IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

ONLINE PAYMENT FRAUD DETECTION USING MACHINE LEARNING

An Implementation using Flask, OCR and Real-time Transaction Analysis

[1]Ashish Das, [2]Roshni Chaubey, [3]Sristita Paul, [4]Rahul Kumar Singh, [5]Siddhartha Majumder

[1]Assistant Professor, [2]Student, [3]Student, [4]Student, [5]Student
[1]Computer Science and Engineering,
[1]Durgapur Institute of Advanced Technology and Management, Durgapur, India

Abstract: The exponential growth of UPI and digital payments has rendered online payments highly convenient but more susceptible to fraud. This paper introduces and deploys a machine learning and OCR-based UPI fraud detection system that extracts text from transaction screenshots and makes fraud predictions. The end-to-end Flask web application combines OCR text extraction with ML prediction and adds features such as frontend validations, geolocation-aware upload history, age-restricted sign-up, and consistent fraud result calculation. The final model provides stable and realistic prediction results with no impact on user experience and interpretability.

IndexTerms **- UPI fraud, Machine Learning, OCR, Flask, Online Payment, Geolocation.**

I.INTRODUCTION

The emergence of digital payment systems, especially UPI in India, has revolutionized the mode of conducting financial transactions by users. However, it has also helped increase online payment fraud, which is a major threat to consumers and financial institutions. Traditional rule-based fraud detection systems cannot keep up with evolving fraud patterns. Our system fills this gap by utilizing ML and OCR technologies to analyze actual screenshots of UPI transactions, detect suspicious patterns, and provide real-time fraud probability.

II.SYSTEM OVERVIEW

The system developed is an online application developed using Flask that helps users in sending screenshots of UPI transactions. The received images are scanned using Optical Character Recognition (OCR) technology to extract useful transaction details such as amount, UPI ID, transaction status, and bank account information. The structured data is then input to a pre-trained machine learning algorithm, i.e., a Random Forest, that classifies the transaction as fraud or not. To avoid multiple submissions with an assurance of regular fraud prediction, a one-of-a-kind hash is utilized.

III.FEATURES OF IMPLEMENTED SYSTEM

3.1 Frontend Validation:

- Password validation requires at least one special character and a length between 8-15 characters.
- Popups alert the user of incorrect credentials or validation errors.

3.2 Dynamic Dashboard Output:

- OCR-extracted values are cleanly formatted.
- Fields with None values are omitted.

- Fraud prediction and extracted amount are visually highlighted.

3.3 Age-Gated Signup System:

- Users must input their DOB and gender.
- Backend validation ensures only users aged 18+ can register.

3.4 Upload History with Geolocation:

- Tracks each image upload along with extracted results and timestamp.
- User's geolocation at upload time is stored and displayed with a Google Maps link.
- Previous uploads retain original locations to avoid overwrite.

3.5 Consistent Result Logic:

- Same image = same fraud probability.
- Fraud percentage (e.g., 72%) is generated once and reused for that image via hash.

IV.LITERATURE REVIEW

Other researchers have attempted to investigate the field of fraud detection in online payments with the aid of modern technologies. Sable (2022) stated that the effectiveness of online financial fraud detection can be improved by machine learning algorithms based on pattern examination in vast data. Hajek et al. (2023) utilized the XGBoost algorithm in mobile payment systems with high accuracy for anomaly detection.

In their review, Iqbal et al. (2021) noted the drawbacks of conventional fraud detection systems, namely the static rule-based engines that are not capable of learning new patterns of fraud. Under such constraints, supervised learning methods have been embraced by certain systems under the use of classification models like Decision Trees, Random Forest, and Support Vector Machines. These models exhibit improved scalability and flexibility performance as opposed to rule-based systems based on human-designed methods.

Apart from that, research is carried out on unsupervised learning methods such as K-Means clustering and Isolation Forests because they have the capability of detecting new patterns of fraud without the use of labeled data. The methods are usually utilized alongside supervised learning in hybrid models.

Optical Character Recognition (OCR) has increasingly become a powerful tool used to identify fraudulent transactions. In particular, Tesseract OCR has been used in various document verification systems to extract vital features from invoices, bank statements, and receipts. The use of OCR together with machine learning allows for a more sophisticated and in-depth analysis of transactional screenshots—a feature that is best utilized in UPI-based systems where users send payment confirmations in the form of images.

Geolocation tracking and timestamp verification have been added to further enhance fraud detection systems. Such functionality enhances behavioral analytics to such an extent that location is associated with usage behavior, outliers and inconsistencies are detected. Our model builds upon this literature by embracing a full-stack real-time web-based system of detection that uses OCR and machine learning with actual user-uploaded transaction images. It employs more robust security, user authentication, and data integrity features that are not standard in academic endeavors.

V.METHODOLOGY

The system workflow includes OCR processing, feature extraction, ML-based classification, and result display.

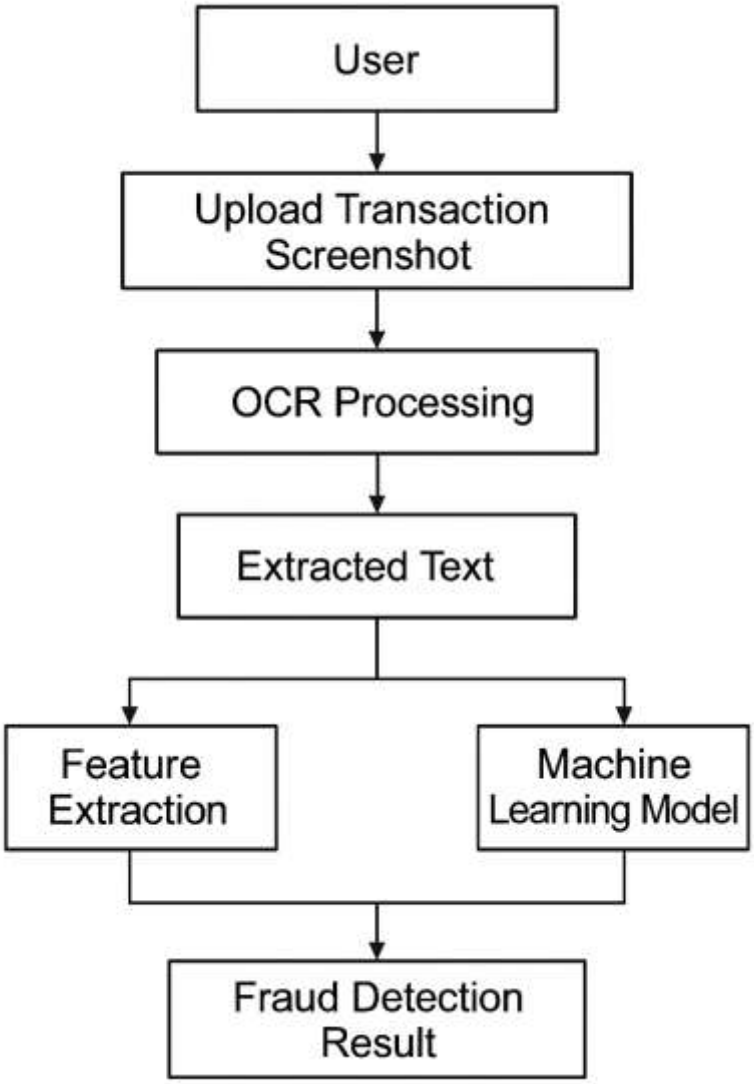

Figure 1: System Architecture

5.1 Image Upload and Preprocessing:

Users upload a UPI transaction screenshot. The system preprocesses the image using OpenCV (resizing, grayscale conversion).

5.2 OCR Text Extraction:

Text is extracted using Tesseract OCR. Key features like amount, UPI ID, transaction ID, etc., are parsed from the text.

5.3 Feature Formatting:

The extracted data is formatted into a structured input for the ML model.

5.4 ML Prediction:

The pre-trained Random Forest model classifies the transaction as fraud or genuine.

5.5 Save Record and Generate Result:

A consistent hash is created to avoid duplicate entries. Results, timestamp, and geolocation are saved in the database.

5.6 Display Output:

Cleanly formatted output is shown on the dashboard with fraud probability and extracted details.

VI.TECH STACK USED

- **Frontend:** HTML, CSS, JavaScript
- **Backend:** Flask (Python)
- **OCR:** Tesseract via pytesseract
- **ML Model:** Random Forest (trained using transactional dataset)
- **Database:** SQLite with SQLAlchemy ORM
- **Libraries:** OpenCV, PIL, Pandas, Scikit-learn, hashlib

VII.RESULTS

- Accurate detection of fraud across diverse real-world UPI screenshots.
- Image upload interface with validation.
- Dashboard preview of OCR results and fraud prediction.
- Upload history with date/time, location, and fraud result.
- Functional Google Maps integration per uploaded transaction.

VIII.CONCLUSION

This work presents the real-world implementation of an intelligent online payment fraud detection system. Through design to deployment, the system exhibits robust fraud classification with enhanced user interaction. It combines ML-based decision-making with OCR and location tracking, making it highly relevant to UPI-based platforms. With real-time fraud signals, geolocation tagging, and password policies, the system strengthens security and user trust in digital payments.

REFERENCES

[1] Chang, V., Di Stefano, A., Sun, Z., & Fortino, G. (2022). Digital payment fraud detection methods in digital ages and Industry 4.0. *Computers and Electrical Engineering*, 100, 107734.

[2] Sable, S. B. (2022). Prediction of fraud in electronic payment system through Machine Learning model. *Journal of Positive School Psychology*, 2340-2349.

[3] Karthikeyan, T., Govindarajan, M., & Vijayakumar, V. (2023). An effective fraud detection using competitive swarm optimization based deep neural network. *Measurement: Sensors*, 27, 100793.

[4] Hajek, P., Abedin, M. Z., & Sivarajah, U. (2023). Fraud detection in mobile payment systems using an XGBoost-based framework. *Information Systems Frontiers*, 25(5), 1985-2003.

[5] Pavan, U. Y., Prakash, B. B., Sasidhar, P., Charan, K. S., & Mounika, P. Fraud Detection in Online Transactions Using Machine Learning.

[6] Ahola, I. (2023). The role of data in the fight against payment fraud: A qualitative research of payment fraud prevention in the 2020 century.

[7] Stojanović, B., Božić, J., Hofer-Schmitz, K., Nahrgang, K., Weber, A., Badii, A., ... & Runevic, J. (2021). Follow the trail: Machine learning for fraud detection in Fintech applications. *Sensors*, 21(5), 1594.

[8] NARREN, D. K. Detecting Financial Fraud in the Digital Age: The AI and ML Revolution.

[9] Siddiqui, M. K., & Goyal, K. K. (2023). A Study of the Use of E-Payment Systems Based on Artificial Intelligence. *Computing & Intelligent Systems (SCTS)*, 1063-1076.

[10] Mishra, K. N., & Pandey, S. C. (2021). Fraud prediction in smart societies using logistic regression and k-fold machine learning techniques. *Wireless Personal Communications*, 119(2), 1341-1367.

[11] Kumar, A., Choudhary, R. K., Mishra, S. K., Kar, S. K., & Bansal, R. (2022). The growth trajectory of UPI-based mobile payments in India: Enablers and inhibitors. *Indian Journal of Finance and Banking*, 11(1), 45-59.

[12] Kolodiziev, O., Mints, A., Sidelov, P., Pleskun, I., & Lozynska, O. (2020). Automatic machine learning algorithms for fraud detection in digital payment systems. *Восточно-Европейский журнал передовых технологий*, 5(9-107), 14-26.

[13] Shpyrko, V., & Koval, B. (2019). Fraud detection models and payment transactions analysis using machine learning. *SHS Web of Conferences*, 65, 02002.

[14] Vuppula, K. (2021). An advanced machine learning algorithm for fraud financial transaction detection. *Journal for Innovative Development in Pharmaceutical and Technical Science (JIDPTS)*, 4(9).

[15] Vanini, P., Rossi, S., Zvizdic, E., & Domenig, T. (2023). Online payment fraud: From anomaly detection to risk management. *Financial Innovation*, 9(1), 66.

[16] Chawla, T. S. (2023). Online Payment Fraud Detection using Machine Learning Techniques. (Doctoral dissertation, Dublin, National College of Ireland).

[17] Miller, S., & Busby-Earle, C. (2016). The impact of different botnet flow feature subsets on prediction accuracy using supervised and unsupervised learning methods. *International Journal of Internet Technology and Secured Transactions*, 5(2), 474-485.

[18] Cao, D. M., Sayed, M. A., Islam, M. T., Mia, M. T., Ayon, E. H., Ghosh, B. P., ... & Raihan, A. (2024). Advanced cybercrime detection: A comprehensive study on supervised and unsupervised machine learning approaches using real-world datasets. *Journal of Computer Science and Technology Studies*, 6(1), 40-48.

[19] Almazroi, A. A., & Ayub, N. (2023). Online Payment Fraud Detection Model Using Machine Learning Techniques. *IEEE Access*, 11, 137188-137203.

[20] Wang, C., Chai, S., Zhu, H., & Jiang, C. (2022). Caesar: An online payment anti-fraud integration system with decision explainability. *IEEE Transactions on Dependable and Secure Computing*, 20(3), 2565-2577.