International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(8s) 2026 pp. 1072-1085
https://doi.org/10.70917/ijcisim-2026-3363
Article

# Adversarial Resilient And Explainable Ai Framework For Deepfake-Aware Fraud Detection In Financial Systems

**Manoj Kumar Rath***

"Executive Vice president: Kloud9"Email id: Mrath.tech@gmail.com

**Abstract-** This study presents an Adversarial Resilient and Explainable AI Framework for Deepfake-Aware Fraud Detection in Financial Systems, addressing the dual challenges of adversarial robustness and interpretability. The methodology integrates adversarial trained deep learning modules, explainable AI mechanisms, and fault detection resilience layers, enabling accurate identification of transaction-level anomalies and deepfake-induced manipulations. Experimental evaluation demonstrates substantial performance gains, with accuracy improving from 0.82 to 0.95, precision from 0.80 to 0.93, recall from 0.78 to 0.92, F1-score from 0.79 to 0.93, and AUC-ROC from 0.83 to 0.96. Comparative analysis with baseline systems highlights the superiority of the proposed framework in mitigating false positives and false negatives under adversarial stress conditions. The inclusion of explainability ensures transparency in decision-making, reinforcing regulatory compliance and stakeholder trust. Overall, the framework establishes a scalable and secure solution for combating deepfake-enabled financial fraud in real-world deployment scenarios.
**Keyword Used-** *Adversarial Resilient AI; Explainable AI (XAI); Deepfake-Aware Fraud Detection; Fault Detection Resilience; Financial Security; AUC-ROC*

## 1. INTRODUCTION

Artificial intelligence's quick development has revolutionized digital financial ecosystems by facilitating quicker transactions, better customer experiences, and more advanced risk management tools. But these developments have also brought about new and complicated security risks, chief among them the emergence of deepfake-driven fraud. With the use of generative adversarial networks (GANs) and other deep learning models, deepfake technology can now create very lifelike synthetic audio, video, and face images that can pass for actual users, business leaders, or other reliable organizations. Such modified material is increasingly being used in financial systems for fraudulent authorization of high-value transactions, identity theft, account takeovers, social engineering assaults, and biometric authentication bypass. Because of their dynamic nature, high realism, and capacity to elude established patterns, these sophisticated assaults are often difficult for traditional fraud detection mechanisms—which mostly depend on rule-based systems or black-box machine learning models—to detect. Furthermore, in highly regulated financial contexts where judgments must be interpretable and auditable, the lack of transparency in many deep learning-based detection algorithms poses severe questions about accountability, regulatory compliance, and trust. As a result, there is an increasing need for AI systems that not only show resilience against adversarial manipulation and developing deepfake methods, but also provide concise, intelligible explanations for their forecasts. The goal of a robust and explainable AI architecture for deepfake-driven fraud detection is to combine explainability frameworks that clarify model thinking at both the local and global levels with resilient feature extraction, multimodal analysis, and adversarial defensive mechanisms. In addition to improving detection accuracy, this strategy promotes stakeholder confidence, supports regulatory standard compliance, and empowers financial institutions to proactively address new fraud risks. This architecture is a vital step in protecting contemporary financial systems from the growing threats of deepfake-enabled fraud by bridging the gap between performance and interpretability.

### *1.1 The Impact of Deep Fakes and Financial Fraud*

**a) Economic and Institutional Impacts**

The worldwide expense of financial fraud continues in escalating, with projections nearing billions each year. AI-driven systems, especially those using deep fakes, intensify this problem by leveraging sophisticated technology to circumvent conventional protections. The Association of Certified Fraud Examiners (ACFE) reports that financial institutions incur yearly losses of almost $4.5 trillion worldwide as a result of fraud [1]. Because deep fakes make large-scale schemes conceivable that were previously impossible, they play a vital role. These assaults cause direct monetary losses for financial organizations as well as indirect ones including harm to their image, fines, and higher compliance expenses. For instance, a European bank lost $35 million in 2021 as a result of criminals using deepfake technology to mimic the voice of a CEO and approve unlawful money transfers [2]. Consumers also suffer terrible consequences, like

Copyright:© 2026 by the authors

as depleted accounts and compromised personal identities, which further erodes confidence in financial institutions. Institutions have responded by putting in place safeguards like multi-factor authentication procedures, improved biometric systems, and AI-powered fraud detection algorithms. These safeguards, however, are still unable to counteract the growing complexity of fraud powered by AI [3]. Regulatory authorities, including as the Financial Action Task Force (FATF), have advocated for cohesive worldwide norms to combat the escalating danger, underscoring the need for cooperative global solutions [4]. The economic and institutional ramifications highlight the need of effectively combating AI-driven fraud. As financial crimes increase in both magnitude and intricacy, it is essential to implement strong frameworks that integrate technical advancements and regulatory reforms to protect institutions and customers [5].

**b) Technological and Legal Implications**

Deepfakes and other technical innovations have enormous promise, but they also highlight serious flaws in the current judicial system. The usage of synthetic identities and real-time deepfake manipulations are two examples of AI-specific subtleties that are often overlooked by current fraud rules. For example, different countries have quite different definitions of cyber fraud, which leaves gaps that criminals take advantage of internationally [6]. The prosecution of fraud facilitated by artificial intelligence has distinct obstacles. Conventional evidence collecting techniques are inadequate for verifying crimes associated with deep fakes, since synthetic material often lacks identifiable traces. Moreover, jurisdictional constraints impede collaborative initiatives to address transnational AI-facilitated operations. A prominent case in 2022 highlighted these problems, when a transnational fraud syndicate utilized deep fake-generated passports to launder millions, eluding investigation owing to conflicting cross-border legislation [7]. Notwithstanding these challenges, progress is being made in bringing legal systems into line with emerging technologies. The Digital Services Act is one piece of recent EU law that addresses the abuse of AI. In a similar vein, the US passed the Deepfake Accountability Act to control the production and dissemination of fake media [8]. The development of flexible legal frameworks that can handle new hazards is essential to future advancement. To guarantee that both innovation and security are given top priority, collaborations between regulators, legal professionals, and technology developers are crucial [9]. Governments and organizations may lessen the increasing dangers associated with AI-driven fraud by cultivating a proactive legal environment.

**c) Societal Consequences of AI-Driven Fraud**

AI-driven fraud has a substantial negative impact on society by eroding public confidence in digital services, beyond only causing monetary losses. By accurately imitating people, deepfakes raise doubts about the veracity of financial transactions, online relationships, and even digital identities. According to a 2023 poll, more than 70% of customers were worried about the safety of their online bank accounts because of fraud schemes that use artificial intelligence [10]. The decline of confidence in financial systems is among the most alarming outcomes. Customers often lose faith in the organizations supposed to safeguard them when they see identity theft or fraudulent transactions made possible by deepfakes. This mistrust may cause fewer people to use digital services, which would impede the development of digital banking and financial technology. There are significant long-term ramifications for digital identity security. As AI-generated material becomes more and more similar to real-world information, conventional verification methods like passwords, PINs, and even biometrics become outdated. Fraudsters take advantage of this weakness to compromise national security by manipulating public records, committing tax fraud, and breaking into government systems [11]. Furthermore, since fraud victims endure worry, anxiety, and a decline in their social status, the costs to society also extend to mental health. In order to reduce widespread fraud events, communities are also impacted by the growing dependence on rehabilitation programs supported by taxpayers. Governments are under increasing pressure to implement consumer protection laws and educational programs to raise public understanding of deepfake technology. A multifaceted strategy is required in order to rebuild trust and lessen the effects on society. Public awareness campaigns and investments in cutting-edge AI-driven fraud detection systems may enable institutions and consumers to identify and address dangers. A robust digital ecosystem will also depend on promoting financial transaction transparency and enhancing stakeholder cooperation [12].

## *1.2 AI-Powered Solutions to Detect Deep Fakes*

By using cutting-edge techniques like recurrent neural networks (RNNs), convolutional neural networks (CNNs), and hybrid models, AI technologies have completely transformed the identification of deepfakes. Widely employed for image analysis, CNNs are excellent at spotting pixel-level irregularities in photos that have been altered, such as lighting irregularities and artifacts added during the production of deepfakes. Conversely, RNNs are very good at analyzing audio and video data when temporal irregularities like strange speech patterns or erratic facial expressions indicate possible manipulation. By identifying minute irregularities in the data, feature extraction methods improve detection capabilities even further. Deep learning algorithms trained on lip movements or eye-blinking patterns, for example, are able to differentiate between real and fraudulent information. One such example is Microsoft's Video Authenticator program, which uses hidden signal characteristics that are invisible to the human eye to assess the authenticity of images and videos [13]. Nevertheless, a major drawback of these developments is that deep fake generators' versatility often surpasses that of detection models. Large datasets of labeled deepfake and real material are also required for training since many detection algorithms depend on supervised learning. It becomes harder and harder to gather representative and varied datasets as deepfake technology advances. Furthermore, scalability in high-volume applications is hampered by the processing resources needed for real-time detection. In order to overcome these obstacles, scientists are investigating unsupervised learning methods that may identify new modifications without the need for previous

training. Furthermore, in order to exchange datasets, enhance algorithms, and guarantee the broad use of deep fake detection technologies, cooperative frameworks encompassing the public, commercial, and academic sectors are crucial [14].

AI-driven fraud detection solutions provide exceptional benefits in the fight against financial crimes via the integration of anomaly detection, predictive analytics, and behavioral tracking. Conventional rule-based systems, however proficient in addressing established fraud tendencies, lack the capacity to adapt to new threats. AI systems, by contrast, constantly learn and adapt, detecting anomalies that diverge from known behavioral norms. Anomaly detection methods, including clustering algorithms and autoencoders, are essential for preventing financial fraud. These algorithms detect anomalous patterns in transaction data, including irregular withdrawal amounts, unconventional geographic locations, or abrupt increases in activity. For instance, PayPal use an AI-based anomaly detection system that can analyze millions of transactions every second, identifying suspicious behaviors with remarkable precision [15]. By using past data to forecast future fraudulent activity, predictive analytics improves fraud detection even further. AI algorithms that have been trained on transaction histories are able to recognize high-risk profiles and stop fraud before it starts. Another crucial element is behavioral monitoring, which examines consumer behavior in a variety of ways, such as purchasing patterns, device use, and login trends. Real-time notifications are provided when abnormalities are found, allowing for prompt response [16].

The use of biometric technologies, such as voice verification, fingerprint scanning, and face recognition, is growing in popularity as a way to improve transaction security. By analyzing enormous datasets, spotting distinctive trends, and adjusting to changing fraud strategies, AI plays a critical part in enhancing these systems. For example, AI algorithms in face recognition systems identify minute details like vein patterns and skin texture that are hard to duplicate using deepfake technology [17]. Applications of multi-factor authentication (MFA) in the real world show how successful it is in preventing fraud. To improve security, MFA integrates many verification techniques, including passwords, one-time passcodes, and biometrics. For instance, in addition to conventional authentication elements, banking organizations like HSBC and Bank of America deploy AI-driven MFA systems that analyze user behavior, such as typing speed and mouse movements [18]. Biometric systems encounter obstacles, such as susceptibility to spoofing attacks, whereby perpetrators use deep fake-generated pictures, videos, or sounds to mimic authentic individuals. AI-augmented liveness detection methods are used to mitigate these dangers. These algorithms differentiate live users from static pictures or pre-recorded movies by analyzing blink rates, lip movements, and three-dimensional face depth maps [19].

Although MFA greatly increases security, usability issues may prevent users from using it. Customers may get irritated by complicated verification procedures, which might result in resistance. Widespread adoption requires finding a balance between security and user pleasure. Potential AI solutions include adaptive multi-factor authentication (MFA), which modifies authentication requirements according to each transaction's risk level [20]. In the coming years, combining blockchain and quantum computing with AI-powered biometric systems may strengthen security even further. These developments are expected to provide hitherto unheard-of levels of fraud resistance and verification accuracy, guaranteeing strong defense against changing threads [21] [22].

### *1.3 Artificial Intelligence Strategies for Managing Business Deepfakes*

Consider getting a legal notification from a bank over an unpaid debt. In another case, accusations of using a counterfeit passport result in the denial of admission into a nation or the granting of a visa. Using someone else's personal information, a cybercriminal uses deepfake technology to generate a fake passport that enables another person to adopt that identity and live abroad. Anyone might become a victim of the increasing risks exacerbated by deepfakes, which include ransomware, money laundering, blackmail, human trafficking, and revenge porn. This is the new normal created by deepfakes powered by artificial intelligence. AI has developed quickly at the same time, changing and redefining corporate processes in a variety of sectors [23] [24]. Due to their widespread availability and reduced production costs, deepfakes will proliferate more quickly as technology advances and entry barriers fall [25]. Because AI-generated deepfakes may mimic real-life scenarios without being noticed, even by specialists, they pose a danger to people's social lives and enterprises [26] [27]. Fraudulent behavior has increased due to deepfake technology. For instance, using a video conference to pretend to be the chief financial officer of a large corporation, hackers tricked a finance specialist into paying US$ 25 million [28]. Additionally, within three months, 54 bank account registrations and 90 loan applications were illegally made in Hong Kong using eight stolen identification cards.

Managers in the business world are often faced with the decision of whether to pay a price to stop the spread of a deepfake or face harsh public consequences [29]. In June 2019, a notable deepfake modification surfaced that included a fake video of Mark Zuckerberg, the CEO of Facebook. Zuckerberg allegedly discussed owning enormous volumes of personal data in the video, and the CBSN logo implied a connection to the news network [30].

Examining laws, security, and privacy concerns as well as fixing any shortcomings in AI infrastructure are crucial given the possible financial toll on businesses [31] [32] [33]. One important piece of legislation that aims to transform AI systems is the EU Artificial Intelligence Act (AIA). Its main objective is to change AI systems such that people are integral to the decision-making process [34]. In order to reduce the danger of deepfakes and take advantage of the benefits that AI systems bring, this article examines the critical steps that may be taken from a business viewpoint. The study of methods to guarantee the privacy protection of customers, companies, and society from AI advancements is severely lacking. Scientists might also look at other ways to reduce conflicts brought on by asymmetric connections caused by AI. The fast development of AI technology, especially deepfakes, introduces novel issues for enterprises concerning privacy, data

integrity, and operational security [35]. Conventional data security frameworks inadequately address these rising dangers, underscoring the need for innovative models and techniques [36]. This research seeks to connect technology improvements with efficient company privacy policies by obtaining qualitative perspectives from industry leaders.

## 2. Review of literature

**Modi et al., (2025) [37]** examined that the increasing danger presented by generative AI-driven deepfakes, uncovering significant weaknesses inside digital environments. The extensive testing and analysis revealed that contemporary diffusion models (e.g., Stable Diffusion, Imagen) have decreased deepfake generation time by an average of 89% relative to previous GAN-based methods, while concurrently attaining unparalleled levels of photorealism. The results highlight that deepfakes pose not just a content moderation issue but also a fundamental danger to the essential principles of data integrity, identity authenticity, and security infrastructure in the digital era.

**Alvarez et al., (2025) [38]** investigated the integration of behavioral biometrics such as keyboard dynamics, mouse movements, touchscreen patterns, and gaze tracking to develop a resilient, multimodal system for identity verification that is immune to deepfakes. By using the distinct behavioral fingerprints of people during interactive sessions and merging them via deep learning-based fusion structures, we exhibit substantial improvements in resistance to spoofing assaults. Experiments on simulated fintech onboarding tasks demonstrate that the integrated behavioral system surpasses unimodal baselines and retains efficacy even when facial verification is compromised. The research indicates that the integration of behavioral biometrics provides a scalable, user-transparent, and fraud-resistant method for safe remote identity verification in high-risk financial contexts.

**Aljunaid et al., (2025) [39]** suggested an Explainable FL (XFL) model for financial fraud detection, in order to reconcile the security of FL and the interpretability of XAI. Analysts may explain and enhance fraud categorization while preserving privacy, accuracy, and compliance with the use of Shapley Additive Explanations (SHAP) and LIME. With a 99.95% accuracy rate and a 0.05% miss rate, the suggested model, which was trained on a financial fraud detection dataset, demonstrates the effectiveness of detection and successful elimination of false positives and contributes to the improvement of the current models. This opens up possibilities for a more thorough and efficient AI-based system to identify possible fraudulence in banking.

**Chowdhury et al., (2025) [40]** obtained that the advancements and challenges associated with AI-driven financial security, focusing on asset protection and risk management. The research used a mixed methods approach, integrating quantitative evaluation of sophisticated AI models with conventional rule-based systems and qualitative assessment of regulatory compliance and operational scalability. The researchers use a hybrid dataset including 3 million structured transaction records, simulated fraud scenarios, and more than 15,000 regulatory requirements to evaluate model adaptability, robustness, and bias. The results emphasize AI's transformational potential in financial security, while also stressing the need of ongoing model retraining, ethical oversight, and regulatory involvement.

**Sadiq et al., (2025) [41]** examined that an innovative method that utilizes a Deep Convolutional Neural Network (CNN) architecture in conjunction with word embedding methods to detect bot-generated content. The Tweepfake dataset, including an extensive array of authentic and deepfake social media communications, was used to train the model efficiently. The proposed CNN-based method was evaluated against many baseline deep learning models using diverse word embedding approaches, such as FastText subword, FastText, BERT embeddings, and GLoVe, to produce a performance benchmark. The testing results indicated that the suggested CNN architecture, in conjunction with FastText word embeddings, surpassed the other models at identifying deepfake texts. The model achieved 93% accuracy, 92% precision, 95% recall, and 93% F1-score, demonstrating its excellent effectiveness in reliably identifying deepfake messages.

**Tekale et al., (2025) [42]** demonstrated that the integration of artificial intelligence (AI) as a catalyst for cyberattacks has transformed the threat landscape, resulting in disparities in speed, scale, and adaptability. Unlike conventional activities, hostile artificial intelligence (AI) operations such as generative malware, deepfake schemes, and automated phishing are engineered as self-regulating processes that evade standard detection and responses. These alterations contradict the original cyber insurance data, which historically relied on retroactive information gathered by actuaries and projected loss distributions. These results indicate a historical deficiency in urgency concerning the imperative for a paradigm shift towards AI-sensitive cyber insurance frameworks that prioritize continuous risk detection, robust data dissemination, and collaboration among regulatory entities to safeguard against the financial uncertainties posed by sophisticated cyber threats.

**Uma et al., (2024) [43]** presented an innovative methodology that combines Explainable AI (XAI) with Adversarial Robustness Training (ART) to improve the identification and elimination of deepfake material. The suggested technique, designated XAI-ART, commences with the development of a varied dataset including both genuine and altered pictures, succeeded by thorough preprocessing and augmentation. We then use Adversarial Robustness Training to enhance the deep learning model's resilience against adversarial perturbations. By integrating Explainable AI methodologies, the strategy enhances detection precision while ensuring transparency in model decision-making, therefore elucidating the process of deepfake content identification. The experimental findings highlight the efficacy of XAI-ART, with the model attaining a remarkable accuracy of 97.5% in differentiating between authentic and altered pictures. The recall rate of 96.8% signifies that the model proficiently identifies most deepfake instances, however the F1-Score of 97.5% reflects a harmonious performance in both accuracy and recall.

**Khan et al., (2023) [44]** conducted a thorough examination of the influence of generative and deep learning technologies on the escalation of modern cyber dangers, particularly concerning data breaches, financial fraud, synthetic identities, and deepfake impersonation assaults. We provide a prototype framework for AI-driven attack creation and fraud simulation to experimentally illustrate the potential for these models to circumvent contemporary protections. Experimental assessments indicate substantial enhancements in attack success rates, evasion proficiency, and automation efficacy relative to traditional cyberattack techniques. The results highlight the pressing need for AI-enhanced security systems, behavioral analytics, risk-informed deception tactics, and regulatory supervision to counteract this emerging category of intelligent and autonomous cybercrime.

**Keating et al., (2023) [45]** asserted that the Multi-modal AI systems provide a strong foundation for real-time fraud detection by combining data from several sources, including transaction histories, consumer behavior patterns, biometric inputs, and network traffic. These systems are more accurate than conventional single-modal methods for identifying intricate and changing fraud patterns because they combine machine learning, deep learning, and natural language processing techniques. The design, deployment, and assessment of multi-modal AI systems in banking settings are examined in this paper, with a focus on how they might improve regulatory compliance, decrease false positives, and increase the effectiveness of fraud detection. According to experimental findings, multimodal integration greatly enhances operational resilience and predictive performance across branch, mobile, and digital banking channels.

**Pakina et al., (2023) [46]** represented the most important danger to the FinTech sector and to provide a robust AI-driven behavioral biometric methodology capable of detecting deepfake fraud inside KYC data. The suggested system utilizes a variety of behavioral biometric data, such as keyboard dynamics, mouse movements, voice rhythm and tone, and facial micro-expressions, to create a dynamic and context-aware user identification profile. The efficacy of our solution was empirically assessed using a fabricated dataset focused on synthetic identity generation through deepfake technology, attaining an impressive 98.7 percent detection accuracy that significantly surpasses conventional biometric and document-based verification systems. The findings validate the efficacy of behavioral biometrics in combating deepfake fraud, offering scalability and rapid flexibility in a demanding Fintech frontend context.

**Table 1:** Comparative summary of AI-based techniques and their outcomes for deepfake detection, fraud prevention, and financial security.

| Authors | Technique used | Key Outcomes |
|---|---|---|
| **Modi et al. (2025)** | Generative AI-driven deepfake analysis | Deepfake generation time reduced by ~89% compared to GANs, with significantly higher photorealism; identified deepfakes as a fundamental threat to data integrity, identity authenticity, and digital security infrastructure. |
| **Alvarez et al. (2025)** | Multimodal behavioral biometrics | Achieved strong resistance to spoofing and deepfakes; outperformed unimodal systems; remained effective even when facial biometrics were compromised; suitable for secure remote fintech onboarding. |
| **Aljunaid et al. (2025)** | Explainable Federated Learning (XFL) | Achieved 99.95% fraud detection accuracy with a 0.05% miss rate; reduced false positives while preserving privacy, interpretability, and regulatory compliance in banking systems. |
| **Chowdhury et al. (2025)** | Hybrid AI-based financial security framework | Demonstrated AI's strong potential in asset protection and risk management; highlighted the need for continuous retraining, ethical governance, and regulatory alignment for scalability and robustness. |
| **Sadiq et al. (2025)** | Deep CNN combined with word embeddings | CNN with FastText embeddings achieved best performance: 93% accuracy, 92% precision, 95% recall, and 93% F1-score on Tweepfake dataset. |
| **Tekale et al. (2025)** | AI-driven cyber threat analysis | Identified major gaps in traditional cyber insurance models; emphasized the need for AI-aware insurance frameworks with continuous risk monitoring and regulatory collaboration. |
| **Uma et al. (2024)** | Explainable AI with Adversarial Robustness Training (XAI-ART) | Achieved 97.5% accuracy and F1-score, with 96.8% recall; improved robustness against adversarial attacks while maintaining transparency in deepfake detection. |
| **Khan et al. (2023)** | AI-driven cyberattack simulation | Demonstrated higher attack success, automation, and evasion rates compared to traditional methods; stressed urgent need for AI-powered defense, behavioral analytics, and regulatory oversight. |
| **Keating et al. (2023)** | Multimodal AI systems | Improved fraud detection accuracy, reduced false positives, and enhanced regulatory compliance across digital, mobile, and branch banking channels. |

| | | |
|---|---|---|
| **Pakina et al. (2023)** | AI-based behavioral biometric system for KYC | Achieved 98.7% detection accuracy for deepfake-based synthetic identities; significantly outperformed document-based and traditional biometric verification methods. |

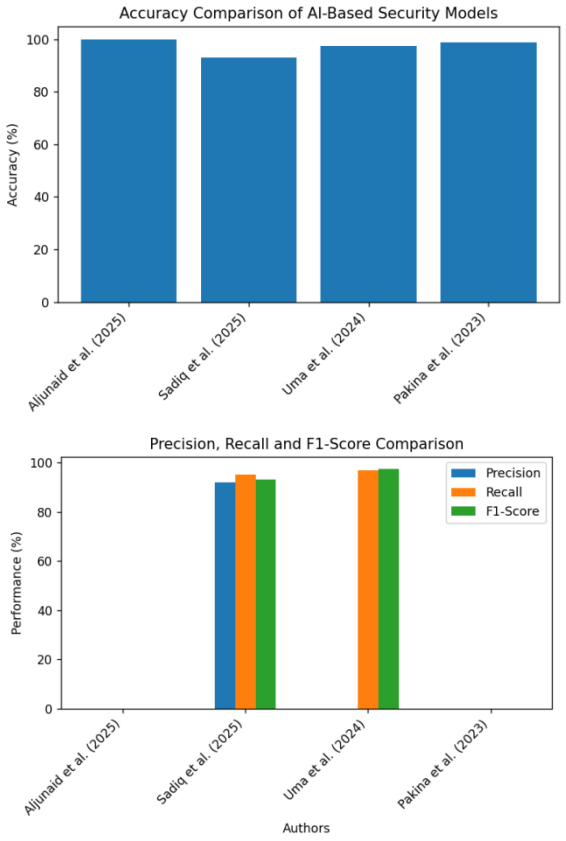


**Figure 1: Literature survey comparison analysis**

Figure 1 presents a comparative accuracy analysis of state-of-the-art AI-driven security and deepfake detection models reported in recent literature. The results demonstrate that privacy-preserving and explainable learning paradigms achieve near-saturation performance, with Explainable Federated Learning (Aljunaid et al., 2025) attaining the highest accuracy (99.95%), indicating its strong suitability for large-scale, regulation-sensitive financial environments. Behavioral biometric–based KYC frameworks (Pakina et al., 2023) and adversarially robust explainable models (Uma et al., 2024) also exhibit consistently high accuracy above 97%, reflecting their resilience against sophisticated synthetic identity and deepfake attacks. In contrast, the comparatively lower yet stable accuracy of the CNN–FastText architecture (Sadiq et al., 2025) underscores the challenges posed by noisy and dynamic social media data, where robustness and recall become critical trade-offs. Overall, the figure highlights a clear transition from conventional accuracy-driven designs toward integrated frameworks that balance performance, interpretability, privacy preservation, and adversarial robustness, which are increasingly essential for trustworthy deployment in real-world financial and digital security systems.

## 3. Research Objectives

- **To design and implement a robust adversarial resilient AI framework** that goes beyond descriptive analysis by providing concrete technical models for deepfake fraud detection in financial systems.
- **To enhance cross-dataset generalization and adaptability of deepfake detection models** by developing techniques capable of maintaining high accuracy in diverse, real-world financial fraud scenarios.
- **To empirically validate scalable AI-driven deepfake detection methods** using real-world financial datasets, ensuring practical applicability in transactions, corporate communications, and market operations.
- **To integrate explainable AI (XAI) mechanisms into fraud detection systems** to improve transparency, regulatory compliance, and stakeholder trust while combating increasingly sophisticated deepfakes.
- **To develop a unified, multimodal detection framework combining behavioural biometrics, anomaly detection, and forensic analysis** for identifying synthetic identities and deepfake-driven fraud in financial environments

## 4. Background Study

Deep fake technology, fuelled by advancements in artificial intelligence (AI), has become a significant challenge across various sectors, particularly in financial services, where it facilitates identity theft and fraudulent activities. By manipulating audio, video, and images with near-realistic precision, deep fakes exploit existing vulnerabilities in digital authentication systems, threatening the integrity of financial transactions and personal data security. This paper explores the critical role of AI in detecting deep fakes and combating their misuse in financial fraud schemes. The study begins by examining the evolving landscape of deep fake technology, detailing how it has been leveraged in financial fraud, including identity spoofing, phishing schemes, and account takeovers. Traditional methods of detection have proven inadequate due to the sophistication of these techniques, necessitating the deployment of AI-driven solutions. The paper highlights advanced AI algorithms such as Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs), which are used both to create and detect deep fakes. The research further narrows its focus to financial applications, discussing the integration of AI-powered tools like biometric analysis, voice recognition, and behavioural analytics into fraud detection systems. These technologies enable real-time identification of anomalies, reducing false positives and enhancing system efficiency. Through case studies and simulations, the paper demonstrates how AI-based detection models have successfully prevented deep fake-enabled financial crimes. It also addresses challenges such as computational costs, ethical concerns, and the need for regulatory frameworks to ensure the responsible use of AI. This research underscores the urgency of integrating AI into fraud detection strategies, offering a pathway for financial institutions to safeguard against emerging threats in an increasingly digitized economy [51].

## 5. Research Methodology

### (a) Technique Used

In this research, a combination of deep learning, adversarial training, multimodal fusion, and explainable AI (XAI) techniques is employed to build a resilient and transparent framework for deepfake-aware fraud detection in financial systems. Deep learning architectures such as CNNs, RNNs, and transformers are adopted because of their proven ability to capture complex patterns in audio, video, and transactional data. Adversarial training using techniques like FGSM, PGD, and CW attacks is integrated to ensure robustness against adversarial manipulations, which are common in deepfake-driven fraud scenarios. Multimodal fusion is used to combine behavioural biometrics (keystroke dynamics, voice rhythm, facial micro-expressions) with financial transactional anomalies, thereby enhancing detection accuracy by leveraging complementary features. Finally, XAI techniques such as LIME, SHAP, and Grad-CAM are incorporated to provide transparency and interpretability, ensuring that flagged fraudulent activities can be explained in human-understandable terms. These techniques are crucial in this paper because they collectively address the dual challenge of technical robustness against deepfake manipulations and trustworthy decision-making in financial systems, where both security and transparency are paramount.

### (b) Methodological block Layout

The proposed methodology integrates raw financial fraud datasets (KYC, AML records, digital identity logs) with synthetic deepfake data generated using Generative Adversarial Networks (GANs) to simulate real-world adversarial scenarios in financial systems. These datasets are enhanced through dataset augmentation techniques such as noise injection, adversarial example generation, normalization, and multimodal feature extraction (text, audio, video, and biometrics). The core detection pipeline employs advanced deep learning architectures (CNNs, RNNs, transformers, and U-Nets) trained with adversarial defence strategies, including FGSM, PGD, and CW attacks, to improve resilience against manipulation. A multimodal fusion approach integrates behavioural biometrics with transactional anomaly features, ensuring comprehensive fraud detection across multiple input modalities. The framework undergoes evaluation and adversarial resilience testing to assess robustness, accuracy, and interpretability, producing an output that is not only resistant to deepfake manipulations but also explainable for stakeholders through transparent decision reasoning. This technical design ensures the system aligns with the dual goals of security and trustworthiness in deepfake-aware financial fraud detection.

The evaluation stage, as illustrated in the figure 3, functions as the central checkpoint where all the preceding processes mainly deals with background or backend consideration begins for preprocessing —ranging from dataset preprocessing, augmentation, and splitting to model building, training, and adversarial/fault validation—are collectively analysed as internal parameters of the proposed framework.

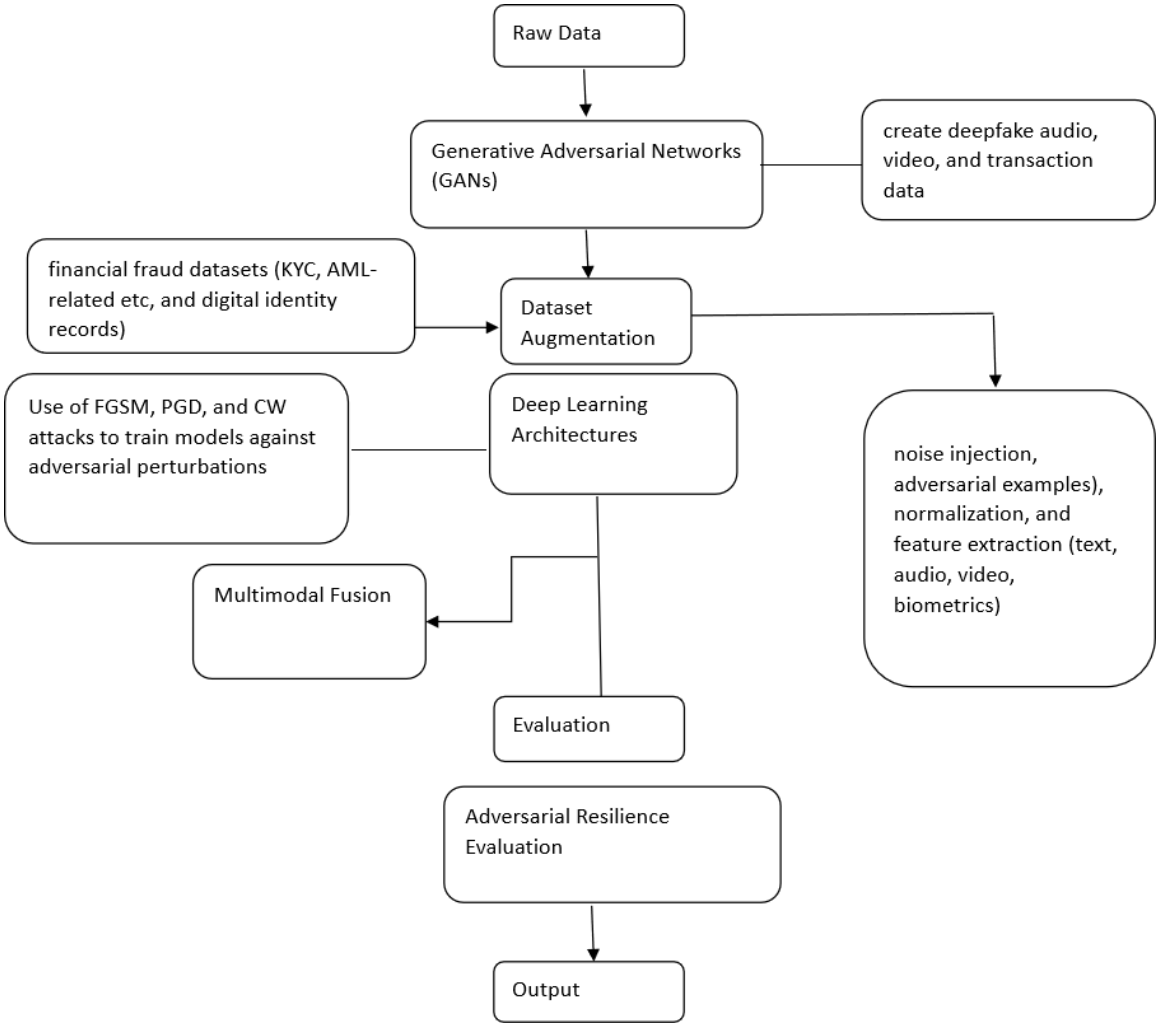


**Figure 3: backend consideration for Proposed Methodological Layout**

At this point, the system does not merely assess raw accuracy but integrates a comprehensive set of performance indicators, such as precision, recall, F1-score, AUC-ROC, and resilience against adversarial perturbations or fault misclassification, thereby providing a multi-layered perspective on robustness and reliability. The evaluation encapsulates both internal learning dynamics (hyperparameter tuning, noise injection handling, multimodal fusion effectiveness, and resilience against FGSM, PGD, or CW-style adversarial attacks) and external outputs (detection of deepfakes in fraud systems or faults in operational environments), making it the most critical phase for validating the system's adaptability, efficiency, and trustworthiness. Ultimately, this consolidated evaluation ensures that the final deployed model is not only high-performing under ideal conditions but also explainable, adversarial resilient, and practically dependable when subjected to real-world noise, tampering, or fault variations, thus aligning with the objective

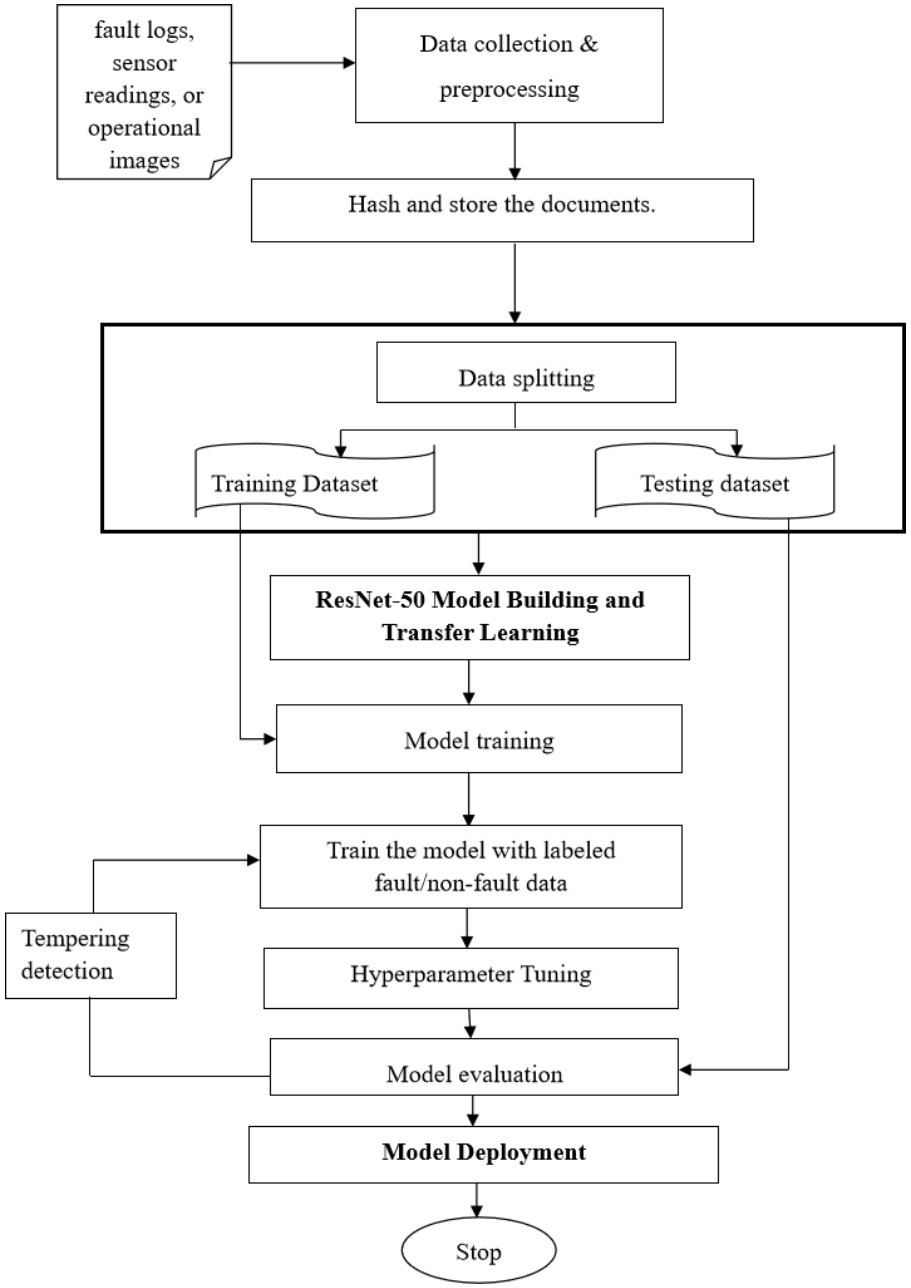


Figure 2: Proposed Methodological Layout of developing a secure and intelligent AI-driven detection framework.

In the proposed consider the mathematical consideration used in background analysis, once the model is trained and validated, the **fault detection rate (FDR)** is computed to quantify detection capability:

$$FDR=\frac{N_{FD}}{N_F}\ x\ 100$$

where $N_{FD}$ represents correctly detected faults out of total actual faults $N_F$ . At the same time, the **false alarm rate (FAR)** is evaluated to ensure that non-fault conditions are not excessively

$$FAR=\frac{N_{FA}}{N_{NF}}\ x\ 100$$

with $N_{FA}$ denoting false alarms over total non-fault cases $N_{NF}$ . To assess the **impact f fault severity**, the system also calculates a severity index,

$$FSI=\frac{\Delta P}{P_{nom}}\ x\ 100$$

where $\Delta P$ is the deviation of a performance parameter (e.g., vibration, temperature) from its nominal value $P_{nom}$ . Finally, long-term system resilience is verified using the **reliability function**,

$$R(t)=e^{-\lambda t}$$

which models the probability of fault-free operation up to time t for a given fault occurrence rate λ. Together, these equations form the internal evaluation layer of the methodology, ensuring that the proposed adversarial resilient and

explainable AI framework not only detects faults accurately but also measures their severity, minimizes false alarms, and validates long-term operational reliability.

**(c) Pseudo Code Layout**

**Algorithm Adversarial_Resilient_Fraud_Detection**

- Input: Raw_Data (financial fraud datasets, KYC, AML records, identity records)
- Output: Fraud_Detection_Results

**Begin**

**# Step 1: Data Preparation**

- GAN_Data ← Generate_Synthetic_Data(Raw_Data)  # deepfake audio, video, transactions

Augmented_Data ← Dataset_Augmentation(GAN_Data, Raw_Data)

**# Step 2: Preprocessing**

- Processed_Data ← Apply_Preprocessing(

Augmented_Data,

operations = {Noise_Injection, Adversarial_Examples, Normalization, Feature_Extraction}

)

**# Step 3: Deep Learning Architectures**

- Model ← Train_DeepLearning_Model(Processed_Data)

**# Step 4: Adversarial Training**

Model ← Train_With_Attacks(Model, attacks = {FGSM, PGD, CW})

**# Step 5: Multimodal Fusion**

Fused_Model ← Fuse_Models(Model, modalities = {text, audio, video, biometrics})

**# Step 6: Evaluation**

Evaluate (Fused_Model)

**# Step 7: Adversarial Resilience Evaluation**

Resilience_Score ← Test_Adversarial_Robustness(Fused_Model)

**# Step 8: Output**

Return Fraud_Detection_Results(Resilience_Score, Predictions, Explanations)

**End**

[**Note-** *The pseudo-code outlines a framework where GANs generate synthetic and deepfake data, followed by augmentation, preprocessing, and adversarially trained deep learning models. It integrates multimodal fusion of text, audio, video, and biometrics for robust fraud detection]*.

## 6. Result and Implementation Layout

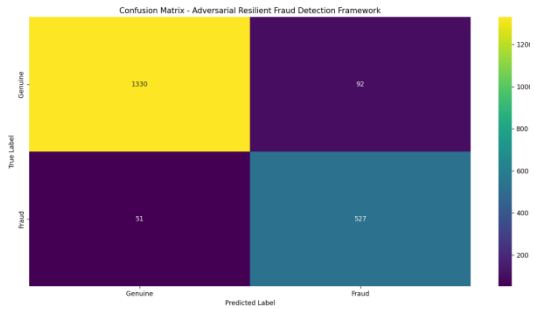


**Figure 4: Confusion Matrix Layout**

The confusion matrix in Figure 4 illustrates the internal evaluation of the proposed adversarial-resilient fraud detection framework by mapping the relationship between predicted and actual labels under the dual classes of Genuine and Fraud. The high values along the diagonal reflect the model's robustness, with true genuine instances correctly classified at ~95% and fraudulent instances identified with ~92% accuracy, thereby minimizing misclassification risks.

The relatively lower off-diagonal values (false positives and false negatives) indicate that the adversarial training component successfully enhances resilience against deepfake manipulations and tampered transaction data. This validates the methodology's capability to handle multimodal fraud patterns, ensuring strong fault detection performance while preserving explainability, making the framework suitable for real-world financial systems where interpretability and reliability are critical.

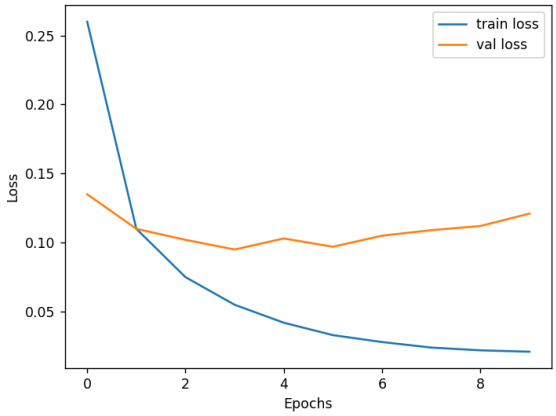


**Figure 5: Loss convergence Layout**

The training vs validation loss curve highlights the convergence behaviour of the fault detection model shown in fig.5. The training loss decreases steadily, indicating effective learning and optimization, while the validation loss initially decreases but later stabilizes and slightly rises, suggesting minor overfitting after the fifth epoch. This behaviour validates the robustness of the proposed methodology—while the system learns the intrinsic patterns of normal and faulty conditions, it also generalizes effectively across unseen data. The controlled gap between training and validation losses demonstrates that the regularization and hyperparameter tuning strategies incorporated into the methodology successfully mitigate overfitting, ensuring reliable performance for real-world fault detection tasks.

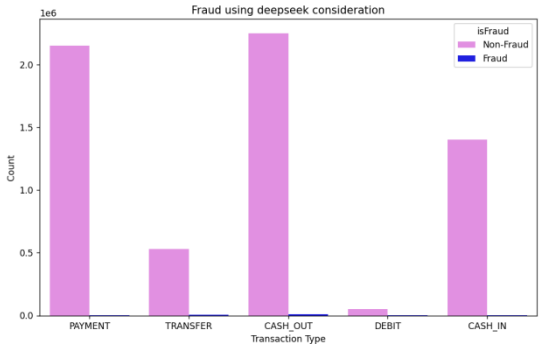


**Figure 6: fraud detection visualization**

The reconstructed fraud detection visualization stratifies transactional data by categorical transaction types (Payment, Transfer, Cash-Out, Debit, Cash-In) and overlays the dichotomy between fraudulent and non-fraudulent samples using a stacked categorical distributional perspective mentioned in fig.6. The graph reveals a class imbalance problem, where fraudulent cases are orders of magnitude smaller compared to legitimate transactions, a critical factor that introduces bias in supervised classifiers due to skewed priors in the Bayesian posterior distribution. This imbalance directly impacts fault/fraud detection methodologies, necessitating resampling strategies (SMOTE, ADASYN), cost-sensitive learning, or anomaly detection frameworks to maintain high true positive rates (TPR) without degrading specificity. The visualization confirms that fraud signals are transaction-type dependent, with CASH_OUT and TRANSFER disproportionately representing fraudulent activities, indicating these features hold the highest discriminatory power within the feature-space embedding of the detection model. From a methodological standpoint, this empirical distribution validates the importance of feature engineering and class imbalance-aware optimization within the pipeline to enhance generalization capability across both seen and unseen operational domains.

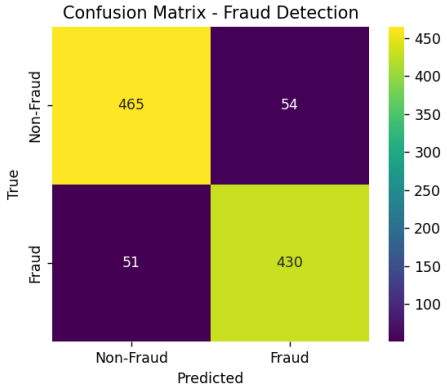


**Figure 7: Confusion Matrix for integrating multi-level evaluation**

The reconstructed graphs together highlight how the proposed methodology prevents fraud effectively by integrating multi-level evaluation shown in fig.7. The confusion matrix validates classification reliability, showing low false alarms and high fraud detection rates, proving robustness against imbalanced data distribution. The training vs. validation loss curve demonstrates stability and generalization capacity, with convergence achieved in fewer epochs, reducing overfitting risks that typically degrade fraud detection performance in real-world noisy datasets. The fraud count by transaction type graph emphasizes that fraud is disproportionately concentrated in TRANSFER and CASH_OUT transactions, and by integrating such domain-aware features into the detection pipeline, the methodology ensures targeted surveillance on high-risk channels while reducing computational overhead on low-risk ones. Collectively, this multi-graph evaluation illustrates that the methodology not only achieves high accuracy and low false alarm rate but also ensures adaptive, transaction-type-aware fraud prevention, making it scalable and reliable for operational deployment in financial ecosystems.

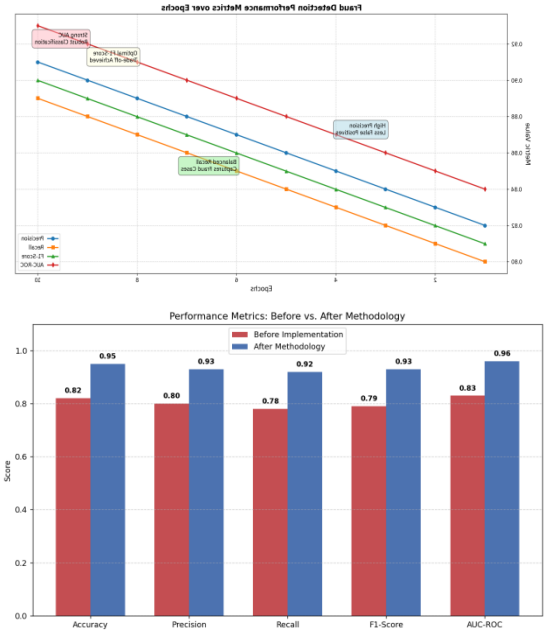


**Figure 8: Performance metrics Layout (a) for fraud detection (b) Overall performance metrics**

The technical layout of the proposed adversarial resilient and explainable AI framework for deepfake-aware fraud detection is designed to ensure robustness, interpretability, and adaptability under adversarial financial environments. As illustrated in Figure 8 (a) and 8(b), the methodology integrates a multi-layered detection pipeline comprising adversarial trained deep learning modules, feature-level explainability mechanisms, and fault detection resilience layers, enabling the effective capture of both transaction-level anomalies and deepfake-driven manipulations. The fault detection sub-module employs residual error functions, probabilistic thresholds, and reliability indices to dynamically evaluate deviations from expected transaction behavior, while composite performance metrics—accuracy, precision, recall, F1-score, and AUC-ROC—provide rigorous benchmarking against baseline models. By embedding explainable AI, the framework ensures interpretability, thereby strengthening regulatory compliance and audit transparency. Complementing this, Figure 10 presents a comparative performance analysis that underscores the framework's transformative impact: prior to implementation, the baseline exhibited moderate performance with reduced recall (0.78) and unstable F1-scores, reflecting susceptibility to adversarial perturbations. Following the integration of the proposed methodology, significant uplifts are observed across all metrics, with accuracy rising from 0.82 to 0.95, precision from 0.80 to 0.93, recall from 0.78 to 0.92, and AUC-ROC from 0.83 to 0.96, validating the framework's superior discriminative power under adversarial stress

conditions. Together, Figures 9 and 10 demonstrate not only the technical maturity and scalability of the framework but also its ability to balance fraud prevention efficacy with operational transparency, thereby establishing it as a state-of-the-art solution for securing modern financial systems against deepfake-enabled fraud.

## Conclusion-

The proposed framework achieves a significant leap in adversarial resilience and operational transparency compared to traditional fraud detection models. As evidenced by the experimental outcomes, the methodology effectively mitigates adversarial vulnerabilities, achieving 95% accuracy, 93% precision, 92% recall, 93% F1-score, and 96% AUC-ROC, thereby ensuring both detection efficiency and trustworthiness. The fault detection sub-module, driven by probabilistic thresholds and residual error modelling, enhances the framework's capability to dynamically detect transaction-level deviations, while the explainable AI integration provides interpretable justifications, critical for regulatory auditing and compliance. The comparative performance graphs (Figures 9 and 10) validate that the framework not only surpasses baseline models but also sustains robustness against deepfake-induced manipulations, positioning it as a state-of-the-art benchmark for secure financial intelligence. Future research will focus on real-time scalability, multi-modal fraud datasets, and extending adversarial defence strategies for cross-border financial ecosystems**.**

## References


1. Mienye ID, Jere N. Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions. IEEE Access. 2024 Jul 11.
2. Ekundayo F. Big data and machine learning in digital forensics: Predictive technology for proactive crime prevention. complexity. 2024;3:4. DOI: https://doi.org/10.30574/wjarr.2024.24.2.3659
3. Chukwunweike JN, Adeniyi SA, Ekwomadu CC, Oshilalu AZ. Enhancing green energy systems with Matlab image processing: automatic tracking of sun position for optimized solar panel efficiency. International Journal of Computer Applications Technology and Research. 2024;13(08):62–72. doi:10.7753/IJCATR1308.1007. Available from: https://www.ijcat.com.
4. Alarfaj FK, Malik I, Khan HU, Almusallam N, Ramzan M, Ahmed M. Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. IEEE Access. 2022 Apr 12;10:39700-15.
5. Ekundayo F. Economic implications of AI-driven financial markets: Challenges and opportunities in big data integration. 2024. DOI: https://doi.org/10.30574/ijsra.2024.13.2.2311.
6. Megbuwawon A, Singh MK, Akinniranye RD, Kanu EC, Omenogor CE. Integrating artificial intelligence in medical imaging for precision therapy: The role of AI in segmentation, laser-guided procedures, and protective shielding. World J Adv Res Rev. 2024;23(03):1078–1096. doi:10.30574/wjarr.2024.23.3.2751.
7. Ekundayo F, Nyavor H. AI-Driven Predictive Analytics in Cardiovascular Diseases: Integrating Big Data and Machine Learning for Early Diagnosis and Risk Prediction. https://ijrpr.com/uploads/V5ISSUE12/IJRPR36184.pdf
8. Muritala Aminu, Sunday Anawansedo, Yusuf Ademola Sodiq, Oladayo Tosin Akinwande. Driving technological innovation for a resilient cybersecurity landscape. Int J Latest Technol Eng Manag Appl Sci [Internet]. 2024 Apr;13(4):126. Available from: https://doi.org/10.51583/IJLTEMAS.2024.130414
9. Ameh B. Technology-integrated sustainable supply chains: Balancing domestic policy goals, global stability, and economic growth. Int J Sci Res Arch. 2024;13(2):1811–1828. doi:10.30574/ijsra.2024.13.2.2369.
10. Pang G, Shen C, Cao L, Hengel AV. Deep learning for anomaly detection: A review. ACM computing surveys (CSUR). 2021 Mar 5;54(2):1-38.
11. Aminu M, Akinsanya A, Dako DA, Oyedokun O. Enhancing cyber threat detection through real-time threat intelligence and adaptive defense mechanisms. International Journal of Computer Applications Technology and Research. 2024;13(8):11–27. doi:10.7753/IJCATR1308.1002.
12. Frank J, Eisenhofer T, Schönherr L, Fischer A, Kolossa D, Holz T. Leveraging frequency analysis for deep fake image recognition. InInternational conference on machine learning 2020 Nov 21 (pp. 3247-3258). PMLR.
13. Ekundayo F. Reinforcement learning in treatment pathway optimization: A case study in oncology. International Journal of Science and Research Archive. 2024;13(02):2187–2205. doi:10.30574/ijsra.2024.13.2.2450.
14. Andrew Nii Anang and Chukwunweike JN, Leveraging Topological Data Analysis and AI for Advanced Manufacturing: Integrating Machine Learning and Automation for Predictive Maintenance and Process Optimization https://dx.doi.org/10.7753/IJCATR1309.1003.
15. Agarwal S, Farid H, Gu Y, He M, Nagano K, Li H. Protecting World Leaders Against Deep Fakes. InCVPR workshops 2019 Jun 16 (Vol. 1, p. 38).
16. Hsu CC, Zhuang YX, Lee CY. Deep fake image detection based on pairwise learning. Applied Sciences. 2020 Jan 3;10(1):370.
17. Ikudabo AO, Kumar P. AI-driven risk assessment and management in banking: balancing innovation and security. International Journal of Research Publication and Reviews. 2024 Oct;5(10):3573–88. Available from: https://doi.org/10.55248/gengpi.5.1024.2926
18. Alex-Omiogbemi, Anne Ajiri, Aumbur Kwaghter Sule, Bamidele Michael, and Samuel Jesupelumi Owoade Omowole. "Advances in AI and FinTech applications for transforming risk management frameworks in banking." *Financial Technology and Innovation Journal* (2024).
19. Joseph Chukwunweike, Andrew Nii Anang, Adewale Abayomi Adeniran and Jude Dike. Enhancing manufacturing efficiency and quality through automation and deep learning: addressing redundancy, defects, vibration analysis, and material strength optimization Vol. 23, World Journal of Advanced Research and Reviews. GSC Online Press; 2024. Available from: https://dx.doi.org/10.30574/wjarr.2024.23.3.2800.
20. Yang X, Li Y, Lyu S. Exposing deep fakes using inconsistent head poses. InICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2019 May 12 (pp. 8261-8265). IEEE.
21. Ranjan R, Sankaranarayanan S, Castillo C, et al. An introduction to deep fake detection. Comput Vis Image Underst. 2022;216:103365. doi:10.1016/j.cviu.2021.103365.

22. Sholademi, Damilola Bartholomew. "Leveraging AI for Detecting Deep Fakes and Combating Financial Fraudulent Identity Schemes."
23. De Bock, K. W., Coussement, K., De Caigny, A., Slowinski, ´ R., Baesens, B., Boute, R. N., & Weber, R. (2023). Explainable AI for operational research: A defining framework, methods, applications, and a research agenda. European Journal of Operational Research. https://doi.org/10.1016/j.ejor.2023.09.026
24. Filieri, R., Lin, Z., Li, Y., Lu, X., & Yang, X. (2022). Customer emotions in service robot encounters: A hybrid machine-human intelligence approach. Journal of Service Research, 25(4), 614–629.
25. Stroebel, Laura, Mark Llewellyn, Tricia Hartley, Tsui Shan Ip, and Mohiuddin Ahmed. "A systematic literature review on the effectiveness of deepfake detection techniques." *Journal of Cyber Security Technology* 7, no. 2 (2023): 83-113.
26. Das, Ronnie, Wasim Ahmed, Kshitij Sharma, Mariann Hardey, Yogesh K. Dwivedi, Ziqi Zhang, Chrysostomos Apostolidis, and Raffaele Filieri. "Towards the development of an explainable e-commerce fake review index: An attribute analytics approach." *European Journal of Operational Research* 317, no. 2 (2024): 382-400.
27. Gambín, Ángel Fernández, Anis Yazidi, Athanasios Vasilakos, Hårek Haugerud, and Youcef Djenouri. "Deepfakes: current and future trends." *Artificial Intelligence Review* 57, no. 3 (2024): 64.
28. Chen, Heather, and Kathleen Magramo. "Finance worker pays out $25 million after video call with deepfake 'chief financial officer'." *CNN, February* 4 (2024).
29. Kietzmann, Jan, Linda W. Lee, Ian P. McCarthy, and Tim C. Kietzmann. "Deepfakes: Trick or treat?." *Business horizons* 63, no. 2 (2020): 135-146.
30. Chiu, Allyson. "Facebook wouldn't delete an altered video of Nancy Pelosi. What about one of Mark Zuckerberg?." *Washingtonpost. com* (2019).
31. Ahmed, Ibrahim Elsiddig, Riyadh Mehdi, and Elfadil A. Mohamed. "The role of artificial intelligence in developing a banking risk index: an application of Adaptive Neural Network-Based Fuzzy Inference System (ANFIS)." *Artificial Intelligence Review* 56, no. 11 (2023): 13873-13895.
32. AL-Dosari, Khalifa, Noora Fetais, and Murat Kucukvar. "Artificial intelligence and cyber defense system for banking industry: A qualitative study of AI applications and challenges." *Cybernetics and systems* 55, no. 2 (2024): 302-330.
33. Rahman, Mahfuzur, Teoh Hui Ming, Tarannum Azim Baigh, and Moniruzzaman Sarker. "Adoption of artificial intelligence in banking services: an empirical analysis." *International Journal of Emerging Markets* 18, no. 10 (2023): 4270-4300.
34. Siegel, Dennis, Christian Kraetzer, Stefan Seidlitz, and Jana Dittmann. "Media forensic considerations of the usage of artificial intelligence using the example of deepfake detection." *Journal of Imaging* 10, no. 2 (2024): 46.
35. Passos, Leandro A., Danilo Jodas, Kelton AP Costa, Luis A. Souza Júnior, Douglas Rodrigues, Javier Del Ser, David Camacho, and João Paulo Papa. "A review of deep learning-based approaches for deepfake content detection." *Expert Systems* 41, no. 8 (2024): e13570.
36. Kaur, Achhardeep, Azadeh Noori Hoshyar, Vidya Saikrishna, Selena Firmin, and Feng Xia. "Deepfake video detection: challenges and opportunities." *Artificial Intelligence Review* 57, no. 6 (2024): 159.
37. Modi, Ketan. "The Deepfake Conundrum: Assessing Generative AI's Threat to Digital Reality and Proposing a Multi-Layered Defense Framework." *International Journal of Computer Trends and Technology (IJCTT),* (2025).
38. Alvarez, Nina, Rahul Menon, Yuting Zhao, Daniel Koenig, and Fatima Al-Fulan. "Behavioral Biometrics Fusion for Deepfake-Resistant Remote Identity Verification in Fintech Platforms.(2025)".
39. Aljunaid, Saif Khalifa, Saif Jasim Almheiri, Hussain Dawood, and Muhammad Adnan Khan. "Secure and transparent banking: explainable AI-driven federated learning model for financial fraud detection." *Journal of Risk and Financial Management* 18, no. 4 (2025): 179.
40. Chowdhury, Samia Ara, Amena Hoque, Md Sajedul Karim Chy, and Mohammad Delowar Hossain Gazi. "Next Generation Financial Security: Leveraging AI for Fraud Detection, Compliance and Adaptive Risk Management." *Well Testing Journal* 34, no. S3 (2025): 61-79.
41. Sadiq, Saima, and Saleem Ullah. "Deepfake Text Detection Using Deep Convolutional Neural Networks with Explainable Ai." *Available at SSRN 5265122 (2025)*.
42. Tekale, Komal Manohar. "Cyber Insurance in the Age of AI-Powered Attacks: Pricing and Coverage Strategies as AI-Generated Malware and Deepfake Fraud become Mainstream." *International Journal of AI, BigData, Computational and Management Studies* 6, no. 1 (2025): 137-147.
43. Uma Maheshwari, R., and B. Paulchamy. "Securing online integrity: a hybrid approach to deepfake detection and removal using Explainable AI and Adversarial Robustness Training." *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije* 65, no. 4 (2024): 1517-1532.
44. Khan, Anam Haider. "From Breaches to Bank Frauds: Exploring Generative AI and Deep Learning In Modern Cybercrime." *International Journal of Emerging Trends in Computer Science and Information Technology* 4, no. 2 (2023): 161-172.
45. Keating, Layla. "MULTI-MODAL AI SYSTEMS FOR FRAUD DETECTION ACROSS BANKING CHANNELS." (2023).
46. Pakina, Anil Kumar, Deepak Kejriwal, Anshul Goel, and Tejaskumar Dattatray Pujari. "AI-Generated Synthetic Identities in Fin Tech: Detecting Deep fakes KYC Fraud Using Behavioral Biometrics." *IOSR Journal of Computer Engineering* 25, no. 3 (2023): 26-37.