World Journal of Innovation and Modern Technology, Vol. 7, Issue 2, (Apr) 2024
ISSN 2682-5910
DOI: 10.53469/wjimt.2024.07(03).01

Machine Learning-Driven Digital Identity Verification for Fraud Prevention in Digital Payment Technologies

Lichen Qin[1,*], Yuqiang Zhong[2], Han Wang[3], Qishuo Cheng[4], Jinxin Xu[5]

[1]Department of Computer Science, University of Rochester, Rochester, NY, USA
[2]Department of Information and Computer Sciences, Henan Agricultural University, Shenzhen, Guangdong, China
[3]Department of Mathematics, University of Southern California, Alhambra, CA, USA
[4]Department of Economics, University of Chicago, Chicago, IL, USA
[5]Department of Cox Business School, Southern Methodist University, Dallas, TX, USA
Corresponding author, lichenqin1@gmail.com USA

Abstract: ***This article explores how machine learning techniques can be used to drive digital authentication to prevent fraud in digital payment technologies. First, it introduces the development trend and fraud risk of digital payment technology, and then analyzes the limitations of traditional authentication methods, focusing on the potential of machine learning in digital authentication. It then explores specific application scenarios of machine learning in digital authentication, including data collection and preparation, feature engineering, model selection and training, as well as real-time monitoring and anti-fraud processing. Finally, current challenges and solutions are discussed, as well as the future of machine learning in digital payment technology. Through in-depth analysis of these contents, the article aims to provide readers with valuable insights to help them better use machine learning technology to improve the security and reliability of digital payments and promote the sustainable development of the digital economy.***

Keywords: Machine learning; Digital authentication; Fraud prevention; Digital payment.

1. INTRODUCTION

In today's digital world, with the rapid development of digital payment technology, people's choice of payment methods is also undergoing great changes. However, with the popularity of digital payments, fraud has also become rampant, bringing serious losses and risks to individuals and businesses. To address this challenge, more and more financial institutions and payment service providers are seeking new solutions to ensure the security and reliability of digital payments. In this context, machine learning technology is gradually becoming one of the important tools for digital identity verification and fraud prevention. By analyzing large amounts of user data and transaction behavior, machine learning algorithms can quickly and accurately identify suspicious activity and take appropriate measures to prevent it. This article will explore how machine learning-driven digital authentication technology can be used to effectively prevent fraud in digital payment technology.

We will start with a look at the trends in the adoption of digital payment technology and the fraud risks that this presents, then discuss the limitations of traditional authentication methods and focus on the potential of machine learning in digital authentication. Next, we will explore the specific application scenarios of machine learning in digital authentication, including data collection and preparation, feature engineering, model selection and training, and real-time monitoring and anti-fraud processing. Finally, we will discuss current challenges and solutions, and look at the future of machine learning in digital payment technology.

Through this research, we aim to provide financial institutions, payment service providers, and relevant stakeholders with valuable insights to help them better leverage machine learning technologies to improve the security of digital payments, build trust, and drive sustainable development of the digital economy.

2. RELATED WORK

2.1 Trends in digital payments

As a concept, Digital Payment has different definitions and understandings in both academia and industry. Domestic scholars have a vague definition of digital payment, and there are different views. Zhang Yang (2019) emphasizes the digitization of the transaction process in digital payments, arguing that any payment method that does not require physical currency delivery can be considered as digital payment. However, Zhong Xuejin (2019) positioned digital payment as a payment mode in the digital economy era independent of physical payment, cash payment and bank card payment. In addition, there are also differences in the industry's understanding of digital payment, such as the research report released by Jingdong Digital Technology and mastercard in 2019 defined digital payment as a digital payment method achieved by digital technology, excluding card-based payment dominated by banks and card organizations. Others argue that digital payments only include payments made using digital yuan, but this definition is clearly too narrow.

In contrast, foreign scholars are more inclined to equate digital Payment with Electronic Payment (E-payment), and believe that any way to transfer funds through electronic terminals without the use of physical media such as paper money and bills can be called digital payment. Taking India as an example, according to the Indian Payments and Settlement Act issued in 2007, digital payment is defined as electronic fund transfer, including various electronic payment methods such as ATM transactions, telephone transfers, POS machine transfers, Internet and credit card payments.

On this basis, this paper proposes a new understanding of digital payment. The digital payment is placed under the digital framework of the entire payment system, and the application of digital technology in the payment field is emphasized. Specifically, digital payment is the application of automatic identification technology, information and communication technology, blockchain, big data, cloud computing and other digital technologies in the payment process. In contrast, the narrow sense of digital payment refers to the use of digital technology to innovate payment methods, focusing on the digitalization of payment methods. It should be emphasized that digital payment and electronic payment are roughly the same in terms of the range of payment methods, but there are subtle differences in concept, digital payment emphasizes the use of digital technology, while electronic payment emphasizes that payment orders are issued by electronic terminals.

For a clearer understanding of digital payments, this paper incorporates it into the conceptual framework of existing payment methods and presents the differences between digital payments and electronic payments. As a non-cash payment method, digital payment is roughly the same as electronic payment in the range of payment methods covered, but it emphasizes the application of digital technology. Overall, the development of digital payments is of great significance in driving innovation and progress in payment methods.

2.2 Digital payment authentication

Digital authentication is a process of confirming and verifying a user's identity through digital technology, designed to ensure the security and authenticity of users when conducting online financial transactions or accessing sensitive information. The development of this field is mainly due to the progress of a variety of related technologies such as biometrics, two-factor authentication, blockchain technology and digital signature technology. In the field of digital authentication, biometrics is a common method, including fingerprint recognition, facial recognition and iris recognition. By scanning and comparing users' unique biometrics, highly accurate authentication can be achieved, greatly improving security. Another important authentication method is two-factor authentication, which combines multiple factors such as passwords, mobile phone verification codes, and fingerprints to improve the security of authentication. Through the combination of multiple factors, even if one factor is cracked, the security of user identity can be guaranteed.

Blockchain technology also plays an important role in digital identity verification. Through distributed ledgers and encryption algorithms, blockchain ensures the transparency and trust of identity authentication, preventing the occurrence of fraud such as data tampering and identity forgery. In addition, digital signature technology is also a key component of digital authentication. Using public key encryption and digital certificate, digital signature technology can ensure the integrity and authenticity of information, thereby preventing the risk of data tampering or forgery, and improving the reliability of authentication.

In general, the continuous development and innovation of digital authentication technology makes it more secure and convenient for users to authenticate when making digital payments and accessing sensitive information, effectively preventing the occurrence of fraud, and protecting users' privacy information.

2.3 Traditional authentication challenges

Traditional digital payment technology has some shortcomings in ensuring user identity and transaction security, mainly reflected in the following aspects:

Security risks: Traditional digital payment technology often relies on a simple password or a single authentication method, such as a username and password combination, which is at risk of being guessed or stolen. Once the user's login information is leaked, it may lead to account theft or illegal transactions, which will bring serious losses to the user and the platform.

Poor user experience: Traditional digital payment technology usually requires users to enter cumbersome information or carry out complex operations in the authentication process, such as entering a long password, multiple SMS verification, etc. These steps are cumbersome and easy to cause user dissatisfaction, reducing the comfort and convenience of payment experience.

High risk of fraud: The relatively weak security of traditional digital payment technology is easily exploited by criminals for fraudulent activities. For example, the use of fake websites and phishing and other means to steal users' account information and payment passwords, illegal transactions or fake sales, resulting in user property losses and trust crisis.

Difficult to solve the identity verification problem: In scenarios such as cross-border payments and international transactions, traditional digital payment technologies are often difficult to solve the differences in identity verification standards and laws and regulations in different countries or regions, resulting in authentication problems in the payment process, increasing the complexity and risk of payment.

Data privacy issues: Traditional digital payment technology usually requires users to provide personal sensitive information and financial information, there is a risk of personal privacy disclosure and abuse. Especially in third-party payment platforms and e-commerce platforms, users' personal information may be abused for commercial purposes such as advertising push and data analysis, causing concerns about user privacy disclosure.
To sum up, there are many shortcomings in traditional digital payment technology. It is necessary to introduce more advanced authentication technology and encryption means to improve payment security and user experience, reduce fraud risk, and protect users' personal privacy and property security.

2.4 Machine learning digital authentication

Machine learning has a wide application prospect in digital authentication, which can improve the accuracy, security and convenience of authentication. Machine learning can be used for user behavior analysis. By analyzing the user's login pattern, activity trajectory, device information and other data, the machine learning model can build a model of the user's behavior characteristics, so as to identify abnormal behavior and possible fraud. For example, when a machine learning algorithm detects a login location or device that doesn't match a user's habits, additional authentication processes can be triggered to ensure account security. Machine learning can be used for face recognition and biometric recognition. By training deep learning models, highly accurate face recognition can be achieved for digital authentication. Biometric identification technology, such as fingerprint recognition, iris recognition, etc., can also be modeled and identified by machine learning algorithms to further improve the accuracy and security of identity verification.

In addition, machine learning can also be used for device fingerprinting. Each device has unique hardware and software characteristics that can be used as a device fingerprint to verify the identity of the device. By monitoring the fingerprint characteristics of devices and using machine learning algorithms to identify them, it is possible to effectively prevent unauthorized devices from accessing user accounts.

Finally, machine learning can also be used to assess the strength of cryptography and detect password cracking. By analyzing the patterns and patterns of password selection, the machine learning model can assess the strength of passwords and provide suggestions for improvement. At the same time, machine learning algorithms can also detect and prevent password cracking, improving the security of accounts. To sum up, machine learning has a wide range of application prospects in digital identity verification, which can improve the accuracy and security of identity verification through user behavior analysis, biometric identification, device fingerprint recognition and cryptographic strength assessment, and provide strong support for digital identity authentication.

3. MACHINE LEARNING AUTHENTICATION MODEL

3.1 Machine learning face recognition model

Face recognition technology belongs to an important research object in the field of image engineering, it is a kind of biometric identification technology based on human facial characteristics information, it is generally known as portrait recognition, facial recognition, with great application value, such as criminal detection, intelligent payment and other fields. Generalized face recognition includes the related technology of establishing face recognition system. It includes face detection, face location, identity recognition, image preprocessing and so on. Face recognition technology relies on the most basic facial data feedback, the first is to establish a database, which stores the amount of data enough to identify a person, for the data transmitted in, the first step will determine whether there is a face, after capturing the face feature extraction, the feature data feedback to the background host, the background uses big data to compare the collected data, At the same time, search the database, whether there is the same identity information as its data, in order to prevent identity impersonation and other problems, as shown in Figure 1 is a company's employee door face recognition system.

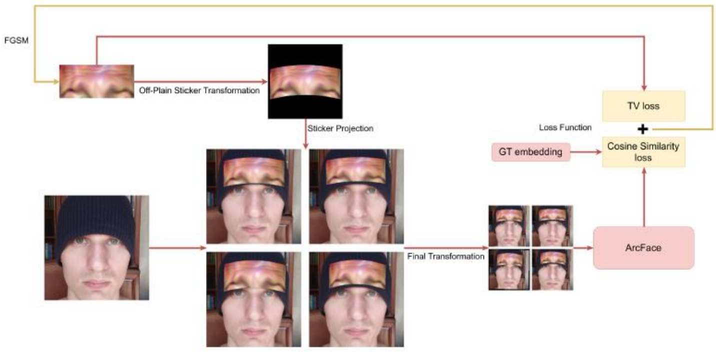

Figure 1: Face against architecture for financial fraud detection

An adversarial attack is a deliberate addition of subtle interference to the input sample that is imperceptible to humans, causing the model to give an incorrect output with a high degree of confidence. Application scenarios include the currently hot CV and NLP directions, for example, misclassifying classifiers by adding carefully prepared perturbing noise to pictures, or misclassifying emotions by substituting synonyms for certain words in a sentence.

To brush face payment for example, in simple terms, the anti-attack technology is to find the vulnerability in the face recognition algorithm by adding some disturbance to the person's face to mistakenly identify someone as another person, so as to achieve behaviors such as stolen brushes. Therefore, the research on anti-attack technology can improve the security of face recognition and avoid the phenomenon of face payment application theft. There are many ways to classify attacks. The most common ones are the attack environment classification and attack purpose classification.

(1) Attack environment classification:

White-box attack (white-box attack), where all the information and parameters inside the model are known, generates adversarial samples based on the gradient of the given model, and attacks the network.
Black-box attack, which does not know the parameters and structural information of the model, only through the input and output of the model, to generate a countersample

(2) Attack purpose classification:

Targeted attack: misclassifies an input to a specified class for a multi-classification network.

Non-target attacks, which only need to generate a countersample to fool the neural network.

3.2 Machine learning for credit card fraud detection

Credit card fraud is a tricky issue for many financial institutions, and although financial institutions have built various fraud detection systems based on machine learning algorithms, some fraudsters can still avoid fraud detection because these systems often do not contain information about the counterparty's knowledge of the fraud detection mechanism. In order to incorporate the ever-expanding knowledge base of fraudsters into an adaptive fraud detection system, based on the Credit Card Fraud Detection Dynamic Model Research Project (hereinafter referred to as the "Project"), this paper explores the best strategies to simulate fraudsters using game theory adversarial learning methods, and preemptively adjusts the fraud detection system to improve its ability to deal with potential threats.

Financial institutions can preemptively retrain their models by modeling the likely strategies of fraudsters, and such dynamic models are better than static models at detecting fraudulent credit card transactions. With the increase of game rounds, the AUC(Area Under Curve) scores of adversarial learning model and static fraud model gradually widen. Thus, it can be proved that by understanding the weaknesses of its own model, adversarial model can preferentially adjust the classifier to provide a better defense mechanism to detect fraud. This provides a direction for financial institutions to improve the performance of credit card fraud detection system.

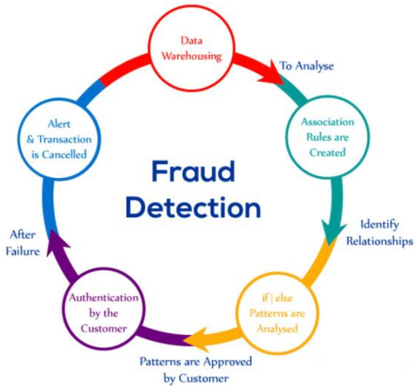

Figure 2: Machine learning credit card fraud architecture

The project uses a logistic regression classifier for fraud detection, initially determining the fraudster's best strategy based on the number of undetected fraudulent transactions, and assuming that it uses that strategy in future transactions to improve the initial classifier. The test results show that this adversarial framework-based dynamic model can produce a higher AUC score on the validation set after multiple iterations, i.e. a higher model accuracy rate, compared to the static model commonly used by financial institutions.

3.3 Machine learning fraud detection benefits

The amount of data is one of the key requirements for conducting AI-based payment fraud detection. In order to train high-quality machine learning models, large amounts of internal historical data are needed. However, few training sets contain an equal number of data samples in both classes (fraud and normal transactions). This means that you may need to use dimensionality reduction or data enhancement techniques to deal with data imbalances before training the model. Ensuring that the amount of data is adequate and representative is critical to the performance of the model. In addition to data volume, data quality is also an important aspect. If the historical data is biased, the results of the model will also be affected. This bias can be due to improper data collection and collation, or mixing information from fraudulent transactions with information from normal transactions. Therefore, it is essential to ensure that data is collected and sorted cleanly and appropriately to reduce bias in model results.

Another key requirement is factor integrity. Even with large amounts of data, models may not be effective at detecting fraudulent transactions if the data does not reflect real transactions or does not match the business logic. Therefore, ensuring the integrity of the data, and that it matches the business logic, is an important step to ensure the validity of the model. To sum up, in order to ensure the best performance of AI-based payment fraud detection models, attention needs to be paid to the key requirements of data volume, data quality, and factor integrity. Only when these requirements are met will the model be able to accurately identify fraudulent transactions and provide effective protection for customers and businesses.

4. CONCLUSION

Machine learning-driven digital identity verification technology, fraud prevention in digital payments has been significantly improved. Machine learning algorithms can analyze large amounts of user data and transaction behavior, quickly and accurately identify suspicious activity, and take appropriate measures to prevent fraud from occurring. Through the analysis of user behavior, biometrics, device fingerprints, etc., machine learning has shown broad application prospects in digital identity verification. The application of this technology not only improves the accuracy and security of identity verification, but also provides important support for the sustainable development of digital payments. Training high-quality machine learning models requires large amounts of historical data and ensuring the quality and integrity of the data to reduce bias in model results. In addition, there is a need to continuously improve and innovate digital authentication technologies to improve payment security, user experience and privacy protection levels. By making full use of machine learning technology, we can effectively prevent fraud in the field of digital payments, enhance user trust, and promote the sustainable development of the digital economy.

As digital payment technology continues to evolve and spread, machine learning-driven digital identity verification technology will play an increasingly important role in the future. With the further maturity and popularization of big data and artificial intelligence technologies, machine learning models will become more accurate and efficient. In the future, we can expect more innovations and breakthroughs, such as combining blockchain technology to achieve secure storage and verification of digital identities, and introducing deep learning and neural network technology to improve the accuracy and reliability of digital identity verification. In addition, with the continuous expansion of the digital payment market, the demand for fraud detection and digital identity verification technology will also increase, which will provide a broader development space for the research and application of related technologies. Therefore, through continuous innovation and improvement of machine learning-driven digital identity verification technology, we are confident that we can meet various challenges that may arise in the future digital payment field, provide users and enterprises with a safer and more convenient payment experience, and promote the healthy development of the digital economy.

ACKNOWLEDGEMENT

I would like to thank professors Yufu Wang, Lianwei Li, Jiufan Wang, Guangze Su, and Xiaonan Xu for their outstanding work and contributions to cloud computing, large-scale resource computing and storage, and machine learning. Their research results and profound insights provide valuable inspiration and guidance for my research. In the course of my research, their works have provided me with valuable references and inspirations, enabling me to better understand and apply cloud computing technologies and machine learning algorithms. Their pioneering spirit and practical experience have encouraged me to pursue innovation and breakthrough, and also encouraged me to apply what I have learned to solve practical problems. Without their work, I would not be able to make the progress I have made in this area today. Thank you again for their service and dedication.

REFERENCES

[1] Chen, B., Zhu, Y., Ye, S., & Zhang, R. (2018). Structure of the DNA-binding domain of human myelin-gene regulatory factor reveals its potential protein-DNA recognition mode. Journal of structural biology, 203(2), 170-178.

[2] Wang, Y., Bao, Q., Wang, J., Su, G., & Xu, X. (2024). Cloud Computing for Large-Scale Resource Computation and Storage in Machine Learning. Journal of Theory and Practice of Engineering Science, 4(03), 163-171.

[3] Chen, B., Liu, Z., Perry, K., & Jin, R. (2022). Structure of the glucosyltransferase domain of TcdA in complex with RhoA provides insights into substrate recognition. Scientific reports, 12(1), 9028.

[4] Yu, L., Liu, B., Lin, Q., Zhao, X., & Che, C. (2024). Similarity Matching for Patent Documents Using Ensemble BERT-Related Model and Novel Text Processing Method. Journal of Advances in Information Technology, 15(3).

[5] Zhang, Y., Liu, B., Gong, Y., Huang, J., Xu, J., & Wan, W. (2024). Application of Machine Learning Optimization in Cloud Computing Resource Scheduling and Management. arXiv preprint arXiv:2402.17216.

[6] Gong, Y., Huang, J., Liu, B., Xu, J., Wu, B., & Zhang, Y. (2024). Dynamic Resource Allocation for Virtual Machine Migration Optimization using Machine Learning. arXiv preprint arXiv:2403.13619.

[7] Chen, Baohua, Kay Perry, and Rongsheng Jin. "Neutralizing epitopes on Clostridioides difficile toxin A revealed by the structures of two camelid VHH antibodies." Frontiers in Immunology 13 (2022): 978858.
[8] Liu, B. (2023). Based on intelligent advertising recommendation and abnormal advertising monitoring system in the field of machine learning. International Journal of Computer Science and Information Technology, 1(1), 17-23.
[9] Xu, J., Zhu, B., Jiang, W., Cheng, Q., & Zheng, H. (2024, April). AI-BASED RISK PREDICTION AND MONITORING IN FINANCIAL FUTURES AND SECURITIES MARKETS. In The 13th International scientific and practical conference "Information and innovative technologies in the development of society"(April 02–05, 2024) Athens, Greece. International Science Group. 2024. 321 p. (p. 222).
[10] Ni, Chunhe, et al. "Enhancing Cloud-Based Large Language Model Processing with Elasticsearch and Transformer Models." arXiv preprint arXiv:2403.00807 (2024).
[11] Zhou, Y., Tan, K., Shen, X., & He, Z. (2024). A Protein Structure Prediction Approach Leveraging Transformer and CNN Integration. arXiv preprint arXiv:2402.19095.
[12] Shi, C., Liang, P., Wu, Y., Zhan, T., & Jin, Z. (2024). Maximizing User Experience with LLMOps-Driven Personalized Recommendation Systems. arXiv preprint arXiv:2404.00903.
[13] Wu, Y., Jin, Z., Shi, C., Liang, P., & Zhan, T. (2024). Research on the Application of Deep Learning-based BERT Model in Sentiment Analysis. arXiv preprint arXiv:2403.08217.
[14] He, Zheng, et al. "Application of K-means clustering based on artificial intelligence in gene statistics of biological information engineering."
[15] Li, L., Xu, K., Zhou, H., & Wang, Y. (2024). Independent Grouped Information Expert Model: A Personalized Recommendation Algorithm Based on Deep Learning.
[16] Li, X., Zheng, H., Chen, J., Zong, Y., & Yu, L. (2024). User Interaction Interface Design and Innovation Based on Artificial Intelligence Technology. Journal of Theory and Practice of Engineering Science, 4(03), 1-8.
[17] Chen, P., Lam, K. H., Liu, Z., Mindlin, F. A., Chen, B., Gutierrez, C. B., ... & Jin, R. (2019). Structure of the full-length Clostridium difficile toxin B. Nature structural & molecular biology, 26(8), 712-719.
[18] Pan, Y., Wu, B., Zheng, H., Zong, Y., & Wang, C. (2024, March). THE APPLICATION OF SOCIAL MEDIA SENTIMENT ANALYSIS BASED ON NATURAL LANGUAGE PROCESSING TO CHARITY. In The 11th International scientific and practical conference "Advanced technologies for the implementation of educational initiatives"(March 19–22, 2024) Boston, USA. International Science Group. 2024. 254 p. (p. 216).
[19] Zhang, X., Wang, C., Chen, B., Wang, Q., Xu, W., Ye, S., ... & Zhang, R. (2019). Crystal structure of refolding fusion core of Lassa virus GP2 and design of Lassa virus fusion inhibitors. Frontiers in microbiology, 10, 1829.
[20] Biology-based AI Predicts T-cell Receptor Antigen Binding Specificity. (2024). Academic Journal of Science and Technology, 10(1), 23-27. https://doi.org/10.54097/wy28c490
[21] Wu, Binbin, et al. "Enterprise Digital Intelligent Remote Control System Based on Industrial Internet of Things." (2024).
[22] Chen, B., Basak, S., Chen, P., Zhang, C., Perry, K., Tian, S., ... & Jin, R. (2022). Structure and conformational dynamics of Clostridioides difficile toxin A. Life science alliance, 5(6).
[23] Xu, J., Wu, B., Huang, J., Gong, Y., Zhang, Y., & Liu, B. (2024). Practical Applications of Advanced Cloud Services and Generative AI Systems in Medical Image Analysis. arXiv preprint arXiv:2403.17549.
[24] Che, C., Zheng, H., Huang, Z., Jiang, W., & Liu, B. (2024). Intelligent Robotic Control System Based on Computer Vision Technology. arXiv preprint arXiv:2404.01116.
[25] Zhou, Hong, et al. "Application of Conversational Intelligent Reporting System Based on Artificial Intelligence and Large Language Models." Journal of Theory and Practice of Engineering Science 4.03 (2024): 176-182.
[26] Xu, Kangming, et al. "Intelligent Classification and Personalized Recommendation of E-commerce Products Based on Machine Learning." arXiv preprint arXiv:2403.19345(2024).
[27] Shen, Xinyu, et al. "Biology-based AI Predicts T-cell Receptor Antigen Binding Specificity." Academic Journal of Science and Technology 10.1 (2024): 23-27.
[28] Zheng, Haotian, et al. "Medication Recommendation System Based on Natural Language Processing for Patient Emotion Analysis." Academic Journal of Science and Technology 10.1 (2024): 62-68.
[29] Li, Lianwei, et al. "Independent Grouped Information Expert Model: A Personalized Recommendation Algorithm Based on Deep Learning." (2024).