WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

The Intersection of Federated Learning and Explainable AI in Financial Crime Detection: A Review of Current Trends, Challenges, and Opportunities

[1]Emmanuel Nannim RAMSON, [2]Abdusalam Ya'u Gital, [3]Fatima Umar Zambuk, [4]Mustapha Abdulrahman Lawal, [5]Odulu Lydia & [6]Ismail Zahraddeen Yakubu*

[1,2,3,4]Department of Computer Science, Abubakar Tafawa Balewa University Bauchi
[5]Ground Receiving Station, National Center for Remote Jos, Plateau State
[6]Department of Computing Technology, SRM Institute of Science and Technology, Chennai, India

A B S T R A C T :

The rapid growth of financial crime over the past decade poses a severe threat to global financial stability, with money laundering alone accounting for an estimated 2–5% of global GDP, amounting to over $2 trillion annually. The increasing digitalization of financial transactions has further exacerbated the rise of fraud and cyber-related crimes. Artificial intelligence (AI), particularly machine learning (ML), has emerged as a crucial tool for detecting financial crime by analyzing vast transaction datasets, identifying patterns, and flagging anomalies. However, challenges such as the lack of transparency in AI decision-making, often termed the "black box" problem, hinder the widespread adoption and trust in AI-driven financial crime detection systems. This review examines the effectiveness of federated learning and explainable AI techniques in financial crime detection. Specifically, it explores the advantages and limitations of federated learning in ensuring data privacy and enabling cross-institutional collaboration. Additionally, it assesses explainable AI approaches in enhancing transparency and trust in AI-driven financial crime detection. By critically evaluating existing literature, this review identifies research gaps and outlines future directions for improving AI-based financial crime detection frameworks.

Keywords: Financial Crime, Federated Learning, Explainable AI, Machine Learning, Fraud Detection

1. Introduction

Financial crime encompasses a wide range of illicit activities aimed at securing financial or economic gain through
deception or unethical practices (Reurink, 2018). These crimes typically include money laundering, terrorism financing, tax evasion, fraud (both digital and traditional), market manipulation, and cybercrimes like identity theft and ransomware attacks (Zagaris & Mostaghimi, 2023). The global scale of financial crime has grown dramatically in the past decade, threatening the integrity of financial markets and the stability of the global economy. For instance, a report by the Financial Action Task Force (FATF) in 2021 estimated that money laundering alone represents around 2–5% of the global GDP, translating to over $2 trillion annually. Fraud and cyber-related crimes, especially digital fraud, have also surged due to the increasing digitalization of financial transactions.

One of the most notable impacts of financial crime is the harm inflicted on financial institutions in terms of fines, legal costs, and reputational damage. For instance, in 2020, Goldman Sachs faced a $2.9 billion fine as part of the 1MDB scandal, an example of massive financial crime involving fraud, bribery, and money laundering (Harvey, 2023). Beyond monetary fines, institutions face reputational damage that can erode consumer trust and loyalty, ultimately affecting their
market position. Additionally, governments and regulatory bodies globally have implemented stricter policies, such as the European Union's General Data Protection Regulation (GDPR) and the U.S. Bank Secrecy Act, to combat financial crime, but these efforts often struggle to keep up with the speed and sophistication of modern crimes (Bent & Bent, 2021). Technological advancements have changed the landscape of financial crime detection and prevention. Traditional methods, such as rule-based systems, expert-driven audits, and manual transaction monitoring, though still valuable, are no longer effective enough in dealing with the sophisticated tactics employed by modern criminals (Mohanty & Mishra, 2023). Rulebased systems, for instance, rely on predefined heuristics or threshold-based alerts to detect suspicious transactions. However, this approach tends to generate a high number of false positives, overwhelming analysts and auditors with unnecessary work. Additionally, these methods struggle to identify new patterns of criminal behavior, especially as criminals adapt their tactics to evade detection systems.

To address these challenges, artificial intelligence (AI) has emerged as a powerful tool for combating financial
crime. Machine learning (ML) models, a subset of AI, are capable of analyzing vast amounts of transaction data, identifying patterns, and detecting anomalies more efficiently than traditional methods (Mohanty & Mishra, 2023). These models can learn from historical data and adapt to changing

patterns of fraudulent behavior over time. More sophisticated AI systems, such as deep learning models, can analyze complex, high-dimensional data, further enhancing their ability to detect fraudulent or illegal activities (Nicholls, Kuppa, & Le-Khac, 2021).

However, despite the promise that AI holds for financial crime detection, significant challenges remain. One of the most notable challenges is the lack of transparency in AI decision-making, often referred to as the "black box" problem. Most AI models, particularly deep learning models, are inherently complex and difficult to interpret. While these models may accurately flag suspicious transactions, they often cannot explain why a particular transaction was flagged, making it difficult for financial institutions and regulatory bodies to trust or justify the decisions made by AI systems (Sharma, Mehta, & Sharma, 2024). This opacity is particularly problematic in high-stakes environments such as financial services, where wrong decisions can have far-reaching legal and financial consequences.

Furthermore, the implementation of AI models in financial crime detection requires access to large datasets, often containing sensitive information such as bank transactions, personal identification details, and customer financial behavior. Sharing such data across organizations or jurisdictions can be a regulatory and ethical minefield. Data privacy laws, including the GDPR, restrict how organizations share and process personal data, posing challenges to AI's widespread deployment in combating financial crime.

Federated Learning (FL) is an emerging AI technique that addresses the data-sharing problem in financial crime detection. Introduced by Google in 2017, federated learning allows multiple parties (e.g., financial institutions) to collaboratively train machine learning models without directly sharing their data (Agrawal et al., 2022). Instead of transferring sensitive data to a central server for model training, federated learning distributes the model to each participating entity, where it is trained locally on the respective datasets. After the local training is complete, only the model updates (e.g., gradients or weight adjustments) are shared with the central server, which aggregates the updates to improve the global model (Shaheen, Farooq, Umer, & Kim, 2022). This decentralized approach offers several advantages in financial crime detection, particularly in terms of privacy and compliance. By keeping the data localized, federated learning ensures that sensitive financial data remains within the institution's control, mitigating the risk of data breaches and complying with data protection regulations. Moreover, it enables collaboration between financial institutions, which is essential for detecting cross-border financial crimes such as money laundering and terrorism financing. Criminals often move money across multiple institutions and jurisdictions to obscure their tracks, making it difficult for a single institution to detect suspicious activity on its own. Federated learning facilitates cross-institutional collaboration by allowing institutions to jointly develop more robust models without sacrificing data privacy (Ahmed & Alabi, 2024).

Despite its potential, federated learning also presents challenges. One of the most significant technical challenges is the heterogeneity of data across different institutions. Financial institutions differ in terms of the types of transactions they process, the demographics of their customers, and the local regulations they must follow. This variation can make it difficult to create a unified, global model that performs well across all institutions. Additionally, federated learning requires secure communication protocols to ensure that model updates are not tampered with during transmission. Ensuring the integrity of the shared model updates is crucial for preventing adversarial attacks, where malicious actors could inject harmful model updates to degrade the model's performance (Sen, Waghela, & Rakshit, 2024).

Explainable Artificial Intelligence (XAI) is another critical development in AI that addresses the transparency issue associated with traditional machine learning models. While most advanced AI models, particularly deep learning models, offer state-of-the-art performance, they often operate as "black boxes," meaning that their internal decision-making processes are opaque to users (Arrieta et al., 2020). This opacity is problematic in sensitive applications like financial crime detection, where financial institutions and regulators need to understand and explain the reasoning behind specific AI-driven decisions. The need for interpretability is particularly critical when AI systems are used to flag transactions as suspicious, as these decisions can lead to further investigations, legal action, or the freezing of customer accounts (Das & Rad, 2020).

XAI techniques aim to make AI models more interpretable and transparent without sacrificing accuracy. There are two primary approaches to achieving explainability in AI systems: model-specific techniques and model-agnostic techniques (Chamola et al., 2023). Model-specific techniques are designed for a particular type of model. For instance, decision trees and linear regression models are inherently interpretable, as they provide clear rules or coefficients that can be easily understood by humans. In contrast, deep learning models are less interpretable, but researchers have developed techniques like Layer-wise Relevance Propagation (LRP) and attention mechanisms to visualize the decision-making process in neural networks (Došilović, Brčić, & Hlupić, 2018).

Model-agnostic techniques, on the other hand, can be applied to any machine learning model, regardless of its complexity. Two of the most widely used model-agnostic XAI techniques are Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) (Zafar & Khan, 2021). LIME works by approximating the complex model with a simpler, interpretable model (such as a linear regression) locally around the prediction. This approximation provides insights into which features contributed most to the model's decision for a specific instance (Salih et al., 2024). SHAP, based on cooperative game theory, assigns each feature a contribution score by calculating its marginal contribution to the model's prediction. This approach provides a global view of feature importance across all instances and a local view of feature importance for individual predictions (Y. Wang, 2024).

The application of XAI in financial crime detection is essential for gaining the trust of financial institutions and regulatory bodies. Without explainability, financial institutions may be reluctant to adopt AI systems for detecting fraud or money laundering due to concerns about accountability (Kute, Pradhan, Shukla, & Alamri, 2021). For instance, if a bank uses an AI model to flag a transaction as suspicious, it must be able to explain the decision to regulators, auditors, and even the affected customer. XAI techniques can help institutions provide these explanations, reducing the risk of regulatory fines or legal disputes.

However, the use of XAI also comes with challenges. There is often a trade-off between model interpretability and performance (Shah & Konda, 2021). Complex models like deep neural networks are generally more accurate than simpler, interpretable models, but they are also harder to explain. Striking the right balance between accuracy and interpretability is critical for financial institutions, where both performance and transparency are paramount. Additionally, there are concerns about the scalability of XAI techniques, particularly when applied to large datasets or real-time financial transactions (Weber, Lapuschkin, Binder, & Samek, 2023). Financial institutions need XAI systems that can provide explanations quickly and efficiently, especially in the context of real-time fraud detection.

1.1 Research Question and Objectives

Given the increasing reliance on AI systems for detecting financial crime, it is essential to critically evaluate the effectiveness of federated learning and explainable AI techniques in this domain. The primary research question guiding this review is as follows: How effective are federated learning and explainable AI techniques in detecting financial crime, and what challenges remain in their practical implementation? The objectives of this review are threefold:

1. To explore the application of federated learning in financial crime detection and assess its advantages and limitations, particularly in data privacy and cross-institutional collaboration.
2. To examine the use of explainable AI techniques in enhancing transparency and trust in AI-driven financial crime detection systems, with a focus on model-agnostic and model-specific approaches.
3. To critically evaluate the existing literature on federated learning and explainable AI in financial crime detection, identify gaps in current research, and propose potential directions for future studies.

This review aims to address these objectives and provide a comprehensive overview of the current state of research on AIdriven financial crime detection and offer insights into the practical challenges and future opportunities in this field.

2. Methodology

The methodology section is crucial in any systematic review, as it ensures that the literature search, selection process, and analysis are transparent, reproducible, and unbiased. For this review on the application of federated learning and explainable AI in detecting financial crime, a systematic approach was followed to gather, evaluate, and synthesize relevant literature. This section outlines the literature search strategy, the criteria for selecting studies, and the framework used for analyzing and synthesizing the findings.

2.1 Literature Search and Selection Process

The literature search was conducted across several electronic databases, including Google Scholar, IEEE Xplore, ACM Digital Library, ScienceDirect, and Scopus. These databases were selected based on their relevance to AI, computer science, and finance. The search spanned articles published between 2015 and 2024, considering that both federated learning and explainable AI are relatively recent developments, with federated learning gaining prominence around 2017. To ensure a comprehensive review, both peer-reviewed articles and conference papers were included. The search queries were designed to cover the key terms related to the topic, specifically targeting the intersection of federated learning, explainable AI, and financial crime detection. Boolean operators were used to combine terms, ensuring that the search results captured studies addressing all the relevant themes.

Table 1. Search Queries and Results

Database	**Search Query**	**Number of Results**	**Filtered (Relevant) Results**
Google Scholar	("Federated Learning" AND "Explainable AI" AND "Financial Crime Detection")	120	15
IEEE Xplore	("Federated Learning" OR "Decentralized AI" AND "Financial Fraud Detection")	95	5
ACM Digital Library	("Explainable AI" AND "Money Laundering" OR "Financial Crime" AND "Fraud Detection")	83	10
ScienceDirect	("Federated Learning" OR "XAI" AND "Anomaly Detection" AND "Financial Crimes" OR "Fraud")	76	8
Scopus	("Explainable AI" AND "Fraud Detection" OR "Financial Crime" AND "Artificial Intelligence" OR "Machine Learning")	150	20
Total		**524**	**58**

Search Strategy

1. **Initial Search**: The initial search across databases returned a total of 524 results. These results included peerreviewed journal articles, conference papers, and some preprints relevant to federated learning, explainable AI, and financial crime detection.
2. **Filtering by Title and Abstract**: The initial pool was screened by titles and abstracts. Duplicates were removed, and studies that did not align with the key focus areas (i.e., federated learning and explainable AI in the context of financial crime) were excluded. This filtering process reduced the number of articles to **58**.
3. **Full-Text Review**: After an in-depth review of the full texts, a further elimination was made based on whether the studies provided substantial empirical or theoretical insights into the application of federated learning and explainable AI in detecting financial crime. Articles focusing solely on general AI applications, without specific reference to financial crime, were excluded, narrowing down the pool to **58** relevant

studies.

4. **Final Selection**: The remaining studies were then classified based on their contributions to the research question, with a focus on studies that directly compared federated learning or explainable AI techniques for financial crime detection. Studies that discussed only one of the two AI approaches, but not their intersection, were still considered valuable if they contributed significant theoretical or empirical data relevant to either field.

2.2 Criteria for Selecting Studies

To ensure that the review is comprehensive, yet focused, the following inclusion and exclusion criteria were applied:

Inclusion Criteria:

1. **Publication Year**: Studies published between **2015 and 2024** were included, as this time frame captures the emergence and development of both federated learning and explainable AI in financial crime detection.
2. **Language**: Only articles written in **English** were considered.
3. **Relevance to Topic**: The studies must address at least one of the core themes: federated learning, explainable AI, or financial crime detection. Preference was given to studies that discuss the intersection of these themes.
4. **Study Type**: Empirical studies, theoretical papers, review articles, and case studies were included to ensure a diverse range of perspectives.
5. **Quality**: Only peer-reviewed journal articles and conference papers were included to maintain a high standard of credibility.

Exclusion Criteria:

1. **Non-technical Studies**: Articles focusing on policy or legal aspects of financial crime without any technical analysis of AI-based detection methods were excluded.
2. **Irrelevant Domains**: Studies discussing federated learning or explainable AI in non-financial domains (e.g., healthcare or education) were excluded unless they offered generalizable insights into AI's application to anomaly detection or data privacy.
3. **Duplicate Publications**: Where the same research appeared in different journals or conferences, only the most comprehensive version was retained.

The literature was classified into two main categories: **federated learning applications** and **explainable AI applications**. Within these categories, studies were further classified based on their focus (e.g., empirical vs. theoretical studies) and their contribution to financial crime detection.

Table 2. Literature Classification

Category	Subcategory	Number of Studies	Percentage of Total (%)
Federated Learning	Empirical Studies	15	25.8
	Theoretical/Review Studies	12	20.6
Explainable AI	Empirical Studies	18	31.0
	Theoretical/Review Studies	13	22.4
Total		**58**	**100**

Most of the literature focuses on empirical studies (31.03%), indicating that a significant portion of the existing research provides practical implementations or case studies of federated learning or explainable AI in financial crime detection. This is valuable for understanding these AI techniques' real-world applications and limitations.

Framework for Analyzing and Synthesizing the Literature

After the literature selection process, the following framework was applied to analyze and synthesize the studies:

1. **Categorization by AI Technique**: The studies were first categorized based on whether they primarily discussed federated learning, explainable AI, or both. Studies that integrated both techniques (e.g., using federated learning models with explainable AI mechanisms) were particularly emphasized.
2. **Analysis of Methodologies**: For each study, the methodology was critically analyzed to assess the type of AI models used, the datasets employed, and the metrics used for evaluation. This allowed for a comparative analysis of the performance and limitations of various AI techniques in detecting financial crime.
3. **Evaluation of Federated Learning Applications**: The analysis of federated learning studies focused on its implementation in financial crime detection, including discussions of data privacy, model aggregation methods, and the challenges posed by data heterogeneity across institutions. Special attention was given to studies that proposed novel aggregation algorithms or tackled adversarial attacks within federated learning systems.
4. **Evaluation of Explainable AI Applications**: For explainable AI studies, the focus was on the techniques used to provide interpretability, such as LIME, SHAP, or attention mechanisms. The trade-offs between model performance and interpretability were analyzed, and the effectiveness of XAI in enhancing trust and regulatory compliance in financial institutions was explored.
5. **Critical Review and Synthesis**: The findings from both federated learning and explainable AI studies were synthesized to highlight areas of convergence, where both techniques offer complementary benefits. For instance, several studies suggested that federated learning could be

enhanced with explainable AI techniques to address both data privacy concerns and the need for transparent decision-making in financial crime detection. Gaps in the literature, such as the lack of standardized evaluation metrics or the under-explored intersection of these AI techniques, were identified and discussed.

6. **Identification of Challenges and Future Research Directions**: Finally, based on the synthesis, the key challenges faced by federated learning and explainable AI in the context of financial crime detection were identified. These challenges include issues such as the high computational cost of federated learning, the difficulty of balancing model performance with interpretability in XAI, and the need for better techniques to handle adversarial attacks. Future research directions were proposed, focusing on how these challenges could be addressed in subsequent studies.

3. Result and Analysis

A. Conventional Machine Learning 1) Statistical Methods in fraud detection

Fraud detection is critical in safeguarding various industries such as banking, healthcare, and finance. The rapid rise of sophisticated fraudulent activities, driven by technology, has rendered traditional detection methods insufficient. To counteract this growing threat, machine learning (ML) algorithms have emerged as effective tools for identifying and mitigating fraudulent behavior. Over the years, numerous studies have explored the application of ML techniques for fraud detection, each contributing unique perspectives and advancements to the field. This literature review delves into significant contributions, highlighting the problems addressed, the methodologies proposed, and the findings achieved while drawing connections between various studies to establish a cohesive narrative on machine learning's role in fraud detection. One of the earlier explorations into fraud detection through statistical methods was provided by Li et al. (2008), who conducted a comprehensive survey on healthcare fraud detection. The study underscores the complexity of healthcare fraud, such as overbilling and false claims, which traditional detection methods struggled to address effectively. By focusing on statistical methods, Li et al. (2008) laid the groundwork for understanding how both unsupervised and supervised learning algorithms can aid in detecting fraudulent activities in the healthcare domain. The study highlighted the limitations of existing systems and suggested that the integration of machine learning with traditional statistics could significantly improve detection outcomes. This early work forms the foundation for the later inclusion of machine learning in fraud detection efforts across other sectors.

Building on these statistical approaches, Bolton and Hand (2002) reviewed various fraud detection techniques, particularly focusing on financial sectors. They highlighted the limitations of traditional rule-based systems, which often struggle to adapt to the ever-evolving nature of fraud schemes, such as money laundering and credit card fraud. Their work pointed out the dynamic and adaptable nature of machine learning techniques compared to the rigidity of traditional statistical methods. They emphasized that combining both statistical methods and machine learning approaches could better capture the complexity of financial fraud. This comparative evaluation provided a crucial shift toward more robust detection systems, which increasingly rely on adaptive learning models to combat evolving fraudulent activities.

While Li et al. (2008) and Bolton and Hand (2002) emphasized statistical techniques, Wang (2010) extended this discourse by focusing on accounting fraud detection through data mining. Wang (2010) provided a critical analysis of how data mining techniques, such as decision trees and support vector machines (SVMs), outperform traditional auditing methods in detecting fraudulent activities within corporate financial statements. However, Wang (2010) also acknowledged the challenges posed by small sample sizes of fraud cases, which limit the generalization capability of ML models. To counteract this, the study proposed the application of semi-supervised learning models that could effectively leverage both labeled and unlabeled data. The innovative approach highlighted the potential for ML techniques to manage imbalanced datasets, a recurring challenge in fraud detection, thus pushing the field toward more advanced methods of managing real-world complexities.

Similarly, the study by Whiting et al. (2012) explored machine learning's ability to detect patterns of management fraud, particularly focusing on ensemble methods. Ensemble models aggregate multiple machine learning techniques to improve accuracy and predictive power. Whiting et al. (2012) demonstrated that these methods, specifically rule ensembles, were adept at identifying complex fraud patterns typically associated with management-level deception. The study found that by leveraging multiple algorithms, ensemble methods could not only detect fraudulent behavior but also preemptively identify potential fraud risks. This proactive approach contrasts with earlier models that were more reactionary in nature, thus marking a significant progression in fraud detection research. The findings of Whiting et al. (2012) further validated Wang's (2010) assertion that machine learning models, especially ensemble methods, can manage the complexity and subtlety of fraud patterns in a way that traditional systems cannot.

The integration of natural language processing (NLP) with machine learning in fraud detection represents another key advancement. Purda (2014) explored the potential of language analysis in detecting fraudulent intent in corporate financial reports. The study developed a model using a bag-of-words approach to assess linguistic patterns within the management discussion and analysis (MD&A) sections of financial reports. Unlike traditional fraud detection methods that relied heavily on numerical data, Purda's (2014) approach focused on the language used in these reports, achieving a fraud detection rate of 82%. This marked a shift toward incorporating NLP techniques into machine learning models, opening new avenues for detecting deception based on textual patterns. The ability to detect fraud through language analysis added another layer of complexity to machine learning's role in fraud detection, proving its versatility across different data types.

West and Bhattacharya (2016) extended the discussion on machine learning techniques in financial fraud detection by providing a comprehensive review of computational intelligence methods, such as genetic algorithms, fuzzy logic, and neural networks. Their work built upon earlier research by examining how these computational methods could be applied to a variety of financial fraud types, such as insider trading and tax evasion. However, West and Bhattacharya (2016) noted that despite the promise of computational intelligence techniques, there remained a significant gap in understanding how different fraud types corresponded with algorithmic performance. This echoed the sentiments of previous studies, like Bolton and Hand (2002), regarding the need for hybrid models that combine multiple techniques to achieve more robust fraud detection. West and Bhattacharya's review reinforced the idea

that machine learning is most effective when combined with other computational approaches, offering a more holistic solution to the complexities of fraud.

Edge and Sampaio (2009) approached fraud detection from a real-time perspective, exploring signature-based methods that analyze account activity to detect anomalous behavior. Their work highlighted the potential of account profiling technologies in financial fraud detection, particularly when enhanced with machine learning algorithms for anomaly detection. Edge and Sampaio's (2009) focus on real-time detection marked a significant departure from previous studies, which primarily focused on post-fraud detection. By leveraging machine learning algorithms, signature-based methods could continuously update their detection parameters to respond to new fraud techniques. This real-time approach became increasingly important as fraudsters developed more sophisticated methods to evade traditional detection systems.

More recent contributions, such as that by Hashemi et al. (2023), have focused on leveraging deep learning techniques in banking fraud detection. Using a combination of class weight-tuning hyperparameters, Bayesian optimization, CatBoost, and XGBoost, Hashemi et al. (2023) developed a machine learning model capable of addressing the imbalanced datasets typically associated with banking fraud. This study emphasized the importance of optimizing hyperparameters to improve the performance of fraud detection models, particularly in dealing with the class imbalance between fraudulent and nonfraudulent transactions. The incorporation of deep learning marked an advancement in the ability to detect complex fraud schemes in banking, particularly those involving large datasets. Hashemi et al.'s (2023) work aligns with West and Bhattacharya's (2016) suggestion of adopting hybrid models, further emphasizing the need for multi-faceted approaches in modern fraud detection.

Domingues et al. (2018) contributed to this evolving body of research by comparing various outlier detection algorithms, such as isolation forests and one-class support vector machines (SVM), across multiple domains including fraud detection. The study revealed that outlier detection techniques, particularly isolation forests, were highly effective in identifying fraudulent behavior due to their robustness and scalability. Domingues et al.'s (2018) work provided empirical evidence supporting the scalability of these algorithms, highlighting their suitability for large datasets often encountered in sectors like banking and telecommunications. Their findings align with Hashemi et al. (2023), demonstrating the effectiveness of machine learning models that can manage imbalanced data while maintaining robust detection capabilities.

Table 3: Summary of Statistical Methods

Authors	**Proposed Study**	**Key Findings**
(J. Li, Huang, Jin, & Shi, 2008)	A survey on statistical methods for health care fraud detection	This paper provides a comprehensive survey of statistical methods for health care fraud detection, highlighting areas for future research and identifying gaps in existing research.
(Bolton & Hand, 2002)	Statistical fraud detection: A review	Statistical fraud detection technologies, such as statistics and machine learning, effectively detect activities like money laundering, e-commerce credit card fraud, telecommunications fraud, and computer intrusion.
(S. Wang, 2010)	A Comprehensive Survey of Data Mining-Based Accounting-Fraud Detection Research	Data mining-based accounting fraud detection models are more effective than auditors alone, but more no-tag data mining algorithms are needed to address the small size of fraud samples.
(Whiting, Hansen, McDonald, Albrecht, & Albrecht, 2012)	Machine learning methods for detecting patterns of management fraud	Ensemble methods, such as rule ensembles, show promise in detecting management fraud patterns, with potential for proactive discovery and mitigation.
(Purda & Skillicorn, 2015)	Accounting Variables , Deception , and a Bag of Words : Assessing the Tools of Fraud Detection	This paper develops a statistical method for analyzing language in management discussion and analysis sections of firm's annual and quarterly reports, demonstrating a highly effective fraud-detection rate of 82%.
(West & Bhattacharya, 2016)	Intelligent financial fraud detection: A comprehensive review	Computational intelligence-based financial fraud detection methods show promise, but a research gap exists in addressing associations between fraud types, algorithms, and their performance.
(Edge & Sampaio, 2009)	A survey of signature-based methods for financial fraud detection	Account signatures, a proactive account profiling technology, can effectively detect financial fraud in real-time by analyzing streaming financial data and identifying ambiguous user behavior.
(Hashemi, Mirtaheri, & Greco, 2022)	Fraud Detection in Banking Data by Machine Learning Techniques	Using class weight-tuning hyperparameters, Bayesian optimization, CatBoost, and XGBoost, combined with deep learning, significantly improves fraud detection in banking transactions.
(Domingues, Filippone, Michiardi, & Zouaoui, 2018)	A comparative evaluation of outlier detection algorithms: Experiments and analyses	Outlier detection algorithms from various fields perform well in fraud detection, intrusion detection, medical diagnoses, and data cleaning, with varying scalability, memory consumption, and robustness.
(West & Bhattacharya, 2016)	Intelligent financial fraud detection: A comprehensive review	Computational intelligence-based financial fraud detection methods show promise, but a research gap exists in addressing associations between fraud types, algorithms, and their performance.

2) Machine Learning

The literature on machine learning techniques for fraud detection has expanded significantly in recent years, addressing the growing complexities of fraudulent activities across various sectors. As fraudsters evolve their strategies, researchers and practitioners have turned to machine learning and deep learning models to improve the detection and prevention of fraud. This review integrates studies from multiple domains, such as credit card fraud, e-commerce fraud, and online financial transactions, highlighting the evolution of techniques and findings in machine learning-based fraud detection.

One of the most notable studies in this domain is by Ali et al. (2022), who conducted a systematic literature review focused on financial fraud detection using machine learning. Their review revealed that machine learning techniques, particularly support vector machines (SVMs) and artificial neural networks (ANNs), are highly effective in detecting financial fraud, especially in cases of credit card fraud, which remains the most common type of financial fraud. The authors found that these machine learning models outperform traditional methods due to their ability to learn from large datasets and detect hidden patterns. However, they also noted some challenges, such as the need for better data preprocessing techniques to enhance model performance. This study provided a strong foundation for subsequent research that explored more sophisticated machine learning techniques for fraud detection across various sectors.

Following this, Xu et al. (2023) introduced deep boosting decision trees (DBDT), a model that combines the strengths of conventional decision trees with the power of deep learning techniques. Their research addressed the problem of poor interpretability often associated with deep learning models while improving fraud detection accuracy. By employing DBDT, Xu et al. (2023) demonstrated significant improvements in detection rates without compromising the model's interpretability. This is particularly important in industries like finance, where regulatory compliance demands a clear understanding of how fraud detection models make decisions. Their findings emphasized that combining traditional methods with deep learning can create models that are both powerful and interpretable, marking a significant advancement in the fraud detection literature.

In the e-commerce sector, Damayanti and Adrianto (2023) examined the effectiveness of machine learning algorithms in detecting fraudulent activities in online transactions. Their study highlighted that machine learning significantly enhances the accuracy of fraud detection in e-commerce, with random forest models performing the best in terms of accuracy. However, they identified poor data distribution as a critical challenge in this domain, which negatively affects the performance of machine learning models. The authors suggested that improving the quality of data used in training these models could further enhance their accuracy and reliability. This research not only confirmed the efficacy of machine learning in detecting ecommerce fraud but also pointed out the limitations that need to be addressed for broader application.

Wang et al. (2022) explored the integration of machine learning algorithms with quantum annealing solvers for online fraud detection, addressing the limitations of traditional machine learning methods in handling highly imbalanced datasets. They found that quantum-enhanced support vector machines (SVMs) outperform conventional machine learning methods in terms of both speed and accuracy, especially when applied to time-sensitive online fraud detection. Their research demonstrated that quantum-enhanced models are particularly useful for detecting fraud in real-time online transactions. However, the study also noted that traditional methods still perform comparably in non-time series data, indicating that quantum computing may offer limited benefits in certain types of fraud detection tasks. This work introduced quantum computing as a potential game-changer in the field of fraud detection, especially for tasks that require high computational power and real-time detection.

Similarly focusing on credit card fraud detection, Mienye and Sun (2023) proposed a hybrid feature selection technique that combines information gain and genetic algorithms to improve detection accuracy. Their approach ensures that only relevant features are selected for the machine learning model, thus improving its performance. The authors demonstrated that this hybrid method significantly enhances the detection of credit card fraud by reducing false positives and improving the model’s efficiency. Their findings suggest that feature selection is a crucial step in improving the accuracy and effectiveness of machine learning models for fraud detection, especially in domains where data is noisy or unbalanced. This study contributes to the growing body of literature that emphasizes the importance of feature engineering in building effective fraud detection models.

Li et al. (2023) introduced a novel approach using graph learning techniques in the context of internet financial fraud detection. Their method, TA-Struc2Vec, improves detection efficiency by learning both topological features and transaction amount features within financial transaction networks. This approach allows for the intelligent classification and prediction of fraudulent activities in financial networks, making it particularly effective in detecting complex fraud schemes that involve multiple entities. The use of graph learning represents a shift toward more sophisticated models that can handle the intricacies of financial fraud networks. Li et al.'s (2023) research contributes to the growing interest in using graph-based models for fraud detection, particularly in domains where fraud occurs within interconnected systems or networks.

Esenogho et al. (2022) proposed a neural network ensemble classifier combined with a hybrid data resampling method to improve credit card fraud detection. Their model achieved outstanding results, with a sensitivity and specificity of 0.996 and 0.998, respectively, outperforming other machine learning algorithms. The authors highlighted that the ensemble approach, which aggregates multiple models, enhances the robustness of fraud detection systems, particularly in handling imbalanced datasets. Their findings reinforce the value of ensemble methods in fraud detection, where multiple models can be combined to achieve superior performance. This research is aligned with other studies that advocate for the use of ensemble learning as a solution to the challenges posed by fraud detection tasks, particularly in the financial domain.

Psychoula et al. (2021) focused on explainable machine learning for fraud detection, which has become increasingly important as machine learning models are deployed in sensitive industries like finance. Their study highlighted that selecting appropriate background data and balancing supervised and unsupervised models can improve the performance of fraud detection systems while maintaining interpretability. The authors argued that for machine learning models to be widely adopted in industries like banking and finance, it is essential that these models provide clear explanations for their predictions. Their work contributes to the growing discourse on explainable AI, emphasizing that fraud detection models must not only be accurate but also interpretable to gain trust and compliance from regulatory bodies.

The work of Btoush et al. (2023) presented a systematic review of credit card cyber fraud detection using machine learning and deep learning techniques. Their findings confirmed that both machine learning and deep learning significantly improve the efficiency and accuracy of fraud detection systems. The authors also identified emerging trends, such as the increasing use of deep learning models like convolutional neural networks (CNNs) and recurrent neural networks (RNNs), which can capture complex patterns in transactional data. Their review provided a comprehensive overview of how machine

learning and deep learning techniques are being used to combat credit card fraud in cyberspace, reinforcing the importance of these technologies in modern fraud detection systems.

Finally, Fiore et al. (2017) explored the use of generative adversarial networks (GANs) to improve classification effectiveness in credit card fraud detection. They demonstrated that GANs, which are typically used for generating synthetic data, can be highly effective in mimicking the minority class in imbalanced datasets, such as fraudulent transactions. By generating synthetic fraudulent data, GANs help improve the performance of fraud detection models by addressing the class imbalance problem. Their study showed that GANs could be an effective tool for enhancing the detection of rare fraud cases, providing an innovative approach to the challenges posed by imbalanced datasets. The list of the most recent review papers in TABLE 4.

Table 4: Summary of ML methods

Authors	Proposed Study	Key Findings
(Ali et al., 2022)	Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review	Machine learning techniques, particularly support vector machines and artificial neural networks, effectively detect financial fraud, with credit card fraud being the most common type addressed.
(Xu, Wang, Liao, & Wang, 2023)	Efficient Fraud Detection using Deep Boosting Decision Trees	Deep boosting decision trees (DBDT) significantly improve fraud detection performance and maintain good interpretability while combining the advantages of conventional methods and deep learning.
(Damayanti & Adrianto, 2023)	Machine learning for e-commerce fraud detection	Machine learning significantly enhances e-commerce fraud detection accuracy, but poorer-quality data distribution remains a challenge; random forests perform best in terms of accuracy.
(H. Wang, Wang, Liu, & Alidaee, 2022)	Integrating Machine Learning Algorithms With Quantum Annealing Solvers for Online Fraud Detection	Quantum enhanced SVM outperforms traditional machine learning methods in speed and accuracy for online fraud detection in highly imbalanced data, while traditional methods perform similarly in non-time series data.
(Mienye & Sun, 2023b)	A Machine Learning Method with Hybrid Feature Selection for Improved Credit Card Fraud Detection	The hybrid feature-selection technique using information gain and genetic algorithms improves credit card fraud detection by ensuring only relevant features are used in machine learning.
(R. Li, Liu, Ma, Yang, & Sun, 2022)	Internet Financial Fraud Detection Based on Graph Learning	TA-Struc2Vec improves Internet financial fraud detection efficiency by learning topological features and transaction amount features in financial transaction network graphs, allowing intelligent classification and prediction.
(Esenogho, Mienye, Swart, Aruleba, & Obaido, 2022)	A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection	The proposed neural network ensemble classifier with hybrid data resampling method effectively detects credit card fraud with a sensitivity and specificity of 0.996 and 0.998, outperforming other algorithms.
(Psychoula et al., 2021)	Explainable Machine Learning for Fraud Detection	Machine learning methods can improve fraud detection by selecting appropriate background data sets and balancing supervised and unsupervised models.
(Btoush et al., 2023)	A systematic review of literature on credit card cyber fraud detection using machine and deep learning	Machine learning and deep learning techniques can effectively detect credit card cyber fraud, improving efficiency and accuracy.
(Fiore, De Santis, Perla, Zanetti, & Palmieri, 2019)	Using generative adversarial networks for Improving Classification Effectiveness in credit card fraud detection	Generative Adversarial Networks (GANs) can improve the effectiveness of credit card fraud detection by mimicking minority classes in imbalanced datasets.

C. DEEP LEARNING

Deep learning techniques have been used in various fields for predictions and found to be giving highly accurate results. There were some challenges with DL such as a need for heavy computing, and more data, however, in today's world these challenges are no longer a big issue. This section describes the existing literature focused on usage of DL methods for detecting money laundering.

1) CNN

Convolutional Neural Networks (CNNs) have been widely adopted across various domains for fraud detection and beyond, thanks to their ability to capture complex patterns in data. This review explores key advancements in CNN approaches, focusing on their application in fraud detection, financial analytics, and beyond, as well as addressing their limitations, potential, and regularization methods.

Cheng et al. (2022) proposed a Graph Neural Network for Fraud Detection using spatial-temporal attention, which outperformed other state-of-the-art methods in detecting fraudulent transactions, uncovering fraud patterns, and mining fraud hotspots in credit card data. Their spatial-temporal attention-based graph network (STAGN) demonstrated superior performance in identifying suspicious activity by considering both spatial and temporal dimensions, making it a critical advancement in using graph-based CNNs for real-time fraud detection.

Zhang et al. (2018) introduced a CNN model for online transaction fraud detection, which showed substantial improvements in precision and recall—26% and 2%, respectively—compared to traditional models. This CNN approach avoids generating derivative features, which often lead to unnecessary complexity, and instead focuses on intrinsic feature extraction, enhancing the detection efficiency of online fraud in transaction data. This innovation is a testament to CNN's ability to detect real-time fraud with minimal feature engineering.

Fu et al. (2016) built a CNN-based framework for credit card fraud detection, showcasing its effectiveness in capturing intrinsic fraud patterns within credit card transaction data. The framework outperformed state-of-the-art methods in realworld settings, making it a robust and scalable solution for financial institutions seeking to improve fraud detection accuracy. This study highlights how CNNs, through their pattern recognition capabilities, are well-suited to the complexities of fraud detection in financial transactions.

Similarly, Gambo et al. (2022) presented a CNN model specifically designed for credit card fraud detection, achieving high accuracy, precision, and recall when compared to existing models. Their work illustrates how CNN architectures can be tailored to specific fraud detection tasks, offering improved performance by leveraging deep learning's ability to learn from vast datasets. This model demonstrated strong adaptability to the unique challenges of credit card fraud detection, a task that involves balancing detection efficiency with minimizing false positives.

In addition to their direct applications in fraud detection, CNNs have been the focus of broader research aimed at improving their general performance across different domains. For example, Li et al. (2020) conducted a comprehensive review of CNNs, analyzing their applications and prospects across multiple fields such as computer vision and natural language processing. They provided insights into 1-D, 2-D, and multidimensional convolutions, paving the way for more advanced CNN architectures that can address increasingly complex tasks in the future.

Regularization, a critical factor in enhancing CNN performance, has also been the subject of thorough investigation. Santos and Papa (2022) conducted a survey on regularization methods for CNNs, focusing on how data augmentation, internal changes, and label transformations can help avoid overfitting, a common problem in machine learning models. Their study provided key strategies for improving the generalizability and robustness of CNNs, ensuring that they perform well not only on training data but also on unseen, real-world data, which is particularly important in fraud detection tasks.

Kiranyaz et al. (2019) provided a survey on 1D Convolutional Neural Networks and their applications, emphasizing their real-time, low-cost solutions in various industries, including engineering. Their work demonstrated that 1D CNNs could offer state-of-the-art performance in a range of applications, showing great potential for scaling CNN technology across domains where data is structured in one dimension, such as time series analysis in fraud detection or signal processing.

In another financial application, Hosaka (2019) applied CNNs to bankruptcy prediction, using financial ratios as input. The CNNs, trained on the GoogLeNet architecture, significantly outperformed traditional methods like decision trees and linear discriminant analysis, demonstrating the power of CNNs in financial forecasting tasks. This shows that CNNs are not limited to fraud detection but can also be leveraged for predicting financial outcomes such as corporate bankruptcy.

A broader meta-analysis by Ghanbari et al. (2021) on CNNs in remote sensing applications highlighted that while CNNs have significantly improved performance in this domain, there are still challenges to be addressed. Their findings are relevant to financial fraud detection, as remote sensing techniques often deal with similarly large, complex datasets, indicating that improvements in CNNs for other domains could have cross-disciplinary applications.

Finally, Gu et al. (2015) reviewed recent advances in CNNs, showing how innovations in this architecture have led to breakthroughs in tasks like visual recognition, speech recognition, and natural language processing. These advances are relevant to the continuous improvement of CNNs in fraud detection, where the ability to process vast amounts of structured and unstructured data can lead to better detection algorithms capable of handling increasingly sophisticated fraud schemes.

Table 5: Summary of CNN methods

Authors	**Proposed Study**	**Key Findings**
(Cheng, Wang, Zhang, & Zhang, 2020)	Graph Neural Network for Fraud Detection via Spatial-Temporal Attention	The spatial-temporal attention-based graph network (STAGN) outperforms other state-of-the-art methods in credit card fraud detection, detecting suspicious transactions, mining fraud hotspots, and uncovering fraud patterns.
(Zhang, Zhou, Zhang, Wang, & Wang, 2018)	A Model Based on Convolutional Neural Network for Online Transaction Fraud Detection	The proposed convolutional neural network model effectively detects online transaction fraud without generating derivative features, improving precision and recall by 26% and 2% compared to existing models.
(Fu, Cheng, Tu, & Zhang, 2016)	Credit Card Fraud Detection Using Convolutional Neural Networks	Our CNN-based fraud detection framework effectively captures intrinsic patterns of fraud behaviors in credit card transactions, outperforming state-of-the-art methods in realworld transactions.
(Gambo, Zainal, & Kassim, 2022)	A Convolutional Neural Network Model for Credit Card Fraud Detection	The proposed Convolutional Neural Network model effectively detects credit card fraud with high accuracy, precision, and recall compared to existing studies.

(Z. Li, Liu, Yang, Peng, & Zhou, 2021)	A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects	This review provides novel ideas and perspectives for convolutional neural networks, covering 1-D, 2-D, and multidimensional convolutions, and explores future work in computer vision and natural language processing.
(Santos & Papa, 2022)	Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks	Regularization methods for Convolutional Neural Networks can significantly improve performance in image processing tasks, with three main areas: data augmentation, internal changes, and label transformations.
(Kiranyaz et al., 2021)	1D Convolutional Neural Networks and Applications: A Survey	1D CNNs have shown state-of-the-art performance in various engineering applications, offering real-time and lowcost solutions for various industries.
(Ghanbari, Mahdianpari, Homayouni, & Mohammadimanesh, 2021)	A Meta-Analysis of Convolutional Neural Networks for Remote Sensing Applications	Convolutional neural networks (CNNs) have significantly improved remote sensing applications, but further advancements are needed to address critical issues and challenges.
(Hosaka, 2019)	Bankruptcy prediction using imaged financial ratios and convolutional neural networks	Convolutional neural networks, trained on GoogLeNet, can effectively predict corporate bankruptcy using financial ratios, outperforming methods like decision trees and linear discriminant analysis.
(Gu et al., 2018)	Recent advances in convolutional neural networks	Recent advances in convolutional neural networks have led to improved performance in various tasks, including visual recognition, speech recognition, and natural language processing.

2) LSTM

Long Short-Term Memory (LSTM) networks have shown significant promise in detecting fraudulent activities by modeling time sequences and capturing temporal dependencies. This review focuses on various applications of LSTM-based approaches in fraud detection, demonstrating how LSTM networks can enhance the accuracy and efficiency of detecting fraud in financial transactions, marketing strategies, and more.

Esenogho et al. (2022) presented a Neural Network Ensemble with Feature Engineering for credit card fraud detection, which incorporates hybrid data resampling techniques. Their model, which achieved a sensitivity of 0.996 and specificity of 0.998, outperformed other algorithms, demonstrating how feature engineering and ensemble learning can enhance the effectiveness of LSTM-based fraud detection methods. The results underscore the potential of integrating LSTMs with resampling techniques to handle the complexities of credit card fraud data.

In a similar study, Mienye and Sun (2023) proposed a Deep Learning Ensemble With Data Resampling, which used the SMOTE-ENN method to improve credit card fraud detection. Their model achieved a sensitivity of 1.000 and specificity of 0.997, outperforming other machine learning methods. This highlights the effectiveness of LSTM-based deep learning ensembles in improving detection accuracy, particularly when combined with advanced data resampling techniques to address class imbalance, a common challenge in fraud detection.

Benchaji et al. (2021) introduced a Credit Card Fraud Detection Model Based on LSTM Recurrent Neural Networks, which effectively captures historical purchase behavior. By leveraging the sequential nature of transaction data, the LSTM model improves fraud detection accuracy on new transactions. This approach demonstrates the strength of LSTM networks in learning from time-series data, allowing them to adapt to changing shopping patterns and identify anomalies in real-time. Jan (2021) explored Financial Statement Fraud Detection Using Deep Learning, where LSTMs outperformed traditional Recurrent Neural Networks (RNNs), achieving an accuracy of 94.88% in detecting financial statement fraud in listed companies. The study highlights how LSTMs can outperform conventional models by effectively modeling temporal dependencies in financial data, making them particularly suited for fraud detection in corporate environments where historical trends are critical.

Zhan (2020) proposed an LSTM-Focalloss neural network model for bank fraud detection, which improved the
detection of fraudulent transactions more effectively than other methods. This model, which incorporates focal loss to address class imbalance, reduced financial losses for both consumers and banks. By combining LSTM with specialized loss functions, the model can better handle highly imbalanced datasets, which is a frequent issue in fraud detection tasks.

Bouzidi et al. (2021) focused on LSTM-Based Automated Learning for marketing fraud detection and financial forecasting. Their LSTM model improved fraud detection by localizing and recognizing different aspects of corporate marketing strategies, all while using fewer parameters and less computational power. This approach demonstrates how LSTMs can be optimized for resource-constrained environments while maintaining high performance in detecting fraud.

Jurgovsky et al. (2018) highlighted Sequence Classification for Credit-Card Fraud Detection, where LSTM networks improved detection accuracy, particularly in offline transactions. They found that manual feature aggregation strategies enhanced both sequential and non-sequential learning,

emphasizing the role of LSTM in handling the temporal aspects of credit card fraud detection. The study confirms that LSTMs excel in tasks where the order of transactions is important, as they can learn the sequential patterns of legitimate versus fraudulent activities.

Bindu et al. (2023) introduced a Credit Card Fraud Detection System Using LSTM and SMOTE-ENN, which combines LSTM deep recurrent neural networks with SMOTE-ENN resampling to effectively capture historical purchase behavior. This approach improves fraud detection accuracy on incoming transactions, further demonstrating the synergy between LSTM models and advanced resampling techniques to address the challenges posed by imbalanced datasets in fraud detection.

Wiese and Omlin (2009) discussed Modelling Time with LSTM Recurrent Neural Networks in credit card transactions. Their system adapts to changing shopping profiles of legitimate cardholders, improving fraud detection by learning temporal patterns in transaction sequences. This early study showcased the potential of LSTMs to model time-based features in fraud detection, laying the groundwork for more recent advances in sequential data processing.

Finally, Mayaki and Riveill (2022) proposed Multiple Inputs Neural Networks for Medicare Fraud Detection, which use LSTM autoencoders to outperform baseline models in detecting Medicare fraud. Their LSTM-based approach showed robustness to class imbalance, further illustrating how LSTM networks can be applied beyond financial fraud to other areas like healthcare fraud detection.

Table 6: Summary of LSTM methods

Authors	**Proposed Study**	**Description**
(Esenogho et al., 2022)	A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection	The proposed neural network ensemble classifier with hybrid data resampling method effectively detects credit card fraud with a sensitivity and specificity of 0.996 and 0.998, outperforming other algorithms.
(Mienye & Sun, 2023a)	A Deep Learning Ensemble with Data Resampling for Credit Card Fraud Detection	The proposed deep learning ensemble with SMOTEENN method achieves a sensitivity and specificity of 1.000 and 0.997, outperforming other widely used machine learning methods in credit card fraud detection.
(Benchaji, Douzi, & El Ouahidi, 2021)	Credit Card Fraud Detection Model Based on LSTM Recurrent Neural Networks	Our credit card fraud detection model using LSTM networks effectively captures historic purchase behavior, improving fraud detection accuracy on new incoming transactions.
(Jan, 2021)	Detection of Financial Statement Fraud Using Deep Learning for Sustainable Development of Capital Markets under Information Asymmetry	The LSTM model outperforms the RNN model in detecting financial statement fraud, with an accuracy of 94.88% in TWSE/TEPx listed companies.
(Zhan, 2020)	Research on bank fraud transaction detection based on LSTM-Focalloss	LSTM-Focalloss neural network model effectively detects bank fraud transactions more effectively than other methods, reducing losses to consumers and banks.
(Bouzidi, Boudries, & Amad, 2021)	LSTM-based automated learning with smart data to improve marketing fraud detection and financial forecasting	The proposed LSTM-based model improves marketing fraud detection and financial forecasting by learning to localize and recognize different aspects of corporate marketing and business strategies using fewer parameters and less computation.
(Jurgovsky et al., 2018)	Sequence classification for credit-card fraud detection	LSTM networks improve credit-card fraud detection accuracy, particularly on offline transactions, and manual feature aggregation strategies enhance both sequential and non-sequential learning approaches.
(Benchaji et al., 2021)	Credit Card Fraud Detection Using LSTM	Our novel system for credit card fraud detection using LSTM deep Recurrent Neural Networks and SMOTEENN effectively captures historic purchase behavior, improving fraud detection accuracy on new incoming transactions.
(Wiese & Omlin, 2009)	Credit Card Transactions, Fraud Detection, and Machine Learning: Modelling Time with LSTM Recurrent Neural Networks	LSTM recurrent neural networks can effectively model time in credit card transaction sequences, improving fraud detection by adapting to changing shopping profiles of legitimate card holders.

(Mayaki & Riveill, 2022)	Multiple Inputs Neural Networks for Medicare fraud Detection	Multiple inputs neural networks with LSTM autoencoders outperform baseline models in detecting Medicare fraud, making them more robust to class imbalance.

3) MLP

Multilayer Perceptron (MLP) approaches have been widely applied in fraud detection due to their ability to model complex relationships in large datasets. This review explores how MLP and related neural network architectures enhance fraud detection, from tax fraud and credit card fraud to financial fraud in cryptocurrency networks.

López et al. (2019) introduced Tax Fraud Detection Through Neural Networks, where neural networks, including MLP, were used to detect tax fraud in personal income tax returns with an efficiency rate of 84.3%. This method outperformed other traditional models, demonstrating how MLP networks can capture hidden patterns in tax data to improve fraud detection accuracy.

Anowar and Sadaoui (2020) presented an Incremental Neural-Network Learning for Big Fraud Data, where an MLPbased incremental learning approach was applied to large-scale credit card data. This model effectively detected fraud by updating itself with new data without retraining from scratch, significantly outperforming non-incremental methods. The study highlights MLP's scalability and adaptability in big data environments, where fraud detection models must handle continuously growing datasets.

Montini et al. (2013) proposed a Sampling Diagnostics Model for Neural System Training Optimization for bank fraud detection. The model integrates hybrid sampling techniques to reduce database size and optimize MLP training. By combining MLP networks with sampling methods, this approach reduced computational cost while maintaining accuracy, making it practical for large financial datasets.

In a case study on the Tehran Stock Exchange**,** Mohammadi and Faramarzi (2016) found that MLP neural networks and radial basis function networks effectively predicted fraud. MLP achieved 936% accuracy, demonstrating its strength in handling financial fraud detection in stock market environments. This case underscores the robustness of MLP in detecting patterns and anomalies in financial systems with complex transactional behaviors.

Menshchikov et al. (2022) conducted a Comparative Analysis of Machine Learning Methods for Financial Fraud Detection and found that MLP was the most suitable algorithm for financial fraud detection in big data systems. Their study highlighted MLP's high classification performance, minimal training, and prediction time, making it the optimal choice for real-time fraud detection in large-scale financial systems.

El hlouli et al. (2020) explored Credit Card Fraud Detection Based on MLP and Extreme Learning Machine (ELM) Architectures, comparing the two models in terms of accuracy and speed. While both architectures effectively detected credit card fraud, ELM outperformed MLP in predicting new fraudulent transactions, proving faster and more accurate. This comparison highlights the strengths and limitations of MLP relative to other architectures in time-sensitive fraud detection scenarios.

Mienye and Sun (2023) introduced a Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection, combining MLP and SMOTE-ENN resampling techniques. Their ensemble achieved 1.000 sensitivity and 0.997 specificity, outperforming other machine learning methods in credit card fraud detection. This approach demonstrates how MLP can be enhanced with resampling techniques to improve performance on imbalanced datasets.

Ghosh et al. (2023) applied Ensemble Deep Learning Models for Fraud Detection in Bitcoin **Networks**, using MLP, Feedforward Neural Network, and Attention LSTM to achieve 99.62% accuracy in detecting fraud. MLP played a crucial role in this ensemble, demonstrating its effectiveness in identifying fraudulent activities in complex cryptocurrency transactions, which often involve multiple data streams and evolving patterns.

Ali et al. (2022), in their Systematic Literature Review on Financial Fraud Detection, noted that artificial neural networks, including MLP, are among the most effective machine learning techniques for detecting financial fraud, with credit card fraud being the most commonly addressed. The study provides further validation of MLP's widespread applicability in different financial fraud contexts.

Moumeni et al. (2021) focused on Machine Learning for Credit Card Fraud Detection, where MLP, along with logistic regression (LR) and principal component analysis (PCA), showed promising performance in detecting fraud using real datasets from an American bank. Their research illustrates MLP's flexibility in integrating with other algorithms to improve credit card fraud detection.

Table 7: Summary of MLP methods

Authors	**Proposed Study**	**Key Findings**
(Pérez López, Delgado Rodríguez, & de Lucas Santos, 2019)	Tax Fraud Detection through Neural Networks: An Application Using a Sample of Personal Income Taxpayers	Neural networks effectively detect tax fraud in personal income tax returns, with an efficiency rate of 84.3%, improving tax fraud detection compared to other models.
(Anowar & Sadaoui, 2020)	Incremental Neural-Network Learning for Big Fraud Data	Our incremental neural-network learning approach effectively detects fraud in large-scale credit-card data, outperforming non-incremental methods.
(Montini et al., 2013)	A Sampling Diagnostics Model for Neural System Training Optimization	The hybrid-sampling model for bank fraud diagnosis reduces database volume and improves training results, presenting similar precisions to traditional methods.
(Mohammadi & Faramarzi, 2016)	The Application of Neural Networks to Predict Fraud: Case Study of Tehran Stock Exchange	Neural networks effectively predict fraud in the Tehran Stock Exchange with 936% and 917% accuracy using Multilayer Perceptron network and radial basis function network, respectively.

(Menshchikov, Perfilev, Roenko, Zykin, & Fedosenko, 2022)	Comparative Analysis of Machine Learning Methods Application for Financial Fraud Detection	Multilayer perceptron (MLP) is the most suitable machine learning algorithm for financial fraud detection in big data systems, with high classification performance and minimal training and prediction time.
(Riffi, Mahraz, El Yahyaouy, & Tairi, 2020)	Credit Card Fraud Detection Based on Multilayer Perceptron and Extreme Learning Machine Architectures	Both Multilayer Perceptron and Extreme Learning Machine classifiers effectively detect credit card fraud, with ELM being faster and more accurate in predicting new fraudulent transactions.
(Mienye & Sun, 2023a)	A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection	The proposed deep learning ensemble with SMOTEENN method achieves a sensitivity and specificity of 1.000 and 0.997, outperforming other widely used machine learning methods in credit card fraud detection.
(Ghosh, Chowdhury, Das, & Sadhukhan, 2023)	Enhancing Financial Fraud Detection in Bitcoin Networks Using Ensemble Deep Learning	Ensemble deep learning models using Multi-Layer Perceptron, Feedforward Neural Network, and Attention LSTM effectively detect financial fraud in Bitcoin networks with 99.62% accuracy and over 99% precision and recall.
(Ali et al., 2022)	Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review	Machine learning techniques, particularly support vector machines and artificial neural networks, effectively detect financial fraud, with credit card fraud being the most common type addressed.
(Sailusha, Gnaneswar, Ramesh, & Rao, 2020)	Machine Learning for Credit Card Fraud Detection	MLP, LR, and PCA algorithms show promising performance in detecting credit card fraud using real datasets from an American bank.

4) AUTOENCODERS

Autoencoders have emerged as a powerful technique for fraud detection, particularly in credit card transactions, by reconstructing data and identifying anomalies that may indicate fraudulent behavior. This review highlights various autoencoder approaches applied to financial and medical fraud detection.
Ding et al. (2023) introduced a Credit Card Fraud Detection Model Based on Improved Variational Autoencoder Generative Adversarial Network (VAEGAN), where an oversampling method enhanced the detection of fraudulent transactions. The model outperformed other fraud detection methods in terms of precision, F1-score, and other key indicators, proving the effectiveness of autoencoders when integrated with adversarial networks for handling imbalanced datasets.
Chen et al. (2023) proposed a Variational AutoEncoder-Based Relational Model (VAERM) for Cost-Effective Medical Fraud Detection. This model not only detects fraudulent medical claims but also aids in fraud investigation, outperforming state-of-the-art methods in both automatic detection and human-led investigations. The study demonstrates how autoencoders can generalize beyond financial fraud, applying to the healthcare industry where relational data is crucial.
Mitra et al. (2022) implemented Autoencoders for Credit Card Fraud Detection using real-time data from European cardholders, significantly improving detection accuracy for banks and insurance companies. This practical approach demonstrates how autoencoders can effectively handle real-time transactional data, enhancing the detection of new fraud patterns.
Du et al. (2023) developed the AutoEncoder and LightGBM (AED-LGB) method for credit card fraud detection, which significantly outperformed traditional machine learning models. The combined use of autoencoders and LightGBM (a gradient boosting framework) improved accuracy, true positive rate, true negative rate, and Matthew's correlation coefficient, highlighting the synergy between autoencoders and advanced machine learning models.
Putrada and Ramadhan (2023) introduced the MDIASE-Autoencoder, a novel anomaly detection method that significantly enhanced the performance of credit card fraud detection models. By combining autoencoders with MLP networks, their model increased the AUC from 0.558 to 0.999, demonstrating the potential of autoencoders to improve detection performance when integrated with other architectures.
Chaquet-Ulldemolins et al. (2022) addressed the Black-Box Challenge for Fraud Detection by using Interpretable Autoencoders. This model provided traceability between inputs and outputs, ensuring compliance with financial regulations and enabling transaction-level inquiries. The interpretability of the autoencoder added transparency to credit card fraud detection systems, an important factor for regulatory adherence and customer trust.
Misra et al. (2020) presented an Autoencoder-Based Two-Stage Model for credit card fraud detection, where the autoencoder was followed by a classifier. This two-stage approach outperformed single-stage systems and other autoencoder models, enhancing the ability to detect fraudulent transactions with greater precision.
Lin and Jiang (2021) proposed the Autoencoder and Probabilistic Random Forest (AE-PRF) method, which
effectively detected credit card frauds, particularly in imbalanced datasets. The integration of a probabilistic random forest model with autoencoders improved fraud detection performance, especially when dealing with skewed data distributions where fraudulent cases are rare.

Zamini and Montazer (2018) developed an Autoencoder-Based Clustering Method for credit card fraud detection, achieving 98.9% accuracy and 81% true positive rate (TPR). This approach demonstrated that autoencoders can effectively capture patterns in clustered data, outperforming other clustering methods in fraud detection.

Mohammad et al. (2023) introduced a novel Autoencoder and Decoder Machine Learning (ML) method for credit card fraud detection. Their unsupervised model detected 284,807 fraudulent transactions from a European dataset, demonstrating the efficiency of autoencoders in processing large-scale transactional data without the need for labelled training data.

Table 8: Summary of AE methods

Authors	Proposed Study	Key Findings
(Ding, Kang, Feng, Peng, & Yang, 2023)	Credit Card Fraud Detection Based on Improved Variational Autoencoder Generative Adversarial Network	The improved VAEGAN oversampling method effectively detects credit card fraud, outperforming other methods in terms of precision, F1_score, and other indicators.
(Chen, Hu, Yi, Alazab, & Li, 2022)	A Variational AutoEncoderBased Relational Model for Cost-Effective Automatic Medical Fraud Detection	The Variational AutoEncoder-based Relational Model (VAERM) effectively detects and investigates medical fraud, outperforming state-of-the-art methods in both automatic and fraud investigation tasks.
(Zou, Zhang, & Jiang, 2019)	Credit Card Fraud Detection using Autoencoders	Autoencoders effectively detect credit card fraud using real-time data from European card holders, improving fraud detection accuracy in banks and insurance companies.
(Du, Lv, Guo, & Wang, 2023)	AutoEncoder and LightGBM for Credit Card Fraud Detection Problems	The AED-LGB algorithm outperforms other machine learning methods in detecting credit card frauds, with higher accuracy, true positive rate, true negative rate, and Matthew's correlation coefficient.
(Putrada & Ramadhan, 2023)	MDIASE-Autoencoder: A Novel Anomaly Detection Method for Increasing The Performance of Credit Card Fraud Detection Models	The MDIASE-autoencoder significantly improves the performance of credit card fraud detection models, particularly MLP, with an AUC increase from 0.558 to 0.999.
(Chaquet-Ulldemolins, Gimeno-Blanes, MoralRubio, Muñoz-Romero, & Rojo-Álvarez, 2022)	On the Black-Box Challenge for Fraud Detection Using Machine Learning (II):	This paper proposes an interpretable and agnostic methodology for credit fraud detection, improving accuracy and enabling traceability between inputs and
	Nonlinear Analysis through Interpretable Autoencoders	outputs, enabling compliance with regulations and transaction-level inquiries.
(Misra, Thakur, Ghosh, & Saha, 2020)	An Autoencoder Based Model for Detecting Fraudulent Credit Card Transaction	The two-stage model using an autoencoder and classifier effectively detects fraudulent credit card transactions, outperforming single-stage systems and other autoencoder-based models.
(Lin & Jiang, 2021)	Credit Card Fraud Detection with Autoencoder and Probabilistic Random Forest	The AE-PRF method effectively detects credit card frauds, with excellent performance even in imbalanced datasets.
(Zamini & Montazer, 2018)	Credit Card Fraud Detection using autoencoder based clustering	This autoencoder-based clustering method accurately detects credit card fraud with 98.9% accuracy and 81% TPR, outperforming other methods.
(Al-Shabi, 2019)	Credit Card Fraud Detection Using Autoencoder & Decoder ML	This project developed an unsupervised credit card fraud detection method using autoencoders, which effectively identified 284,807 fraudulent transactions from a European dataset.

5) GNN

Graph Neural Networks (GNNs) have gained significant attention for fraud detection due to their ability to model complex relationships and interactions in graph-structured data. These approaches focus on improving accuracy, handling large-scale data, and enhancing model interpretability. Below is a summary of various GNN approaches applied to fraud detection tasks.

Liu et al. (2022) introduced the Hierarchical Attention-based Graph Neural Network (HAGNN), which improves fraud detection by incorporating both relation and neighborhood attention modules. This approach is particularly effective at detecting camouflaged nodes, which are often missed by traditional methods, making it highly efficient for fraud detection in dynamic environments.

Cheng et al. (2022) proposed the Graph Neural Network with Spatial-Temporal Attention (STAGN), which outperforms other models in credit card fraud detection. This method detects suspicious transactions by mining fraud hotspots and uncovering underlying fraud patterns, showcasing the importance of incorporating temporal and spatial features in graph networks.

Li et al. (2022) introduced the Dual-Augment Graph Neural Network (DAGNN), which focuses on augmenting disparities between target nodes and their heterogeneous neighbors while reinforcing similarities with homogenous neighbors. This method improves fraud detection accuracy, especially in complex, heterogeneous networks where traditional approaches may struggle.

Liu et al. (2020) developed GraphConsis, a GNN framework that addresses the inconsistency problem in fraud detection tasks by combining context embeddings with node features. This approach resolves feature inconsistencies and enhances relation attention weights, offering more accurate and robust fraud detection.

Zeng and Tang (2021) presented RLC-GNN, a spatial-based GNN architecture for fraud detection that improves performance by learning layer by layer and continuously correcting errors. This method outperforms single-layer CAREGNN models, particularly in detecting fraud on platforms like Yelp and Amazon, where fraudsters can camouflage their activities.

Lu et al. (2022) developed BRIGHT, a real-time fraud detection framework using GNNs. This system efficiently detects fraud in e-commerce platforms, reducing latency by over 75% compared to baseline models. Its success highlights the importance of real-time processing in dynamic environments like online retail.

Pereira and Murai (2021) examined how effective GNNs are for fraud detection in large-scale network data. While GNNs show potential, challenges such as class imbalance and concept drift remain critical obstacles to achieving consistent performance across different datasets.

Liu et al. (2021) introduced a Graph Neural Network and Transaction Graph Model specifically for credit card fraud detection. This model effectively represents transaction characteristics, reducing interference from abnormal samples and improving overall detection accuracy.

Qin et al. (2022) proposed the Neural Meta-Graph Search (NGS), which improves graph-based fraud detection while maintaining model explainability. By using meta-graphs to search for fraud patterns, NGS outperforms state-of-the-art methods and ensures transparency, an essential feature for regulatory compliance and user trust.

Dou et al. (2020) presented the CARE-GNN, which focuses on detecting camouflaged fraudsters by addressing feature and relation camouflages in graph networks. CARE-GNN outperforms other GNN-based fraud detection models, proving especially useful in environments where fraudsters actively attempt to conceal their actions.

Table 9: summary of GNN methods

Authors	**Proposed Study**	**Key Findings**
(Y. Liu, Sun, & Zhang, 2023)	Improving Fraud Detection via Hierarchical Attentionbased Graph Neural Network	The Hierarchical Attention-based Graph Neural Network (HAGNN) effectively detects fraud by incorporating relation and neighborhood attention modules, improving detection of camouflaged nodes.
(Cheng et al., 2020)	Graph Neural Network for Fraud Detection via Spatial-Temporal Attention	The spatial-temporal attention-based graph network (STAGN) outperforms other state-of-the-art methods in credit card fraud detection, detecting suspicious transactions, mining fraud hotspots, and uncovering fraud patterns.
(Q. Li et al., 2022)	Dual-Augment Graph Neural Network for Fraud Detection	The Dual-Augment Graph Neural Network (DAGNN) outperforms state-of-the-art models in fraud detection tasks by augmenting disparity between target node and heterogenous neighbors and similarity between target node and homogenous neighbors.
(Z. Liu, Dou, Yu, Deng, & Peng, 2020)	Alleviating the Inconsistency Problem of Applying Graph Neural Network to Fraud Detection	GraphConsis, a new GNN framework, effectively addresses inconsistencies in fraud detection tasks by combining context embeddings with node features, addressing feature inconsistencies, and learning relation attention weights.
(Zeng & Tang, 2021)	RLC-GNN: An Improved Deep Architecture for Spatial-Based Graph Neural Network with Application to Fraud Detection	The RLC-GNN improves fraud detection performance by learning layer by layer and continuously correcting mistakes, outperforming single-layer CARE-GNNs on Yelp and Amazon datasets.

(Lu et al., 2022)	BRIGHT - Graph Neural Networks in Real-time Fraud Detection	The BRIGHT framework efficiently enables real-time fraud detection in e-commerce by outperforming baseline models and reducing latency by over 75%.
(Pereira & Murai, 2021)	How effective are Graph Neural Networks in Fraud Detection for Network Data?	Graph Neural Networks show promise in detecting financial fraud in large-scale data, but face challenges like imbalance between positive and negative classes and concept drift.
(G. Liu, Tang, Tian, & Wang, 2021)	Graph Neural Network for Credit Card Fraud Detection	Our Graph Neural Network and Transaction Graph model effectively detect credit card fraud by comprehensively representing transaction characters and reducing interference from abnormal samples.
(Qin, Liu, He, & Ao, 2022)	Explainable Graphbased Fraud Detection via Neural Meta-graph Search	NGS (Neural meta-Graph Search) improves graphbased fraud detection performance while maintaining model explainability, outperforming state-of-the-art methods on real-world datasets.
(Dou et al., 2020)	Enhancing Graph Neural Network-based Fraud Detectors against Camouflaged Fraudsters	The CARE-GNN model effectively detects fraud by addressing feature and relation camouflages, outperforming state-of-the-art GNN-based fraud detectors.

C. XAI

Explainable AI (XAI) techniques have become essential for enhancing trust and transparency in fraud detection models, enabling both users and regulatory bodies to understand the reasoning behind predictions. The approaches highlighted here offer diverse methods for improving explainability while maintaining high detection accuracy.

Cirqueira et al. (2020) discuss scenario-based requirements elicitation for user-centric explainable AI (XAI) in fraud detection. By tailoring explanations to stakeholder needs, this approach builds trust among experts in the banking sector, helping to make complex AI systems more accessible and transparent. Vyas (2023) emphasizes the role of Java-based AI systems in fraud detection and prevention in financial and e-commerce sectors, where large-scale data analysis and suspicious pattern detection improve real-time detection rates.

Psychoula et al. (2021) present a comprehensive overview of explainable machine learning techniques for fraud detection. By carefully selecting background datasets and balancing supervised and unsupervised models, they highlight how explainability can be integrated into fraud detection models without sacrificing performance. Xu et al. (2023) introduce deep boosting decision trees (DBDT), which significantly improve fraud detection performance while maintaining a balance between interpretability and accuracy, combining deep learning with conventional decision tree approaches.

In an evaluation of post-hoc explanation methods, Jesus et al. (2021) reveal that popular XAI tools such as LIME, SHAP, and TreeInterpreter, while widely used, may not be the best option for decision-making tasks. Their findings indicate that simpler data-driven explanations may yield more accurate results, prompting further evaluation of current XAI techniques. Fawcett et al. (1998) also explore AI approaches for fraud detection and risk management, showcasing how these techniques have evolved over time to better detect fraud and manage risk across multiple application domains.

Mill et al. (2023) propose an XAI research agenda for real-time credit card fraud detection, addressing the growing regulatory requirements and the need for transparency in financial transactions. Wu and Wang (2021) present a locally interpretable anomaly detection framework for credit card fraud, outperforming existing methods while providing clear insights into feature importance and transaction behavior.

Finally, Gupta (2023) highlights the benefits of machine learning and artificial intelligence in fraud prevention, where largescale data processing, pattern recognition, and anomaly detection play pivotal roles in identifying fraudulent activities and preventing financial losses. These XAI approaches demonstrate the evolving focus on making fraud detection more transparent and interpretable while maintaining high accuracy and adaptability to various industries.

Table 5: summary of XAI methods

Authors	**Proposed Study**	**Key Foundings**
(Cirqueira, Nedbal, Helfert, & Bezbradica, 2020)	Scenario-Based Requirements Elicitation for User-Centric Explainable AI - A Case in Fraud Detection	Scenario-based requirements elicitation for explainable AI in fraud detection can enhance stakeholder trust and help tailor explanations to their needs, benefiting experts in banking fraud.
(Cirqueira et al., 2020)	Java in Action: AI for Fraud Detection and Prevention	Java-based AI systems can effectively detect and prevent fraud in financial and e-commerce sectors by analyzing large amounts of data and detecting suspicious patterns.

(Psychoula et al., 2021)	Explainable Machine Learning for Fraud Detection	Machine learning methods can improve fraud detection by selecting appropriate background data sets and balancing supervised and unsupervised models.
(Xu et al., 2023)	Efficient Fraud Detection using Deep Boosting Decision Trees	Deep boosting decision trees (DBDT) significantly improve fraud detection performance and maintain good interpretability while combining the advantages of conventional methods and deep learning.
(Jesus et al., 2021)	How can I choose an explainer? An applicationgrounded Evaluation of Posthoc Explanations	The XAI Test reveals that popular explainable AI methods, like LIME, SHAP, and TreeInterpreter, generally have worse impact on decisionmaking tasks than desired, with Data Only being the most accurate and least preferred option.
(Fawcett, Haimowitz, Provost, & Stolfo, 1998)	AI Approaches to Fraud Detection and Risk Management	AI techniques can effectively detect fraud, computer intrusions, and risk scoring in various application domains.
(Mill, Garn, Ryman-Tubb, & Turner, 2023)	Opportunities in Real Time Fraud Detection: An Explainable Artificial Intelligence (XAI) Research Agenda	This paper advocates for the adoption of Explainable Artificial Intelligence (XAI) techniques for credit card fraud detection, highlighting regulatory changes and the operating environment for transactions.
(Wu & Wang, 2021)	Locally Interpretable OneClass Anomaly Detection for Credit Card Fraud Detection	Our novel anomaly detection framework for credit card fraud detection outperforms existing methods and provides clear explanations for feature importance in transaction data.
(Gupta, 2023)	Leveraging Machine Learning and Artificial Intelligence for Fraud Prevention	Machine learning and artificial intelligence significantly enhance fraud prevention efforts by effectively analyzing massive volumes of data, identifying patterns, and detecting abnormal behaviors.

D. FEDERATED LEARNING

Federated Learning (FL) has proven to be a valuable approach for enhancing fraud detection across various domains, providing improved performance and privacy-preserving features. Here's a summary of key approaches and their findings: Ferrari et al. (2020) introduced federated meta-learning for credit card fraud detection, which leverages local data from multiple banks. This approach significantly enhances detection performance compared to traditional methods, highlighting the benefits of decentralized learning.

Yang et al. (2019) developed FFD, a federated learning-based method for credit card fraud detection. This method achieves an average test AUC of 95.5%, demonstrating superior performance over traditional fraud detection systems. Talluri et al. (2023) proposed a cloud-native federated learning architecture for telecom fraud detection, which improves detection accuracy by up to 23% compared to locally trained models, showcasing the effectiveness of cloud-based federated learning solutions.

Myalil et al. (2021) addressed the challenges of detecting fraudulent transactions in non-IID settings with active adversaries. Their robust collaborative fraud detection approach using federated learning overcomes data imbalance and enhances model robustness.

Bian and Zheng (2023) introduced the FedAvg-DWA algorithm, which enhances credit card fraud detection by addressing class imbalance and small class samples in a federated learning environment, leading to improved detection rates.

Supriya et al. (2023) combined federated learning with genetic algorithms and particle swarm optimization to create a hybrid model for insurance fraud detection. This model achieves an accuracy of 94.47% while preserving privacy. Ferdous Aurna et al. (2023) explored federated learning-based credit card fraud detection using various deep learning algorithms, including CNN, MLP, and LSTM. Their models achieved high detection rates of 99.51%, 98.77%, and 98.20%, respectively, while maintaining data privacy.

Qiu et al. (2023) presented an efficient vertical federated learning approach with secure aggregation, which improves training performance without compromising security compared to homomorphic encryption methods.

Kang et al. (2019) proposed a federated learning scheme for mobile networks that uses reputation and consortium blockchain for worker selection, enhancing the reliability of federated learning tasks in these networks.

Singh et al. (2020) focused on anomaly detection using federated learning in neural network autoencoder models. Their approach enables smart edge devices to detect anomalies with reduced bandwidth, lower latency, and power consumption, effectively combating online theft and scams. These studies demonstrate the growing effectiveness of federated learning in improving fraud detection across various sectors, emphasizing the advantages of privacy preservation, robustness, and enhanced performance.

Table 5: summary of Federated Learning methods

Authors	Proposed Study	Key Findings
(Zheng, Yan, Gou, & Wang, 2021)	Federated Meta-Learning for Fraudulent Credit Card Detection	Federated meta-learning enables banks to detect credit card fraud using local data, achieving significantly higher performance compared to traditional methods.
(Yang, Zhang, Ye, Li, & Xu, 2019)	FFD: A Federated Learning Based Method for Credit Card Fraud Detection	Federated learning-based fraud detection system (FFD) achieves an average of 95.5% test AUC, improving credit card fraud detection compared to traditional systems.
(Talluri, Zhang, & Chen, 2023)	A Cloud-Native Federated Learning Architecture for Telecom Fraud Detection	The cloud-native federated learning architecture improves telecom fraud detection by up to 23% compared to locally trained models.
(Myalil, Rajan, Apte, & Lodha, 2021)	Robust Collaborative Fraudulent Transaction Detection using Federated Learning	Federated learning can effectively detect fraudulent transactions in a non-IID setting with active adversaries, overcoming data imbalance and ensuring robust models.
(Bian & Zheng, 2023)	FedAvg-DWA: A Novel Algorithm for Enhanced Fraud Detection in Federated Learning Environment	The FedAvg-DWA algorithm improves credit card fraud detection by considering small class samples and addressing class imbalance in a federated learning environment.
(Supriya, Victor, Srivastava, & Gadekallu, 2023)	A Hybrid Federated Learning Model for Insurance Fraud Detection	The hybrid Federated Learning model combining Genetic Algorithm and Particle Swarm Optimization effectively detects insurance fraud with an accuracy of 94.47%, improving privacy preservation.
(Aurna, Hossain, Taenaka, & Kadobayashi, 2023)	Federated Learning-Based Credit Card Fraud Detection: Performance Analysis with Sampling Methods and Deep Learning Algorithms	Federated Learning-based fraud detection systems with CNN, MLP, and LSTM models achieve high detection rates of 99.51%, 98.77%, and 98.20%, preserving sensitive data privacy.
(Qiu et al., 2023)	Efficient Vertical Federated Learning with Secure Aggregation	Our novel design for vertical federated learning secures data exchange and improves training performance without impacting performance compared to homomorphic encryption.
(Kang et al., 2020)	Reliable Federated Learning for Mobile Networks	The proposed worker selection scheme using reputation and consortium blockchain improves the reliability of federated learning tasks in mobile networks.
(Kang et al., 2020)	Anomaly Detection Using Federated Learning	Federated learning in neural network autoencoder models enables smart edge devices to detect anomalies in data streams, reducing online theft and scams by utilizing less bandwidth, lower latency, and power consumption.

D. The Intersection of Federated Learning and Explainable AI in Financial Crime Detection

(Cheng et al., 2020) proposed Graph Neural Network for Fraud Detection via Spatial-Temporal Attention. The spatial-temporal attention-based graph network (STAGN) outperforms other state-of-the-art methods in credit card fraud detection, detecting suspicious transactions, mining fraud hotspots, and uncovering fraud patterns.

(Long, Fang, Luo, Wei, & Weng, 2023) proposed MS_HGNN: a hybrid online fraud detection model to alleviate graph-based data imbalance. The hybrid graph neural network model effectively detects online fraud by considering feature, category, and relation imbalances, outperforming best-in-class models on Amazon and Yelp datasets.

(R. Li et al., 2022) proposed an internet Financial Fraud Detection Based on Graph Learning. TA-Struc2Vec improves Internet financial fraud detection efficiency by learning topological features and transaction amount features in financial transaction network graphs, allowing intelligent classification and prediction.

(Y. Liu et al., 2023) improved Fraud Detection via Hierarchical Attention-based Graph Neural Network. the Hierarchical Attention-based Graph Neural Network (HAGNN) effectively detects fraud by incorporating relation and neighborhood attention modules, improving detection of camouflaged nodes.

(Supriya et al., 2023) proposed a Learning model combining Genetic Algorithm and Particle Swarm Optimization effectively detects insurance fraud with an accuracy of 94.47%, improving privacy preservation.

(Esenogho et al., 2022) proposed a Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection. The proposed neural network ensemble classifier with hybrid data resampling method effectively detects credit card fraud with a sensitivity and specificity of 0.996 and 0.998, outperforming other algorithms.
(Rao et al., 2020) proposed xFraud: Explainable Fraud Transaction Detection. The xFraud effectively predicts fraud transactions in online retail platforms using a graph neural network, outperforming various baseline models and providing meaningful explanations for business analysis.
(Jinyin Chen et al., 2022) proposed Graph-Fraudster: Adversarial Attacks on Graph Neural Network-Based Vertical Federated Learning Graph-Fraudster is a novel adversarial attack method that achieves state-of-the-art attack performance in GNN-based vertical federated learning, even with two defense mechanisms applied.
(Qin et al., 2022) proposed an Explainable Graph-based Fraud Detection via Neural Meta-graph Search NGS (Neural meta-Graph Search) improves graph-based fraud detection performance while maintaining model explainability, outperforming state-of-the-art methods on real-world datasets.
(Zeng & Tang, 2021) proposed RLC-GNN: An Improved Deep Architecture for Spatial-Based Graph Neural Network with Application to Fraud Detection The RLC-GNN improves fraud detection performance by learning layer by layer and continuously correcting mistakes, outperforming single-layer CARE-GNNs on Yelp and Amazon datasets.

Table 6: Summary of Federated Learning and Explainable AI in Financial Crime Detection

Authors	**Proposed Study**	**Key Findings**
(Cheng et al., 2020)	Graph Neural Network for Fraud Detection via Spatial-Temporal Attention	The spatial-temporal attention-based graph network (STAGN) outperforms other state-of-the-art methods in credit card fraud detection, detecting suspicious transactions, mining fraud hotspots, and uncovering fraud patterns.
(Long, Fang, Luo, Wei, & Weng, 2023)	MS_HGNN: a hybrid online fraud detection model to alleviate graphbased data imbalance	The hybrid graph neural network model effectively detects online fraud by considering feature, category, and relation imbalances, outperforming best-in-class models on Amazon and Yelp datasets.
(R. Li et al., 2022)	Internet Financial Fraud Detection Based on Graph Learning	TA-Struc2Vec improves Internet financial fraud detection efficiency by learning topological features and transaction amount features in financial transaction network graphs, allowing intelligent classification and prediction.
(Y. Liu et al., 2023)	Improving Fraud Detection via Hierarchical Attention-based Graph Neural Network	The Hierarchical Attention-based Graph Neural Network (HAGNN) effectively detects fraud by incorporating relation and neighborhood attention modules, improving detection of camouflaged nodes.
(Supriya et al., 2023)	A Hybrid Federated Learning Model for Insurance Fraud Detection	The hybrid Federated Learning model combining Genetic Algorithm and Particle Swarm Optimization effectively detects insurance fraud with an accuracy of 94.47%, improving privacy preservation.
(Esenogho et al., 2022)	A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection	The proposed neural network ensemble classifier with hybrid data resampling method effectively detects credit card fraud with a sensitivity and specificity of 0.996 and 0.998, outperforming other algorithms.
(Rao et al., 2020)	xFraud: Explainable Fraud Transaction Detection	xFraud effectively predicts fraud transactions in online retail platforms using a graph neural network, outperforming various baseline models and providing meaningful explanations for business analysis.
(Jinyin Chen et al., 2022)	Graph-Fraudster: Adversarial Attacks on Graph Neural NetworkBased Vertical Federated Learning	Graph-Fraudster is a novel adversarial attack method that achieves state-of-the-art attack performance in GNN-based vertical federated learning, even with two defense mechanisms applied.
(Qin et al., 2022)	Explainable Graph-based Fraud Detection via Neural Meta-graph Search	NGS (Neural meta-Graph Search) improves graphbased fraud detection performance while maintaining model explainability, outperforming state-of-the-art methods on real-world datasets.

(Zeng & Tang, 2021)	RLC-GNN: An Improved Deep Architecture for Spatial-Based Graph Neural Network with Application to Fraud Detection	The RLC-GNN improves fraud detection performance by learning layer by layer and continuously correcting mistakes, outperforming single-layer CARE-GNNs on Yelp and Amazon datasets.

4. Discussion

This reviewed paper showcases the continuous evolution of machine learning techniques in the fight against fraud. Early contributions, such as those by Li et al. (2008) and Bolton and Hand (2002), laid the groundwork for incorporating statistical and machine-learning methods into fraud detection systems. Over time, more sophisticated models, such as ensemble methods (Whiting et al., 2012), NLP-based models (Purda, 2014), and deep learning frameworks (Hashemi et al., 2023), have been developed to address the growing complexity and diversity of fraudulent activities. The shift toward hybrid models, integrating various machine learning and computational intelligence techniques, highlights the necessity of multifaceted approaches to detect and mitigate fraud effectively. As fraudulent schemes continue to evolve, so too must the algorithms designed to combat them, ensuring that machine learning remains at the forefront of fraud detection research.

From the review, machine learning techniques for fraud detection have evolved from traditional statistical methods to more sophisticated approaches involving deep learning, quantum computing, and ensemble methods. Studies by Ali et al. (2022), Xu et al. (2023), and others have demonstrated that machine learning models such as support vector machines, deep boosting decision trees, and neural network ensembles can significantly improve fraud detection accuracy. However, challenges such as imbalanced datasets, poor data distribution, and the need for interpretability remain. Future research may benefit from further exploring quantum computing, graph learning, and GANs to address these challenges and enhance fraud detection in various industries.

Furthermore, CNNs have demonstrated substantial potential across a variety of applications, particularly in fraud detection where they have been successfully applied to online transaction fraud, credit card fraud, and even corporate bankruptcy prediction. The development of novel architectures like spatial-temporal attention-based networks and the use of advanced regularization methods ensure that CNNs continue to evolve, offering real-time, scalable, and accurate solutions to fraud detection and beyond. Future research is likely to further enhance CNNs' applicability across domains, making them a critical tool for addressing the complexities of modern fraud and financial analytics.

It was also notice that LSTM-based approaches have proven to be highly effective in detecting fraud across various
domains, from credit card transactions to financial statements and Medicare fraud. Their ability to model temporal sequences, adapt to changing patterns, and integrate with data resampling techniques makes them particularly suited for fraud detection tasks, where time-series data and class imbalance pose significant challenges. As research continues to evolve, LSTMs will likely remain a key tool for combating fraud in increasingly complex financial and commercial environments.

Additionally, MLP-based approaches have consistently proven to be effective in fraud detection, particularly in
financial systems involving large datasets and complex patterns. MLP networks excel at capturing hidden relationships within transactional data, and when combined with techniques like data resampling, incremental learning, and ensemble methods, they offer robust and scalable solutions for detecting fraud in real-time. As fraud detection systems continue to evolve, MLP remains a critical tool for addressing the increasing complexity of financial fraud.

Similarly, autoencoder-based approaches have proven highly effective in various fraud detection tasks, from credit card fraud to medical and financial fraud. Autoencoders, especially when integrated with advanced techniques such as GANs, LightGBM, and clustering, provide robust solutions for anomaly detection in imbalanced datasets. Moreover, their ability to enhance interpretability and compliance in fraud detection systems makes them a valuable tool for real-time financial systems, where both accuracy and transparency are essential.

Recently, GNNs have proven highly effective in fraud detection across various domains, including credit card transactions, e-commerce, and large-scale network data. These approaches address challenges like class imbalance, temporal dynamics, and camouflaged fraudsters, showcasing the adaptability of GNNs in handling complex, interconnected datasets. Additionally, advancements like explainable GNNs ensure that these models can meet regulatory standards while maintaining high detection accuracy.

In recent years, the intersection of Federated Learning (FL) and Explainable AI (XAI) has gained considerable momentum in the domain of financial crime detection. This is driven by the growing need to safeguard sensitive data while improving transparency in fraud detection systems. As financial crimes become more sophisticated, leveraging these technologies has the potential to significantly enhance detection capabilities, privacy preservation, and trust in AI systems. The interest in integrating FL and XAI for fraud detection has seen a steady increase over the past few years, particularly as the financial industry increasingly turns to AI for automating fraud detection. Based on the papers reviewed, publication activity surged from 2019 to 2023. This trend mirrors the growing adoption of AI technologies in the financial sector, influenced by regulatory pressures to ensure data privacy and explainability in machine learning models. To visualize this trend, Fig. depict the publication counts per year from 2019 to 2024 highlight.

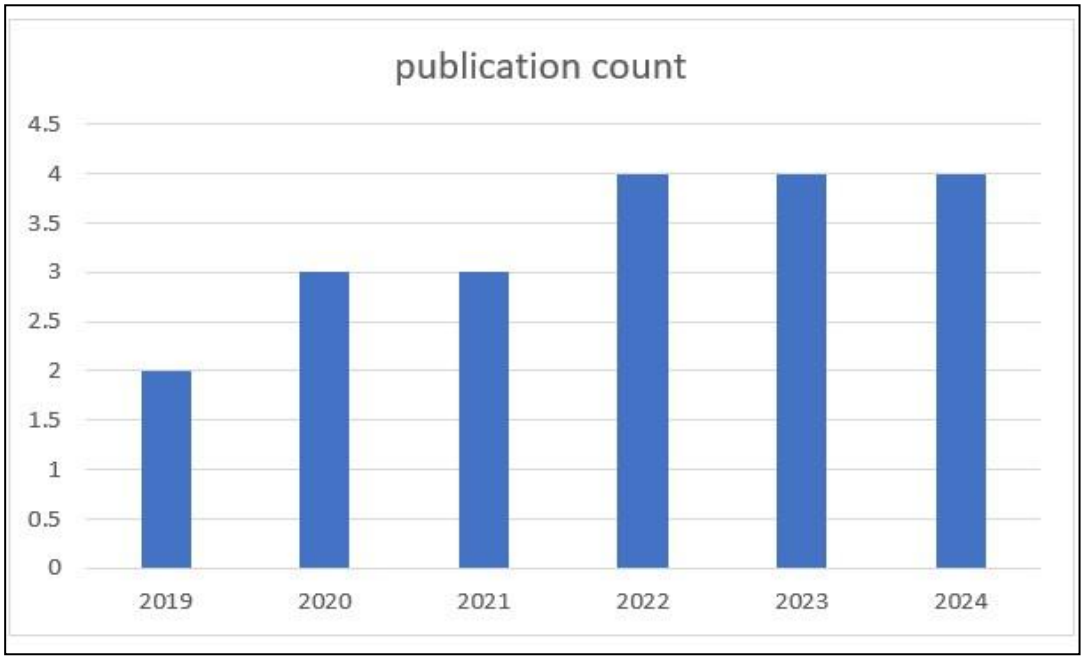

Fig. Publication Trend in FL and XAI for Financial Crime Detection (2019-2024)

The Figure shows an upward trend, with notable publications in 2020 and beyond, reflecting the convergence of FL and XAI research. shows the increase in publication activity from 2019 to 2024, reflecting the growing interest in applying FL and XAI for fraud detection. At the core of this intersection lies the need for financial institutions to not only detect fraud effectively but also to ensure privacy and transparency in the decision-making process. FL allows multiple institutions to collaboratively train models on local data without sharing sensitive information, ensuring data privacy. However, such blackbox models often lack interpretability, which is where XAI comes into play. XAI techniques, such as Neural Meta-Graph Search (NGS) and xFraud, provide the necessary tools to explain decisions, allowing users to understand why certain transactions are flagged as fraudulent.

The intersection of Federated Learning and Explainable AI in financial crime detection offers a powerful solution to modern challenges in the field. These technologies enable privacy-preserving collaboration across institutions while ensuring that fraud detection systems remain interpretable and accountable. The growing body of research reflects this synergy, with a steady increase in publications and advancements in GNNs, hybrid models, and robust defenses. Moving forward, the key challenges will involve scaling these models for larger datasets, improving robustness against adversarial threats, and ensuring that explanations remain understandable to non-technical users.

5. Research Gap

Despite advancements in AI and machine learning, existing fraud detection methods continue to face significant challenges. Traditional models often fail to handle the complex relationships between transactions and entities in financial networks, treating transactions as isolated events rather than leveraging the contextual information that could reveal fraudulent patterns (Xu & Ke, 2020). This limitation is compounded by the issue of imbalanced data, where fraudulent transactions are significantly rarer than legitimate ones. Many models are biased towards the majority class, leading to poor detection rates for fraudulent activities (Buda et al., 2018; Johnson & Khoshgoftaar, 2019).

The complex nature of financial networks necessitates the modeling of intricate relationships and dependencies between transactions, accounts, and entities. Traditional machine learning models, which often treat data points independently, lack the capability to capture these complexities, resulting in suboptimal performance in fraud detection tasks (Xu & Ke, 2020). Furthermore, imbalanced datasets exacerbate this issue, as the scarcity of fraudulent transactions compared to legitimate ones causes models to become biased towards the majority class, leading to a high rate of false negatives (Buda et al., 2018). Effective fraud detection requires robust strategies to manage this imbalance and accurately identify rare fraudulent activities.

Moreover, the "black-box" nature of many AI models hinders their adoption in the financial sector, where interpretability and trust in model decisions are paramount (Rudin, 2019). The lack of transparency makes it difficult for stakeholders to understand, trust, and justify the decisions made by these models, which is crucial for regulatory compliance and operational confidence. Explainable AI (XAI) methods are needed to bridge this gap by providing clear and interpretable insights into the model's decision-making process, enhancing stakeholder trust and facilitating the adoption of AI-based solutions (Lundberg & Lee, 2020).

Additionally, models trained on historical data often struggle to generalize to new, unseen data, especially in the dynamic and evolving landscape of financial fraud (Kim et al., 2020). The continuously changing tactics of fraudsters mean that models must adapt quickly to new patterns and anomalies. Federated Learning (FL) presents a promising approach by enabling collaborative learning across institutions while preserving data privacy, thereby enhancing the model's generalization capabilities and robustness (Kairouz et al., 2021). However, FL also introduces challenges such as communication overhead and heterogeneity in data distribution across different entities, which need to be addressed for effective implementation.

To address these limitations, there is a need for a novel approach that integrates the strengths of Graph Neural Networks (GNNs), Explainable AI (XAI), and Federated Learning (FL). GNNs can model the complex relationships between transactions and entities, providing a more holistic view of fraud patterns. By effectively capturing the relational and contextual information inherent in financial data, GNNs can significantly improve fraud detection performance. However, their application in fraud detection is still nascent and requires further development to ensure scalability and efficiency in handling large-scale financial networks (Wu et al., 2020).

Effective techniques for handling imbalanced data, such as oversampling, undersampling, and cost-sensitive learning, need to be developed and integrated into the model training process to enhance the detection of rare fraudulent transactions (Buda et al., 2018; Johnson & Khoshgoftaar, 2019). Incorporating state-of-the-art explainable AI methods will make the model's decisions more transparent and interpretable, enabling stakeholders to understand and trust the system's outputs (Lundberg & Lee, 2020; Ribeiro et al., 2016). Moreover, employing federated learning will enhance the model's generalization capabilities and ensure data privacy, improving the robustness of fraud detection systems (Kairouz et al., 2021).

In general, despite the progress made in AI and machine learning for fraud detection, significant gaps remain. The integration of GNNs, XAI, and FL presents a promising avenue to address these gaps by improving the accuracy, interpretability, and privacy of fraud detection systems. Hence, developing a comprehensive framework that leverages these advanced methodologies to create a robust, interpretable, and privacy-preserving solution for financial fraud detection is an intriguing opportunity that needs to be explored.

6. Conclusion and Future Work

This review paper highlights the significant strides made in recent years, with an observable upward trend in research
focused on the intersection of FL and XAI. The integration of Federated Learning (FL) and Explainable AI (XAI) in financial crime detection represents a crucial advancement in addressing modern fraud challenges. As financial institutions increasingly adopt AI systems for detecting fraudulent activities, the need for privacy-preserving and transparent solutions becomes more pronounced. Federated Learning allows for the secure, decentralized training of models across multiple organizations without exposing sensitive data, thereby addressing privacy concerns in compliance with stringent regulations. On the other hand, Explainable AI ensures that these models remain interpretable, offering stakeholders insights into how and why specific decisions—such as flagging suspicious transactions—are made. The emergence of Graph Neural Networks (GNNs) and hybrid FL models has been especially impactful in improving fraud detection accuracy, while methods such as Neural Meta-Graph Search (NGS) and attention-based mechanisms have demonstrated how explainability can be integrated without sacrificing performance. Research into adversarial attacks also underscores the importance of building robust and reliable FL models, particularly in environments prone to malicious activities. However, challenges remain in scaling FLXAI systems to handle large-scale, real-time financial data and in ensuring that explanations provided by AI systems are easily interpretable by non-technical users. Future research should focus on improving model scalability, enhancing defense mechanisms against adversarial threats, and optimizing the user-friendliness of explanations, especially for stakeholders in banking, insurance, and financial services. In conclusion, the fusion of Federated Learning and Explainable AI in financial crime detection offers a promising pathway to creating effective, transparent, and privacy-conscious fraud detection systems. This evolving research area holds the potential to transform how financial institutions combat fraud, bolstering trust and accountability while maintaining the highest standards of data security.

REFERENCES

1. Agrawal, S., Sarkar, S., Aouedi, O., Yenduri, G., Piamrat, K., Alazab, M., . . . Gadekallu, T. R. (2022). Federated learning for intrusion detection system: Concepts, challenges and future directions. *Computer Communications, 195*, 346-361.
2. Ahmed, A. A., & Alabi, O. (2024). Secure and scalable blockchain-based federated learning for cryptocurrency fraud detection: A systematic review. *Ieee Access*.
3. Al-Shabi, M. (2019). Credit card fraud detection using autoencoder model in unbalanced datasets. *Journal of Advances in Mathematics and Computer Science, 33*(5), 1-16.
4. Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., . . . Saif, A. (2022). Financial fraud detection based on machine learning: a systematic literature review. *Applied Sciences, 12*(19), 9637.
5. Anowar, F., & Sadaoui, S. (2020). *Incremental neural-network learning for big fraud data*. Paper presented at the 2020 IEEE international conference on systems, man, and cybernetics (SMC).
6. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., . . . Benjamins, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion, 58*, 82-115.
7. Aurna, N. F., Hossain, M. D., Taenaka, Y., & Kadobayashi, Y. (2023). *Federated Learning-Based Credit Card Fraud Detection: Performance Analysis with Sampling Methods and Deep Learning Algorithms.* Paper presented at the 2023 IEEE International Conference on Cyber Security and Resilience (CSR).
8. Benchaji, I., Douzi, S., & El Ouahidi, B. (2021). Credit card fraud detection model based on LSTM recurrent neural networks. *Journal of Advances in Information Technology, 12*(2).
9. Bent, C. C., & Bent, K.-A. (2021). The Intersection of GDPR, US Discovery and Technology in the Financial Crime Discipline *The GDPR Challenge* (pp. 179-203): CRC Press.
10. Bian, K., & Zheng, H. (2023). *FedAvg-DWA: A Novel Algorithm for Enhanced Fraud Detection in Federated Learning Environment.* Paper presented at the 2023 4th International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE).
11. Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical science, 17*(3), 235-255.
12. Bouzidi, Z., Boudries, A., & Amad, M. (2021). LSTM-based automated learning with smart data to improve marketing fraud detection and financial forecasting. *EMAN 2021–Economics & Management: How to Cope with Disrupted Times*, 191.

13. Btoush, E. A. L. M., Zhou, X., Gururajan, R., Chan, K. C., Genrich, R., & Sankaran, P. (2023). A systematic review of literature on credit card cyber fraud detection using machine and deep learning. *PeerJ Computer Science*, *9*, e1278.
14. Chamola, V., Hassija, V., Sulthana, A. R., Ghosh, D., Dhingra, D., & Sikdar, B. (2023). A review of trustworthy and explainable artificial intelligence (xai). *Ieee Access*.
15. Chaquet-Ulldemolins, J., Gimeno-Blanes, F.-J., Moral-Rubio, S., Muñoz-Romero, S., & Rojo-Álvarez, J.-L. (2022). On the black-box challenge for fraud detection using machine learning (ii): nonlinear analysis through interpretable autoencoders. *Applied Sciences*, *12*(8), 3856.
16. Chen, J., Hu, X., Yi, D., Alazab, M., & Li, J. (2022). A Variational AutoEncoder-Based Relational Model for Cost-Effective Automatic Medical Fraud Detection. *IEEE Transactions on Dependable and Secure Computing*, *20*(4), 3408-3420.
17. Chen, J., Huang, G., Zheng, H., Yu, S., Jiang, W., & Cui, C. (2022). Graph-fraudster: Adversarial attacks on graph neural network-based vertical federated learning. *IEEE Transactions on Computational Social Systems*, *10*(2), 492-506.
18. Cheng, D., Wang, X., Zhang, Y., & Zhang, L. (2020). Graph neural network for fraud detection via spatial-temporal attention. *IEEE Transactions on Knowledge and Data Engineering*, *34*(8), 3800-3813.
19. Cirqueira, D., Nedbal, D., Helfert, M., & Bezbradica, M. (2020). *Scenario-based requirements elicitation for user-centric explainable AI: A case in fraud detection.* Paper presented at the International cross-domain conference for machine learning and knowledge extraction.
20. Damayanti, R., & Adrianto, Z. (2023). MACHINE LEARNING FOR E-COMMERCE FRAUD DETECTION. *Jurnal Riset Akuntansi dan Bisnis Airlangga Vol*, *8*(2), 1562-1577.
21. Das, A., & Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371*.
22. Ding, Y., Kang, W., Feng, J., Peng, B., & Yang, A. (2023). Credit card fraud detection based on improved Variational Autoencoder Generative Adversarial Network. *Ieee Access*.
23. Domingues, R., Filippone, M., Michiardi, P., & Zouaoui, J. (2018). A comparative evaluation of outlier detection algorithms: Experiments and analyses. *Pattern recognition*, *74*, 406-421.
24. Došilović, F. K., Brčić, M., & Hlupić, N. (2018). *Explainable artificial intelligence: A survey.* Paper presented at the 2018
25. 41st International convention on information and communication technology, electronics and microelectronics (MIPRO).
26. Dou, Y., Liu, Z., Sun, L., Deng, Y., Peng, H., & Yu, P. S. (2020). *Enhancing graph neural network-based fraud detectors against camouflaged fraudsters.* Paper presented at the Proceedings of the 29th ACM international conference on information & knowledge management.
27. Du, H., Lv, L., Guo, A., & Wang, H. (2023). AutoEncoder and LightGBM for credit card fraud detection problems. *Symmetry*, *15*(4), 870.
28. Edge, M. E., & Sampaio, P. R. F. (2009). A survey of signature based methods for financial fraud detection. *Computers & security*, *28*(6), 381-394.
29. Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., & Obaido, G. (2022). A neural network ensemble with feature engineering for improved credit card fraud detection. *Ieee Access*, *10*, 16400-16407.
30. Fawcett, T., Haimowitz, I., Provost, F., & Stolfo, S. (1998). Ai approaches to fraud detection and risk management. *AI Magazine*, *19*(2), 107-107.
31. Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, *479*, 448-455.
32. Fu, K., Cheng, D., Tu, Y., & Zhang, L. (2016). *Credit card fraud detection using convolutional neural networks.* Paper presented at the Neural Information Processing: 23rd International Conference, ICONIP 2016, Kyoto, Japan, October 16–21, 2016, Proceedings, Part III 23.
33. Gambo, M. L., Zainal, A., & Kassim, M. N. (2022). *A convolutional neural network model for credit card fraud detection.* Paper presented at the 2022 International Conference on Data Science and Its Applications (ICoDSA).
34. Ghanbari, H., Mahdianpari, M., Homayouni, S., & Mohammadimanesh, F. (2021). A meta-analysis of convolutional neural networks for remote sensing applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 3602-3613.
35. Ghosh, C., Chowdhury, A., Das, N., & Sadhukhan, B. (2023). *Enhancing Financial Fraud Detection in Bitcoin Networks Using Ensemble Deep Learning.* Paper presented at the 2023 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS).
36. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., . . . Cai, J. (2018). Recent advances in convolutional neural networks. *Pattern recognition*, *77*, 354-377.
37. Gupta, P. (2023). Leveraging Machine Learning and Artificial Intelligence for Fraud Prevention. *SSRG International Journal of Computer Science and Engineering*, *10*(5), 47-52.
38. Harvey, W. S. (2023). *Reputations at stake*: Oxford University Press.
39. Hashemi, S. K., Mirtaheri, S. L., & Greco, S. (2022). Fraud detection in banking data by machine learning techniques. *Ieee Access*, *11*, 3034-3043.
40. Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert systems with applications*, *117*, 287-299.
41. Jan, C.-L. (2021). Detection of financial statement fraud using deep learning for sustainable development of capital markets under information asymmetry. *Sustainability*, *13*(17), 9879.
42. Jesus, S., Belém, C., Balayan, V., Bento, J., Saleiro, P., Bizarro, P., & Gama, J. (2021). *How can I choose an explainer? An application-grounded evaluation of post-hoc explanations.* Paper presented at the Proceedings of the 2021 ACM conference on fairness, accountability, and transparency.
43. Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert systems with applications*, *100*, 234-245.
44. Kang, J., Xiong, Z., Niyato, D., Zou, Y., Zhang, Y., & Guizani, M. (2020). Reliable federated learning for mobile networks. *IEEE Wireless Communications*, *27*(2), 72-80.

45. Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. *Mechanical systems and signal processing, 151*, 107398.
46. Kute, D. V., Pradhan, B., Shukla, N., & Alamri, A. (2021). Deep learning and explainable artificial intelligence techniques applied for detecting money laundering–a critical review. *Ieee Access, 9*, 82300-82317.
47. Li, J., Huang, K.-Y., Jin, J., & Shi, J. (2008). A survey on statistical methods for health care fraud detection. *Health care management science, 11*, 275-287.
48. Li, Q., He, Y., Xu, C., Wu, F., Gao, J., & Li, Z. (2022). *Dual-augment graph neural network for fraud detection.* Paper presented at the Proceedings of the 31st ACM International Conference on Information & Knowledge Management.
49. Li, R., Liu, Z., Ma, Y., Yang, D., & Sun, S. (2022). Internet financial fraud detection based on graph learning. *IEEE Transactions on Computational Social Systems, 10*(3), 1394-1401.
50. Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems, 33*(12), 6999-7019.
51. Lin, T.-H., & Jiang, J.-R. (2021). Credit card fraud detection with autoencoder and probabilistic random forest. *Mathematics, 9*(21), 2683.
52. Liu, G., Tang, J., Tian, Y., & Wang, J. (2021). *Graph neural network for credit card fraud detection.* Paper presented at the 2021 International Conference on Cyber-Physical Social Intelligence (ICCSI).
53. Liu, Y., Sun, Z., & Zhang, W. (2023). Improving fraud detection via hierarchical attention-based graph neural network. *Journal of Information Security and Applications, 72*, 103399.
54. Liu, Z., Dou, Y., Yu, P. S., Deng, Y., & Peng, H. (2020). *Alleviating the inconsistency problem of applying graph neural network to fraud detection.* Paper presented at the Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval.
55. Long, J., Fang, F., Luo, C., Wei, Y., & Weng, T.-H. (2023). MS_HGNN: a hybrid online fraud detection model to alleviate graph-based data imbalance. *Connection Science, 35*(1), 2191893.
56. Lu, M., Han, Z., Rao, S. X., Zhang, Z., Zhao, Y., Shan, Y., . . . Jiang, J. (2022). *Bright-graph neural networks in real-time fraud detection.* Paper presented at the Proceedings of the 31st ACM International Conference on Information & Knowledge Management.
57. Mayaki, M. Z. A., & Riveill, M. (2022). Multiple inputs neural networks for medicare fraud detection. *arXiv preprint arXiv:2203.05842*.
58. Menshchikov, A., Perfilev, V., Roenko, D., Zykin, M., & Fedosenko, M. (2022). *Comparative analysis of machine learning methods application for financial fraud detection.* Paper presented at the 2022 32nd Conference of Open Innovations Association (FRUCT).
59. Mienye, I. D., & Sun, Y. (2023a). A deep learning ensemble with data resampling for credit card fraud detection. *Ieee Access, 11*, 30628-30638.
60. Mienye, I. D., & Sun, Y. (2023b). A machine learning method with hybrid feature selection for improved credit card fraud detection. *Applied Sciences, 13*(12), 7254.
61. Mill, E. R., Garn, W., Ryman-Tubb, N. F., & Turner, C. (2023). Opportunities in real time fraud detection: An explainable artificial intelligence (XAI) research agenda. *International Journal of Advanced Computer Science and Applications, 14*(5), 1172-1186.
62. Misra, S., Thakur, S., Ghosh, M., & Saha, S. K. (2020). An autoencoder based model for detecting fraudulent credit card transaction. *Procedia Computer Science, 167*, 254-262.
63. Mohammadi, S., & Faramarzi, K. (2016). The Application of Neural Networks to Predict Fraud: Case Study of Tehran Stock Exchange. *Asian Journal of Research in Banking and Finance, 6*(6), 43-50.
64. Mohanty, B., & Mishra, S. (2023). Role of Artificial Intelligence in Financial Fraud Detection. *Academy of Marketing Studies Journal, 27*(S4).
65. Montini, D. A., Matuck, G. R., Da Cunha, A. M., Dias, L. A. V., Ribeiro, A. L. P., & Montini, A. A. (2013). *A sampling diagnostics model for neural system training optimization.* Paper presented at the 2013 10th International Conference on Information Technology: New Generations.
66. Myalil, D., Rajan, M., Apte, M., & Lodha, S. (2021). *Robust collaborative fraudulent transaction detection using federated learning.* Paper presented at the 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA).
67. Nicholls, J., Kuppa, A., & Le-Khac, N.-A. (2021). Financial cybercrime: A comprehensive survey of deep learning approaches to tackle the evolving financial crime landscape. *Ieee Access, 9*, 163965-163986.
68. Pereira, R. D., & Murai, F. (2021). How effective are Graph Neural Networks in Fraud Detection for Network Data? *arXiv preprint arXiv:2105.14568*.
69. Pérez López, C., Delgado Rodríguez, M. J., & de Lucas Santos, S. (2019). Tax fraud detection through neural networks: An application using a sample of personal income taxpayers. *Future Internet, 11*(4), 86.
70. Psychoula, I., Gutmann, A., Mainali, P., Lee, S. H., Dunphy, P., & Petitcolas, F. (2021). Explainable machine learning for fraud detection. *Computer, 54*(10), 49-59.
71. Purda, L., & Skillicorn, D. (2015). Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemporary Accounting Research, 32*(3), 1193-1223.
72. Putrada, A. G., & Ramadhan, N. G. (2023). *MDIASE-Autoencoder: A Novel Anomaly Detection Method for Increasing The Performance of Credit Card Fraud Detection Models.* Paper presented at the 2023 29th International Conference on Telecommunications (ICT).
73. Qin, Z., Liu, Y., He, Q., & Ao, X. (2022). *Explainable graph-based fraud detection via neural meta-graph search.* Paper presented at the Proceedings of the 31st ACM International Conference on Information & Knowledge Management.
74. Qiu, X., Pan, H., Zhao, W., Ma, C., de Gusmão, P. P. B., & Lane, N. D. (2023). Efficient Vertical Federated Learning with Secure Aggregation. *arXiv preprint arXiv:2305.11236*.

75. Rao, S. X., Zhang, S., Han, Z., Zhang, Z., Min, W., Chen, Z., . . . Zhang, C. (2020). xFraud: explainable fraud transaction detection. *arXiv preprint arXiv:2011.12193*.
76. Reurink, A. (2018). Financial fraud: A literature review. *Journal of Economic Surveys, 32*(5), 1292-1325.
77. Riffi, J., Mahraz, M. A., El Yahyaouy, A., & Tairi, H. (2020). *Credit card fraud detection based on multilayer perceptron and extreme learning machine architectures.* Paper presented at the 2020 International Conference on Intelligent Systems and Computer Vision (ISCV).
78. Sailusha, R., Gnaneswar, V., Ramesh, R., & Rao, G. R. (2020). *Credit card fraud detection using machine learning.* Paper presented at the 2020 4th international conference on intelligent computing and control systems (ICICCS).
79. Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2024). A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. *Advanced Intelligent Systems*, 2400304.
80. Santos, C. F. G. D., & Papa, J. P. (2022). Avoiding overfitting: A survey on regularization methods for convolutional neural networks. *ACM Computing Surveys (CSUR), 54*(10s), 1-25.
81. Sen, J., Waghela, H., & Rakshit, S. (2024). Privacy in Federated Learning. *arXiv preprint arXiv:2408.08904*.
82. Shah, V., & Konda, S. R. (2021). Neural Networks and Explainable AI: Bridging the Gap between Models and Interpretability. *INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY, 5*(2), 163-176.
83. Shaheen, M., Farooq, M. S., Umer, T., & Kim, B.-S. (2022). Applications of federated learning; taxonomy, challenges, and research trends. *Electronics, 11*(4), 670.
84. Sharma, R., Mehta, K., & Sharma, P. (2024). Role of Artificial Intelligence and Machine Learning in Fraud Detection and Prevention *Risks and Challenges of AI-Driven Finance: Bias, Ethics, and Security* (pp. 90-120): IGI Global.
85. Supriya, Y., Victor, N., Srivastava, G., & Gadekallu, T. R. (2023). *A Hybrid Federated Learning Model for Insurance Fraud Detection.* Paper presented at the 2023 IEEE International Conference on Communications Workshops (ICC Workshops).
86. Talluri, S., Zhang, Q., & Chen, R. (2023). *A Cloud-Native Federated Learning Architecture for Telecom Fraud Detection.* Paper presented at the NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium.
87. Wang, H., Wang, W., Liu, Y., & Alidaee, B. (2022). Integrating machine learning algorithms with quantum annealing solvers for online fraud detection. *Ieee Access, 10*, 75908-75917.
88. Wang, S. (2010). *A comprehensive survey of data mining-based accounting-fraud detection research.* Paper presented at the 2010 International Conference on Intelligent Computation Technology and Automation.
89. Wang, Y. (2024). A Comparative Analysis of Model Agnostic Techniques for Explainable Artificial Intelligence. *Research Reports on Computer Science*, 25–33-25–33.
90. Weber, L., Lapuschkin, S., Binder, A., & Samek, W. (2023). Beyond explaining: Opportunities and challenges of XAI-based model improvement. *Information fusion, 92*, 154-176.
91. West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: a comprehensive review. *Computers & security, 57*, 47-66.
92. Whiting, D. G., Hansen, J. V., McDonald, J. B., Albrecht, C., & Albrecht, W. S. (2012). Machine learning methods for detecting patterns of management fraud. *Computational Intelligence, 28*(4), 505-527.
93. Wiese, B., & Omlin, C. (2009). Credit card transactions, fraud detection, and machine learning: Modelling time with LSTM recurrent neural networks *Innovations in neural information paradigms and applications* (pp. 231-268): Springer.
94. Wu, T.-Y., & Wang, Y.-T. (2021). *Locally interpretable one-class anomaly detection for credit card fraud detection.* Paper presented at the 2021 International Conference on Technologies and Applications of Artificial Intelligence (TAAI).
95. Xu, B., Wang, Y., Liao, X., & Wang, K. (2023). Efficient fraud detection using deep boosting decision trees. *Decision Support Systems, 175*, 114037.
96. Yang, W., Zhang, Y., Ye, K., Li, L., & Xu, C.-Z. (2019). *Ffd: A federated learning based method for credit card fraud detection.* Paper presented at the Big Data–BigData 2019: 8th International Congress, Held as Part of the Services Conference Federation, SCF 2019, San Diego, CA, USA, June 25–30, 2019, Proceedings 8.
97. Zafar, M. R., & Khan, N. (2021). Deterministic local interpretable model-agnostic explanations for stable explainability. *Machine Learning and Knowledge Extraction, 3*(3), 525-541.
98. Zagaris, B., & Mostaghimi, A. (2023). Cybercrime and Transnational Organized Crime. *IELR, 39*, 90.
99. Zamini, M., & Montazer, G. (2018). *Credit card fraud detection using autoencoder based clustering.* Paper presented at the 2018 9th International Symposium on Telecommunications (IST).
100. Zeng, Y., & Tang, J. (2021). Rlc-gnn: An improved deep architecture for spatial-based graph neural network with application to fraud detection. *Applied Sciences, 11*(12), 5656.
101. Zhan, F. (2020). *Research on bank fraud transaction detection based on LSTM-Focalloss.* Paper presented at the Proceedings of the 2020 3rd International Conference on Algorithms, Computing and Artificial Intelligence.
102. Zhang, Z., Zhou, X., Zhang, X., Wang, L., & Wang, P. (2018). A model based on convolutional neural network for online transaction fraud detection. *Security and Communication Networks, 2018*(1), 5680264.
103. Zheng, W., Yan, L., Gou, C., & Wang, F.-Y. (2021). *Federated meta-learning for fraudulent credit card detection.* Paper presented at the Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on
104. Artificial Intelligence.
105. Zou, J., Zhang, J., & Jiang, P. (2019). Credit card fraud detection using autoencoder neural network. *arXiv preprint arXiv:1908.11553*.