

Taylor & Francis
Taylor & Francis Group

Applied Artificial Intelligence
An International Journal

ISSN: 0883-9514 (Print) 1087-6545 (Online) Journal homepage: www.tandfonline.com/journals/uaai20

AI-based Identity Fraud Detection: A Systematic Review

ChuoJun Zhang , Asif Q. Gill , Bo Liu & Memoona Anwar

To cite this article: ChuoJun Zhang , Asif Q. Gill , Bo Liu & Memoona Anwar (2026) AI-based Identity Fraud Detection: A Systematic Review, Applied Artificial Intelligence, 40:1, 2731995, DOI: 10.1080/08839514.2026.2731995

To link to this article: https://doi.org/10.1080/08839514.2026.2731995

© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC.

View supplementary material

Published online: 01 Oct 2026.

Submit your article to this journal

Article views: 970

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=uaai20

APPLIED ARTIFICIAL INTELLIGENCE
2026, VOL. 40, NO. 1, e2731995 (35 pages)
https://doi.org/10.1080/08839514.2026.2731995

Taylor & Francis
by informa

OPEN ACCESS Check for updates

AI-based Identity Fraud Detection: A Systematic Review

ChuoJun Zhang, Asif Q. Gill, Bo Liu, and Memoona Anwar

University of Technology Sydney, Sydney, Australia

ABSTRACT

With the rapid development of digital services, personally identifiable information (PII) is increasingly exposed to identity fraud (IDF). The growing use of artificial intelligence (AI)-enabled deepfake technologies has further amplified this threat, enabling fraudsters to produce highly sophisticated counterfeit documents, photos, and videos. There is a pressing need to systematically review IDF detection methods, their limitations, and potential solutions. This paper presents a systematic literature review examining AI-based IDF detection and prevention methods. The review identifies two principal categories of detection approaches, consolidates findings into a taxonomy of AI-based IDF methods, and highlights key insights, open challenges, and emerging trends. This work provides a foundational knowledge base for researchers and practitioners advancing this critical area.

Introduction

Identity is used to identify an individual or entity (ISO/IEC 2011; Rountree 2013; Segovia Domingo and Enríquez Á 2018). Identity fraud (IDF) is a criminal act involving the use of another person's identity or the creation of a fake identity for fraudulent purposes (Irvin-Erickson 2024; United Nations Office on Drugs and Crime 2011). Fraudsters' main goal or motive could be to obtain money and public resources, evade obligation, or commit other crimes (Ahmed 2020). With the rapid development of digital and online services, a large amount of personally identifiable information (PII) is scattered and generally stored in public databases. The monetary benefits of selling personal information on the black market have driven cyber-criminals to covet this personal data. Every year, billions of PII records are breached or stolen due to cyber-attacks and security vulnerabilities (OAIC 2020a, 2020b, 2021a, 2021b, 2022a, 2022b, 2023a, 2023b, 2024), providing the means for committing IDF. Between 2020 and 2024, multiple data breaches have occurred, compromising personal information, such as names, dates of birth, social security numbers, addresses, e-mails, phone numbers, and even bank account passwords (OAIC 2020a, 2020b, 2021a, 2021b, 2022a, 2022b, 2023a, 2023b, 2024). In recent years, the development of Generative Artificial Intelligence (GAI) technology has made IDF feasible with the possibility of generating fake identities at a scale. Several fraud cases involving deepfake technology have emerged recently. For instance, deepfake allowed criminals to bypass identity verification systems and conduct fraud through live video (Chen and Magramo 2025). The consequences of identity theft and the use of fake identities are far-reaching, influencing both individuals and organizations across a wide range of sectors (Irvin-Erickson 2024). The fraud ring or individuals obtain true identity information from others or fabricate fake identities through black market channels, using them for targeted scams, extortion, insurance and loan fraud, and even participating in illegal activities like money laundering (Figure 1) (Ahmed 2020).

Artificial intelligence (AI) or GAI can be used for both enabling IDFs via generating fake identities and fraud detection methods. There is a need to review and consolidate knowledge on this topic to facilitate the understanding of AI and IDF. Thus, this paper aims to address the following main research question:

RQ: What is known about AI-based IDF methods?

CONTACT ChuoJun Zhang Chuojun.Zhang@student.uts.edu.au
Supplemental data for this article can be accessed online at https://doi.org/10.1080/08839514.2026.2731995
© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

This review paper focused on the study of AI-based detection methods covering IDF prevention and detection stages. To the best of our knowledge, there is no similar Systematic Literature Review (SLR) study that examines the AI and integrated IDF prevention and detection processes. The main contributions and reflections of this paper are as follows:

(1) A novel taxonomy for IDF prevention and detection methods, categorized into two key phases: authentication and continuous authentication. This taxonomy provides a structured approach to understanding the strategies and challenges employed in each phase.
(2) Key insights about the increasing importance of biometric authentication, especially facial recognition.
(3) Continuous authentication using user behavior analysis, which is becoming more important for improving fraud detection after the initial authentication.

This paper is organized as follows. First, it discusses the research motivation and method. Second, it presents the analysis of the IDF detection methods based on a review of selected papers. Third, it discusses the open challenges before concluding.

A preprint version of this manuscript has been deposited on arXiv (Zhang et al. 2025).

Research background and motivation

This section first establishes the problem context by defining identity-related crimes, their processes, and current AI-based defenses, then reviews related work to pinpoint the research gaps this SLR will address.

Problem context

Identity, which serves to identify an individual and entity, is widely used daily to access various activities, such as public services, education, medication, job applications, financial transactions, social media, etc. (Rountree 2013). Identity theft refers to the unauthorized acquisition of personal identity attributes such as PII, biometric data, whereas IDF refers to the subsequent misuse of those identity attributes to obtain value, commit deception, or evade controls (Ahmed 2020; United Nations Office on Drugs and Crime 2011). These offenses are referred to as identity-related crimes, which cover all forms of illicit conduct involving identity, including acquiring, manipulating, producing, transferring, possessing, or using identity information for unlawful activity. While the two concepts were often conflated historically, the distinction was formalized in 2011 by the United Nations Commission on Crime Prevention and Criminal Justice (United Nations Office on Drugs and Crime 2011).

Identity theft is often the first step of IDF, which primarily involves the unlawful acquisition of identity information, such as PII (ACT Legislation Register 2002). As shown in Figure 1, identity thieves initiate malicious attacks, including malware, phishing, and brute force, to intrude into various institutions such as financial institutions, hospitals, governments, and enterprises to steal identity information (OAIC 2020a, 2020b, 2021a, 2021b, 2022a, 2022b, 2023a, 2023b, 2024). These stolen identities are either directly sold or transferred to accomplices, such as identity forgers, who produce synthetic or fabricated identities for resale to criminal individuals or groups (ACT Legislation Register 2002; Blauth, Gstrein, and Zwitter 2022). With the assistance of AI technologies, particularly GAI, the manipulation and creation of both real and synthetic identities have become significantly easier (FedPayments Improvement 2022a). Identity fraudsters then use this identity information to create new identities or take over other identities to gain either financial or non-financial benefits at the cost of other accounts (Ahmed 2020). For example, defraud financial institutions to receive payouts or services (FedPayments Improvement 2022b). This identity information can also be used to facilitate a means to an end as part of other illegal acts such as human trafficking, money laundering, and illegal immigration (Shafique et al. 2017). Beyond financial losses, the consequences of IDF extend, encompassing psychological distress, reputation damage, and potential legal repercussions for victims (Button and Cross 2017; Button, Lewis, and Tapley 2014; Cavaliere et al. 2021; Green et al. 2021; Modic and Anderson 2015; Norris, Brookes, and Dowell 2019).

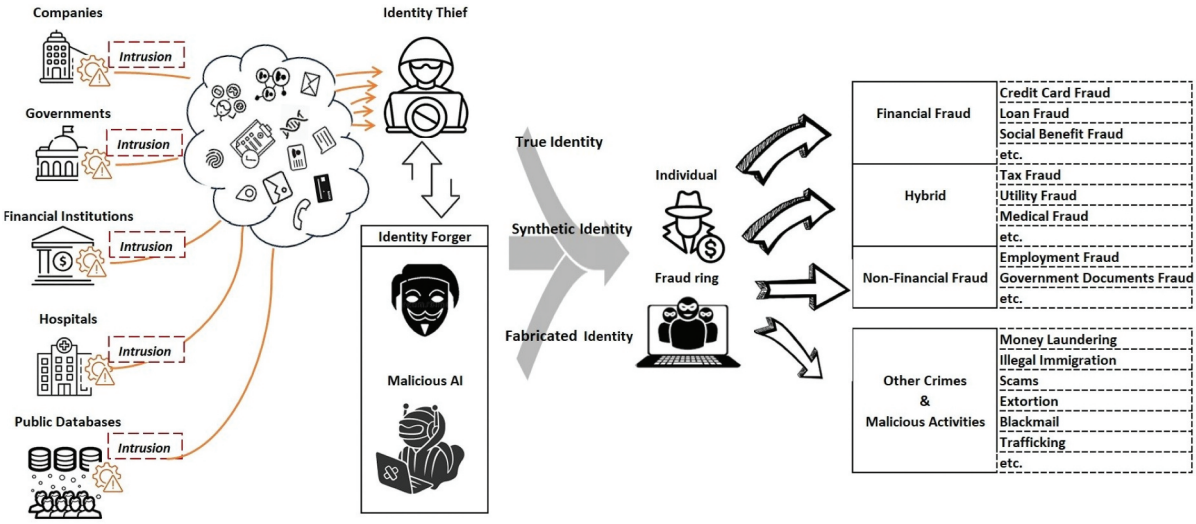


Figure 1. The IDF chain: from data theft to fraudulent use cases (image adapted from Ahmed 2020, 59).

The current defense against IDF primarily relies on two phases: authentication and continuous authentication. Authentication is the initial phase of validating a claimed identity to ensure its authenticity and legitimacy. Authentication methods are categorized into three categories based on the type of information being validated: knowledge, possession, and inherence (Velásquez, Caro, and Rodríguez 2017). Knowledge-based authentication involves confirming a user's identity through information they know, such as a one-time password, answers to security questions, or a personal identification number (Lalchand, Srinivas, and Gregorie 2023). Possession-based authentication verifies identity through physical items the user owns, such as mobile devices, smart cards, or hardware tokens (Luxton, Kayl, and Mishkind 2012). Inherence-based authentication confirms a user's identity through their biometric or behavioral characteristics. This type of biometric authentication includes, but is not limited to, facial recognition, voice recognition, and fingerprint recognition (Kodituwakku 2015; Masood et al. 2023; Velásquez, Caro, and Rodríguez 2017).

However, initial authentication alone may not be enough, as data breaches and advanced attack techniques (e.g., biometric spoofing and deepfakes) still enable criminals to bypass verification systems relatively easily (Ayeswarya and Norman 2018; Galbally, Marcel, and Fierrez 2014; Heusch et al. 2020). Therefore, defending against IDF also requires continuous authentication. Continuous authentication is the phase of ongoing identity validation and anomaly detection, achieved by monitoring user behavior and contextual data in real-time (Baig and Eskeland 2021). It reduces vulnerabilities to unauthorized access and enhances the overall security of online interactions and transactions (Gill 2012).

AI, which originated in 1956 to replicate human thinking and problem-solving (Russell and Norvig 2021), has evolved to effectively combat IDF in both authentication and continuous authentication. By integrating machine learning and deep learning models with technologies such as computer vision, natural language processing (NLP), and graph technology, AI empowers systems to perform tasks like identification, recognition, and anomaly detection (Russell and Norvig 2021). For example, in biometric authentication, AI leverages deep learning and computer vision technologies to authenticate users by first recognizing and then comparing their biometric features (Jain and Kumar 2012; Kodituwakku 2015). In continuous authentication, AI incorporates graph technology for relationship modeling and NLP for analyzing textual data to detect fraud patterns across structured data (e.g., transaction records) and unstructured data (e.g., social media content), which enables real-time and adaptive authentication (Baig and Eskeland 2021).

This review, motivated by an industrial project to develop an AI fraud engine for identity verification systems, investigates how AI detects IDF during the two authentication phases and identifies gaps in current methods to guide the development of more robust solutions. It introduces a novel framework by adopting authentication processes as a consolidated structure, systematically deconstructing and analyzing each phase of IDF detection methods and its AI integration. This hierarchical "top-down" approach provides a

rigorous and structured exploration of methodologies. Furthermore, this review addresses existing research gaps and provides insights and future directions for researchers and practitioners.

Related work and gaps

Existing literature exhibits two major gaps. First, many prior reviews focus on financial fraud detection rather than IDF as a distinct category (Ali et al. 2022; Hilal, Gadsden, and Yawney 2021; Jáuregui-Velarde et al. 2024; Rodrigues et al. 2022). Second, existing studies typically examine specific technical aspects of fraud detection but do not position these methods within the two-phase authentication life cycle, and in particular have not integrated UEBA-based continuous authentication, where IDF is operationally materialized (Ali et al. 2022; Hilal, Gadsden, and Yawney 2021; Jáuregui-Velarde et al. 2024).

In contrast, this review consolidates authentication and continuous authentication methods into a single taxonomy grounded in IDF processes, which addresses the cross-phase fragmentation prevalent in existing reviews.

Research methodology

This review conducts a systematic literature review based on the detailed approach described in Kitchenham et al. (2009). It begins with research planning, where we define research questions, keywords, databases, and selection criteria. Then, we conduct the data search and selection based on the inclusion and exclusion criteria and quality assessment criteria. Finally, we extract data from the selected documents according to the defined fields. This systematic review was reported in accordance with the PRISMA 2020 statement. The completed checklist and flowchart are provided in Appendix A Figure A1 and Appendix F Table F1. To ensure reproducibility, Appendix C Table C1 summarizes all 43 included studies with publication venue, year, and primary analytical method. The review protocol was not prospectively registered on PROSPERO or OSF. Without such a record, readers cannot independently verify that the review questions, search strategy, and eligibility criteria were specified in advance and applied without post-hoc adjustment, and we acknowledge this as a limitation of the review.

Research questions

The primary aim of this SLR is to discover all AI-based methods to detect and prevent IDF during authentication processes. The methods are compared to evaluate their effectiveness and identify open challenges and future development points in combating IDF.

This SLR will address the main research question:

RQ: What is known about AI-based IDF detection methods?

It has been further divided into the following sub-questions:

- RQ1: what are the current AI-based IDF detection methods?
- RQ2: what are the key open challenges in addressing IDF?

Search strategy

The search strategy includes selecting the databases, keywords, search strings, and selection criteria in the search.

Four databases were selected to ensure comprehensive coverage of peer-reviewed technology and computer science literature. *ACM Digital Library* and *IEEE Xplore* provide the most authoritative coverage of core computer science and electrical engineering conferences and journals. While ScienceDirect contributes high-impact literature in AI and engineering, Scopus was utilized as a broad multidisciplinary index to capture high-quality research from other major publishers such as Taylor & Francis, Wiley, and Springer.

Preprints and industry reports were not included, as they typically follow non-uniform evaluation standards and are therefore unsuitable as a systematic evidence base.

Two limitations are acknowledged: First, Web of Science was not searched separately, as Scopus was relied upon for broad multidisciplinary coverage. Second, the reliance on published literature means the review may be subject to publication bias, a common limitation of systematic reviews of this kind.

The databases are listed as

- ACM Digital Library (https://dl.acm.org/);
- IEEE Xplore (https://ieeexplore.ieee.org/);
- ScienceDirect (https://www.sciencedirect.com/);
- Scopus (https://www.scopus.com/).

Based on the research questions defined in Section II-A, we divided the research questions into three search categories and collected a series of related concepts for each (Table 2). We then formulated the search query as Table 1.

We searched the titles and abstracts for articles published from 2020 to 2024. This range was chosen to capture the most recent generation of AI-based methods to IDF detection, reflecting a broad scope of AI-based methods rather than the underlying methods' novelty. The period is marked by a notable rise in AI research activity of which the introduction of GPT-3 by OpenAI in 2020 is one prominent example, though AI and its related technologies have been studied and applied in fraud detection for decades (Heaven 2020).

We defined the following inclusion and exclusion criteria to restrict the selection of articles from the returned results (Table 3).

This review only includes technical papers on IDF detection and excludes articles that address cybercrimes and cyberattacks, as mentioned in Inclusion Criteria 5. The reason for designing Inclusion Criteria 5 this way is that the focus of detecting IDF that involves using false or stolen identities is on validating the authenticity of the identity being used (i.e., whether it is true or false). However, as the search query includes "identity theft," articles related to defending against cyberattacks, such as malware detection, are also included in the search results. This is not aligned with the research problem of detecting IDF. Therefore, to more accurately select articles relevant to

Table 1. Search query string.

("identity" OR "ID")
AND
("fraud" OR "theft" OR "scam")
AND
("detect" OR "prevent" OR "verify" OR "authenticate")
AND
("artificial intelligence" OR "machine learning" OR "deep learning")

Table 2. Search categories and relevant concepts.

Search Categories	Relevant Concepts
IDF detection	identity, ID, fraud, theft, scam, detect, prevent, verify, authenticate
AI-related technologies	machine learning, deep learning, artificial intelligence

Table 3. Inclusion and exclusion criteria.

No.	Inclusion	Exclusion
1	Articles from 2020 to 2024	Book chapters, books, datasets, newspapers, magazines, newsletters
2	Peer-reviewed articles	Non-computer science articles
3	English-language articles	Articles on cybercrimes or cyberattacks
4	Conference and journal papers	Duplicate studies
5	Focus on identity fraud	–

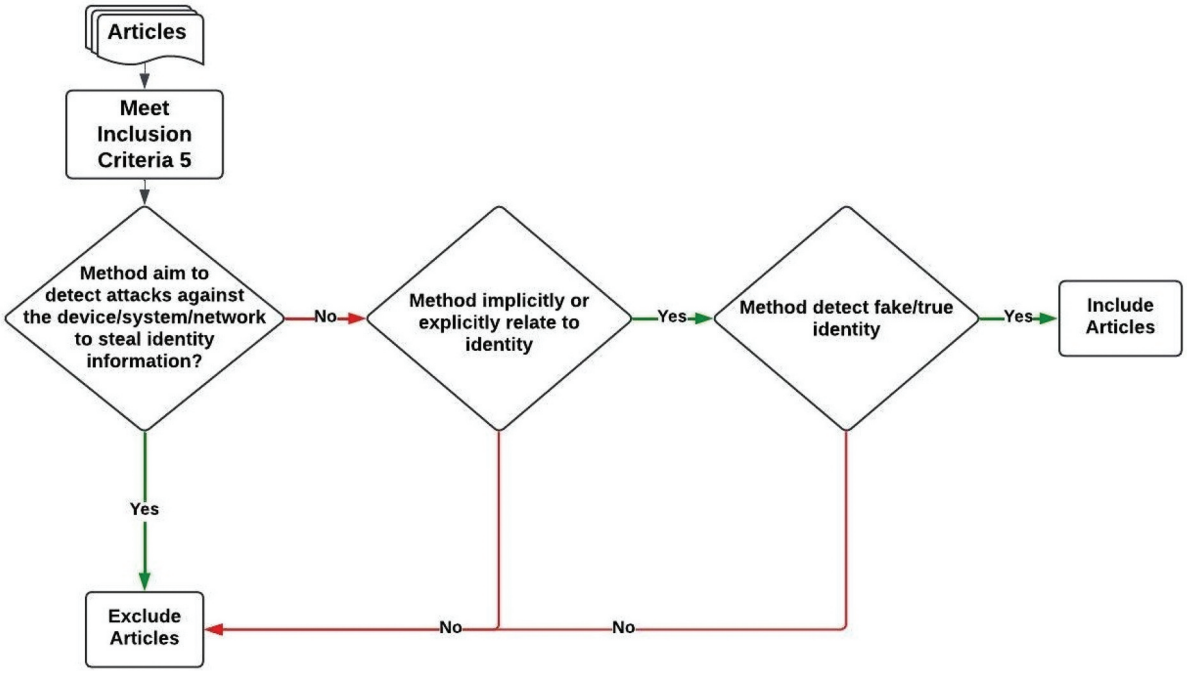


Figure 2. Decision rules for Inclusion Criteria 5 (Table 4).

this topic, Inclusion Criteria 5 filters the results using the process outlined in Figure 2. If an article does not focus on detecting cyberattacks and proposes a method related to "identity" used for validating true or false identities, it will be included.

The five stages of filtering, as described in Table 4, ensure a thorough selection of relevant articles. First, duplicates are removed, followed by the exclusion of reviews and case studies. Next, papers are filtered based on keywords in their titles and abstracts. The fourth stage applies inclusion and exclusion criteria to refine the selection further. Finally, the remaining full texts are evaluated using six Quality Assessment Criteria (Table 5) to confirm their relevance to the research topic.

Table 4. Filtering rules.

Filter	Description
Filter 1	Duplicate papers
Filter 2	Remove reviews, surveys, and case studies
Filter 3	Title includes keywords in the search string
Filter 4	Title and abstract filtering using Exclusion Criteria 2 and 3 and Inclusion Criteria 5
Filter 5	Text filtering

Table 5. Quality assessment criteria.

ID	Quality Assessment
1	Does the study clarify its purpose?
2	Is the study relevant to the research questions?
3	Are the techniques presented and implemented?
4	Are the results evaluated adequately?
5	Does the study provide clear findings and discussion?
6	Is the future direction clearly stated?

Table 6. Data extraction form.

Information Extracted	Research Questions
The identified problems in IDF detection	RQ1
The techniques used for IDF detection	RQ1
The strengths and weaknesses of the techniques	RQ1
The AI technologies used in the techniques	RQ1
The evaluation metrics used to assess the effectiveness of the techniques	RQ1
The existing challenges, gaps, trends, and future direction of the study	RQ1 and RQ2

Data extraction form

This section gathers and extracts data from the primary studies addressing the review questions outlined in Section II-A. The extracted data were organized in a spreadsheet, and the critical information retrieved from these studies was summarized using the Data Extraction Form in Table 6.

Search results

To ensure a rigorous and transparent selection process, initial screening was conducted using Rayyan, which facilitated automated de-duplication and a systematic title and abstract review. As detailed in the PRISMA flow calculations (Appendix A, Note 1 and Note 2), 1257 records were initially identified, with 137 duplicates removed (Filter 1). During the title and abstract screening, Rayyan's keyword-highlighting feature was used to manually exclude 81 non-primary papers, i.e., survey, literature review, systematic literature review, and case report (Filter 2) and 269 papers lacking core keywords (Filter 3). Subsequently, 588 papers were excluded (Filter 4) for falling outside the scope of IDF or focusing primarily on prevention against cyberattacks, based on Exclusion Criterion 3 and Inclusion Criterion 5 (Figure 2). Following this, EndNote was employed for full-text management and quality assessment, where 139 additional reports were excluded (Filter 5) based on a predefined quality score (3–6). This multi-stage process resulted in a final inclusion of 43 studies.

The quality assessment criteria are in Table 5. If the paper meets the requirements, it will score 1; otherwise, it will score 0. Papers in a score range from 3 to 6 were selected. The specific rationale for each criterion is as follows:

- **QA1 and QA2 (Relevance and Alignment)**: Beyond keyword screening, QA1 evaluates if the study explicitly states its objectives by clearly defining the research challenges and proposed solutions. QA2 assesses the alignment between these objectives and the experiments; a score of 1 is awarded only if the methodology and validation metrics directly address the research questions.
- **QA3 and QA4 (Implementation and Evaluation)**: These criteria focus on technical rigor. QA3 requires a transparent description of data characteristics and a logically verifiable workflow. QA4 requires robust evaluation through standard metrics and comparative analysis (e.g., ablation studies).

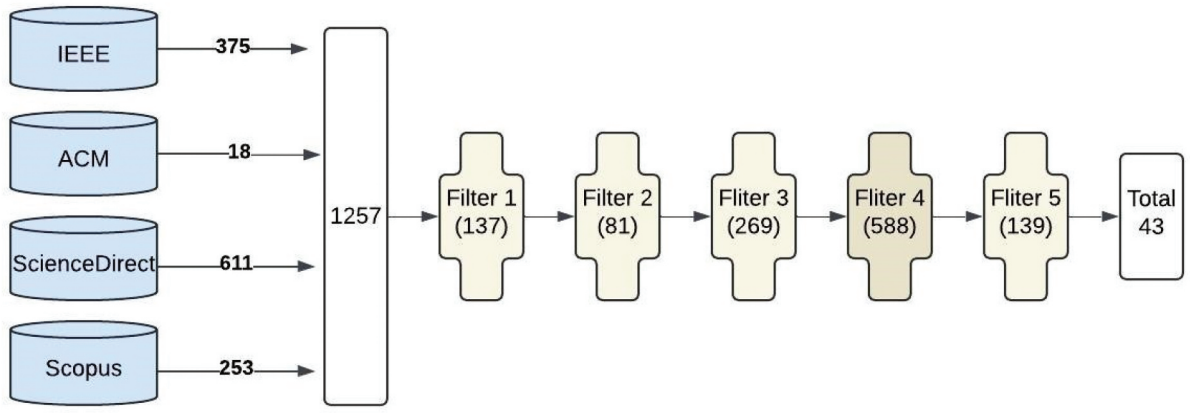


Figure 3. Search results.

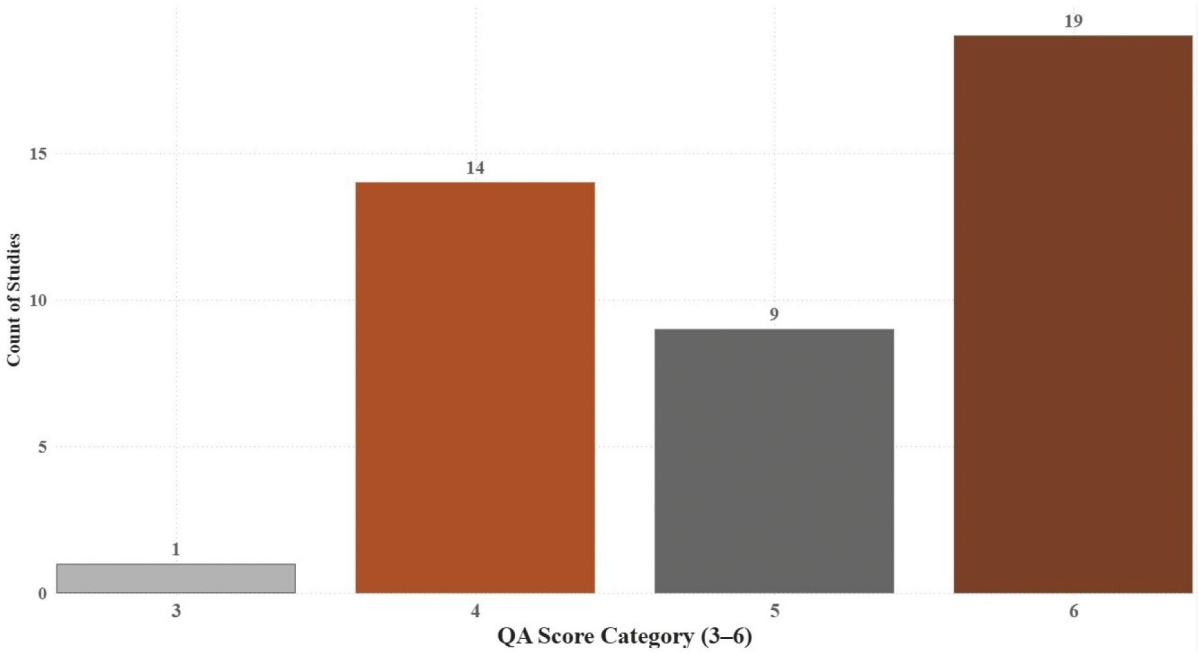


Figure 4. Quality assessment score distribution ($n = 43$).

- **QA5 and QA6 (Discussion and Future Directions)**: These evaluate the presence of in-depth results analysis and forward-looking perspectives.

The selection threshold was set at score 3, but with a primary focus on the technical core (QA1–QA4). Under this rule, a paper must satisfy the majority of the first four criteria to be included. For example, even if a study provides a good discussion and future work (QA5 and QA6), it would still be excluded if it only satisfies one of the first four criteria (QA1–QA4). This ensures that we only include papers with a solid foundation in problem definition, methodology, and evaluation, rather than those that only provide descriptive analysis. Figure 3 presents the selection outcomes.

The distribution of quality assessment scores is shown in Figure 4. In total, 19 studies (44.2%) scored 6, 9 studies (20.9%) scored 5, 14 studies (32.6%) scored 4, and 1 study (2.3%) scored 3. Most score-4 studies primarily lost points on the findings and/or directions criteria, which indicate weaker articulation of implications and contributions rather than deficiencies in purpose, relevance, presentation, or evaluation. Sensitivity analysis confirmed that the key findings remained stable after removing lower-quality studies. Appendix B Table B1 provides the full per-paper item-level scoring.

Validity and rigor

To ensure the validity and rigor of the research methodology, the search terms and keywords derived from the research questions were tested across the pre-selected databases, and the search process was reviewed for coverage prior to documentation. Study selection followed a layered set of pre-defined criteria (Filter 1–5), applied sequentially and consistently so that each inclusion and exclusion decision could be traced to a specific criterion; the resulting counts at every stage are documented in the PRISMA flow diagram (Appendix A). The quality assessment criteria (QA1–QA6) were applied by the lead researcher to reduce subjectivity in the final selection. Given that human error can introduce bias in a review of this scale, regular meetings were held between the lead researcher and the supervising researchers, in which screening and scoring decisions were reviewed, discussed, and revised where necessary. As studies were not scored independently by two reviewers, inter-rater reliability statistics could not be computed for the quality assessment, which we note as a limitation of this procedure.

Results

This section presents the results of the SLR, addressing RQ1 and RQ2. The findings are organized into five main areas: (A) taxonomy of AI-based IDF detection methods (Figure 5), (B) framework and frequency analysis for open challenges, (C) observed trends, and (D–E) an in-depth analysis of individual methods, including the identification of open challenges across technical and nontechnical dimensions.

Taxonomy

The review of the literature resulted in the identification of three main types of AI-based IDF detection methods: biometric recognition, visual anomaly detection, and user-behavior anomaly detection. These methods are organized into a taxonomy (Figure 5) and broadly classified into two authentication phases: authentication and continuous authentication. The categorization was based on the operational purposes, process, and core principles of the methods.

Conceptually, this two-phase structure aligns with established theories of identity assurance and trust management. Authentication corresponds to the initial trust establishment phase, in which an entity's claimed identity is verified against trusted credentials (Grassi, Garcia, and Fenton 2017; ISO/IEC 2011). Continuous authentication reflects the principle of sustained trust evaluation, where trust is not treated as a one-time decision but as a dynamic, session-long process that must adapt to evolving risk signals (ISO/IEC 2011; Litan 2015). This framing draws on risk-based authentication models, in which the level of scrutiny is proportional to the assessed risk at each interaction point (Grassi, Garcia, and Fenton 2017). By grounding the taxonomy in these broader conceptual foundations, the classification moves beyond a purely technical inventory toward a theoretically informed understanding of how AI-based methods address different stages and dimensions of the identity–trust lifecycle.

In this taxonomy, authentication and continuous authentication methods are first distinguished by the timing of identity validation. Authentication typically validates identity as a discrete event at the entry point, whereas continuous authentication involves ongoing monitoring of user behavior, device characteristics, and environmental factors throughout a session (ISO/IEC 2011; Litan 2015; Sadowski et al. 2019). Authentication methods were further classified based on their operational purposes. For example, methods designed to defend against attacks during authentication, such as fake biometrics or identity forgery, were categorized as authentication defense methods. Conversely, methods that solely address authentication system functionality were classified under authentication. Following this second-round classification, a detailed examination of each method's operational principles, functional components, and implementation techniques was conducted to ensure clear and systematic categorization.

This review revealed that biometric recognition is the sole method applied for authentication. Methods targeting authentication attacks rely on visual anomaly detection. In this review, we use the term "Visual

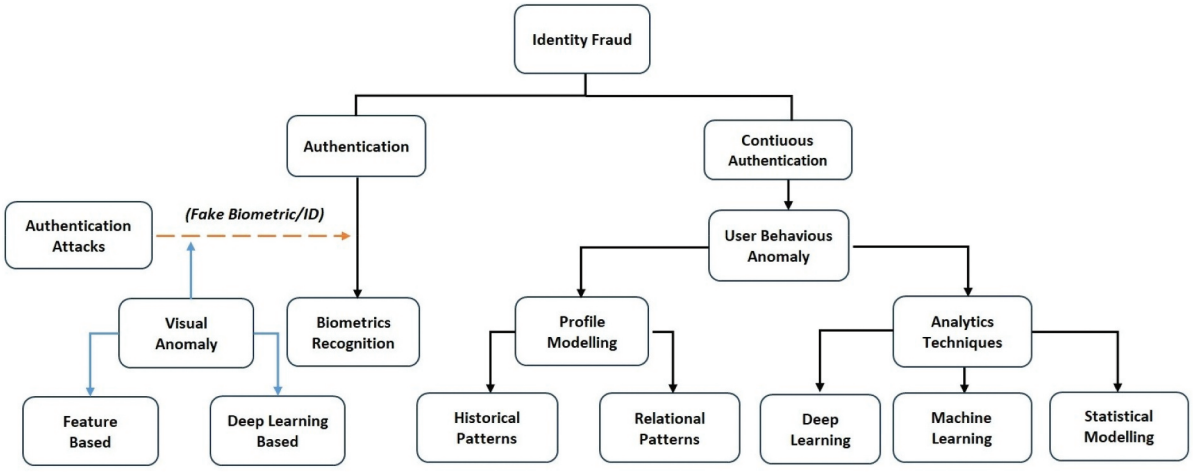


Figure 5. Taxonomy of AI-based IDF detection and prevention methods.

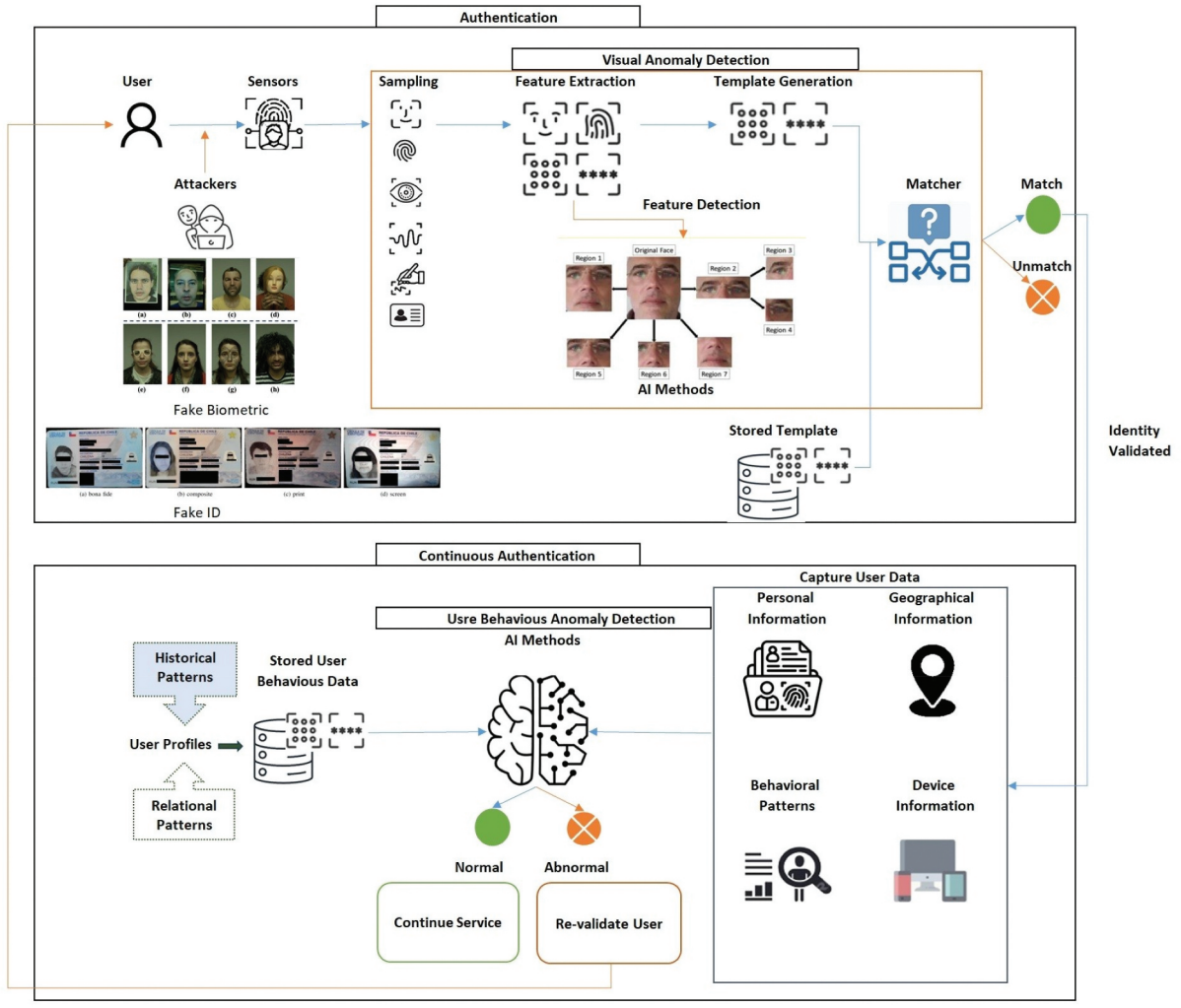


Figure 6. Operational process of authentication and continuous authentication with IDF detection and prevention methods (source of image data: Benalcazar et al. 2023; Heusch et al. 2020; V. Sharma et al. 2023). Note: The IDF detection and prevention process begins with identity authentication, where users' biometric data or official ID documents are captured. Key features are extracted to create a user-specific template, which is stored and matched against records to validate identity. Visual anomaly detection with advanced feature extraction algorithms analyzes provided samples in detail to safeguard the process. Continuous authentication monitors users' behavioral patterns, relying on baseline profiles to identify anomalies. If anomalies are detected, the system suspends access and prompts for additional identity verification.

Anomaly Detection" to describe the methods used to detect these authentication attacks. These attacks include biometric spoofing, deepfakes, and ID document forgery. Here, biometric spoofing (also known as a Presentation Attack) involves presenting a fake biometric artifact such as a photo or mask to the sensor. Deepfakes are AI-generated media where a person's likeness is synthesized or replaced, and ID document forgery refers to the alteration or complete fabrication of identity documents. These visual anomaly detection methods are further divided into feature-based and deep learning-based approaches.

For continuous authentication, all methods fall under user behavior anomaly detection, where User and Entity Behavior Analytics (UEBA) serves as the foundational framework. This framework was further examined through two key technical components: profiling models and analytical techniques (Litan 2015). The review articles were categorized accordingly to analyze the methods used for constructing behavioral profiles and the analytical techniques applied for anomaly detection, offering a comprehensive understanding of the underlying technologies.

Figure 6 further illustrates how these methods are implemented within two key authentication phases: authentication and continuous authentication. It maps the IDF detection and prevention methods from the taxonomy to their respective roles in these phases, providing a detailed view of their functional

implementation. For example, in customer on-boarding (authentication), selfie-liveness and ID document images are processed by biometric recognition and visual anomaly models to determine whether the extracted features match a trusted reference. Onboarding is rejected only when a mismatch or anomaly is detected. During the session (continuous authentication), behavioral signals are analyzed by user-behavioral anomaly detection models to detect deviations from the user's historical or relational profile. When anomalies are detected, the system triggers a step-up identity validation request to re-confirm the user's identity.

Framework and frequency analysis for open challenges

The open challenges identified in this study were examined using the Adaptive Enterprise Architecture (EA) framework (Gill 2012), which addresses key aspects critical to industrial projects and enterprise decision-making. This structured approach facilitates the categorization and analysis of challenges in IDF detection methods. This review selected three key dimensions from the framework: Technology, Human, and Environment.

The Technology dimension was evaluated based on quality characteristics such as accuracy, robustness, resiliency, and effectiveness (Kharchenko, Fesenko, and Illiashenko 2022), while the Human and Environment dimensions were derived and coded to capture nontechnical challenges. To quantify and highlight prominent issues across methods, frequency analysis was applied, aligning the findings with key themes identified in the selected literature. In this synthesis, papers were not treated as mutually exclusive; for example, a study reporting both accuracy and robustness challenges was tallied under both categories. This approach was used to examine both individual open challenges for each method and broader challenges across the entire review.

Tables 7 and 8 present evaluation criteria and frequency analysis results for the open challenges in this study. The findings highlight four key challenges in AI-based IDF detection methods: data quality and diversity, data privacy and security, dynamic and evolving fraud patterns, and optimizing training time and method efficiency. These challenges will be examined in detail in the discussion section, guiding the exploration of future research directions and advancements. The evaluation categories are nonmutually exclusive; for example, a study might report both accuracy and robustness metrics, contributing to multiple evaluation categories.

Trends in method distribution

The distribution of AI-based IDF detection methods reveals important trends in the focus and application of these techniques. As shown in Table 9, 55.81% of the methods are predominantly applied to authentication, whereas 44.19% are designed for continuous authentication, reflecting a growing emphasis on adaptive and ongoing monitoring mechanisms.

As shown in Figure 7, biometric authentication is the most commonly used method in authentication, accounting for 79% (19/24) of authentication-related articles. Among these, 36.8% (7/19) focus on performance improvements in biometric recognition systems, while 63.2% (12/19) address their vulnerabilities. Notably, facial recognition emerges as the most discussed biometric method, appearing in 12 articles within this category.

Visual anomaly detection is identified as the primary defense mechanism in authentication processes, representing 39.35% of the total articles reviewed. In contrast, continuous authentication methods exclusively rely on user behavior anomaly detection, which accounts for 44.19% of the articles.

Deep learning techniques dominate both visual anomaly detection and user behavior anomaly detection, with 76% of visual anomaly detection and 53% of user behavior anomaly detection articles employing this approach. These findings indicate a growing research emphasis on deep learning as a means to address the evolving challenges of IDF.

This distribution provides a foundation for analyzing the operational principles, related techniques, and specific challenges of each method, as elaborated in the subsequent sections.

Table 7. Technology's challenges and analysis.

Technology	Definition	Open Challenge	Reference	Freq.	%
Accuracy	The capability of a method to deliver results from given data that closely align with their true values (Kharchenko, Fesenko, and Illiashenko 2022)	**Data quality**	Benalcazar et al. (2023); Chang, Shih, and Lin (2024); Chapel, Al-Ghadi, and Burie (2023); Claveria et al. (2021); Islam and Elwarfalli (2023); Izdihar et al. (2023); Magee, Sheridan, and Thorpe (2023); Mazaheri and Roy-Chowdhury (2022); V. Sharma et al. (2023); Srivastva, Singh, and Singh (2022); Vinod and Senthil Kumar (2023)	11	**26%**
		Inconsistent performance across different environments	Claveria et al. (2021); M. Hu and Hu (2023)	2	5%
		Accuracy trade-off with technique's efficiency	S. Hu et al. (2021); Izdihar et al. (2023)	2	5%
		Threshold misconfiguration	Chang, Shih, and Lin (2024); Singhai, Aanjankumar, and Poonkuntran (2023)	2	5%
		Encryption techniques	H. Yang et al. (2023)	1	2%
		Poor data preprocessing	Chang, Shih, and Lin (2024)	1	2%
Robustness	The capacity of a method to function correctly across diverse input data and operating conditions (Kharchenko, Fesenko, and Illiashenko 2022)	**Inconsistent performance with diverse datasets**	Agarwal et al. (2020); Al-Ghadi et al. (2023); Cao et al. (2023); Guo et al. (2023); Hanok and Shankaraiah (2023); M. Hu and Hu (2023); Jaiswal and Srivastava (2023); Jayanshre and Amsaprabhaa (2024); Krishna and Bhuvaneswari (2023); Safara et al. (2020); Srivastva, Singh, and Singh (2022); Uliyan, Sadeghi, and Jalab (2020); Wang et al. (2021); Ye et al. (2022)	14	**33%**
		Inconsistent performance with diverse features	Agarwal et al. (2020); Al-Ghadi et al. (2023); Hanok and Shankaraiah (2023); Jayanshre and Amsaprabhaa (2024); Umbrani et al. (2024)	5	12%
		Inconsistent performance with low-quality data samples	Chapel, Al-Ghadi, and Burie (2023)	1	2%
Resiliency	The ability of a method to maintain functionality across changing requirements or under dynamic conditions (Kharchenko, Fesenko, and Illiashenko 2022).	**Limited transferability across conditions**	Chapel, Al-Ghadi, and Burie (2023); Claveria et al. (2021); M. Hu and Hu (2023); Islam and Elwarfalli (2023); Levi et al. (2022); Pujol et al. (2019); V. Sharma et al. (2023); Trabelsi, Pic, and Dugelay (2022); Uliyan, Sadeghi, and Jalab (2020)	9	**21%**
		Lack of data diversity	Al-Ghadi et al. (2023); Dornadula and Geetha (2019); Hong, Connie, and Ong Goh (2023); Krishna and Bhuvaneswari (2023)	4	9%
		Limited transferability of the techniques	Agarwal et al. (2020); Aguilera et al. (2023); Al-Ghadi et al. (2023); Cao et al. (2023); Chapel, Al-Ghadi, and Burie (2023); Dong et al. (2023); Guo et al. (2023); Hong, Connie, and Ong Goh (2023); Jaiswal and Srivastava (2023); Pujol et al. (2019); Umbrani et al. (2024); Vinod and Senthil Kumar (2023); Zhao and Wang (2021)	13	**30%**
		Evolving technology	Dong et al. (2023); Magee, Sheridan, and Thorpe (2023); Mazaheri and Roy-Chowdhury (2022); Trabelsi, Pic, and Dugelay (2022)	4	9%
		Inconsistent performance over time	Dornadula and Geetha (2019); Wang et al. (2021)	2	5%

(*Continued*)

Table 7. (Continued).

Technology	Definition	Open Challenge	Reference	Freq.	%
Effectiveness	The efficiency of the method, measured by its ability to achieve intended outcomes relative to the resources utilized (Kharchenko, Fesenko, and Illiashenko 2022)	Techniques complexity	Benalcazar et al. (2023); Dornadula and Geetha (2019); Priyaradhikadevi et al. (2023); Pujol et al. (2019); Safara et al. (2020); Zhao and Wang (2021)	6	14%
		Lengthy training time for algorithms/ models	Agarwal et al. (2020); Cao et al. (2023); Chapel, Al-Ghadi, and Burie (2023); Dong et al. (2023); M. Hu and Hu (2023); Islam and Elwarfalli (2023); Jaiswal and Srivastava (2023); Khan et al. (2023); Pujol et al. (2019); Uliyan, Sadeghi, and Jalab (2020); Vinod and Senthil Kumar (2023)	11	**26%**
		Resource demands for model update	Dornadula and Geetha (2019); Levi et al. (2022); Lv et al. (2021)	3	7%
		Inconsistent performance over time	Dornadula and Geetha (2019); Wang et al. (2021)	2	5%
		Performance dependency on key modules	Safara et al. (2020); Ye et al. (2022)	2	5%
		Lack data diversity	Jaiswal and Srivastava (2023); Mazaheri and Roy-Chowdhury (2022)	2	5%
		Limited datasize	Hanok and Shankaraiah (2023); Hong, Connie, and Ong Goh (2023); V. Sharma et al. (2023); Srivastva, Singh, and Singh (2022)	4	9%

Table 8. Human's and environment's challenges and analysis.

Category	Definition	Challenge	Reference	Frequency	Percentage
Human	Variations in application environments, or user behaviors that affect the performance of methods	Dynamic user states/ behavior	Claveria et al. (2021); Hanok and Shankaraiah (2023); Islam and Elwarfalli (2023); Levi et al. (2022); Wang et al. (2021); H. Yang et al. (2023); Zhao and Wang (2021)	7	16%
		Evolving fraud tactics	Chapel, Al-Ghadi, and Burie (2023); Dong et al. (2023); Gangan et al. (2023); Guo et al. (2023); S. Hu et al. (2021); Mazaheri and Roy-Chowdhury (2022); V. Sharma et al. (2023); Trabelsi, Pic, and Dugelay (2022); Vinod and Senthil Kumar (2023)	9	21%
Environment	The influence of legal and regulatory frameworks on the functionality and performance of methods	Data privacy regulation	Al-Ghadi et al. (2023); Benalcazar et al. (2023); Chang, Shih, and Lin (2024); Kakkar et al. (2023); Lv et al. (2021); Singhai, Aanjankumar, and Poonkuntran (2023); H. Yang et al. (2023)	7	16%

AI-based IDF prevention and detection method: authentication

This section focuses on two AI-based authentication methods: biometric recognition and visual anomaly detection. Each method is analyzed in detail based on its operational principles, including core components and related implementation techniques. Using the categorization framework from Section B, the open challenges of current techniques are systematically evaluated and elaborated in the subsequent sections.

AI-based authentication method: biometrics recognition

This section analyzes the biometric recognition system in terms of its operational components and examines the open challenges identified through frequency analysis.

Operational principle. Biometrics refers to the process of identifying individuals by capturing and analyzing their unique physical or behavioral characteristics using electronic devices, such as face, fingerprints, signature, voice, electrocardiogram (ECG) and chirp sound (Claveria et al. 2021; Islam and Elwarfalli 2023;

Table 9. Overview of authentication and continuous authentication's IDF methods.

Main Process	Substage	IDF Methods	Subcategories	Reference	Sub-Freq.	Sub-%	Total Freq.	%	Total %
Authentication	Authentication	Biometrics recognition	Biometrics recognition	Claveria et al. (2021); M. Hu and Hu (2023); Islam and Elwarfalli (2023); Izdihar et al. (2023); Krishna and Bhuvaneswari (2023); Srivastva, Singh, and Singh (2022); Vinod and Senthil Kumar (2023)	7	100%	7	16.28%	55.81%
	Authentication defense	Visual anomaly	Feature base	Avcibas (2024); Chapel, Al-Ghadi, and Burie (2023); Izdihar et al. (2023); Magee, Sheridan, and Thorpe (2023); Pujol et al. (2019); V. Sharma et al. (2023); Vinod and Senthil Kumar (2023)	7	41%	17	39.53%	
			Deep learning base	Agarwal et al. (2020); Al-Ghadi et al. (2023); Benalcazar et al. (2023); Cao et al. (2023); Guo et al. (2023); Izdihar et al. (2023); Jaiswal and Srivastava (2023); Jayanshre and Amsaprabhaa (2024); Mazaheri and Roy-Chowdhury (2022); V. Sharma et al. (2023); Trabelsi, Pic, and Dugelay (2022); Uliyan, Sadeghi, and Jalab (2020); Vinod and Senthil Kumar (2023)	13	76%			
			Deep learning base	Agarwal et al. (2020); Al-Ghadi et al. (2023); Benalcazar et al. (2023); Cao et al. (2023); Guo et al. (2023); Izdihar et al. (2023); Jaiswal and Srivastava (2023); Jayanshre and Amsaprabhaa (2024); Mazaheri and Roy-Chowdhury (2022); V. Sharma et al. (2023); Trabelsi, Pic, and Dugelay (2022); Uliyan, Sadeghi, and Jalab (2020); Vinod and Senthil Kumar (2023)	13	76%			
			Profile modeling: historical pattern	Aguilera et al. (2023); Dornadula and Geetha (2019); Gangan et al. (2023); Hanok and Shankaraiah (2023); Hong, Connie, and Ong Goh (2023); Kakkar et al. (2023); Levi et al. (2022); Lv et al. (2021); Priyaradhikadevi et al. (2023); Safara et al. (2020); Singhai, Aanjankumar, and Poonkuntran (2023); Umbrani et al. (2024); Wang et al. (2021); H. Yang et al. (2023); Zhao and Wang (2021)	15	79%			

(*Continued*)

Table 9. (Continued).

Main Process	Substage	IDF Methods	Subcategories	Reference	Sub-Freq.	Sub-%	Total Freq.	%	Total %
Continuous authentication	Continuous authentication	User behavior anomaly detection	Profile modeling: relational pattern	Chang, Shih, and Lin (2024); S. Hu et al. (2021); Khan et al. (2023); Ye et al. (2022)	4	21%	19	44.19%	44.19%
			Analytics techniques: deep learning	Chang, Shih, and Lin (2024); Gangan et al. (2023); Hong, Connie, and Ong Goh (2023); S. Hu et al. (2021); Kakkar et al. (2023); Khan et al. (2023); Levi et al. (2022); Safara et al. (2020); H. Yang et al. (2023); Ye et al. (2022)	10	53%			
			Analytics techniques: machine learning	Aguilera et al. (2023); Dornadula and Geetha (2019); Gangan et al. (2023); Levi et al. (2022); Lv et al. (2021); Priyaradhikadevi et al. (2023); Singhai, Aanjankumar, and Poonkuntran (2023); Umbrani et al. (2024); Wang et al. (2021)	9	47%			
			Analytics techniques: statistical modeling	Hanok and Shankaraiah (2023); Zhao and Wang (2021)	2	11%			

Note:
[a]The Reference column lists all the papers that include methods within the subcategories, including duplicates.
[b]Subfrequency represents the number of papers for each subcategory.
[c]Subpercentage is calculated as the ratio of sub-frequency to total frequency, showing the relative contribution of each subcategory.
[d]Total frequency represents the total number of papers across the two substages.
[e]Percentage is calculated as total frequency divided by the overall frequency of the main process (e.g., for authentication, 7/(7 + 17)).
[f]Total percentage indicates the proportion of the main process's total frequency to the total number of selected articles (e.g., for authentication, 24/43).

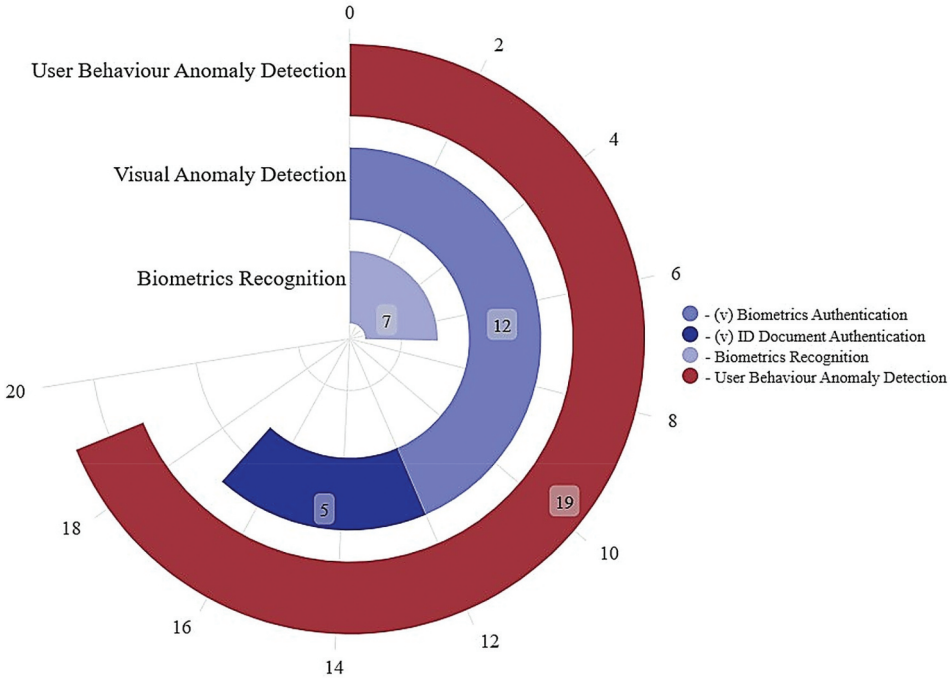


Figure 7. Distribution of IDF detection and prevention articles (2020–2024): continuous authentication methods account for the largest share (44%). Note: Each article is assigned to a single method category under the proposed taxonomy (mutually exclusive). Total article count = 43. Blue = authentication (inner circle: authentication methods; outer circle: authentication defense methods), red = continuous authentication.

Izdihar et al. 2023; Pujol et al. 2019; Srivastva, Singh, and Singh 2022; Vinod and Senthil Kumar 2023), as shown in Figure 8. Characterized by their inherent uniqueness and reliability, biometrics offer an effective method for validating identity (Claveria et al. 2021; Jain and Kumar 2012). As illustrated in Figure 6, a biometric recognition system operates through a pipeline of four modules, including sensor capture, feature extraction, template matching, and decision-making (Jain and Kumar 2012), progressing from raw data acquisition to a binary accept/reject outcome. Of particular note across the reviewed studies is the distinction between one-to-one verification (e.g., smartphone unlock, M. Hu and Hu 2023; Krishna and Bhuvaneswari 2023; Vinod and Senthil Kumar 2023) and one-to-many identification (e.g., watchlist screening in financial onboarding, Izdihar et al. 2023), which presents different scalability and accuracy trade-offs.

Open challenges. Appendix D Table D1 summarizes the key challenges identified in biometric recognition systems based on the findings of this review. These challenges highlight critical limitations in the current state of the technology, including data quality, system resilience, implementation effectiveness and security vulnerabilities, which are detailed below.

- Data quality is the most frequently cited challenge affecting model accuracy, due to variability in data-sampling conditions and sample collection devices (see Table D1). Factors, such as lighting and angle (M. Hu and Hu 2023), are common sources of variability in biometric recognition. As noted in Uliyan, Sadeghi, and Jalab (2020), different fingerprint sampling devices produce varying biometric templates, which also increases the difficulty of detection. In facial recognition systems, dataset diversity is another critical factor influencing performance. For example, changes in a person's age or facial expressions in facial recognition systems can impact verification accuracy (Izdihar et al. 2023). If the dataset lacks diverse samples for each user, the verification results may be significantly affected.
- High-quality data relies on advanced sampling devices, particularly when employing more specialized biometric features like Chirp Sound and ECG (Claveria et al. 2021; Islam and Elwarfalli 2023; Srivastva, Singh, and Singh 2022). These features have been validated for their role in biometric recognition (Srivastva, Singh, and Singh 2022). However, implementing these methods requires specialized equipment for capturing high-quality signals. Similarly, noise during voice data collection remains a significant challenge, as highlighted by Claveria et al. (2021) and Nugroho and Winarno (2022).
- A significant challenge to achieving the effectiveness of biometric recognition systems lies in their inherent complexity. Long training times (M. Hu and Hu 2023; Islam and Elwarfalli 2023; Vinod and Senthil Kumar 2023), intricate biometric data structures such as ECG signal sequences, and extended data collection for specialized biometrics (Claveria et al. 2021), all contribute to delays in achieving practical and timely implementation.
- Common biometric systems (e.g., facial recognition, fingerprints, voice recognition) are vulnerable to potential security breaches from evolving fraud technologies, such as morphing attacks and presentation attacks (Avcibas 2024; Guo et al. 2023; M. Hu and Hu 2023; V. Sharma et al. 2023) and deepfake attacks (Agarwal et al. 2020; Cao et al. 2023; Dong et al. 2023; Jayanshre and Amsaprabhaa 2024; Mazaheri and Roy-Chowdhury 2022). Criminals can bypass these systems using simple forgeries, like 2D or 3D-printed masks or by displaying images on a phone screen (Guo et al. 2023; V. Sharma et al. 2023). In high-security areas such as border control, criminals may use morphing techniques by submitting a fused-feature passport photo during visa applications (M. Hu and Hu 2023). This changes the stored template in the database, complicating the detection of fraudulent activities.

AI-based authentication defense method: visual anomaly detection

This section analyzes the visual anomaly detection methods in terms of their key implementation techniques. Additionally, it summarizes the detection techniques, targeted features, addressed attack types, and the algorithms or models in table and examines the open challenges identified through frequency analysis to provide insights into the limitations of current methods.

Operational principle. Visual anomaly detection is a method designed to detect abnormal or unexpected patterns in images and videos that are different from the expected form as defined (J. Yang et al. 2022). This technique is implemented within recognition systems to enhance the

matching accuracy with more detailed features, as highlighted in the designated region of the authentication process (Figure 6). Among the articles discussing vulnerabilities in authentication systems, their main focus is to counter biometric spoofing and deepfake and ID document forgery attacks. A common pitfall for these attacks is that low-resolution video and compression artifacts (e.g., from social media uploads) can destroy the subtle digital on which visual anomaly detection models depend.

Appendix G Table G1 summarized articles for visual anomaly detection along with the types of attacks they counter. Aside from Dong et al. (2023), Guo et al. (2023), and V. Sharma et al. (2023), which use adversarial learning to counter these attacks, other studies (Avcibas 2024; Cao et al. 2023; Izdihar et al. 2023; Jayanshre and Amsaprabhaa 2024; Trabelsi, Pic, and Dugelay 2022) rely on techniques in feature detection or deep learning models to achieve successful defense.

Based on the collected articles, visual anomaly detection can be implemented using feature detection and deep learning. The reviewed studies employ two complementary approaches. *Feature-based methods* use handcrafted descriptors, such as HOG (Pujol et al. 2019; V. Sharma et al. 2023), LBP (Chapel, Al-Ghadi, and Burie 2023), SIFT and ORB (Vinod and Senthil Kumar 2023), Gabor filters (Izdihar et al. 2023), and statistical co-occurrence matrices (Avcibas 2024), to extract texture and spatial anomalies from images. These methods focus on critical regions of change, such as facial landmarks in paper-cut attacks (Guo et al. 2023) or subtle statistical differences in morphing attacks. *Deep learning methods* automate feature extraction using architectures, such as CNN (Izdihar et al. 2023; V. Sharma et al. 2023; Vinod and Senthil Kumar 2023), Inception/XceptionNet (Cao et al. 2023; Jayanshre and Amsaprabhaa 2024; Mazaheri and Roy-Chowdhury 2022), and encoder-decoder networks (e.g., U-Net, Jaiswal and Srivastava 2023). Encoder–decoder architectures are particularly effective for pixel-level localization of tampered regions (Mazaheri and Roy-Chowdhury 2022; V. Sharma et al. 2023). Several studies combine both approaches, using handcrafted descriptors to augment deep network features (Cao et al. 2023; V. Sharma et al. 2023). Table G1 maps each reviewed study to its detection approach, targeted features, attack type, and model architecture.

Open challenges. Appendix D Table D2 provides an overview of the open challenges of the visual anomaly detection method identified in this review. This method faces several critical challenges, including data–quality issues, limited robustness and resilience, operational inefficiencies, data privacy constraints, and increasingly sophisticated fraud tactics driven by technological advancements as shown below:

- Detection accuracy is influenced by the quality of the data, such as image resolution (Mazaheri and Roy-Chowdhury 2022) and the quality of data processing (Benalcazar et al. 2023; Chapel, Al-Ghadi, and Burie 2023; Magee, Sheridan, and Thorpe 2023). For instance, Benalcazar et al. (2023) applied generative adversarial networks (GAN), StyleGAN2, and CycleGAN to generate synthetic ID cards. Though the method performs well on trained datasets, the synthetic data lack real-world fidelity, compromising the model's performance in real-world scenarios (Chapel, Al-Ghadi, and Burie 2023). The study then highlights the impact of document misalignment during video processing, which leads to detection errors.
- Lack of data diversity adversely affects robustness and transferability, hindering the method's performance in real-world scenarios (Agarwal et al. 2020; Al-Ghadi et al. 2023; Cao et al. 2023; Guo et al. 2023; Jaiswal and Srivastava 2023; Pujol et al. 2019; Uliyan, Sadeghi, and Jalab 2020). As mentioned in Al-Ghadi et al. (2023), Cao et al. (2023), Pujol et al. (2019), Uliyan, Sadeghi, and Jalab (2020), the model can only distinguish seen patterns in the trained images and may fail to adapt to diverse scenarios, such as real-time sample capturing with complex lighting, angles, backgrounds, etc.
- The use of deep learning methods results in a slower process during both training and testing phases, particularly when handling large datasets (Agarwal et al. 2020; Dong et al. 2023; Guo et al. 2023; Mazaheri and Roy-Chowdhury 2022; V. Sharma et al. 2023).
- Data privacy regulations have restricted access to private data, affecting the implementation and effectiveness of methods. Both Al-Ghadi et al. (2023) and Benalcazar et al. (2023) highlight the prolonged data collection process for real ID document samples due to regulatory constraints.

- Fast-evolving fraud tactics, such as the use of advanced technologies like deepfake, also challenge the ability of these methods to effectively handle new scenarios (Dong et al. 2023; Magee, Sheridan, and Thorpe 2023; Mazaheri and Roy-Chowdhury 2022; Trabelsi, Pic, and Dugelay 2022). For instance, while Benalcazar et al. (2023) demonstrates promising performance on previously unseen deepfake models, the high transferability of evolving adversarial examples across diverse model architectures and real-world conditions poses significant challenges.

AI-based IDF prevention and detection method: continuous authentication

This section focuses on continuous authentication, primarily implemented through user behavior anomaly detection. The analysis of these methods is conducted within the framework of UEBA, which serves as the foundational operational principle to guide their implementation. Additionally, the open challenges were identified through frequency analysis.

AI-based continuous authentication method: user behavior anomaly detection

User behavior anomaly detection utilizes AI techniques, such as machine learning and deep learning, to identify deviations from normal behavior in user profiles. In the reviewed articles, the analysis reveals the application of methods based on UEBA, a widely adopted approach in industrial practices (Sadowski et al. 2019). These methods construct user profiles by analyzing user behavior patterns and their interactions with entities such as devices and applications (Litan 2015). This profile serves as a baseline for the users and is compared with real-time user data to identify potential anomalies (Litan 2015; Sadowski et al. 2019). Figure 9 shows the three pillars that underpin the functionality of the UEBA tool.

The user behavior anomaly detection methods are analyzed using the dimensions of the UEBA framework. The analysis is organized into five key aspects: (a) core data for user profile creation, (b) profile modeling techniques, (c) analytics techniques, (d) use cases, and (e) open challenges.

Core data for user profile construction. User profiles provide an important baseline for analyzing user behavior patterns and detecting anomalies (Sadowski et al. 2019). Figures 8 and 10 illustrate the data and feature elements identified in the reviewed articles. Across the 19 studies, the attributes used for user profile construction fall into four categories: personal information (e.g., name, date of birth, Chang, Shih, and Lin 2024; S. Hu et al. 2021; Khan et al. 2023; Ye et al. 2022), behavioral patterns (including device interaction, Levi et al. 2022; Wang et al. 2021; H.

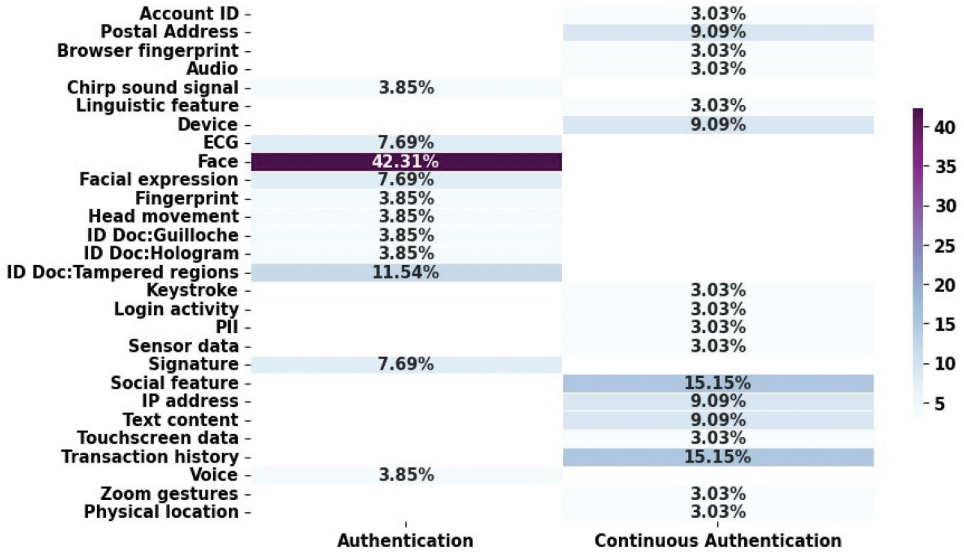


Figure 8. Data and feature elements across authentication and continuous authentication: face-related features dominate authentication (42%), while continuous authentication relies on diverse behavioral signals. *Note: because some articles used multiple data types, the results reflect an aggregated analysis of the overall data.*

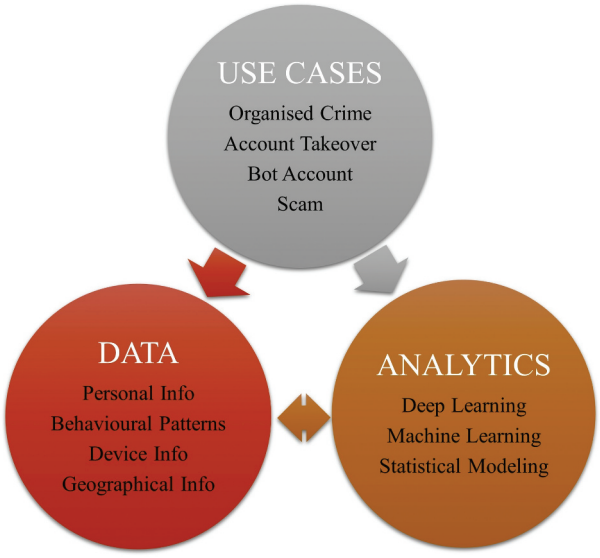


Figure 9. The three pillars of UEBA (adapted from Sadowski et al. 2019).

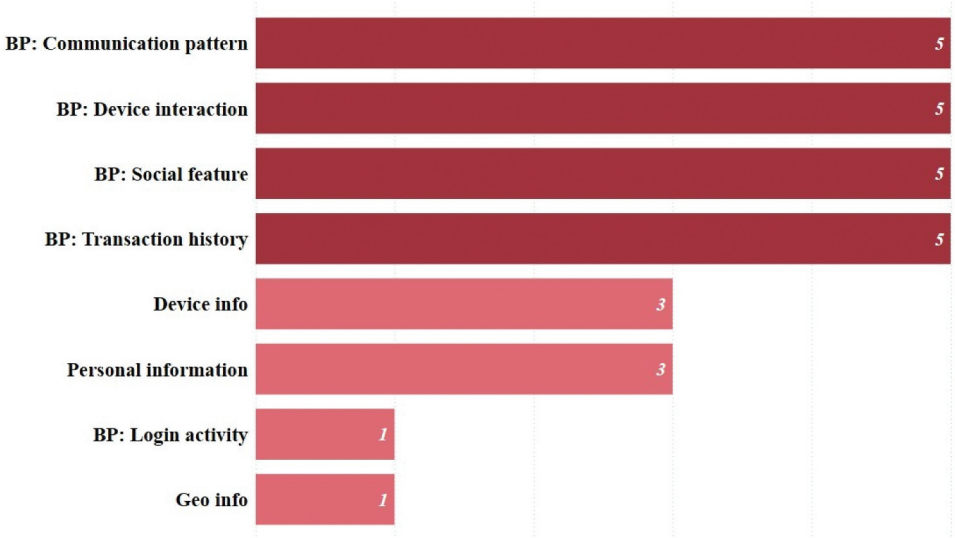


Figure 10. Distribution of attributes for building user profiles. Four behavioral attributes are equally prevalent, suggesting multisource data is central to effective profiling. Note: Because some articles used multiple data types, the results reflect an aggregated analysis of the overall data. (Abbreviation refers to Table G1).

Yang et al. 2023; Zhao and Wang 2021), social features (Aguilera et al. 2023; Gangan et al. 2023; Hanok and Shankaraiah 2023; Lv et al. 2021; Umbrani et al. 2024), transaction history (Dornadula and Geetha 2019; Kakkar et al. 2023; Lv et al. 2021; Priyaradhikadevi et al. 2023; Singhai, Aanjankumar, and Poonkuntran 2023), and communication patterns (Aguilera et al. 2023; Hanok and Shankaraiah 2023; Hong, Connie, and Ong Goh 2023; Safara et al. 2020; Umbrani et al. 2024), device information, such as IP and MAC addresses (Chang, Shih, and Lin 2024; S. Hu et al. 2021; Ye et al. 2022), and geographical information (S. Hu et al. 2021). As shown in Figure 10, communication patterns, device interaction, social features, and transaction history are equally prevalent, each utilized in five studies, suggesting that multisource behavioral data, rather than any single attribute, is central to effective profile construction.

User profile modeling techniques. Due to differences in modeling purposes, the modeling techniques of user profiles can be broadly categorized into two types: historical patterns modeling and relational patterns modeling. Their workflows for anomaly detection differ slightly.

- *Historical patterns modeling*: this is the most common approach, utilized in 15 articles (Aguilera et al. 2023; Dornadula and Geetha 2019; Gangan et al. 2023; Hanok and Shankaraiah 2023; Hong, Connie, and Ong Goh 2023; Kakkar et al. 2023; Levi et al. 2022; Lv et al. 2021; Priyaradhikadevi et al. 2023; Safara et al. 2020; Singhai, Aanjankumar, and Poonkuntran 2023; Umbrani et al. 2024; Wang et al. 2021; H. Yang et al. 2023; Zhao and Wang 2021). First, behavior data of users and entities is collected from various data sources (e.g., logs, devices, application activities, etc.) and preprocessed. Next, feature extraction is performed. Due to the diversity of behavioral data, this step is particularly crucial. Feature engineering, a method used to extract more representative features from the data, has become prevalent in user behavior anomaly detection (Aguilera et al. 2023; Gangan et al. 2023; Hanok and Shankaraiah 2023; S. Hu et al. 2021; Kakkar et al. 2023; Levi et al. 2022; Lv et al. 2021; Safara et al. 2020; Umbrani et al. 2024; Wang et al. 2021). It helps models focus on more relevant features and effectively represent user profiles. Subsequently, historical behavior patterns are modeled. The model learns the normal behavior of users as a baseline. When new user data is received, the model automatically identifies whether the new behavior is expected or anomalous.

- *Relational patterns modeling*: This method first analyzes labeled business data to identify fraudster behavior patterns and then constructs these relational patterns using graph techniques (Chang, Shih, and Lin 2024; S. Hu et al. 2021; Khan et al. 2023; Ye et al. 2022). It highlights the interrelations among users, which support the mapping of networks that underpin organized crime. Therefore, the model detects anomalies by analyzing whether new data points align with the characteristics of the identified anomaly group.

Analytics techniques. This review identified analytics techniques, including machine learning, deep learning, and statistical modeling, to identify and analyze abnormal behavior patterns (Sadowski et al. 2019). Figure 11 presents the distribution of analytics techniques in this review.

As shown in Figure 11, three analytics approaches are employed. *Deep learning* is the most prevalent (10 studies), with supervised methods dominating (Gangan et al. 2023; Hong, Connie, and Ong Goh 2023; Kakkar et al. 2023; Levi et al. 2022; Safara et al. 2020; H. Yang et al. 2023; Ye et al. 2022), supplemented by unsupervised (Chang, Shih, and Lin 2024; S. Hu et al. 2021), and semisupervised (Khan et al. 2023) approaches. Deep learning's key advantage lies in processing unstructured data, such as graphs, text, and audio (Hong, Connie, and Ong Goh 2023; Safara et al. 2020; H. Yang et al. 2023; Ye et al. 2022) and in reducing reliance on manual labeling through unsupervised techniques (Chang, Shih, and Lin 2024; Khan et al. 2023). *Machine learning* appears in nine studies (Aguilera et al. 2023; Dornadula and Geetha 2019; Gangan et al. 2023; Levi et al. 2022; Lv et al. 2021;

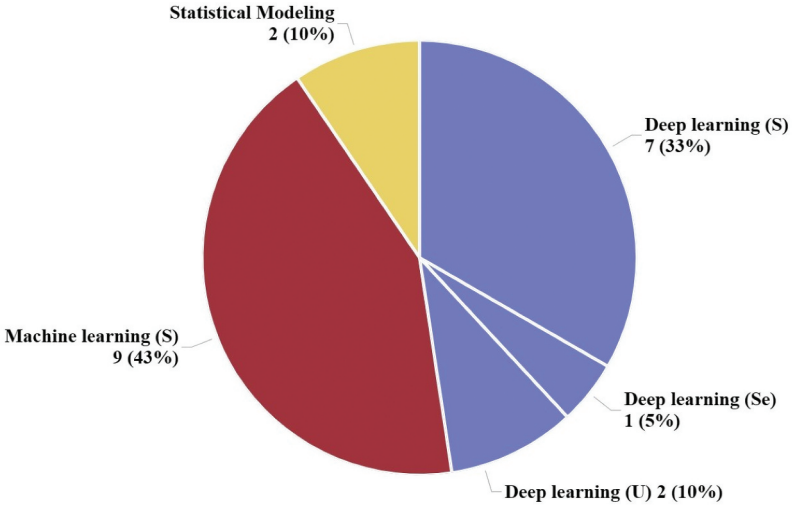


Figure 11. Analytics techniques for user behavior anomaly detection: machine learning (43%) and deep learning (33%) dominate, together comprising three-quarters of approaches.

Priyaradhikadevi et al. 2023; Singhai, Aanjankumar, and Poonkuntran 2023; Umbrani et al. 2024; Wang et al. 2021), exclusively using supervised classification: SVM (Aguilera et al. 2023; Dornadula and Geetha 2019; Priyaradhikadevi et al. 2023; Umbrani et al. 2024; Wang et al. 2021) and Random Forest (Aguilera et al. 2023; Dornadula and Geetha 2019; Umbrani et al. 2024; Wang et al. 2021) are the most frequently applied algorithms, demonstrating strong performance in establishing behavioral baselines (Appendix C). *Statistical modeling* is employed in two studies using *F*-tests (Hanok and Shankaraiah 2023) and Gaussian Process Models (Zhao and Wang 2021), proving effective with limited samples. Notably, the choice of technique appears driven by data structure: graph-based and textual data favor deep learning, tabular behavioral logs suit machine learning classifiers, and sparse feature spaces benefit from statistical approaches.

Use cases. A use case represents a specific scenario or problem that a method is designed to address, such as detecting fraud. User behavior anomaly detection is applied to defend against various IDF cases. Among the reviewed articles, account takeovers are the most frequently addressed use case, with 11 articles on this topic (Dornadula and Geetha 2019; Hanok and Shankaraiah 2023; Kakkar et al. 2023; Levi et al. 2022; Lv et al. 2021; Priyaradhikadevi et al. 2023; Safara et al. 2020; Singhai, Aanjankumar, and Poonkuntran 2023; Wang et al. 2021; H. Yang et al. 2023; Zhao and Wang 2021), followed by organized crime (Chang, Shih, and Lin 2024; S. Hu et al. 2021; Khan et al. 2023; Ye et al. 2022) and bot accounts (Aguilera et al. 2023; Gangan et al. 2023; Umbrani et al. 2024), while scam detection is the least discussed (Hong, Connie, and Ong Goh 2023).

Open challenges. The open challenges in user behavior anomaly detection reveal a complex interplay of technical and nontechnical factors, with nontechnical issues, such as changing user behavior and data privacy regulations, being the most frequently highlighted. Key technical challenges include the limited diversity of datasets and the lack of transferability of techniques to real-world use cases. The following section elaborates on these challenges derived from Appendix D Table D3:

- Dynamic user behavior, shaped by factors, such as changing life events, priorities, and preferences, introduces significant variability that impacts the detection accuracy of methods (Hanok and Shankaraiah 2023; Levi et al. 2022; Wang et al. 2021; H. Yang et al. 2023; Zhao and Wang 2021). Consequently, anomalies in data do not necessarily indicate actual abnormal behavior. As noted in Chang, Shih, and Lin (2024), detected anomalies may include cases involving normal users. To mitigate this issue, incorporating expert opinions and business-specific rules has been suggested in Chang, Shih, and Lin (2024) and Singhai, Aanjankumar, and Poonkuntran (2023) to refine models and reduce false positives.
- Threshold misconfiguration is a frequently cited challenge affecting accuracy. While the rules outlined in Chang, Shih, and Lin (2024) and Singhai, Aanjankumar, and Poonkuntran (2023) effectively distinguish normal users from anomalies, improper threshold settings can still lead to false negatives or false positives. Achieving an optimal balance in rule configuration remains a critical challenge, as it directly impacts detection performance and system reliability.
- Restricted access to privacy data and user consent emerge as two key highlights identified in the review. Studies such as Chang, Shih, and Lin (2024), Singhai, Aanjankumar, and Poonkuntran (2023), and H. Yang et al. (2023) have raised concerns about accessing private user data for training purposes. Additionally, Kakkar et al. (2023) and Lv et al. (2021) discuss the impact of user consent policies on model performance and validation, particularly when users withdraw their data after training, posing challenges to both compliance and model reliability.
- Inconsistent performance across diverse datasets is a frequently observed issue, particularly in cases involving imbalanced datasets (Hong, Connie, and Ong Goh 2023; Wang et al. 2021). This challenge is further compounded by the scarcity of labeled true fraud cases in user behavior anomaly detection, which limits the model's ability to learn and generalize suspicious features due to insufficient data representation (Hong, Connie, and Ong Goh 2023). Additionally, studies, such as Safara et al. (2020), Ye et al. (2022), and Hanok and Shankaraiah (2023) identify challenges associated with real-time datasets and datasets tailored to varying use cases. For instance, while some methods effectively detect shared device relationships, they often lack sensitivity to subtle anomalies in individual behavioral characteristics. These findings highlight the difficulties in achieving consistent and reliable performance across diverse and dynamic datasets.

- Technique complexity is a significant factor impacting method effectiveness, particularly for those requiring intricate data processing and modeling (Dornadula and Geetha 2019; Priyaradhikadevi et al. 2023; Safara et al. 2020; Zhao and Wang 2021). For instance, Zhao and Wang (2021) describe a method that processes multiple sources of sensor data through unification, sequence conversion, and subsequent model training and classification. This multistep pipeline is resource-intensive and poorly suited for large-scale datasets, creating significant barriers to real-time applications.

Appendix G Table G2 summarizes the articles on user-behavior anomaly detection. Table 10 summarizes the three methods at each phase of the analysis, detailing the data type, application, and applied context. This table also serves as a comparative summary of the strengths and limitations of the reviewed

Table 10. Summary of IDF detection and prevention methods.

Method	Role	Data Type	Applied Techniques	Application Scenario	Strengths	Limitations
Biometric recognition	Verify identity	Biometric data	Feature extraction, matching	Identity verification	• Unique biological traits confirm identity. • Technological advancements ease authentication. • Diversity of biometric traits allows further research.	• Accuracy influenced by external factors. • Requires high-quality devices. • Complexity affects effectiveness. • System vulnerabilities may allow spoofing.
Visual anomaly detection	Detect tampered regions	Image, video	Feature detection deep learning	Biometric spoofing, Deepfake, ID document forgery	• Feature-based and deep-learning methods are easy to implement. • Achieves high performance.	• Performance depends on data quality and preprocessing. • Lacks generalization across datasets. • Data sampling restricted by privacy regulations. • Long processing time for deep learning methods.
User behavior anomaly detection	Detect anomaly behaviors	Behavioral patterns, Geographical Info, Device Info, Personal Info	**Profile modeling**: historical pattern, relational pattern **Analytics techniques**: deep learning, machine learning, statistical modeling	Account takeover, bot account, organized crime, scam	• Works for both supervised and unsupervised learning. • Detects imperceptible events. • Applicable for behavior and transaction monitoring.	• Variability in user behavior leads to false positives. • Hard to balance thresholds. • Insufficient labeled data. • Poorgeneralizationto real-world scenarios. • Resource-intensive, limiting real-time use.

approaches. Appendix C Table C1 summarizes AI's application and presents the performance metrics of the best models identified across all methods.

Discussion

This section addresses the research questions by analyzing the findings and their broader implications. It explores how the results contribute to understanding the topic, suggests directions for future research, and identifies the limitations of this review.

Before proceeding, readers should note that the confidence of the synthesized findings varies across subtopics. As detailed in the quality assessment (Research methodology), most included studies met the majority of quality criteria; however, gaps in reported evaluation results and future directions among lower-scoring studies may limit the depth of evidence in certain areas.

Research questions, key findings, and open challenges

This section addresses the research questions by summarizing the key findings derived from the analysis. Each research question is discussed individually to ensure clarity and relevance.

RQ1: What are the current AI-based IDF detection methods?

Given the impact of authentication and continuous authentication on combating IDF, this SLR carefully selected and reviewed 43 articles that applied AI-based methods to detect and prevent IDF in the two authentication processes. There is a noticeable increase in the number of articles that use AI-based methods to address IDF for the two authentication processes, as shown in Figure 12.

In biometric recognition, facial recognition is the most frequently studied modality in this review. As shown in Figure 13, the number of studies addressing facial recognition has increased year on year, with biometric spoofing and deepfake attacks being the two most commonly targeted threats. The increasing deployment of facial recognition in secure access systems and mobile authentication is driven by its high accuracy and ease of use (Claveria et al. 2021; Uliyan, Sadeghi, and Jalab 2020). However, the emergence of GAI technologies has simultaneously accelerated the sophistication of attacks against these systems, with several studies reporting vulnerabilities to morphing, presentation, and deepfake attacks (Agarwal et al. 2020; Cao et al. 2023; Jayanshre and Amsaprabhaa 2024). This convergence of growing adoption and escalating threats

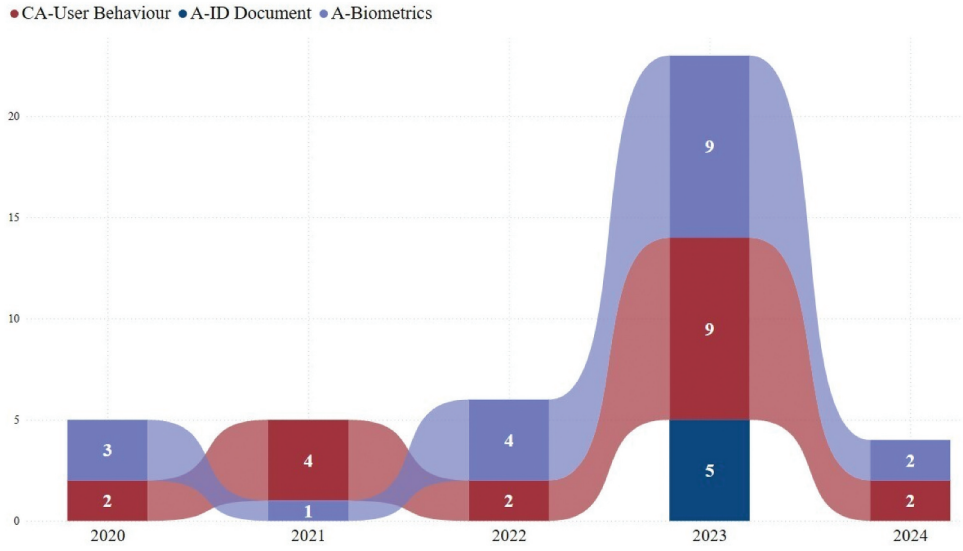


Figure 12. Trends in AI-based methods for authentication and continuous authentication. Note: Abbreviations are defined as follows. (1) A: Authentication. (2) CA: Continuous authentication. (3) A-Biometrics: Biometrics authentication. (4) A-ID Document: ID document authentication.

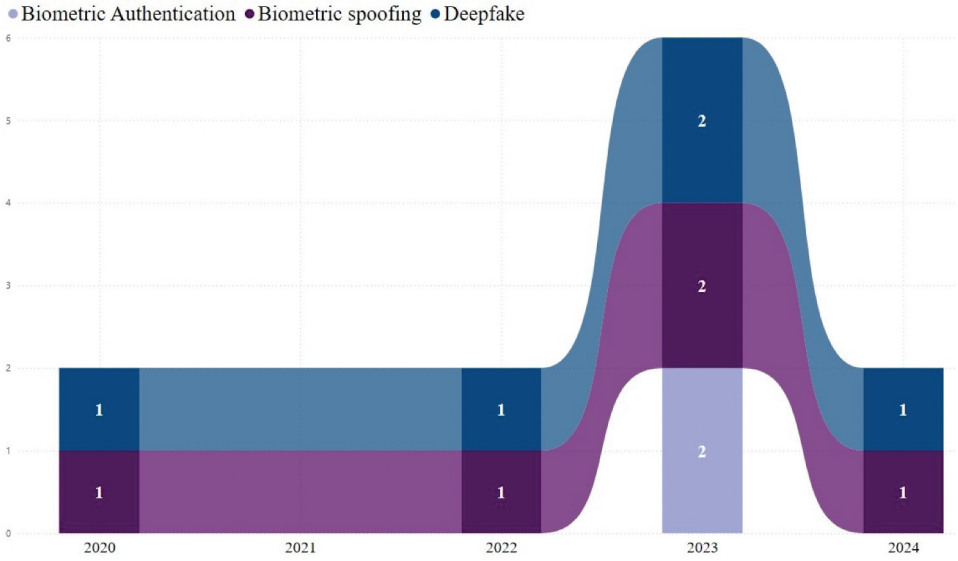


Figure 13. Trends in the number of articles on face recognition over the years.

suggests that facial recognition has become both the primary line of defense and the primary target in IDF, underscoring the need for more resilient detection mechanisms.

Continuous authentication has become increasingly important after initial authentication to combat IDF on a broader scale (H. Yang et al. 2023) (see Figure 12). As shown in Figure 10, communication patterns are now utilized as frequently as device interactions, social features, and transaction history for user profile construction (each appearing in five studies). Among the reviewed studies, H. Yang et al. (2023) demonstrated that unique writing styles, shaped by personal background and education, offer distinguishing characteristics for continuous identity verification. This growing prominence of communication patterns is notable in the context of advancing GAI capabilities. As language models improve in generating and analyzing semantic content, communication-based profiling may become both a more powerful detection tool and a more vulnerable attack. It is a duality that warrants further investigation.

Deep learning is the most widely adopted technique across all three detection methods in this review. In visual anomaly detection, CNNs and their variants (e.g., Inception, XceptionNet) are predominantly used for detecting subtle visual irregularities (Cao et al. 2023; Izdihar et al. 2023; Jayanshre and Amsaprabhaa 2024; Mazaheri and Roy-Chowdhury 2022; V. Sharma et al. 2023; Vinod and Senthil Kumar 2023). In user behavior anomaly detection, GNNs are employed to model relational profiles through graph-based representations (Chang, Shih, and Lin 2024; S. Hu et al. 2021; Khan et al. 2023; Ye et al. 2022). A key advantage reported across these studies is the ability to process diverse data modalities, images, graphs, text, and audio, with reduced reliance on manual feature engineering (Chang, Shih, and Lin 2024; Khan et al. 2023). However, this wide applicability comes with notable limitations: deep learning models depend on large scale and diverse annotated datasets, which are scarce in domains, such as user behavior analysis (Al-Ghadi et al. 2023; Benalcazar et al. 2023), and their high computational requirements often hinder deployment in resource-constrained environments, such as real-time mobile applications (Dong et al. 2023). These findings indicate that while deep learning offers the broadest capability across IDF detection tasks, its practical deployment remains constrained by data availability and computational cost.

RQ2: What are the key open challenges in addressing identity fraud?

The results of this research indicate that certain challenges are repeatedly mentioned in the literature. The findings emphasize the need to address these recurring challenges to advance research and improve practical implementations.

Challenge 1: technical limitations. The results of this SLR indicate that certain technical challenges are repeatedly mentioned in the literature. Our statistical analysis in Table 7 confirms that the four most frequently cited technical challenges are as below: first, poor robustness, largely due to limited data diversity; second, poor resilience, reflecting limited transferability to new scenarios; third, reduced accuracy impacted by poor data quality and noisy samples; and forth, limited effectiveness driven by lengthy training times and computational cost.

While these challenges span different dimensions of model performance, the first three share a common origin: the data bottleneck, specifically low data quality and low data diversity. Its effects are visible in researchers' data choices: as Figure 14 shows, self-collected ($n = 12$) and synthesized ($n = 4$) data are often used to work around the limitations of public benchmarks. Complementing this, our analysis identifies seven studies explicitly combined public datasets with custom data to mitigate these deficiencies. Details of the public/nonpublic dataset classifications and associated studies are provided in Appendix E Tables E2 and E4.

Table 11 provides quantitative evidence of these pitfalls. Most public datasets systematically lack diversity either in scale, scenario coverage, or real-world fidelity.

- For real-world public datasets, the most frequent pitfalls are limited scenarios ($n = 8$), small-scale samples ($n = 5$), outdated data ($n = 5$), or lab-controlled capture ($n = 3$).
- For synthetic datasets, the dominant pitfalls are limited scenario types ($n = 9$) and specific generator artifacts ($n = 5$).

These pitfalls directly explain the two top challenges and the researcher practices observed in this review:

- **Low quality → Poor robustness**: The pitfalls of artifacts bias, outdated and small-scale data explain why robustness is poor. This has led researchers to construct custom or synthetic datasets (e.g.,

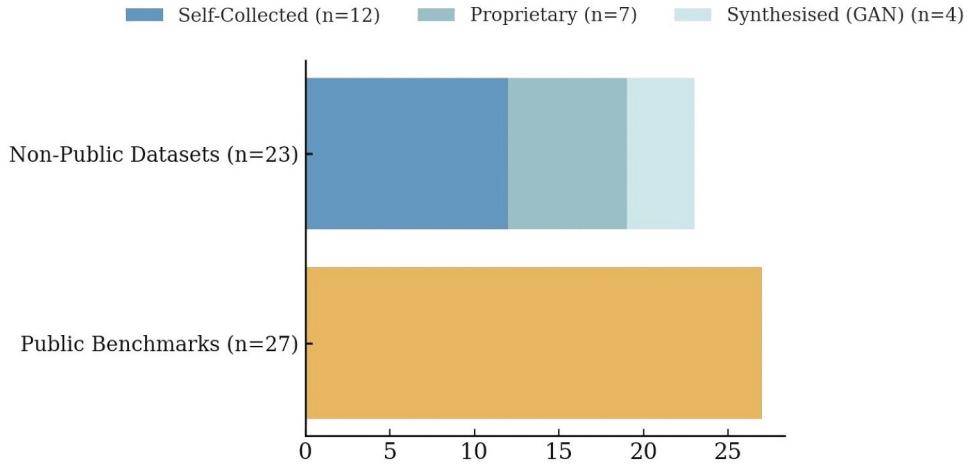


Figure 14. Distribution of data source strategies in 43 reviewed papers.

Table 11. Summary of dataset pitfalls in 45 public datasets.

Dataset Type	Common Pitfalls	Freq.	Example Dataset(s)
Real world ($n = 27$)	Limited scenarios	8	URME, CelebA, B-ID Dataset, ChiSig
Real world ($n = 27$)	Small scale	5	AT&T face dataset, Tufts Face, ECG ID
Real world ($n = 27$)	Outdated data	5	Enron E-mail (pre-2002), URME (2007–2009)
Real world ($n = 27$)	Controlled lab environment	3	University at Buffalo (UB) Datasets, Tufts Face, FRGC
Synthetic ($n = 18$)	Limited scenarios type	9	DFDC, FMIDV, CASIA FASD, MFSD
Synthetic ($n = 18$)	Specific generator bias	5	DFDC, DigiFace-1 M, Face2Face, FaceForensics++

Agarwal et al. 2020; Guo et al. 2023; Jaiswal and Srivastava 2023) to compensate for the low quality of public benchmarks.

- **Low diversity → Poor resilience**: The pitfalls of limited scenarios and task then explain why resilience (i.e. transferability) is poor. Consequently, researchers increasingly conduct cross-dataset evaluation (e.g., Jaiswal and Srivastava 2023; Mazaheri and Roy-Chowdhury 2022; Pujol et al. 2019; V. Sharma et al. 2023) to demonstrate that their models generalize beyond a single, flawed benchmark.

Challenge 2: contextual variability. Although significant research has been conducted on IDF, the understanding of its complex dynamics and mechanisms remains limited. One of the primary challenges is the continuous evolution of fraud tactics driven by advancing fraudulent technologies and emerging real-world threats, such as the rise of GAI. Business data from companies can assist in identifying suspicious patterns and labeling data for machine learning; however, this often occurs after fraud has already taken place. When new patterns emerge, models frequently struggle to make accurate predictions or adapt to novel situations.

Dynamic user behaviors further complicate the issue. As user behavior constantly evolves, baseline patterns established by these methods may no longer hold true. Additionally, anomalies identified by detection systems do not always indicate actual fraud. Further analysis is often required to distinguish true fraud from benign outliers (Chang, Shih, and Lin 2024; Singhai, Aanjankumar, and Poonkuntran 2023).

Challenge 3: data privacy and security issues. Data privacy and security remain critical in IDF detection. A core challenge is the conflict between regulation and performance: while AI models require rich, contextual data for high accuracy, data privacy regulation has restricted the use of privacy data (GDPR). Although data encryption has long been used, encrypted features often compromise model learnability (Lv et al. 2021; Ye et al. 2022), especially where contextual data are used by an increasing number of researchers. Centralized storage, a common practice among many organizations, increases the risk of identity theft, when AI models rely on such data for training (Lv et al. 2021).

Privacy-preserving techniques provide partial mitigation. For instance, Federated Learning (FL) directly addresses the risk of centralized storage (Levi et al. 2022; Lv et al. 2021) by allowing models for decentralized training without exposing PII. Hash functions, meanwhile, help meet regulatory requirements by masking identifiers during training, so that individual identities are not directly exposed (Lv et al. 2021; Ye et al. 2022). Synthetic identity data can reduce direct reliance on PII, but it may fail to capture the full spectrum of real-world anomalies (Aguilera et al. 2023; Benalcazar et al. 2023; Dong et al. 2023; Guo et al. 2023).

Beyond these technical mitigation, this review argues that governance frameworks are equally important. Specifically, auditability mechanisms are needed to support regulatory supervision and transparent decision-making, and fairness assurance is particularly essential for biometric systems, where models must perform equitably across demographic groups to avoid discriminatory outcomes.

Research implications and future direction

The findings and challenges outlined in the previous sections serve as a foundation for identifying potential research directions. By addressing these challenges, IDF detection can achieve significant advancements in the following areas.

Insufficient quality data samples and lack of data diversity for model training

The scarcity and homogeneity of data samples, driven by privacy regulations and limited access to real-world scenarios, significantly hinder effective model training. Although synthetic data generated using advanced techniques can increase sample sizes, it often falls short in representing diverse and unforeseen conditions.

This gap highlights the need for more innovative methods to improve both the quality and diversity of data. Research should focus on designing and validating synthetic data approaches that better mimic realistic variations. Furthermore, future studies should explore Self-Supervised Learning (SSL) as a key strategy to address labeled data scarcity. By generating auxiliary supervision signals from massive unlabeled datasets, SSL can significantly enhance feature representation and diversity, thereby improving model robustness and adaptability to the dynamic and evolving nature of IDF.

Limited understanding of IDF dynamics

Despite progress in understanding IDF, current methods rely heavily on post-incident data, limiting their ability to adapt to evolving fraud patterns. The scarcity of labeled fraud cases further compounds this issue, as supervised models cannot learn patterns absent from their training sets.

Future research should therefore advance anomaly detection techniques that do not depend on pre-labeled data, and prioritize adversarial robustness through techniques such as adversarial training to ensure models remain reliable against attacks designed to bypass authentication systems.

Data privacy and security challenges

Ensuring data privacy and security remains a critical issue. Future discussions should address two key questions:

- How can privacy-preserving methods be implemented during data processing without compromising data utility for analytical purposes?
- How can data lifecycle management be optimized to address IDF at a cross-industry scale while balancing privacy, security, and data utility?

Research should emphasize the development of frameworks and tools that uphold data security while supporting effective detection. This involves a shift toward privacy-preserving identity verification (PIV) such as leveraging advanced cryptographic techniques (e.g., Zero-Knowledge Proofs and Homomorphic Encryption) to verify identity claims without exposing the underlying raw sensitive data. Looking ahead, FL offers a promising path for collective defense. It allows multiple institutions to collaboratively train fraud detection models on distributed datasets without exchanging sensitive information, which significantly enhances the effectiveness of IDF detection by leveraging diverse fraud patterns while strictly adhering to privacy regulations.

Application of multimodal data in IDF detection

As AI's ability to understand semantics improves, multimodal data have been increasingly utilized in IDF detection. Methods employing image data, behavioral data, and time-series data are already well-established; however, context-based detection techniques remain less commonly implemented. Future research should focus on developing effective methods to integrate context-based data, exploring strategies to preserve critical information for model learning, such as temporal or spatial relationships. To address the privacy concerns inherent in multimodal fusion, future research should integrate the PIV mentioned in Data privacy and security challenges, ensuring that high-dimensional contextual data can be utilized without compromising user confidentiality.

Explainability of deep learning models in IDF detection

As established in our findings, deep learning is the most widely adopted technique in the reviewed literature. While these methods deliver strong detection performance, they remain limited by their opacity that their internal decision logic is difficult to interpret, which constrains their use in production settings. This is especially consequential in regulated domains such as finance and healthcare, where automated decisions increasingly need to be explainable to regulators and affected individuals. Article 13 of the EU AI Act (European Parliament and Council of the European Union 2024), for instance, requires high-risk AI systems to be "sufficiently transparent" for deployers to interpret their outputs. Despite this, explainability receives little attention in the studies reviewed, which largely prioritize predictive accuracy over explainability.

Future research should therefore treat explainability as a design objective rather than an afterthought, enabling practitioners to justify individual fraud determinations rather than relying on opaque outputs. In particular, evaluating which explainability methods are effective in IDF detection settings, and designing explanations understandable to nontechnical stakeholders, would strengthen both the deployability and regulatory compliance of these systems.

Summary of findings and emerging frontiers

Building on the analysis presented above, this section summarizes the key findings from the reviewed literature (Summary of findings), then extends the discussion to encompass developments through 2024 and 2025 (Emerging frontiers), offering a more forward-looking perspective on the field's trajectory.

Summary of findings. The findings of this SLR underscore the importance of adopting advanced AI-driven methods, such as anomaly detection and behavior modeling, to enhance fraud detection and prevention systems. Adaptive, scalable, and efficient models capable of handling dynamic user behaviors and evolving fraud tactics are essential for both industrial applications and academic advancements.

Additionally, there is a pressing need to explore unsupervised learning and robust anomaly detection techniques that do not rely on prelabeled data, alongside interdisciplinary approaches to address challenges in data quality and real-world applicability.

A further implication concerns explainability. Deep learning dominates the reviewed literature, yet its opacity remains a persistent constraint in production settings. Across the studies reviewed, explainability is rarely treated as a design consideration, pointing to a gap between current research emphasis and the demands of regulated deployment.

Emerging frontiers. Looking beyond the review period (2020–2024), we now turn to forward-looking observations that fall outside the systematic findings but signal the field's near-term trajectory.

Extending our analysis to developments through 2024–2026, the weaponization of GAI and deepfake technologies stands out as the most disruptive emerging frontier. Recent studies report a shift toward foundation and diffusion models capable of hyperrealistic, cross-modal manipulation, such as synchronizing fabricated audio with facial expressions (S. Sharma and Selwal 2026). The resulting prospect of real-time, zero-shot identity cloning challenges the assumptions underlying static biometric systems. In response, defense paradigms are moving from pixel-level detection toward logical domain checking and cross-modal synergy. Rather than searching for artifacts in the generated output, recent work prioritizes intrinsic authenticity by integrating physical and cognitive principles, such as anatomical constraints, lighting consistency, and physiological signals (Xie et al. 2026), to detect violations of natural laws that current generative models fail to simulate perfectly.

Alongside these algorithmic defenses, prevention at the system-architectural level is equally important. Identity proofing standards such as NIST SP 800-63A-4 (Temoshok et al. 2025) require strict binding between a claimed identity and the person presenting it in high-risk scenarios. FIDO2 passkeys (FIDO Alliance 2024) implement this cryptographically, tying the physical human, digital identity, and device together at the point of authentication, enforcing absolute uniqueness. However, achieving real-world scalability of this technology requires resolving adoption and cross-device synchronization barriers inherent in hardware-backed zero-trust systems (Würsching et al. 2023). Future interdisciplinary research bridging security engineering, usability, and policy design will be essential to translate these architectural principles into deployable solutions.

Recommendations for practitioner

Based on the systematic findings of this review, it is evident that transitioning from laboratory AI models to industrial-grade fraud engines requires addressing beyond mere static accuracy. To assist practitioners in navigating real-world deployment toward robust industrial IDF detection and prevention, the following checklist highlights key considerations across data privacy, model robustness, and model drift and adversarial:

- **Data privacy constraints**: As stringent regulations significantly limit the volume and diversity of identity-related data (e.g., PII), practitioners must first evaluate how data privacy constraints impact data acquisition, feature engineering, and overall modeling capacity. To bridge this gap, the adoption of PIV and FL is recommended, as they enable model training without requiring access to raw sensitive data. Additionally, practitioners should explore alternative modeling approaches, such as

spatiotemporal relationship modeling, to preserve critical fraud patterns without relying on raw personal identifiers, thereby balancing detection efficacy with data minimization principles.

- **Model robustness**: While practitioners possess vast operational datasets, the extreme scarcity of labeled fraud samples remains a primary bottleneck. Therefore, we recommend using synthetic data generation to augment training sets with diverse fraud patterns, alongside SSL to improve feature extraction from unlabeled data. Additionally, practitioners should evaluate model performance under varied in-the-wild conditions, such as low resolution or poor lighting, to ensure stable generalization beyond curated laboratory datasets.
- **Model drift and adversarial**: Given the dynamic nature of IDF, practitioners should incorporate proactive adversarial training during the development phase to anticipate emerging threats like deepfakes. Furthermore, establishing a real-time feedback loop via advanced anomaly detection is essential to identify model drift. Such monitoring ensures the system can detect when incoming fraud patterns deviate from historical training data, allowing for timely incremental learning or manual intervention to maintain long-term efficacy in a rapidly evolving threat landscape.

Limitations

Like any other SLR study, this research has limitations that should be considered when interpreting the findings. First, it only relies on four databases, which are limited in number; however, they provide sufficient coverage of the existing literature. It is worth mentioning that the review was conducted within a specific time frame (2020–2024) and was limited to English-only publications. This may overlook recent advancements or relevant work published in other languages. Additionally, our focus on peer-reviewed literature means we acknowledge potential omissions from preprints and industry reports.

However, this SLR study provides a solid foundation for future studies by other researchers. The screening and selection process is documented in the PRISMA flow diagram (Appendix A). Title/abstract screening was conducted in Rayyan. Due to account expiration, the tags used to categorize records were not retrievable, though aggregate screening counts and all subsequent full-text and quality-assessment records (conducted in EndNote) remain fully intact. A reconciliation of screening counts across all stages, supported by Rayyan and EndNote screenshots, is provided in Appendix A (Figures A2, A3, A4) and the supplementary dataset archived on Zenodo (DOI: [https://doi.org/10.5281/zenodo.21482806]).

Second, due to the scope of this research, a specific set of search strings was employed, which may have influenced the breadth of the results. While efforts were made to balance specificity and comprehensiveness, it is possible that some relevant studies were omitted due to variations in terminology or indexing practices.

Third, the review focuses on the domain of IDF detection and may have overlooked insights from adjacent fields such as cybersecurity which could provide valuable complementary perspectives. However, it was important to keep the focus on the IDF detection rather a broad SLR across cybersecurity. Here, it is important to mention that the research team met regularly to review the results and discuss any individual research or researcher bias with a view to maintain the study focus and improve the quality of results.

Fourth, the frequency tallies and method categorizations presented in our synthesis are interpretive in nature and reflect analytical judgment rather than definitive classifications. Studies were categorized based on their primary method; however, many employ multiple approaches, and a different classification scheme could alter the reported distributions. Readers are encouraged to consult the publicly available extraction dataset to verify or recategorize as needed.

Finally, the rapid pace of advancements in AI and related technologies poses a challenge in maintaining the relevance of the findings. While the review provides a comprehensive snapshot of the field, some trends or methods discussed may already be evolving. We therefore propose a future review roadmap, for instance in 2025–2026, to specifically analyze the surge in GAI and its impact on both creating and detecting IDF, which requires future studies to continuously update the state of knowledge.

Conclusion

IDF can have significant impacts on individuals and organizations, making it essential to understand and investigate methods for its early detection. AI has demonstrated its potential in various tasks, including computer vision, natural language processing, and graph processing, and our systematic literature review highlights its effectiveness in IDF detection. From our findings, anomaly detection, either on image or behavior data, is the main method in IDF detection. Deep learning has higher capabilities in handling complex scenarios that require the input of various data forms, from images to structured data, while machine learning has higher computation efficiency. Privacy data concerns with the use of AI also require attention to be extracted while implementing IDF detection.

Overall, our systematic review highlights the transformative role of AI in combating IDF and identifies key areas for future research. We also provide a detailed list of current gaps and unresolved issues to guide future research in this area. Specifically, the next steps include:

- Addressing data quality and diversity issues through synthetic data and SSL.
- Advancing unsupervised learning techniques and adversarial-robust models for evolving fraud patterns.
- Advancing PIV and FL to balance privacy with utility and enable secure cross-industry collaboration.
- Enhancing detection through multimodal data integration.
- Optimizing resource efficiency for scalable systems.

Author contributions

CRediT: **ChuoJun Zhang:** Conceptualization, Data curation, Formal analysis, Investigation, Project administration, Visualization, Writing – original draft; **Asif Q. Gill:** Conceptualization, Supervision, Writing – review & editing; **Bo Liu:** Supervision, Writing – review & editing; **Memoona Anwar:** Supervision.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author acknowledges support from an industry-linked scholarship administered by the University of Technology Sydney, which provided stipend support during this study.

ORCID

ChuoJun Zhang http://orcid.org/0009-0000-0336-9573
Asif Q. Gill http://orcid.org/0000-0001-6239-6280
Bo Liu http://orcid.org/0000-0002-3603-6617
Memoona Anwar http://orcid.org/0000-0001-5438-8811

Data availability statement

The dataset supporting this review, including screening records, bibliographic information, extraction coding, and quality assessment scores, is publicly available on Zenodo at https://doi.org/10.5281/zenodo.21482806 (Zhang et al. 2026)

Ethical approval

Ethical review and approval were not required for this study, as it is a systematic literature review based on publicly available data and does not involve direct human participants or identifiable personal information.

References

ACT Legislation Register. 2002. "Criminal Code 2002 | Acts." https://www.legislation.act.gov.au/a/2002-51/.

Agarwal, S., T. El-Gaaly, H. Farid, and S. N. Lim 2020. "Detecting Deep-Fake Videos from Appearance and Behavior." In *2020 IEEE International Workshop on Information Forensics and Security (WIFS)*, 1–6. New York, NY, USA. https://doi.org/10.1109/WIFS49906.2020.9360904.

Aguilera, A., P. Quinteros, I. Dongo, and Y. Cardinale. 2023. "CrediBot: Applying Bot Detection for Credibility Analysis on Twitter." *IEEE Access* 11:108365–108385. https://doi.org/10.1109/ACCESS.2023.3320687.

Ahmed, S. R. 2020. *Preventing Identity Crime: Identity Theft and Identity Fraud: An Identity Crime Model and Legislative Analysis with Recommendations for Preventing Identity Crime*. Leiden, The Netherlands: BRILL. https://brill.com/display/title/54591language=en&srsltid=AfmBOopONfpct9vip8jukrr340VeCMFXdKAyW0tyIk5xuh1jRfA1VbQk.

Al-Ghadi, M., Z. Ming, P. Gomez-Krämer, J. C. Burie, M. Coustaty, and N. Sidere. 2023. "Guilloche Detection for ID Authentication: A Dataset and Baselines." In *2023 IEEE 25th International Workshop on Multimedia Signal Processing (MMSP)*, 1–6. France: Poitiers. https://doi.org/10.1109/MMSP59012.2023.10337681.

Ali, A., S. Abd Razak, S. H. Othman, Taiseer Abdalla Elfadil. Eisa, A. Al-Dhaqm, M. Nasser, T. Elhassan, H. Elshafie, and A. Saif. 2022. "Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review." *Applied Sciences* 12 (19): 9637. https://doi.org/10.3390/app12199637.

Avcibas, I. 2024. "Morphed Face Detection with Wavelet-Based Co-Occurrence Matrices." *IEEE Signal Processing Letters* 31:1344–1348. https://doi.org/10.1109/LSP.2024.3397726.

Ayeswarya, S., and J. Norman. 2018. "A Survey on Different Continuous Authentication Systems." *International Journal of Biometrics* 11 (1): 67. https://doi.org/10.1504/ijbm.2019.096574.

Baig, A. F., and S. Eskeland. 2021. "Security, Privacy, and Usability in Continuous Authentication: A Survey." *Sensors* 21 (17): 5967. https://doi.org/10.3390/s21175967.

Benalcazar, D. E., J. E. Tapia, S. A. Gonzalez, and C. Busch. 2023. "Synthetic ID Card Image Generation for Improving Presentation Attack Detection." *IEEE Transactions on Information Forensics and Security* 18:1814–1824. https://doi.org/10.1109/TIFS.2023.3255585.

Blauth, T. F., O. J. Gstrein, and A. Zwitter. 2022. "Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI." *IEEE Access* 10:77110–77122. https://doi.org/10.1109/access.2022.3191790.

Button, M., and C. Cross. 2017. "Technology and Fraud: The 'Fraudogenic' Consequences of the Internet Revolution." In *The Routledge Handbook of Technology, Crime and Justice*, edited by M. R. McGuire and T. J. Holt, 78–95. London, UK: Routledge. https://doi.org/10.4324/9781315743981-5.

Button, M., C. Lewis, and J. Tapley. 2014. "Not a Victimless Crime: The Impact of Fraud on Individual Victims and Their Families." *Security Journal* 27 (1): 36–54. https://doi.org/10.1057/sj.2012.11.

Cao, Y., J. Chen, L. Huang, T. Huang, and F. Ye. 2023. "Three-Classification Face Manipulation Detection Using Attention-Based Feature Decomposition." *Computers and Security* 125:103024. https://doi.org/10.1016/j.cose.2022.103024.

Cavaliere, L. P. L., S. Subhash, P. V. D. Rao, P. Mittal, K. Koti, M. K. Chakravarthi, R. Duraipandian, et al. 2021. "The Impact of Internet Fraud on Financial Performance of Banks." *Turkish Online Journal of Qualitative Inquiry* 12 (6): 8489–8503. https://tojqi.net/index.php/journal/article/view/3260/2209.

Chang, Y. W., H. Y. Shih, and T. N. Lin. 2024. "AI-URG: Account Identity-Based Uncertain Graph Framework for Fraud Detection." *IEEE Transactions on Computational Social Systems* 11 (3): 3706–3728. https://doi.org/10.1109/TCSS.2023.3325739.

Chapel, M. N., M. Al-Ghadi, and J. C. Burie. 2023. "Authentication of Holograms with Mixed Patterns by Direct LBP Comparison." In *2023 IEEE 25th International Workshop on Multimedia Signal Processing (MMSP)*, 1–6. France: Poitiers. https://doi.org/10.1109/MMSP59012.2023.10337669.

Chen, H., and K. Magramo. 2025. "Finance worker pays out $25 million after video call with deepfake 'chief financial officer'." *CNN*. Accessed September 10, 2025. https://edition.cnn.com/2024/02/04/asia/deepfake-cfoscam-hong-kong-intl-hnk/.

Claveria, A., R. Maldonado, A. Mistry, Y. Chen, K. Song, S. Venkatesh, S. Parikh, et al. 2021. "Effects of Pressure on Acoustic Hand Biometric Authentication." In *2021 IEEE MIT Undergraduate Research Technology Conference (URTC)*, 1–5. Cambridge, MA, USA. https://doi.org/10.1109/URTC54388.2021.9701647.

Dong, J., Y. Wang, J. Lai, and X. Xie. 2023. "Restricted Black-Box Adversarial Attack Against DeepFake Face Swapping." *IEEE Transactions on Information Forensics and Security* 18:2596–2608. https://doi.org/10.1109/TIFS.2023.3266702.

Dornadula, V. N., and S. Geetha. 2019. "Credit Card Fraud Detection Using Machine Learning Algorithms." *Procedia Computer Science* 165:631–641. https://doi.org/10.1016/j.procs.2020.01.057.

European Parliament and Council of the European Union. 2024. "Regulation (Eu) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)." *[Official Journal of the European Union, L Series, 12 July 2024]*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

FedPayments Improvement. 2022a. *Toolkit Module 1: Synthetic Identity Fraud: The Basics*. https://fedpaymentsimprovement.org/synthetic-identity-fraud-mitigation-toolkit/synthetic-identity-fraud-basics/.

FedPayments Improvement. 2022b. *Toolkit Module 2: How Synthetic Identities are Used*. https://fedpaymentsimprovement.org/synthetic-identity-fraud-mitigation-toolkit/how-synthetic-identities-are-used/.
FIDO Alliance. 2024. "FIDO2: Web Authentication (WebAuthn) and Client to Authenticator Protocol (CTAP) [Fido Alliance]." https://fidoalliance.org/fido2/.
Galbally, J., S. Marcel, and J. Fierrez. 2014. "Biometric Antispoofing Methods: A Survey in Face Recognition." *IEEE Access* 2:1530–1552. https://doi.org/10.1109/ACCESS.2014.2381273.
Gangan, J., A. Gaikwad, and R. Yadav. 2023. "Detection of Fake Twitter Accounts Using Ensemble Learning Model." In *2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, 1–6. Pune, India. https://doi.org/10.1109/ICCUBEA58933.2023.10391985.
GDPR (General Data Protection Regulation). "Art. 4 GDPR – definitions - general data protection regulation (GDPR); 2018." https://gdpr-info.eu/art-4-gdpr/.
Gill, A. Q. 2012. "The Gill Framework: Adaptive Enterprise Architecture Toolkit." *CreateSpace Independent Publishing Platform*. https://books.google.com.au/books?id=37_ZkgEACAAJ.
Grassi, P. A., M. E. Garcia, and J. L. Fenton. 2017. *Digital Identity Guidelines*. Gaithersburg, MD, USA: National Institute of Standards and Technology. NIST SP 800-63-3.
Green, B., S. Gies, A. Bobnis, N. Leeper Piquero, A. R. Piquero, and E. Velasquez. 2021. "Exploring Identity-Based Crime Victimizations: Assessing Threats and Victim Services Among a Sample of Professionals." *Deviant Behavior* 42 (9): 1086–1099. https://doi.org/10.1080/01639625.2020.1720938.
Guo, Q., S. Liao, Y. Chen, S. Xiao, J. Ma and T. Zhang. 2023. "Patch-Swap Based Approach for Face Anti-Spoofing Enhancement." In *2023 IEEE Region 10 Symposium (TENSYMP)*, 1–5. Canberra, Australia. https://doi.org/10.1109/TENSYMP55890.2023.10223630.
Hanok, S., and S. Shankaraiah. 2023. "In Loco Identity Fraud Detection Model Using Statistical Analysis for Social Networking Sites: A Case Study with Facebook." *International Arab Journal of Information Technology* 20 (2): 209–217. https://doi.org/10.34028/iajit/20/2/15.
Heaven, W. D. 2020. "OpenAI's New Language Generator GPT-3 is Shockingly Good – and Completely Mindless." *MIT Technology Review*. https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/.
Heusch, G., A. George, D. Geissbuhler, Z. Mostaani, and S. Marcel. 2020. "Deep Models and Shortwave Infrared Information to Detect Face Presentation Attacks." *IEEE Transactions on Biometrics, Behavior, and Identity Science* 2 (4): 399–409. https://doi.org/10.1109/tbiom.2020.3010312.
Hilal, W., S. A. Gadsden, and J. Yawney. 2021. "Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances." *Expert Systems with Applications* 193:116429. https://doi.org/10.1016/j.eswa.2021.116429.
Hong, B., T. Connie, and M. K. Ong Goh. 2023. "Scam Calls Detection Using Machine Learning Approaches." In *2023 11th International Conference on Information and Communication Technology (ICoICT)*, 442–447. Melaka, Malaysia. https://doi.org/10.1109/ICoICT58202.2023.10262695.
Hu, M., and Y. H. Hu. 2023. "Machine Learning with Adaptive Image Colorization for Improving Facial Recognition." In *2023 Congress in Computer Science, Computer Engineering, and Applied Computing (CSCE)*, 2020–2024. Las Vegas, NV, USA. https://doi.org/10.1109/CSCE60160.2023.00332.
Hu, S., X. Zhang, J. Zhou, S. Ji, J. Yuan, Z. Li, Z. Wang, et al. 2021. "Turbo: Fraud Detection in Deposit-Free Leasing Service via Real-Time Behavior Network Mining." In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, 2583–2594. Greece: Chania. https://doi.org/10.1109/ICDE51399.2021.00289.
Irvin-Erickson, Y. 2024. "Identity Fraud Victimization: A Critical Review of the Literature of the Past Two Decades." *Crime Science* 13 (1). https://doi.org/10.1186/s40163-024-00202-0.
Islam, M. S., and I. Elwarfalli. 2023. "Deep Learning-Powered ECG-Based Biometric Authentication." In *2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM)*, 1–6. Gazipur, Bangladesh. https://doi.org/10.1109/NCIM59001.2023.10212845.
ISO/IEC. 2011. "ISO/IEC 24760-1:2011: Information Technology – Security Techniques – a Framework for Identity Management – Part 1: Terminology and Concepts [International Standard]." Accessed March 12, 2024. https://www.iso.org/standard/57914.html.
Izdihar, Z. N., S. Y. Prasetyo, S. Philip, and S. M. Isa 2023. "Age Identification Through Facial Images Using Gabor Filter and Convolutional Neural Network (CNN)." *International Journal of Intelligent Systems and Applications in Engineering* 11 (2s): 164–169. https://ijisae.org/index.php/IJISAE/article/view/2607.
Jain, A. K., and A. Kumar. 2012. "Biometric Recognition: An Overview." In *Second Generation Biometrics: The Ethical, Legal and Social Context. (The International Library of Ethics, Law and Technology*, edited by E. Mordini and D. Tzovaras, 49–79. Vol. 11. Dordrecht, The Netherlands: Springer. https://doi.org/10.1007/978-94-007-3892-8_3.
Jaiswal, A. K., and R. Srivastava. 2023. "Fake Region Identification in an Image Using Deep Learning Segmentation Model." *Multimedia Tools & Applications* 82 (25): 38901–38921. https://doi.org/10.1007/s11042-023-15032-6.
Jáuregui-Velarde, R., L. Andrade-Arenas, P. Molina-Velarde, and C. Yactayo-Arias. 2024. "Financial Revolution: A Systemic Analysis of Artificial Intelligence and Machine Learning in the Banking Sector." *International Journal of Electrical and Computer Engineering (IJECE)* 14 (1): 1079–1090. https://doi.org/10.11591/ijece.v14i1.pp1079-1090.

Jayanshre, K., and A. Amsaprabhaa. 2024. "Safeguarding Media Integrity: A Hybrid Optimized Deep Feature Fusion Based Deepfake Detection in Videos." *Computers and Security* 142:103860. https://doi.org/10.1016/j.cose.2024.103860.

Kakkar, S., P. Singh, R. Singh, S. Pal, S.C. Dimri and N. Nishant. 2023. "Analysis of Discovering Fraud in Master Card Based on Bidirectional GRU and CNN Based Model." In *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, 50–55. Erode, India. https://doi.org/10.1109/ICSSAS57918.2023.10331770.

Khan, I., M. U. Bokhari, and B. Hanafi. 2023. "Synthetic Identity Detection Using Inductive Graph Convolutional Networks." In *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, 973–978. New Delhi, India. https://ieeexplore.ieee.org/document/10112528.

Kharchenko, V., H. Fesenko, and O. Illiashenko. 2022. "Quality Models for Artificial Intelligence Systems: Characteristic-Based Approach, Development and Application." *Sensors* 22 (13): 4865. https://doi.org/10.3390/s22134865.

Kitchenham, B., O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman. 2009. "Systematic Literature Reviews in Software Engineering – a Systematic Literature Review." *Information and Software Technology* 51 (1): 7–15. https://doi.org/10.1016/j.infsof.2008.09.009.

Kodituwakku, S. R. 2015. "Biometric Authentication: A Review." *International Journal of Trend in Research and Development* 2 (4): 113–123. https://www.ijtrd.com/papers/IJTRD28.pdf.

Krishna, C. A., and R. Bhuvaneswari. 2023. "Offline Signature Forgery Detection Using Multi-Layer Perceptron." In *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*, 1–4. Pune, India. https://doi.org/10.1109/ASIANCON58793.2023.10269874.

Lalchand, S., V. Srinivas, and J. Gregorie. 2023. "Using Biometrics to Fight Back Against Rising Synthetic Identity Fraud." *Deloitte Insights*. https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2023/financial-institutions-synthetic-identity-fraud.html.

Levi, M., I. Hazan, N. Agmon, and S. Eden. 2022. "Behavioral Embedding for Continuous User Verification in Global Settings." *Computers and Security* 119:102716. https://doi.org/10.1016/j.cose.2022.102716.

Litan, A. 2015. "Market Guide for User and Entity Behavior Analytics [Gartner Research, Id g00276088, Archived]." Accessed November 22, 2024. https://www.gartner.com/document/3134524.

Luxton, D. D., R. A. Kayl, and M. C. Mishkind. 2012. "MHealth Data Security: The Need for HIPAA- Compliant Standardization." *Telemedicine and E-Health* 18 (4): 284–288. https://doi.org/10.1089/tmj.2011.0180.

Lv, B., D. Li, and Y. Sun. 2021. "Research on Modeling of E-Banking Fraud Account Identification Based on Federated Learning." In *2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, 611–618. Calgary, AB, Canada. https://doi.org/10.1109/DASC-PICom-CBDCom-CyberSciTech52372.2021.00105.

Magee, J., S. Sheridan, and C. Thorpe. 2023. "Classifying Recaptured Identity Documents Using the Biomedical Meijering and Sato Algorithms." *Proceedings of Research Papers at APWG.EU TECH 2023; (CEUR Workshop Proceedings; Vol. 3631)*, Dublin, Ireland Medina M. Dmytryshyna Z. Freitas A. (CEUR-WS.org.). https://ceur-ws.org/Vol-3631/paper2.pdf.

Masood, M., M. Nawaz, K. M. Malik, A. Javed, A. Irtaza, and H. Malik. 2023. "Deepfakes Generation and Detection: State-Of-The-Art, Open Challenges, Countermeasures, and Way Forward." *Applied Intelligence* 53 (4): 3974–4026. https://doi.org/10.1007/s10489-022-03766-z.

Mazaheri, G., and A. K. Roy-Chowdhury. 2022. "Detection and Localization of Facial Expression Manipulations." In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2773–2783. Waikoloa, HI, USA. https://doi.org/10.1109/WACV51458.2022.00283.

Modic, D., and R. Anderson. 2015. "It's All Over but the Crying: The Emotional and Financial Impact of Internet Fraud." *IEEE Security & Privacy* 13 (5): 99–103. https://doi.org/10.1109/MSP.2015.107.

Norris, G., A. Brookes, and D. Dowell. 2019. "The Psychology of Internet Fraud Victimisation: A Systematic Review." *Journal of Police and Criminal Psychology* 34 (3): 231–245. https://doi.org/10.1007/s11896-019-09334-5.

Nugroho, K., and E. Winarno. 2022. "Spoofing Detection of Fake Speech Using Deep Neural Network Algorithm." In *2022 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 56–60. Indonesia: Semarang. https://doi.org/10.1109/iSemantic55962.2022.9920401.

OAIC (Office of the Australian Information Commissioner). 2020a. "Notifiable Data Breaches Report: January–June 2020." *Office of the Australian Information Commissioner*. https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breaches-publications/notifiable-data-breaches-report-januaryjune-2020.

OAIC (Office of the Australian Information Commissioner). 2020b. "Notifiable Data Breaches Report: July–December 2020." *Office of the Australian Information Commissioner*. https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breaches-publications/notifiable-data-breaches-report-julydecember-2020.

OAIC (Office of the Australian Information Commissioner). 2021a. "Notifiable Data Breaches Report: January–June 2021." *Office of the Australian Information Commissioner*. https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breachespublications/notifiable-data-breaches-report-januaryjune-2021.

OAIC (Office of the Australian Information Commissioner). 2021b. "Notifiable Data Breaches Report: July–December 2021." *Office of the Australian Information Commissioner*. https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breaches-publications/notifiable-data-breaches-report-july-to-december-2021.

OAIC (Office of the Australian Information Commissioner). 2022a. "Notifiable Data Breaches Report: January–June 2022." *Office of the Australian Information Commissioner*. https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breaches-publications/notifiable-data-breaches-report-january-to-june-2022.

OAIC (Office of the Australian Information Commissioner). 2022b. "Notifiable Data Breaches Report: July–December 2022." *Office of the Australian Information Commissioner*. https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breaches-publications/notifiable-data-breaches-report-july-to-december-2022.

OAIC (Office of the Australian Information Commissioner). 2023a. "Notifiable Data Breaches Report: January–June 2023." *Office of the Australian Information Commissioner*. https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breaches-publications/notifiable-data-breaches-report-january-to-june-2023.

OAIC (Office of the Australian Information Commissioner). 2023b. "Notifiable Data Breaches Report: July–December 2023." *Office of the Australian Information Commissioner*. https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breaches-publications/notifiable-data-breaches-report-july-to-december-2023.

OAIC (Office of the Australian Information Commissioner). 2024. "Notifiable Data Breaches Report: January–June 2024." *Office of the Australian Information Commissioner*. https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breaches-publications/notifiable-data-breaches-report-january-to-june-2024.

Priyaradhikadevi, T., S. Vanakovarayan, and E. Praveena. 2023. "Credit Card Fraud Detection Using Machine Learning Based on Support Vector Machine." In *2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, 1–6. Chennai, India. https://doi.org/10.1109/ICONSTEM56934.2023.10142247.

Pujol, F. A., M. J. Pujol, C. Rizo-Maestre, and M. Pujol. 2019. "Entropy-Based Face Recognition and Spoof Detection for Security Applications." *Sustainability* 12 (1): 85. https://doi.org/10.3390/su12010085.

Rodrigues, V. F., L. M. Policarpo, D. E. da Silveira, R. da Rosa Righi, C. A. da Costa, J. L. V. Barbosa, R. S. Antunes, R. Scorsatto, and T. Arcot. 2022. "Fraud Detection and Prevention in E-Commerce: A Systematic Literature Review." *Electronic Commerce Research and Applications* 56:101207. https://doi.org/10.1016/j.elerap.2022.101207.

Rountree, D. 2013. "Introduction to Identity." *Federated Identity Primer Syngress*: 1–11. https://doi.org/10.1016/b978-0-12-407189-6.00001-7.

Russell, S., and P. Norvig. 2021. *Artificial Intelligence: A Modern Approach*. Pearson Education, Limited. https://www.pearson.com/en-gb/subjectcatalog/p/artificial-intelligence-a-modern-approach-global-edition/P200000003310/9781292401133.

Sadowski, G., J. Care, and N. MacDonald. 2019. "Market Guide for User and Entity Behavior Analytics [Gartner Research, Id g00361156, Archived]." Accessed November 22, 2024. https://www.gartner.com/document/3917096.

Safara, F., A. S. Mohammed, M. Potrus, S. Ali, Q. T. Tho, A. Souri, F. Janenia, and M. Hosseinzadeh. 2020. "An Author Gender Detection Method Using Whale Optimization Algorithm and Artificial Neural Network." *IEEE Access* 8:48428–48437. https://doi.org/10.1109/ACCESS.2020.2973509.

Segovia Domingo, A. I., and M. Enríquez. 2018. "Digital Identity: The Current State of Affairs." Accessed March 12, 2024. https://www.bbvaresearch.com/wp-content/uploads/2018/02/Digital-Identity_the-current-state-of-affairs.pdf.

Shafique, U., H. Khan, Sabah-ud-din. Waqar, A. Sher, A. Zeb, U. Shafi, R. Ullah, and R. Ullah. 2017. "Modern Authentication Techniques in Smart Phones: Security and Usability Perspective." *International Journal of Advanced Computer Science & Applications* 8 (1). https://doi.org/10.14569/IJACSA.2017.080142.

Sharma, S., and A. Selwal. 2026. "Potential of Artificial Intelligence in Deepfake Media: From Generation to Detection Mechanisms, State-Of-The-Art, and Challenges." *Computer Science Review* 60:100866. https://doi.org/10.1016/j.cosrev.2025.100866.

Sharma, V., S. Sharma, and V. Batra. 2023. "U-Architecture for Face Recognition to Prevent Cyber and Spoofing Attacks in IoT." In *2023 International Conference on Data Science and Network Security (ICDSNS)*, 1–8. Tiptur, India. https://doi.org/10.1109/ICDSNS58469.2023.10245245.

Singhai, A., S. Aanjankumar, and S. Poonkuntran. 2023. "A Novel Methodology for Credit Card Fraud Detection Using KNN Dependent Machine Learning Methodology." In *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, 878–884. Salem, India. https://doi.org/10.1109/ICAAIC56838.2023.10141427.

Srivastva, R., Y. N. Singh, and A. Singh. 2022. "Statistical Independence of ECG for Biometric Authentication." *Pattern Recognition* 127:108640. https://doi.org/10.1016/j.patcog.2022.108640.

Temoshok, D., C. Abruzzi, and Y. Y. Choong. 2025. *Digital Identity Guidelines: Identity Proofing and Enrollment*. Gaithersburg, MD, USA: National Institute of Standards and Technology (NIST). NIST SP 800-63A-4. https://doi.org/10.6028/NIST.SP.800-63A-4.

Trabelsi, A., M. Pic, and J. L. Dugelay. 2022. "Recapture Detection to Fight Deep Identity Theft." In *Proceedings of the 2022 4th International Conference on Video, Signal and Image Processing*, 37–42. Shanghai, China. https://doi.org/10.1145/3577164.3577170.

Uliyan, D. M., S. Sadeghi, and H. A. Jalab. 2020. "Anti-Spoofing Method for Fingerprint Recognition Using Patch Based Deep Learning Machine." *Engineering Science and Technology, an International Journal* 23 (2): 264–273. https://doi.org/10.1016/j.jestch.2019.06.005.

Umbrani, K., D. Shah, and A. Pile. 2024. "Fake Profile Detection Using Machine Learning." In *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*, 966–973. Bahrain: Manama. https://doi.org/10.1109/ICETSIS61505.2024.10459570.

United Nations Office on Drugs and Crime. 2011. "United Nations Handbook on Identity- Related Crime." Accessed March 18, 2024. https://www.unodc.org/documents/treaties/UNCAC/Publications/Handbook_on_ID_Crime/10-57802_ebooke.pdf.

Velásquez, I., A. Caro, and A. Rodríguez. 2017. "Authentication Schemes and Methods: A Systematic Literature Review." *Information and Software Technology* 94:30–37. https://doi.org/10.1016/j.infsof.2017.09.012.

Vinod, B., and A. M. Senthil Kumar. 2023. "Handwritten Signature Identification and Fraud Detection Using Deep Learning and Computer Vision." In *2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, 196–200. Erode, India. https://doi.org/10.1109/ICSCDS56580.2023.10104929.

Wang, L., M. S. Hossain, J. Pulfrey, and L. Lancor. 2021. "The Effectiveness of Zoom Touchscreen Gestures for Authentication and Identification and Its Changes Over Time." *Computers and Security* 111:102462. https://doi.org/10.1016/j.cose.2021.102462.

Würsching, L., F. Putz, and S. Haesler. 2023. "FIDO2 the Rescue? Platform vs Roaming Authentication on Smartphones." In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–16. Hamburg, Germany: ACM. https://doi.org/10.1145/3544548.3580993.

Xie, S., T. Qiao, S. Li, X. Zhang, J. Zhou, and G. Feng. 2026. "DeepFake Detection in the AIGC Era: A Survey, Benchmarks, and Future Perspectives." *Information Fusion* 127:103740. https://doi.org/10.1016/j.inffus.2025.103740.

Yang, H., X. Meng, X. Zhao, Y. Wang, Y. Liu, X. Kang, J. Shen, and W. Huang. 2023. "CKDAN: Content and Keystroke Dual Attention Networks with Pre-Trained Models for Continuous Authentication." *Computers and Security* 128:103159. https://doi.org/10.1016/j.cose.2023.103159.

Yang, J., R. Xu, Z. Qi, and Y. Shi. 2022. "Visual Anomaly Detection for Images: A Systematic Survey." *Procedia Computer Science* 199:471–478. https://doi.org/10.1016/j.procs.2022.01.057.

Ye, Q., Y. Gao, and Z. Zhang. 2022. "Modeling Access Environment and Behavior Sequence for Financial Identity Theft Detection in E-Commerce Services." In *2022 International Joint Conference on Neural Networks (IJCNN)*, 1–8. Italy: Padua. https://doi.org/10.1109/IJCNN55064.2022.9892383.

Zhang, C. J., A. Q. Gill, and B. Liu. 2025. "Ai-Based Identity Fraud Detection: A Systematic Review [Arxiv Preprint arxiv:2501.09239]." https://doi.org/10.48550/arXiv.2501.09239.

Zhang, C. J., A. Q. Gill, and B. Liu. 2026. "Dataset For: AI-Based Identity Fraud Detection: A Systematic Review." https://doi.org/10.5281/zenodo.21482806.

Zhao, P., and M. Wang. 2021. "Mobile Behavior Trusted Certification Based on Multivariate Behavior Sequences." *Neurocomputing* 419:203–214. https://doi.org/10.1016/j.neucom.2020.08.003.