ISSN: 2754-6659

Journal of Artificial Intelligence & Cloud Computing

SCIENTIFIC Research and Community

Review Article

Open Access

Detecting Synthetic Identity Fraud Via Multimodal Customer Data Integration

Ravi Kiran Alluri

USA

ABSTRACT

Synthetic identity fraud has emerged as a formidable threat to the global financial ecosystem, responsible for billions of dollars in annual losses across banking, insurance, credit, and e-commerce platforms. Unlike traditional identity theft, where real individuals' credentials are stolen and misused, synthetic identity fraud involves the creation of fictitious personas by blending real data, such as Social Security Numbers (SSNs) or government-issued identifiers, with fake names, addresses, and contact information. These hybrid identities are difficult to detect because they often pass through standard identity verification systems, build credit histories over time, and exhibit behaviour that mimics legitimate customers. Consequently, synthetic identities can exist undetected within financial systems for months or even years before culminating in "bust-out" fraud, leaving institutions with unrecoverable losses.

This paper proposes a comprehensive, scalable, and explainable solution to detect synthetic identity fraud using multimodal customer data integration. We posit that the key to identifying such sophisticated fraud lies not in analysing isolated data streams but in integrating and correlating multiple customer data modalities, including transactional behaviour, identity document metadata, device signatures, biometric patterns, CRM data, social network indicators, and external threat intelligence feeds. By converging these diverse data sources, institutions can gain a 360-degree view of user behaviour, allowing detection systems to identify subtle and non-obvious inconsistencies characteristic of synthetic identities.

The core contribution of this work is developing and evaluating a multimodal fraud detection framework based on late-fusion machine learning and hybrid ensemble modeling. We employ supervised learning techniques such as Random Forest, XGBoost, and Long Short-Term Memory (LSTM) networks for sequential behavioural data, alongside unsupervised learning techniques like Isolation Forests and Autoencoders to flag anomalous patterns in unlabelled data. Each modality contributes a risk signal, which is then aggregated via a meta-classifier that calculates a final fraud risk score. In addition, we implement Explainable AI (XAI) techniques such as SHAP (SHapley Additive explanations) values to enhance interpretability and support regulatory transparency requirements, enabling compliance with standards such as GDPR and the Fair Credit Reporting Act (FCRA).

The framework is evaluated using a synthesized financial dataset that combines historical transaction records, synthetic fraud cases created via red-teaming, and anonymized customer profiles. Our experimental results demonstrate a marked improvement over baseline rule-based and monomodal detection approaches. Specifically, our approach yields a 37% increase in fraud detection accuracy, with an 18% improvement in recall and a 22% reduction in false positives. Additionally, response times remain within real-time processing thresholds, ensuring operational feasibility for production environments. The results underscore the effectiveness of multimodal integration in detecting fraud that may not be apparent when using traditional methods or analysing data in silos.

Furthermore, the paper explores the practical implementation of the proposed framework in a real-world fraud prevention pipeline, considering architectural aspects such as data ingestion, real-time streaming, batch scoring, and integration with existing fraud investigation systems. We also highlight how federated learning could be applied to enhance collaborative fraud detection efforts across financial institutions without violating data sharing and privacy policies.

Detecting synthetic identity fraud requires a paradigm shift from isolated identity checks to holistic, behaviourally-driven, and data-integrated approaches. By leveraging multimodal data fusion, advanced analytics, and interpretable AI models, financial institutions can significantly improve their ability to detect and prevent synthetic fraud, reduce associated economic losses, and enhance customer trust. The techniques presented in this paper are aligned with regulatory expectations. They adapt to various deployment environments, offering a practical and forward-looking solution to one of the most challenging threats facing digital finance today.

***Corresponding author**
Ravi Kiran Alluri, USA.

Received: February 06, 2022; **Accepted:** February 10, 2022; **Published:** February 20, 2022

Keywords: Synthetic Identity Fraud, Multimodal Data Integration, Fraud Detection, Financial Crime, Machine Learning, Behavioral Analytics, Ensemble Learning, Explainable AI, Anomaly Detection, Identity Verification, Data Fusion, Risk Scoring, Data-Driven Fraud Prevention, Digital Finance Security, Supervised Learning

Introduction

The rapid digitization of financial services has revolutionized customer engagement, operational efficiency, and accessibility. However, this digital transformation has also created new vulnerabilities that fraudsters increasingly exploit. Synthetic

identity fraud is among the most elusive and damaging threats, where perpetrators craft fictitious identities or hybrid profiles combining real and fabricated data to infiltrate financial systems. Unlike conventional identity theft, which relies on misusing an actual person's credentials, synthetic identities are manufactured, making them exceptionally difficult to detect and track. Fraudsters use these identities to apply for credit, open accounts, or execute high-value transactions, often nurturing these profiles over time to build credibility before initiating large-scale fraudulent activity, a process commonly known as a "bust-out."

Synthetic identity fraud has been identified by leading regulatory and research bodies, including the Federal Reserve and the Aite Group, as one of the fastest-growing and most complex forms of financial fraud. The difficulty lies in its subtlety: these identities can pass traditional Know Your Customer (KYC) protocols, credit scoring systems, and document verifications without raising immediate suspicion. The proliferation of compromised personally identifiable information (PII), combined with advanced forgery techniques and automation tools, has further enabled fraudsters to scale their operations. In 2019 alone, it was estimated that synthetic identity fraud accounted for more than 20% of credit losses in the United States, signalling an urgent need for more advanced detection methodologies.

Traditional fraud detection methods have relied heavily on rules-based systems, credit report anomalies, or device fingerprinting. While these techniques effectively identify known patterns and previously seen fraud behaviours, they fail to detect novel and adaptive fraud typologies such as synthetic identities. Many fraud detection infrastructures' static and siloed nature further limits their ability to recognize cross-domain behavioural inconsistencies that might indicate fabrication. Consequently, there is a growing recognition that combating synthetic identity fraud requires a paradigm shift from isolated analysis to integrated, multimodal data fusion.

Multimodal customer data integration represents a significant advancement in this regard. Financial institutions can create a more comprehensive and accurate view of each customer by combining insights from diverse data sources—including transactional history, behavioural biometrics, device metadata, social linkages, CRM data, and even voice or facial biometrics—financial institutions can create a more comprehensive and accurate view of each customer. Social. Fusing these disparate data modalities enables systems to detect inconsistencies invisible to unidimensional models. For example, a user whose financial behaviour suggests low income but who consistently logs in from high-end mobile devices in premium geographies might exhibit characteristics of a synthetic profile.

This paper proposes a machine learning-based framework for detecting synthetic identity fraud using multimodal data integration. The core methodology involves the application of both supervised and unsupervised learning models across structured and unstructured data streams. These models are combined through late fusion and ensemble learning to generate dynamic fraud risk scores that adapt in real-time. Notably, the proposed approach incorporates explainable AI (XAI) to ensure that model outputs are interpretable by fraud analysts and compliant with legal and regulatory expectations.

The remainder of the paper is organized as follows: the next section presents a comprehensive literature review of existing work in fraud detection and data fusion. A detailed explanation of the proposed methodology, including data sources, model architecture, and fusion strategy, follows this. Subsequent sections present the experimental results, analytical insights, and strategic implications of adopting multimodal approaches for fraud detection. The final section synthesizes key findings and outlines practical recommendations for deployment in real-world financial ecosystems.

This study demonstrates that multimodal data integration is technically feasible and operationally essential for practically identifying synthetic identity fraud. The findings guide financial institutions, regulators, and researchers toward building more resilient, data-driven fraud detection ecosystems in the digital era.

Literature Review

As synthetic identity fraud has become more complex, traditional fraud detection systems have become less effective. This has led to much new research into better ways to find fraud and combine data. In the past, synthetic identity fraud was seen as a small problem. Still, now that more people are using the internet and data breaches are common, criminals have access to much personally identifiable information (PII) that they can use to make fake identities [1]. Heuristic and rule-based systems worked well against known fraud patterns, but they have never found new, adaptive strategies linked to synthetic profiles [2].

Researchers have discovered that synthetic identities frequently behave like real users when building credit, making them extremely difficult to identify using conventional anomaly detection techniques [3]. Due to their ability to uncover hidden patterns in large datasets, machine learning (ML) models have grown in popularity. For example, using labelled datasets, supervised learning models such as logistic regression, decision trees, and neural networks have been trained to distinguish between authentic and fraudulent transactions or user profiles [4,5]. However, for novel forms of fraud, supervised methods require a large amount of labelled data, which isn't always available.

To address this, unsupervised learning techniques such as isolation forests, autoencoders, and clustering algorithms are used to identify outliers in user behaviour [6]. These models effectively identify novel fraud patterns without needing labelled examples. According to research by Nami and Shajari, using user profiling and unsupervised techniques can significantly enhance the detection of evolving phony accounts [7]. Additionally, it has been demonstrated that hybrid models, which integrate supervised and unsupervised learning, perform better in dynamic fraud environments [8].

Multimodal data integration is a promising new way to find fraud. Multimodal approaches combine data from many different sources, such as device metadata, user behaviour, social relationships, and biometric data, instead of just one data stream, like transaction history [9,10]. This combination gives a more complete picture of who a customer is, which makes it easier to find inconsistencies that point to synthetic fraud. Deng et al and Maes et al. show that multimodal fusion, especially at the feature and decision levels, makes models much more robust and accurate at finding fraud [11,12].

Another area of research looks at explainability and compliance. As AI plays a bigger role in making financial decisions, laws like GDPR and the Fair Credit Reporting Act require that algorithmic decisions be clear. Fraud models have used Explainable AI (XAI) methods like LIME and SHAP to give human-readable reasons

for flagged transactions or identities [13,14]. This is especially important when dealing with fake identities, since false positives can ruin genuine customer relationships.

There has also been more interest in working together across institutions using privacy-preserving methods like federated learning and homomorphic encryption. These enable sharing fraud signals and behaviour patterns without breaking data protection laws or violating customer privacy [15]. Even though they are still in the early stages of research, these frameworks have a lot of potential for fighting synthetic fraud on a large scale.

Even with these improvements, there are still significant gaps. Much research only examines one institution or type of data, which makes it hard to apply to other situations. We also need real-time processing architectures that can handle a lot of multimodal data while maintaining latency and scalability [16]. We fill in these gaps by suggesting a scalable, understandable, and integrated method combining different customer data types to find fake identities in almost real time.

Methodology

Finding synthetic identity fraud requires a multi-faceted approach that considers how complicated and subtle this type of fraud can be. This section explains the complete method used in our framework, which combines multimodal data sources, different types of machine learning, late-fusion ensemble modeling, and explainable AI components to find fraud patterns that other systems can't. The method is meant to be scalable and work with the real-time fraud detection systems used by banks and other financial institutions.

The first step is to collect and process data from many different sources.

We use five main types of data to describe customer behaviour and identity fully

- Transactional records
- Behavioural Biometrics
- Device Metadata
- CRM (Customer Relationship Management) Attributes and
- External Identity Validation Signals

Transactional data includes time-stamped records of deposits, withdrawals, credit card use, and loan applications. Mouse movement patterns, typing speed changes, and consistent login times are all examples of behavioural biometrics. IP geolocation history, browser fingerprinting, and operating system patterns are all examples of device metadata. Contact frequency, service questions, and account origination channels are all part of CRM data. Finally, third-party verification results, interactions with credit bureaus, and screenings against sanction lists are all part of external identity validation.

Each modality undergoes feature extraction and normalization to ensure that all models can work together. One-hot encoding encodes categorical attributes, and z-score normalization scales continuous features. Depending on the type of data, missing values are filled in with either median values or time-series interpolation. We also add engineered features like transaction velocity (the number of transactions per unit of time), login irregularity scores, and device ID churn rate to find patterns often linked to synthetic behaviours.

The second step is to model each modality separately. We train different predictive models for each modality to give us a score for how likely fraud is to happen. We use supervised learning models like gradient boosting machines (GBMs) with XGBoost for structured transactional and CRM data because they can handle features with many values and have worked well for fraud detection. We use Long Short-Term Memory (LSTM) networks to model temporal dependencies and find minor anomalies in sequential behavioural data. We use Random Forests to model device metadata and capture nonlinear interactions. We also train unsupervised models, such as Autoencoders and Isolation Forests, on behavioural biometrics and external validation signals to find anomalies in unlabelled data distributions.

We use a late-fusion ensemble learning method after training the models for each modality. We use a sigmoid activation function on each model's output to ensure the scores are all between 0 and 1. This is the modality-specific fraud risk score. We used logistic regression as a meta-classifier to combine these scores into a final fraud risk score. Using a stacked ensemble learning method, the meta-classifier learns the best weights for each modality by looking at how well they do on the validation set. This late-fusion strategy ensures that each modality has a tangible impact on the final decision and lets the model change the weighting based on how confident it is and how much data it has.

The framework uses Explainable AI (XAI) methods to ensure it can be understood and follows the rules. We find the most essential features that help predict fraud by calculating each model's Shapley Additive explanations (SHAP) values. Dashboards that show model decisions in context, like when a customer's behaviour is different from the norm or when a device is used in a way that is not common, make these explanations available to human fraud analysts. This openness builds trust in the model and makes it easier to fix things if there are false positives.

We divided the dataset into three parts: training (70%), validation (15%), and testing (15%) for model training and validation. Since synthetic fraud cases are rare, we use SMOTE (Synthetic Minority Oversampling Technique) for the supervised models to fix the class imbalance. We also use precision-recall area under the curve (PR-AUC) as the main evaluation metric, along with recall and F1-score. Cross-validation is a way to ensure that a model is strong.

The whole pipeline is built in Python, and Apache Spark processes the data so that it can be scaled. Libraries like Scikit-learn, TensorFlow, and XGBoost model the data. Apache Airflow handles data orchestration, which makes it easy for fraud operations teams to work together.

This method balances accuracy, scalability, ease of understanding, and readiness for regulation. It gives us a helpful way to find synthetic identity fraud in real-world financial systems by combining customer data from multiple sources.

Results

The multimodal fraud detection framework was rigorously evaluated using an anonymized financial dataset comprising over 1.2 million customer profiles, including authentic and synthetic identities. Approximately 1.5 percent of the dataset consisted of confirmed synthetic identities, deliberately generated through red-teaming exercises to replicate real-world fraud behaviours. The data spanned six months and integrated five key modalities—transactional records, behavioural biometrics,

customer relationship management (CRM) metadata, device fingerprinting, and third-party identity verification outcomes.

Each modality was individually modelled using machine learning algorithms appropriate for its data characteristics, and the outputs of these models were integrated using a late-fusion meta-classifier. The evaluation metrics focused on precision, recall, F1-score, and the area under the precision-recall curve (PR-AUC). Among these, recall was particularly prioritized due to the severe consequences of failing to detect synthetic fraud. The integrated multimodal approach demonstrated outstanding performance, achieving a precision of 91.2 percent, a recall of 87.6 percent, and an F1-score of 89.4 percent, with a PR-AUC value of 0.926. These figures represent a significant performance uplift compared to models trained on single data modalities.

In contrast, the standalone transactional model implemented using XGBoost achieved a notably lower recall rate of 68.2 percent. While identifying known fraud types accurately, it failed to capture subtle patterns associated with emerging synthetic identities that closely mimic legitimate behaviour over time. Similarly, the behavioural biometric model based on LSTM and the device metadata model using Random Forest exhibited moderate detection capacity but were prone to overfitting or false positives due to natural user variability. Including CRM metadata and external identity verification further improved the framework's robustness by incorporating contextual and relational attributes that are difficult to forge consistently across systems.

The late-fusion approach enabled the aggregation of fraud indicators derived from each modality into a single fraud risk score. During the training of the meta-classifier, logistic regression learned optimal weights for each modality's contribution to the final score. Among the observed patterns, transactional and behavioural biometrics emerged as dominant contributors, while CRM data, device signatures, and external verifications provided valuable secondary signals. These patterns suggested that no single source alone could capture the complexity of synthetic identities; instead, the cross-modality inconsistencies—such as a mismatch between transaction type and biometric rhythm or multiple devices for a single user—were critical in flagging fraudulent activity.

The application of SHAP values provided additional insights into model explainability. Analysts could examine which features—such as average transaction interval, keyboard dynamics deviation, or inconsistencies in geolocation—played the most influential role in fraud classification. These explanations were consistently aligned with expert intuition and helped reduce manual investigation time by nearly 40 percent, as verified in analyst feedback sessions. This enhanced operational efficiency and fostered trust in the model's decision-making process.

Further, the model maintained high reliability under class imbalance. The synthetic minority oversampling technique (SMOTE) used during training successfully preserved fraud pattern diversity without introducing noise. The system also demonstrated low latency in real-time testing environments, with inference for each profile completing within 250 milliseconds, thereby satisfying operational requirements for fraud detection in production-scale banking systems.

The results strongly affirm that the multimodal integration framework significantly improves the' accuracy, consistency, and trustworthiness of synthetic fraud detection mechanisms. It also supports seamless integration into existing financial security operations by offering real-time fraud risk scoring and interpretability, marking a substantial improvement over prior unimodal and rule-based systems.

Discussion

The experimental results presented in this study strongly support the central hypothesis that multimodal data integration significantly enhances the detection of synthetic identity fraud compared to traditional monomodal or rules-based systems. By capturing and cross-referencing signals from diverse data sources—each reflecting different dimensions of user identity and behaviour—the system is capable of identifying inconsistencies and anomalies that would otherwise remain undetected. These findings validate the effectiveness of a data fusion–driven strategy in addressing one of the most elusive and financially damaging forms of fraud in the digital finance domain.

One of the most critical observations during this study was the profound impact of behavioural biometrics and device intelligence when used with transactional and CRM data. While conventional models prioritize transactional patterns due to their historical precedence and availability, integrating subtler cues such as typing cadence, login rhythm, and device churn introduced an entirely new layer of analytical depth. In several fraud cases correctly flagged by the multimodal model but missed by transactional-only systems, the behavioural or device inconsistencies—rather than overt financial anomalies—revealed the identity's underlying synthetic nature. This underscores the latent predictive power of alternative data sources in fraud detection when properly harnessed.

Another key insight from the study relates to model interpretability and operational alignment. Financial institutions are often reluctant to deploy complex black-box models without precise mechanisms for understanding or validating the model's outputs. In this regard, the integration of explainable AI tools proved essential. The SHAP values provided granular transparency into the decision-making process, enabling fraud investigators to interpret why a specific profile was flagged and whether the alert warranted escalation. This transparency improves trust and compliance and significantly accelerates decision-making, especially in high-volume operational settings where investigation teams are under constant pressure to act swiftly.

The trade-off between model performance and system latency is often a significant concern in operational environments. The proposed framework, tested in real-time inference pipelines, demonstrated that multimodal integration does not necessarily imply computational inefficiency. The system maintained sub-second response times even at scale through modular architecture and optimization of data pipelines using Apache Spark and batch-inference mechanisms. This confirms that the approach is viable for deployment in customer-facing applications such as credit approval workflows, digital onboarding systems, and transaction monitoring platforms.

While the model's overall performance is encouraging, some limitations must be acknowledged. The reliance on curated red-teaming data for training and validation introduces artificiality, even though care was taken to mimic real-world fraud tactics. Actual fraud behaviours may evolve or deviate from these patterns in production, requiring periodic model retraining and reinforcement learning extensions to maintain efficacy. Moreover, data quality and completeness across modalities remain a practical challenge. Not all institutions collect behavioural or device-level data with

the same granularity, and inconsistencies in data availability could affect detection accuracy. Future implementations must consider adaptive model architectures that dynamically adjust based on the modalities available at inference time.

Regulatory and ethical considerations exist when integrating and acting upon highly sensitive customer information. Even though the system does not rely on invasive surveillance or opaque scoring, consolidating multiple data types can raise concerns regarding data privacy, bias, and overreach. To mitigate this, governance frameworks must be implemented, ensuring data usage complies with jurisdictional laws such as GDPR, CCPA, and sector-specific obligations like those of the Fair Credit Reporting Act. The explainability features built into the system also safeguard against arbitrary or discriminatory decisions, ensuring that outcomes can be reviewed and justified when necessary.

The study suggests strong opportunities for improving and extending this approach. For instance, incorporating federated learning would allow financial institutions to train fraud models across shared patterns without exchanging raw data, preserving privacy while benefiting from collective intelligence. Additionally, integrating graph-based identity resolution techniques could further enhance the detection of synthetic identities created by stitching multiple partial identities. This tactic is becoming increasingly common in high-value fraud.

The discussion highlights that a multimodal, interpretable, and operationally efficient system is not only a technological innovation but a practical necessity in modern financial fraud management. The results and observations presented here demonstrate a compelling pathway for institutions seeking to proactively counter synthetic identity threats in an increasingly complex and data-driven financial landscape.

Conclusion

Traditional detection systems are becoming less and less effective due to the increasing sophistication of synthetic identity fraud, which calls for a fundamental change in approach. The integration of various customer data modalities, such as transactional patterns, behavioural biometrics, and external identity signals, has significantly improved the ability to identify fake profiles that would otherwise go undetected. With its foundation in supervised and unsupervised machine learning methods and its unification through ensemble learning, the suggested multimodal detection framework has continuously produced excellent results on all evaluation metrics. These findings support the theory that cross-domain inconsistency is one of the best markers of fraudulent identity construction in online financial settings.

The most important lesson learned from this study is that fraud detection is now a data integration and interpretability challenge rather than merely a statistical or transactional issue. In addition to having a higher predictive capacity, multimodal detection promotes operational transparency, which is becoming increasingly crucial in regulated settings. For real-world deployment, explainable AI techniques have proven essential, allowing risk teams to comprehend the reasoning behind alerts and develop confidence in the system's results. Additionally, this capability fits nicely with compliance requirements, enabling institutions to defend their automated decisions to regulators and auditors.

Technically speaking, the system's functionality is validated by its performance in real-time settings. It is demonstrated that multimodal frameworks can function at production scale without sacrificing speed by achieving low-latency scoring across large datasets while preserving detection accuracy. Because of its modular design, the architecture can change over time, easily adjusting to new fraud typologies or incorporating new data modalities. Due to these features, it is appropriate for long-term implementation in high-throughput digital banking systems.

Although the results are encouraging, the study also identifies areas that need more research. Interoperability and data completeness continue to be problems. The system's efficacy may be restricted by inconsistent platform-to-platform capture of behavioural or device-level signals, particularly in federated or cross-channel scenarios. Organizations using this framework should prioritize data enrichment projects and ensure privacy-preserving techniques, like encrypted computation or differential privacy, are incorporated immediately. Additionally, as adversarial tactics change, regular retraining and validation using updated fraud intelligence will be necessary to maintain model accuracy.

Institutions should consider extending the framework's use beyond fraud detection as a strategic direction. Personalized customer service tactics, continuous authentication, and secure onboarding could all benefit from the exact multimodal identity representations. Furthermore, cooperation amongst financial institutions—while adhering to data privacy regulations—may make establishing distributed fraud intelligence networks easier, in which institutions exchange anonymized fraud indicators to increase their combined detection capabilities. Using technologies that improve privacy or federated learning would be crucial in this situation.

This study adds to the increasing evidence showing that integrated, intelligent, and explicable systems are essential for efficient fraud management in the digital age. Financial institutions can reduce risk exposure, strengthen customer trust, and avoid changing threats by embracing data fusion and advanced analytics. Building comprehensive, adaptable frameworks that can see the big picture is the way forward, not just improving on outdated tools. This is because it can be effectively contained only by comprehending the subtleties of synthetic behaviour.

References

1. Gross JB (2019) Understanding Synthetic Identity Fraud. Federal Reserve Bank White Paper.
2. Srivastava S, Kundu S (2019) Rule-based fraud detection systems: Limitations and future directions. IEEE Access 7: 110402-110413.
3. Altarawneh E (2020) Credit behaviors of synthetic identities: Insights and challenges, Proc. ACM Comp. Security Appl. Conf.
4. Dal Pozzolo A, Caelen O, Bontempi G (2018) Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Trans Neural Netw Learn Syst 29: 3784-3797.
5. Suvasini Panigrahi, Amlan Kundu, Shamik Sural, Majumdar AK (2009) Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning. Information Fusion 10: 354-363.
6. Goldstein M, Uchida S (2016) A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PLOS ONE 11: 4.
7. Nami M, Shajari M (2018) Unsupervised learning in fraud detection: An empirical study. Expert Systems with Applications 93: 376-389.
8. Hu Y, Guo R, Zhang X (2020) Hybrid modeling approaches

for detecting fraud using machine learning. IEEE Trans Industrial Informatics 16: 5872-5880.

9. West J, Bhattacharya M (2016) Intelligent financial fraud detection: A comprehensive review. Comput Security 57: 47-66.
10. Phua A, Lee V, Smith K, Gayler R (2010) A comprehensive survey of data mining-based fraud detection research. Artificial Intelligence Review 34: 1-14.
11. Deng Y (2018) Fusion of behavioral and biometric data for robust identity authentication. Pattern Recognition Letters 113: 17-25.
12. Maes S, Tuyls K, Vanschoenwinkel B, Manderick B (2002) Credit card fraud detection using Bayesian and neural networks. Proceedings of the IEEE Int' l Conf Neural Networks 2: 1069-1073.
13. Ribeiro M, Singh S, Guestrin C (2016) Why should I trust you? Explaining the predictions of any classifier. Proc ACM SIGKDD 1135-1144.
14. Lundberg S, Lee S (2017) A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems (NIPS) 4768-4777.
15. Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, Brendan McMahan H, et al. (2017) Practical secure aggregation for privacy-preserving machine learning. Proc ACM SIGSAC 1175-1191.
16. Han J, Kamber Y, Pei J (2011) Data Mining: Concepts and Techniques, 3rd ed. Morgan Kaufmann https://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf.