deep learning for cybersecurity intrusion detection: Approaches, datasets, and comparative study,” Journal of Information Security and Applications, vol. 50, p. 102419, 2020.
[8] R. K. Mbura, A. K. Benedicto, and R. Sinde, “A Novel Hybrid Approach for Identification of Discriminative Features in Phishing Emails,” IEEE Access, vol. 14, 2026, doi: 10.1109/ACCESS.2025.3649636.
[9] A. S. K. Joseph and S. Srinivasan, “Anti-Phishing Adaptive AI Systems: Efficiently Countering Social Engineering Attacks by Real-Time Analysis of Email Content,” in Proc. 2025 Int. Conf. Computational Innovations and Engineering Sustainability (ICCIES), 2025.
[10] A. Kafi, S. Saha, and N. Shahriar, “Weighted Reciprocal Rank Fusion RAG for Context-Aware DoS Attack Mitigation,” in Proc. 2026 IEEE 23rd Consumer Communications & Networking Conf. (CCNC), 2026.