SECUREPAY: An Explainable Cross-Modal AI Framework for UPI Payment Evidence Verification Using Screenshot Forensics, QR Intent Analysis, and Transaction Anomaly Detection

Mayur Dnyaneshwar Kapadi, Aditya Ramdas Naikade, Nikhil Balu Kapuskar, Yash Kantaram Hembade
Department of Master of Computer Applications, D.Y. Patil Institute of Master of Computer Applications and Management, Akurdi, Pune, India
Guide - Prof. Rahul Chaudhari

Abstract— The widespread use of the Unified Payments Interface (UPI) has made digital payments in India faster and more convenient. At the same time, it has created new opportunities for fraud, including edited payment screenshots, malicious QR codes, phishing attempts, and false claims of successful payment. Many existing methods examine image manipulation, QR-code content, or unusual transactions separately. As a result, they may miss inconsistencies that become clear only when these sources are considered together. This paper presents SECUREPAY, an explainable cross-modal artificial intelligence framework that brings these forms of analysis into one process. It combines screenshot forensics, Optical Character Recognition (OCR), QR-code intent analysis, and machine-learning-based transaction anomaly detection. A cross-modal consistency mechanism compares details such as payment amounts, beneficiary information, transaction identifiers, payment status, and payment intent across the evidence that is available. A weighted risk model then classifies the evidence as low, medium, or high risk and provides reasons for the assessment. The framework is intended to be evaluated through module-wise and integrated experiments, comparative analysis, ablation studies, robustness testing, and cross-dataset validation. Since experimental validation has not yet been completed, no performance improvement is claimed. SECUREPAY is designed to help users and merchants assess suspicious payment evidence; it does not replace independent confirmation that a transaction has been settled.

Keywords— UPI, Digital Payment Fraud, Artificial Intelligence, Computer Vision, OCR, QR Code Security, Image Forensics, Transaction Anomaly Detection, Cross-Modal Verification, Explainable AI.

I. INTRODUCTION

The Unified Payments Interface (UPI) has become one of India's widely used digital payment methods. It enables convenient, interoperable, and real-time transfers through mobile applications and QR codes. However, its growing use has also created opportunities for fraudulent activity. For example, an attacker may edit a payment screenshot by changing the amount, beneficiary name, timestamp, reference identifier, or transaction status. Such an image can persuade a merchant or another individual to accept a payment claim even when no corresponding transaction has taken place.

QR codes present another type of risk. A malicious or substituted code may contain an unintended UPI address, an altered payment amount, unauthorised beneficiary details, or a phishing URL. If users scan a code without checking its contents, they may unknowingly initiate a payment to the wrong recipient.

Each existing technique has a different role and a different limitation. Image forensics can identify signs of visual editing, but it cannot confirm that a payment was completed. OCR reads the text shown in an image, although that text may itself have been fabricated. QR decoding reveals the information encoded in a code, but does not establish whether the beneficiary is genuine. Transaction models can identify unusual patterns in structured records, but may not connect those records to the screenshot or QR code presented as payment evidence. SECUREPAY addresses this gap by examining screenshots, decoded QR information, and available transaction-related records together.

The proposed framework contains four main analysis components: screenshot forensics, OCR-based information extraction, QR intent analysis, and transaction anomaly detection. Their outputs are brought together through cross-modal consistency verification and an explainable risk assessment stage. When a particular source of evidence is missing, the system records it as unavailable rather than treating its absence as proof of fraud. Similarly, neither a screenshot nor a QR code alone is considered confirmation that money has been transferred.

II. RELATED WORK

Research on digital payment fraud has explored different approaches, including image forensics, Optical Character Recognition (OCR), QR-code security, and machine-learning-based fraud detection. These approaches help identify manipulated images, examine QR-code information, and recognise unusual transaction activities. However, most of these techniques focus on individual aspects of payment security rather than examining different sources of payment evidence together.

1. Image Forensics and OCR

Digital image forensics is used to identify possible modifications in images, such as copy-move forgery, image splicing, resampling, and compression inconsistencies. Several deep-learning-based techniques have been studied to detect image manipulation and identify affected regions [1], [2]. Such techniques are relevant to UPI payment screenshots, where details such as transaction amounts, beneficiary names, timestamps, and payment statuses can be modified to create misleading payment evidence.

OCR is another useful technique for extracting payment-related information from screenshots. Smith [3] introduced the Tesseract OCR engine, which provides a method for recognising text in images and documents. In payment verification, OCR can help extract transaction details and convert them into structured information for further analysis. However, extracting text does not guarantee that the

displayed information is genuine. Likewise, image-forensic analysis alone cannot confirm whether a payment has actually been completed.

2. QR Code Security

QR codes have made digital payments easier by allowing users to access payment information through scanning. However, they can also be misused to redirect users to phishing websites or unintended payment destinations. Previous studies have examined security vulnerabilities in QR-code-based applications and proposed methods to improve their safe usage [4].

Sharevski et al. [5] studied phishing attacks involving malicious QR codes, showing how attackers can use QR codes to direct users towards deceptive destinations. Similarly, Njuguna and Ndia [6] reviewed different QR-code security attacks and their countermeasures. Mossano et al. [7] explored the development of SEQR, a secure QR-code scanning application designed to improve protection against unsafe QR-code content.

These studies highlight the importance of checking the information stored within a QR code before proceeding with a payment. In UPI transactions, decoding a QR code can reveal details such as the beneficiary's VPA, payee name, and payment amount. However, successfully decoding a QR code does not necessarily mean that its beneficiary is genuine or that a transaction has been completed.

3. Financial Fraud Detection and Machine Learning

Machine learning has become an important approach in financial fraud detection because it can analyse transaction data and identify patterns that may indicate suspicious activity. Al-Hashedi and Magalingam [8] reviewed data-mining techniques used for financial fraud detection, while Hilal et al. [9] examined different anomaly-detection methods and their applications in financial systems. Hernandez Aros et al. [10] also discussed the use of machine-learning techniques for identifying fraudulent financial activities.

Research has also focused specifically on detecting fraudulent UPI transactions [11]–[14]. These studies demonstrate how data-driven methods can support the identification of suspicious payment activities. However, SECUREPAY focuses on cross-checking payment details across screenshots, QR-code payloads, and available transaction records to provide an explainable risk assessment based on inconsistencies between these evidence sources. In this context, Random Forest can be used for supervised classification when labelled transaction data are available, whereas Isolation Forest can help identify unusual transaction patterns without requiring labelled fraud examples.

Although these methods are useful for analysing transaction behaviour, they may not identify differences between the information displayed in a payment screenshot, the details encoded in a QR code, and the corresponding transaction records.

4. Research Gap

Despite progress in these areas, payment evidence is often analysed separately. SECUREPAY proposes cross-checking available screenshot, QR-code, and transaction details to identify inconsistencies and provide an explainable risk assessment.

For example, a payment screenshot may display a successful transaction even when the corresponding transaction details are inconsistent with the available records. Similarly, the beneficiary information encoded in a QR code may differ from the details displayed in the screenshot. Such situations highlight the need for a method that can examine these sources together.

To address this gap, SECUREPAY proposes an integrated framework combining screenshot forensics, OCR-based information extraction, QR intent analysis, and transaction anomaly detection. It further introduces cross-modal verification to compare available payment details and an explainable risk-assessment mechanism to highlight suspicious inconsistencies. The proposed framework will require experimental evaluation to establish its effectiveness

III. METHODOLOGY

SECUREPAY adopts a modular research approach to examine different forms of UPI payment evidence. The framework brings together screenshot forensics, Optical Character Recognition (OCR), QR intent analysis, and transaction anomaly detection. The results obtained from these modules are combined through cross-modal consistency verification to produce an explainable risk assessment.

The proposed methodology is divided into five stages: dataset preparation, preprocessing and feature extraction, individual module analysis, cross-modal verification, and experimental evaluation.

1. Dataset Preparation

The dataset planned for SECUREPAY includes three types of payment evidence: UPI payment screenshots, QR-code images, and transaction-related records. Authentic screenshots will be collected with appropriate permission, while synthetic screenshots will be generated separately to represent realistic payment layouts. Additional screenshots will be prepared by making controlled changes to details such as transaction amounts, beneficiary names, timestamps, and payment status.

The QR-code dataset will contain legitimate payment information as well as suspicious examples with modified parameters, unexpected URLs, or potentially harmful content. Transaction records will include anonymised or synthetically generated details such as transaction amounts, frequency, and time-related attributes. Public datasets related to image manipulation and financial fraud may be used for module-specific evaluation. However, they will not be treated as paired UPI payment evidence or as a substitute for an integrated UPI dataset. To prevent data leakage, related screenshots, QR-code samples, and transaction records from the same source or transaction will be kept within the same dataset split during training, validation, and testing.

2. Preprocessing and Feature Extraction

Before analysis, the collected data will undergo preprocessing to make it suitable for the individual modules. Screenshot images will be resized and standardised, followed by image processing to examine possible visual inconsistencies.

Tesseract or EasyOCR will be used to extract payment details such as transaction amounts, beneficiary names, UPI IDs, timestamps, payment status, and reference numbers.

QR-code images will be decoded to obtain their embedded payment information. Transaction records will also be examined for missing values, duplicate entries, and inconsistent data formats. Sensitive information will be anonymised or transformed before being used for further processing.

3. Individual Analysis Modules

SECUREPAY consists of four main modules, each responsible for analysing a different aspect of payment evidence.

- Screenshot Forensics: This module uses OpenCV-based image processing and forensic features to examine compression patterns, visual differences, and areas that may have been modified. Particular attention will be given to important payment details, including transaction amounts, beneficiary names, timestamps, and payment status.

- OCR-Based Information Extraction: Tesseract or EasyOCR will be used to read payment-related text from screenshots. The extracted information will be organised into structured fields, while OCR confidence scores will be retained to help identify uncertain or incorrectly recognised details.

- QR Intent Analysis: QR-decoding libraries will be used to inspect the information encoded in UPI payment QR codes. The analysis will cover beneficiary VPA, payee name, transaction amount, currency, and other payment parameters. Malformed payloads and suspicious URLs will also be examined. UPI payment URIs and ordinary web URLs will be analysed separately because they represent different types of QR-code content.

- Transaction Anomaly Detection: Random Forest will be explored as a supervised learning method using labelled transaction data to classify suspicious activity. Isolation Forest will be used as an unsupervised anomaly-detection method to identify unusual transaction patterns based on attributes such as transaction amount, frequency, and time-related information.

4. Cross-Modal Consistency Verification and Risk Assessment

After the individual modules complete their analysis, their outputs will be brought together to check whether the available payment information is consistent. The framework will compare transaction amounts, beneficiary details, reference identifiers, payment status, and payment intent across screenshots, QR-code information, and transaction records.

To combine the results, SECUREPAY proposes a weighted risk-scoring mechanism:

$$R = w_1 S + w_2 Q + w_3 T + w_4 C$$

Where:

- (R) represents the overall risk score.
- (S) represents screenshot manipulation risk.
- (Q) represents QR intent risk.
- (T) represents transaction anomaly risk.
- (C) represents cross-modal inconsistency risk.
- (w_1,w_2,w_3,w_4) represent the weights assigned to the individual components.

The component scores will first be normalised before being combined. The weights and classification thresholds will be determined through validation experiments. Based on the resulting score, the framework will classify the payment evidence into three categories: Low Risk, Medium Risk, and High Risk. In the proposed design, all component weights will be non-negative and their sum will equal one. The weights and risk-classification thresholds will be selected using validation data before final testing.

If any payment information is unavailable, the framework will record it as missing rather than automatically considering it suspicious. It will also provide explanations highlighting the indicators that influenced the final risk classification.

5. Experimental Evaluation

The proposed framework will be evaluated through four configurations: screenshot forensics combined with OCR, QR intent analysis, transaction anomaly detection, and the complete SECUREPAY framework. Their performance will be compared to understand whether combining the different modules provides additional value in identifying suspicious payment evidence.

The evaluation will consider accuracy, precision, recall, F1-score, false-positive rate, and detection latency. Additional tests will examine how the framework performs when screenshots undergo resizing, JPEG compression, blurring, and noise. Ablation studies will also be conducted by removing individual components to understand their contribution to the overall framework. Confusion matrices and appropriate statistical tests will be used to support the evaluation.

IV. DISCUSSION AND LIMITATIONS

SECUREPAY is currently in the proposed research and development stage, and the complete experimental results are not yet available. For this reason, the paper does not include unverified accuracy values or make claims about performance improvements. Instead, the evaluation is designed to examine how each analysis module performs and whether combining them through cross-modal verification can provide a more complete assessment of suspicious payment evidence.

The screenshot-forensics module will be examined to determine how effectively it identifies possible modifications in payment screenshots, particularly changes involving transaction amounts, beneficiary details, timestamps, and payment status. The OCR module will also be assessed based on how accurately it extracts payment information from screenshots under different image conditions.

The QR intent analysis module will be tested for its ability to decode payment information, recognise incorrectly structured parameters, and identify potentially suspicious payment destinations. Similarly, Random Forest and Isolation Forest will be evaluated to examine their ability to identify unusual patterns in transaction records.

The complete SECUREPAY framework will be compared with the individual analysis modules to understand the contribution of combining different sources of payment evidence. The evaluation will focus on identifying inconsistencies in transaction amounts, beneficiary information, reference identifiers, and payment status across the available sources.

The weighted risk-assessment mechanism will also be examined to determine whether it can provide clear reasons for assigning Low Risk, Medium Risk, and High Risk categories. Ablation studies will further help assess the contribution of each module to the overall framework.

Once implementation and experimental testing are completed, the findings will be presented through performance tables, confusion matrices, and comparative graphs. These results will help assess the usefulness of the proposed framework and identify areas that require further improvement.

A. Limitations

The performance of SECUREPAY depends largely on the quality and availability of payment screenshots, QR-code information, and transaction records. Poor image resolution, compression, variations in fonts, and advanced image manipulation may affect the reliability of screenshot analysis and OCR extraction.

QR-code analysis also has certain limitations, as successfully decoding a payment QR code does not guarantee that the beneficiary is genuine. Similarly, transaction anomaly detection depends on the availability of suitable datasets and reliable labels. Another challenge is obtaining independent transaction records, as access to such information may be restricted by privacy requirements and financial regulations

V. FUTURE WORK

Further development of SECUREPAY will focus on making the proposed framework more reliable, accurate, and suitable for practical use. Several areas can be explored to improve its capabilities.

1. Advanced Image Forensics: Future work can explore deep-learning-based techniques to identify more complex screenshot manipulations, including changes that are difficult to detect using conventional image-processing methods.

2. Improved QR Security Analysis: QR-code analysis can be extended to examine malicious redirects, suspicious payment parameters, and unauthorised changes in beneficiary information.

3. Enhanced Explainable AI: Techniques such as SHAP can be explored to provide better explanations of how transaction features influence anomaly detection and the final risk classification.

4. Real-Time Payment Verification: Integration with authorised transaction-verification APIs can be investigated, subject to the required permissions and applicable financial regulations. This could help strengthen the verification of payment-related evidence.

5. Cross-Platform Evaluation: The framework can be tested across different UPI applications, QR-code formats, image conditions, and transaction scenarios to understand how well it performs in different environments.

VI. CONCLUSION

This paper introduced SECUREPAY, an explainable cross-modal AI framework proposed for UPI payment evidence verification. It brings together screenshot forensics, OCR-based information extraction, QR intent analysis, and machine-learning-based transaction anomaly detection within a single framework.

The main focus of SECUREPAY is to examine payment information across different evidence sources rather than analysing each source separately. It compares details such as transaction amounts, beneficiary information, reference identifiers, payment status, and payment intent. Based on these comparisons, a weighted risk-assessment mechanism is proposed to classify payment evidence into Low Risk, Medium Risk, and High Risk categories, along with explanations for the identified indicators.

SECUREPAY aims to identify suspicious payment evidence by examining inconsistencies across available sources. However, its actual effectiveness and reliability still need to be established through implementation and experimental evaluation.

Future research will focus on developing the framework, preparing suitable datasets, conducting comparative experiments, and improving the individual analysis modules. SECUREPAY is intended to assist in assessing potentially suspicious payment evidence; it does not independently confirm whether a financial transaction has been successfully completed.

VII. REFERENCES

[1] M. Zanardelli, F. Guerrini, R. Leonardi, and N. Adami, “Image forgery detection: A survey of recent deep-learning approaches,” Multimedia Tools and Applications, vol. 82, pp. 17521–17566, 2023, doi: 10.1007/s11042-022-13797-w.

[2] V. Kadha, S. Bakshi, and S. K. Das, “Unravelling digital forgeries: A systematic survey on image manipulation detection and localization,” ACM Computing Surveys, vol. 57, no. 12, Art. no. 323, 2025, doi: 10.1145/3731243.

[3] R. Smith, “An overview of the Tesseract OCR engine,” in Proc. 9th Int. Conf. Document Analysis and Recognition (ICDAR), 2007, pp. 629–633, doi: 10.1109/ICDAR.2007.4376991.

[4] K. Krombholz, P. Frühwirt, T. Rieder, I. Kapsalis, J. Ullrich, and E. Weippl, “QR code security: How secure and usable apps can protect users against malicious QR codes,” in Proc. 10th Int. Conf. Availability, Reliability and Security (ARES), 2015, pp. 230–237, doi: 10.1109/ARES.2015.84.

[5] F. Sharevski, A. Devine, E. Pieroni, and P. Jachim, “Phishing with malicious QR codes,” in Proc. European Symp. Usable Security (EuroUSEC), 2022, doi: 10.1145/3549015.3554172.

[6] D. Njuguna and J. Ndia, “Quick response code security attacks and countermeasures: A systematic literature review,” Journal of Cyber Security, vol. 7, no. 1, pp. 1–20, 2025, doi: 10.32604/jcs.2025.059398.

[7] M. Mossano, M. F. Veit, T. Länge, B. M. Berens, F. Sharevski, and M. Volkamer, “Development, evaluation, and implementation of SEQR—A usable secure QR code

scanner," in Proc. 2026 CHI Conf. Human Factors in Computing Systems (CHI '26), Art. no. 1306, 2026, doi: 10.1145/3772318.3793213.

[8] K. G. Al-Hashedi and P. Magalingam, "Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019," Computer Science Review, vol. 40, Art. no. 100402, 2021, doi: 10.1016/j.cosrev.2021.100402.

[9] W. Hilal, S. A. Gadsden, and J. Yawney, "Financial fraud: A review of anomaly detection techniques and recent advances," Expert Systems with Applications, vol. 193, Art. no. 116429, 2022, doi: 10.1016/j.eswa.2021.116429.

[10] L. Hernandez Aros, L. X. Bustamante Molano, F. Gutierrez-Portela, J. J. Moreno Hernandez, and M. S. Rodríguez Barrero, "Financial fraud detection through the application of machine learning techniques: A literature review," Humanities and Social Sciences Communications, vol. 11, Art. no. 1130, 2024, doi: 10.1057/s41599-024-03606-0.

[11] M. Tamilselvi, R. Begum, K. Krishnamoorthi, J. Giri, D. Sheela, and M. O. Sabri, "Experimental evaluation unified payment interface (UPI) fraud detection system using elevated deep learning methodology," in Proc. 2nd Int. Conf. Emerging Trends in Engineering and Medical Sciences (ICETEMS), 2024, doi: 10.1109/ICETEMS64039.2024.10965083.

[12] A. Khedekar and G. Manerkar, "Machine learning-driven detection of UPI fraudulent activities," in Proc. Global Conf. Information Technology and Communication Networks (GITCON), 2025, doi: 10.1109/GITCON65266.2025.11376733.

[13] A. Garande, S. Battul, A. Gunjal, V. Kshirsagar, and V. Shigwan, "Secure UPI: Intelligent Fraud Detection System (AI-Powered UPI Fraud Detection System using Hybrid Learning and Behavioural Analytics)," in Proc. 2nd Int. Conf. Information, Implementation and Innovation in Technology (I2ITCON), 2026, doi: 10.1109/I2ITCON69735.2026.11747583.

[14] A. Das, N. Das, P. Biswas, and N. Kar, "Machine learning based fraud detection in UPI transactions," in Proc. 4th IEEE Int. Conf. Technology, Engineering, Management for Societal Impact using Marketing, Entrepreneurship and Talent (TEMSMET), 2025, doi: 10.1109/TEMSMET65536.2025.11467122.