IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A REVIEW ON UNIFIED MACHINE LEARNING APPROACHES FOR SPAM, PHISHING, QR CODE, AND UPI FRAUD DETECTION

Mohd Zaid, Shreya Jain, Deepanshu Bhargav, Krishna Sharma, Aqib Akhtar Zia

[1]Student, [2]Student,[3] Student, [4]Student, [5]Teacher

Department of Data Science,

Meerut Institue of Engineering and Technology, Meerut, India

Abstract: —With the introduction of digital payments and the likes of QR transactions and web-based services, individuals
become increasingly vulnerable to falling prey to cyber fraud.This can happen if one receives emails from unidentified sources,
visits harmful links, or incorrectly uses QR codes and UPI transactions. As per some research papers, it has been found
that despite being able to control frauds in some areas, we still cannot align ourselves well enough.
In this literature review, I will discuss Sotrian, an online platform using computer algorithms such as logistic regression,
random forest, and XGBoost for predicting fraudulent behavior in email and URL links, QR code scans, and UPI transactions.
The developers behind Sotrian utilized technologies like Next.js and Python to create this web app and include a dashboard for
viewing analytics of detected fraudulent activity. In this context, Sotrian plays a vital role in bridging the existing gap in anti-fraud measures. Specifically, it combines unique aspects of the discussed frameworks and implements
them in a unified approach for monitoring and analyzing online behavior, making it more efficient and user-friendly. In summary,
Sotrian represents a web platform focused on fraud detection within digital payments and QR transactions, web-based services,
and email communications.

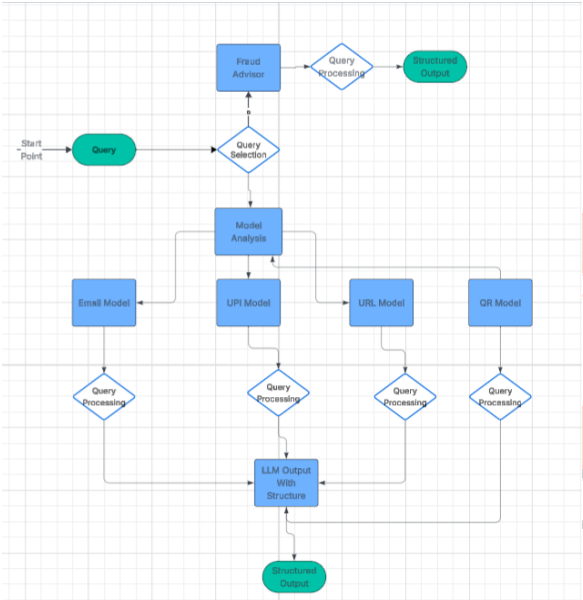
Index Terms - Fraud Detection, Machine Learning, Phishing Detection, QR Code Fraud, UPI Fraud, Sotrian Platform, Multi domain Security

I. INTRODUCTION

The digitalization of communication and finance has made it such that people regularly deal with emails, web pages, QR codes, and even immediate digital payments. However, the use of these technologies comes with an increased risk of being victims of scams and fraud. One of the key areas in which the latter is being carried out is using UPI payments, which involves the use of fake payment instructions, impersonating legitimate accounts, and altering QR codes to make sure that the transfer cannot be reversed. Another way of conducting fraudulent activities is to lure people into phishing web pages and

emails to steal credentials and financial data. The QR code should be paid particular attention to since people cannot inspect its content. Historical fraud detection approaches have primarily been based on rule-checking, static blacklisting, or domain-specific classifiers, and the papers reviewed illustrate time and again that they cannot effectively handle scenarios where the evolution of fraud techniques is rapid, labeled data is scarce, or even where adversaries begin to modify the structures and delivery channels used to transmit messages. Recent research efforts have moved towards approaches based on machine learning, ensemble learning, anomaly detection, and even deep learning, where domain-specific features of emails (TF-IDF), lexical URLs, QR code structure/decoded content, and transaction behavior have been considered. But the papers reviewed point to an issue of fragmentation – one paper looks at phishing emails, another at phishing websites, another at security issues related to QR codes, and yet another at UPI fraud. But in reality, an adversary may deploy multiple vectors in a single fraud attack. Thus, it becomes important to integrate the fragmented evidence from the papers provided in order to design an integrated system architecture. The goal of this review is to consolidate the major findings from the provided literature and to derive a coherent, research-ready description of Sotrian, a proposed multi-domain fraud detection framework that brings together email, URL, QR, and UPI fraud analytics within a common platform..

Fig. 1. Architecture of the unified multi-domain fraud detection system.



II. LITERATURE REVIEW

A. **Summary of Reviewed Studies** The literature survey highlights that the detection of phishing websites has reached an advanced level of maturity and constitutes a mature research field that employs numerous methodologies for detection, such as list-based detection, heuristic detection, visual similarity detection, machine learning detection, and deep learning detection. The systematic survey also revealed that machine learning is the most popular methodology being employed by the majority of studies; PhishTank and Alexa are some of the most common datasets used, while Random Forest is the most prevalent algorithmic approach. Detection of phishing emails, which are part of the common collection of papers, focuses more on public dataset reliability, text preprocessing, vectorization, and implementation in real-world scenarios. Email fraud detection discussed in the attached paper involves treating this problem as a module using public datasets of spam and phishing emails, where the preprocessing stages include tokenization, removing stop words, and TF-IDF vectorization to transform text data into machine learning features. The QR-code fraud literature exposes a distinct challenge because the encoded destination is hidden from the user at scan time. The attached paper explains that QR code images are decoded to extract embedded URLs or UPI payment links, and the extracted content is then analyzed using features similar to URL detection models. These findings support the idea that both decoded-payload analysis and broader QR-centric inspection are useful for pre-scan protection. This becomes quite clear from the research work conducted on the issues related to the problem of payment fraud in UPI systems as well as in other digital payment methods. According to the research presented in

the attached paper, the fraud detection module of the UPI system depends on datasets that contain information such as transaction amount, timestamp, sender and receiver IDs, and frequency of transactions. In the provided instance, SMOTE method is used to overcome the problem of class imbalance. There is one recurring concept that can be found in all the literature that has been provided above: high accuracy can be achieved within the context of specific sub-problems, but there has not been much work done on creating a holistic approach to prevent fraud. The attached paper specifically takes into account this problem and presents a holistic model involving separate models for the email, UPI, URL, and QR modules in one web platform.

TABLE I
SUMMARY OF REVIEWED STUDIES

Reference	Main Fo-cus	Method/Model	Key Insight
Secure QR Code Scanner	QR and malicious URLs	ML-based URL classification	QR codes can conceal malicious links effectively.
Phishing and QR Link for UPI Code	UPI phish-ing and QR fraud	ML on trans-action/link fea-tures	Behavioral analysis improves fraud detection.
Phishing site SLR Web-	Phishing websites	ML and survey DL	Random Forest is widely used in phishing de-tection.
Hybrid Phishing Email Detection	Email phishing	LR, DT, RF	Combining ML improves detection accuracy.
AI-Driven Fraud UPI	UPI fraud	RF, XGBoost, IF	Ensemble methods outperform traditional approaches.
Multimodal Fraud Detection	Multi-domain fraud	NLP, CV, En-semble	Multi-modal data enhances performance.

B. Datasets and Feature Engineering for Fraud Modules :

III. SYSTEM ARCHITECTURE AND METHODOLOGY

Fraud Domain	Common Dataset/Data Type	Main Features	Typical Use
Email phish-ing	Spam/phishing email corpora	TF-IDF, word patterns, suspicious phrases	Detecting spam and phishing emails
Phishing websites	PhishTank, Alexa	URL length, do-main, keywords, structure	Detecting ma-licious URLs
QR fraud	QR-linked URL datasets	Embedded links, decoded content	Identifying malicious QR codes
UPI fraud	Transaction datasets	Amount, time, frequency, sender/receiver patterns	Detecting suspicious payments

Based on the synthesis of the above-discussed articles, it can be recommended that the adopted approach to designing

A solution involves using the principle of modularity where each fraud vector is processed by its own pipeline, but at the same time there is a unified orchestration layer, interface, and dashboard. Based on this approach, the four sources of information used for inputting into the platform include emails, URLs, QR code scanning, and transaction data in UPI. After domain-specific preprocessing, the information goes through the classification/anomaly detection stage, with final aggregation at the user end. For email

fraud detection, the literature consistently supports a text pipeline built on cleaning, tokenization, stop-word removal, normalization, and vectorization using TF-IDF or related NLP features, followed by supervised models such as logistic regression, random forest, SVM, or Naïve Bayes. For URL fraud detection, the methodology centers on lexical and structural attributes, including URL length, token composition, suspicious keywords, domain-related properties, and in some systems external reputation signals, which are then supplied to classifiers such as Random Forest or deep models. For QR fraud detection, one methodological branch decodes the QR code and analyzes the extracted link or payment payload using URL-like features, while another branch inspects QR image structure more directly. For detecting UPI frauds, the use of behavioral factors such as the amount of the transaction, timestamp, sender-receiver factors, frequency-based analysis, and anomaly detection techniques are applied to detect frauds where imbalanced class problems can be handled using techniques such as SMOTE and classified using algorithms like XGBoost, Random Forest, LSTM, or hybrid anomalies. According to the paper, experiments were conducted by using the programming language Python, as well as the libraries Scikit-learn and XGBoost. Datasets were divided into two sets, namely, training and testing sets. Evaluation was performed using accuracy, precision, recall, and AUC, which indicates the need for Sotrian's model not only in terms of improving the prediction but also for allowing modularity, explanation, inference, and metrics in the process.

TABLE III
PROPOSED UNIFIED FRAUD DETECTION ARCHITECTURE

Module	Input Type	Model	Output
Email fraud de-tection	Email text	TF-IDF + Logistic Regression	Spam/phishin classifica- tion
URL fraud tection de-	URL structure	Random Forest	Safe/fraud
QR fraud tection de-	Decoded QR content	Random Forest / anomaly de-tection	Fraud/legitima QR
UPI fraud tection de-	Transaction records	XGBoost SMOTE +	Fraud/non-fraud transaction

IV. PROBLEM STATEMENT

Although there have been advancements in the literature on phishing and fraud detection, the presented articles highlight a structural weakness in the existing mechanisms used to protect against these attacks. A user could be exposed to phishing email that contains an embedded URL leading to a spoofed website, or the same user could scan a QR code that launches an illicit transaction via the UPI method, but traditional approaches fail to account for the overall sequence of events that lead to such incidents. This makes it harder to prevent attacks as separate modules come up with independent decisions regarding the legitimacy of an event. A related issue is the heterogeneity of input data representations, including unstructured text data like emails, structured textual representation like URLs, visual codes such as QR codes, and time-stamped transactional records like UPI fraud cases, which means it would be impractical to rely on a monolithic approach. Moreover, some additional concerns addressed in the accompanying paper include imbalanced datasets, disparate pre-processing steps per module, and the common use of web interfaces to integrate customized models into one coherent platform. Thus, the overall research question revolves around how a universal fraud detection tool could be developed for processing heterogeneous inputs while maintaining high detection rates, enabling near real-time inferences, and providing a seamless user experience for prevention, monitoring, and remediation actions. Sotrian addresses this particular challenge.

V. PROPOSED SOLUTION

mplementation of the solution is through a web-based intelligent application referred to as Sotrian. The application employs a modular approach to detect any frauds that may happen across all online mediums using the same model and platform. On the platform, individuals can enter the relevant data, including emails, websites, QR code scans, and transactions that have been made via UPI on one and the same platform for detection of any fraud. This corresponds to the model of modular fraud detection presented in the attached research paper.

Sotrian is structured as a multi-module fraud detection platform in which each module is responsible for a specific fraud domain. Email fraud detection is performed using textual preprocessing and machine learning classification, URL fraud detection is based on lexical and structural analysis, QR-code fraud detection operates by decoding and inspecting embedded content, and UPI fraud detection relies on behavioral analysis of transaction attributes. The attached paper describes this architecture as a modular system with four autonomous fraud detection modules combined within one web interface.

The frontend framework of Sotrian comprises modern web application frameworks such as Next.js, TypeScript, and Tailwind CSS. The mentioned frameworks provide an extremely professional and intuitive user interface for interacting purposes. In the case of the backend, the process consists of integrating machine learning models that are implemented through Python programming language and Scikit-learn and XGBoost libraries. In order to communicate between the frontend and fraud detection models, the REST API framework is utilized

The logs, actions performed, and results of the fraud detection procedure can be stored using database technologies such as PostgreSQL and MongoDB. Furthermore, security of the platform is guaranteed by use of authentication and access control methods to ensure safety despite handling sensitive data associated with fraud cases. All these implementation components are fully aligned with the platform design described in the accompanying paper. All in all, Sotrian is a feasible cross-domain solution for detecting fraud. All of the specialized fraud detection modules are aggregated into a web platform. This design helps to simplify its utilization and also minimizes fragmentation while analyzing fraud thereby enabling central monitoring with scalability options for future system expansion.

VI. RESULTS AND DISCUSSION

Results The performance results reported in the literature review indicate that machine learning approaches are able to attain high effectiveness on all four targeted fraud domains (email phishing, website phishing, QR code fraud and UPI transactions), although results vary substantially by data source quality and feature representation strategies as well as modeling techniques utilized.

Developed web-based systems in phishing email detection universally provide solid classification performance when aided by thorough preprocessing and supported with sufficiently representative datasets, demonstrating that this domain has reached relative maturity. Similar to email and webmail phishing detection, the superiority of machine learning approaches (largely based on convolutional neural network models) in controlled comparisons is widely reported [33], though interpretation is complicated due to markedly different datasets used across studies.

The use of QR codes for detecting fraud results in a more complex set of observations, yet at the same time shows strong potential. The safe QR scan example proves the superiority of deep learning compared to machine learning methods, while the quishing attacks case highlights the importance of directly analyzing QR codes using gradient boosted models.

The paper on the unified fraud detection framework shows very good results in terms of module-wise performance within the integrated framework, where email, UPI, URL, and QR modules all demonstrate strong classification behavior. Although these results reflect individual module performance rather than a full end-to-end multi-domain evaluation, they still confirm the feasibility of a modular unified approach for fraud detection tasks.

However, certain challenges remain evident from the literature. High performance in controlled environments does not always translate to robustness in real-world conditions due to issues such as distribution shifts, adversarial attacks, lack of labeled data, and continuously evolving fraud techniques. Therefore, the findings emphasize the importance of modularity, explainability, adaptability, and deployability in designing effective fraud detection systems.

VII. CONCLUSION AND FUTURE WORK

The literature review points out that a number of advances have been made with regard to various frauds such as email phishing, phishing website attack, QR code attack, and transaction fraud within Unified Payments Interface (UPI). However, the critical issue that must be addressed in relation to cybersecurity is the lack of an efficient platform that will incorporate these methods of fraud detection into a cohesive system.

It would appear that Sotrian aims to address this issue in some ways. Given its modularity and integration features, it would seem reasonable that it should be used to address fraud detection issues. In any case, there is no loss of comprehensive coordination in applying specialized optimization for specific models within this system.

This brings me to what I think to be the crucial aspect of Sotrian: combining its strengths regarding specialized optimization – the advantages of having such optimizations – and practical solutions such as integrated monitoring and real-time reactions to them. This part may be a bit difficult to explain in detail. According to the literature, several directions worth exploring further to bring significant benefits to research in the field include, first of all, the choice of the dataset. Indeed, while current research focuses predominantly on fraudulent activities and cases in their isolation, moving toward more complex cross-channel attacks can significantly increase evaluation capabilities for unified platform solutions, which seem to be critical under adversarial conditions.

Second, standardizing and promoting explainable AI in all detection modules appears crucial from both technical and security perspectives. This will help develop user trust and facilitate monitoring, allowing security decisions to remain transparent. Otherwise, implementation and reliance on such an approach will hardly be possible.

Regarding deployment strategies, reinforcement of real-time cloud-based solutions to allow more efficient scalable inference and log/alerting from multiple data flows seems highly necessary and promising. In addition, federated learning and privacy-preserving model updating in case of certain restrictions (such as financial and communications frauds where regulatory bodies prohibit centralization of data) seem valuable directions.

In any case, prioritizing these directions may accelerate advancements in the field, although timeframes for it are hard to predict due to limitations imposed on research.

REFERENCE

[1] R. A. Kumar et al., “AI-Driven Detection Mechanism for UPI Fraud and QR Code Tampering,” Proc. ICICV, 2025.

[2] G. R. Charan and K. D. Thilak, “Detection of Phishing Link and QR Code of UPI Transaction using Machine Learning,” Proc. ICIMIA, 2023.

[3] N. Lingareddy et al., “Enhancing Digital Payment Security using Machine Learning Algorithms,” Proc. GINOTECH, 2025.

[4] A. Pawar et al., “Secure QR Code Scanner to Detect Malicious URL using Machine Learning,” Proc. ASIANCON, 2022.

[5] F. Trad and A. Chehab, “Detecting Quishing Attacks with Machine Learning Techniques Through QR Code Analysis,” arXiv, 2025.

[6] A. Khalid et al., “LogiTriBlend: Hybrid Model for Phishing Email Detection,” IEEE Access, 2024
.

[7] A. Safi and S. Singh, “Systematic Review on Phishing Website Detection Techniques,” Journal of King Saud University, 2023.

[8] P. Bhinge et al., “Threat Guard AI: Multi-Domain Fraud Detection System,” IJRASET, 2025.

[9] A. S. Pundir et al., “Fraud Fortify: Multi-Domain Fraud Detection System,” JETIR, 2024.

[10] O. I. Odufisan et al., “Harnessing Artificial Intelligence for Fraud Detection,” Journal of Economic Criminology, 2025.

[11] “Multi-Modal Fraud Detection in Digital Transactions,” IJIRT, vol. 11, no. 12, 2025.

[12] “A hybrid approach to phishing email detection: leveraging machine learning and explainable artificial intelligence,” Int. J. Elect. Comput. Eng., vol. 15, no. 5, pp. 1234–1245, 2025.

[13] “Auto Insurance Fraud Detection with Multimodal Learning,” Data ntelligence, vol. 5, no. 2, pp. 388–405, 2023.

[14] A. H. M. Aburbeian et al., “Credit Card Fraud Detection Using Enhanced Random Forest,” arXiv:2303.06514, 2023.