Journal of
Risk and Financial Management

MDPI

Article

Secure and Transparent Banking: Explainable AI-Driven Federated Learning Model for Financial Fraud Detection

Saif Khalifa Aljunaid [1], Saif Jasim Almheiri [1], Hussain Dawood [1] and Muhammad Adnan Khan [1,2,3,*]

[1] School of Computing, Skyline University College, University City Sharjah, Sharjah 1797, United Arab Emirates; saif.aljunaid@fcd.sharjah.ae (S.K.A.); saif.jasim@shjmun.gov.ae (S.J.A.); dawood.hussain@skylineuniversity.ac.ae (H.D.)
[2] Riphah School of Computing & Innovation, Faculty of Computing, Riphah International University, Lahore Campus, Lahore 54000, Pakistan
[3] Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura 140401, Punjab, India
* Correspondence: adnan.khan@riphah.edu.pk or muhammad.adnan@skylineuniversity.ac.ae; Tel.: +971-558599421

Abstract: The increasing sophistication of fraud has rendered rule-based fraud detection obsolete, exposing banks to greater financial risk, reputational damage, and regulatory penalties. Financial stability, customer trust, and compliance are increasingly threatened as centralized Artificial Intelligence (AI) models fail to adapt, leading to inefficiencies, false positives, and undetected detection. These limitations necessitate advanced AI solutions for banks to adapt properly to emerging fraud patterns. While AI enhances fraud detection, its black-box nature limits transparency, making it difficult for analysts to trust, validate, and refine decisions, posing challenges for compliance, fraud explanation, and adversarial defense. Effective fraud detection requires models that balance high accuracy and adaptability to emerging fraud patterns. Federated Learning (FL) enables distributed training for fraud detection while preserving data privacy and ensuring legal compliance. However, traditional FL approaches operate as black-box systems, limiting the analysts to trust, verify, or even improve the decisions made by AI in fraud detection. Explainable AI (XAI) enhances fraud analysis by improving interpretability, fostering trust, refining classifications, and ensuring compliance. The integration of XAI and FL forms a privacy-preserving and explainable model that enhances security and decision-making. This research proposes an Explainable FL (XFL) model for financial fraud detection, addressing both FL's security and XAI's interpretability. With the help of Shapley Additive Explanations (SHAP) and LIME, analysts can explain and improve fraud classification while maintaining privacy, accuracy, and compliance. The proposed model is trained on a financial fraud detection dataset, and the results highlight the efficiency of detection and successful elimination of false positives and contribute to the improvement of the existing models as the proposed model attained 99.95% accuracy and a miss rate of 0.05%, paving the way for a more effective and comprehensive AI-based system to detect potential fraudulence in banking.

Keywords: financial fraud detection; secure and transparent banking; artificial intelligence (AI); FL; XAI; XFL; SHAP

Academic Editors: Larry Crumbley and Dimitrios Koutmos

Received: 5 March 2025
Revised: 15 March 2025
Accepted: 24 March 2025
Published: 27 March 2025

Citation: Aljunaid, S. K., Almheiri, S. J., Dawood, H., & Khan, M. A. (2025). Secure and Transparent Banking: Explainable AI-Driven Federated Learning Model for Financial Fraud Detection. *Journal of Risk and Financial Management*, *18*(4), 179. https://doi.org/10.3390/jrfm18040179

1. Introduction

The increasing evolution of financial transactions has led to the expansion of fraud schemes, such as e-commerce and digital money transfers. These advancements have significantly enhanced corporate management, reduced operational costs, and improved

overall productivity (Al-dahasi et al., 2025). With the growing reliance on electronic financial transactions, businesses and organizations have transitioned to digital platforms, fundamentally transforming financial operations. However, this shift has also exposed financial systems to new risks, particularly in the form of cybercrime and fraudulent activities. Among these, electronic banking services have emerged as prime targets for malicious attackers, resulting in annual losses amounting to billions of dollars and a sharp surge in financial fraud, with global losses caused by payment fraud increasing from USD 9.84 billion in 2011 to USD 32.39 billion in 2020 and expected to reach USD 40.62 billion by 2027 (Ramachandran et al., 2023).

Financial fraud not only causes substantial economic losses for businesses, governments, and private individuals but also poses severe risks to global financial stability (Choi & Lee, 2018). According to the Association of Certified Fraud Examiners (ACFE), fraud-related crimes such as payment fraud, identity theft, and embezzlement collectively contribute to a staggering USD 4 trillion in annual losses, amounting to approximately 5% of global business revenues Barnes (2020). In 2020 alone, the FBI's Internet Crime Complaint Center (IC3) recorded over 790,000 cybercrime-related complaints, with reported losses exceeding USD 4.2 billion (Johnson, 2022; Ruposky, 2022). Similarly, a report by the International Monetary Fund (IMF) highlights that financial crimes, including money laundering, corruption, and tax evasion, account for 2–5% of the global GDP, translating to trillions of dollars annually (Abolarin, 2025; Al-dahasi et al., 2025; Barker, 2024). This complexity underscores the need for a structured classification framework to better understand the various forms of financial fraud and their institutional impact Zhu et al. (2021), as illustrated in Figure 1.

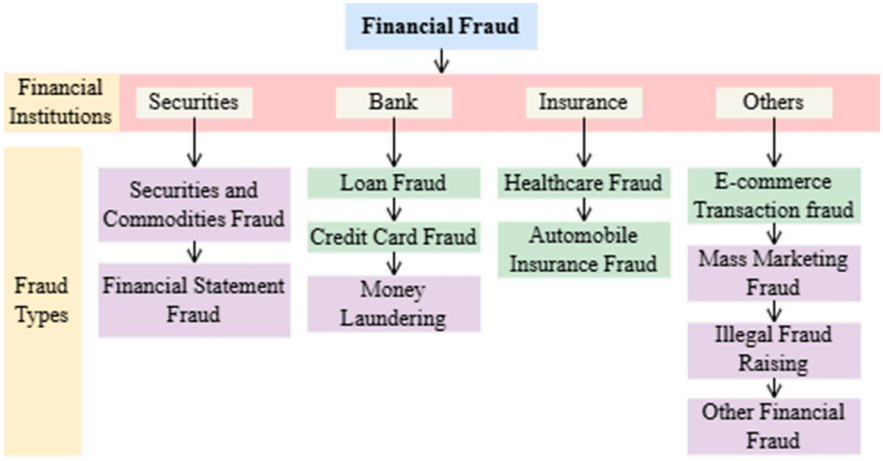

Figure 1. Financial fraud classification framework across financial sectors (Zhu et al., 2021).

Figure 1 presents a structured classification of financial fraud, categorizing it into four key sectors—securities, banking, insurance, and other financial frauds—based on their institutional association and impact. Securities fraud includes financial statement fraud, while banking fraud covers loan fraud, credit card fraud, and money laundering. Insurance fraud involves healthcare and automobile insurance fraud, whereas other financial frauds—such as e-commerce fraud and illegal fund-raising—exploit digital financial systems (Kose et al., 2015; Modi & Dayma, 2018; Yan et al., 2020; Al-Hashedi & Magalingam, 2021). Fraud also takes place at both the customer and business levels where customer fraud involves individual customers and business fraud involves money laundering and tax evasion normally associated with organized financial crime. This subclass is essential for fraud identification so that AI-based technologies can be tailored to particular fraud schemes, enabling institutions to minimize risk and enhance financial security.

High levels of financial fraud have called for more stringent and efficient anti-fraud systems than at any other time in the past. Nevertheless, current methods of fraud detection are less effective with the modern threats of cyberspace. The advancements in the use of online or electronic payment and financial services have made traditional rule-based fraud detection methods inefficient and irrelevant. These systems work under the normative regime using fixed thresholds and are unable to learn the new tactics commonly used by hackers (Ikemefuna et al., 2024; Nicholls et al., 2021). Furthermore, the sheer volume and complexity of financial transactions make manual fraud detection unfeasible, necessitating the development of highly efficient, automated fraud detection systems.

Cybercriminals continuously refine their evasion tactics, making fraudulent activities increasingly difficult to detect. Many fraudsters now camouflage their illicit transactions within legitimate financial activities, further complicating the identification process (Javadpour et al., 2024; Mahboubi et al., 2024). Therefore, fraud detectors need to be capable of identifying complex fraudulent schemes without generating many false alarms that will interfere with normal business. As digital payments continue to gain popularity, it has become important to enhance security measures against fraud operations in operation risk areas (Gandomi et al., 2022). The increasing sophistication of fraud has led researchers to develop more advanced fraud detection methodologies, particularly through the integration of AI and Machine Learning (ML).

ML has emerged as a powerful tool in the development of fraud detection systems, leveraging data-driven insights to enhance fraud identification. By analyzing vast financial datasets, ML algorithms can uncover hidden patterns, detect anomalies, and classify suspicious transactions with remarkable accuracy. The application of ML in fraud detection has led to significant advancements across multiple domains, including banking, healthcare, and cybersecurity Khetani et al. (2023). Cross-domain analysis plays a vital role in refining ML algorithms, ensuring their robustness and adaptability across different industries (Farooq et al., 2024; Ghazal et al., 2024; Khan et al., 2024). However, despite its advantages, ML-based fraud detection models are not without limitations.

One of the principal issues associated with ML-based FD is that of explanation and interpretability. Most conventional ML architectures work in a 'black box' manner to a certain extent, limiting the easy interpretation of their decision-making process by financial analysts and regulatory authorities. The lack again also raises issues of reliability, responsibility, and conformity especially in sectors that are sensitive, such as finance (Jessica et al., 2023). Furthermore, centralized ML-based fraud detection approaches pose additional risks related to data privacy and security, as they require financial institutions to share sensitive transaction data with third-party servers. These concerns highlight the urgent need for fraud detection models that not only offer high accuracy but also ensure transparency, interpretability, and compliance with privacy regulations.

Presently, XAI has been introduced as a way of handling the interpretability of issues that are characteristic of ML-based fraud detection systems. It increases the trustworthiness of the AI-based fraud detection systems and equips users with insights into the models' decisions. Some of the approaches that render interpretable features include SHAP and LIME which allow the financial analyst to understand and verify the accurate classification of fraudulent cases by decreasing dependence on black box models (Rehman & Hashim, 2020; Abbas et al., 2024; Saleem et al., 2024; Shahzad et al., 2024). Moreover, XAI facilitates regulatory compliance by allowing organizations to justify fraud detection decisions to auditors and legal authorities.

While XAI enhances fraud detection transparency, FL offers a decentralized approach to address privacy concerns (Khan et al., 2025). FL enables multiple financial institutions to collaboratively train fraud detection models without sharing raw transaction data, thereby

preserving data privacy and security. Unlike traditional centralized ML models that require data aggregation in a single server, FL ensures that sensitive financial information remains within individual organizations while still contributing to the development of a robust fraud detection model (Khetani et al., 2023). This decentralized approach is further illustrated in Figure 2, which depicts the FL architecture (Lim et al., 2020; Sabuhi et al., 2024) and its iterative training process for secure and privacy-preserving fraud detection.

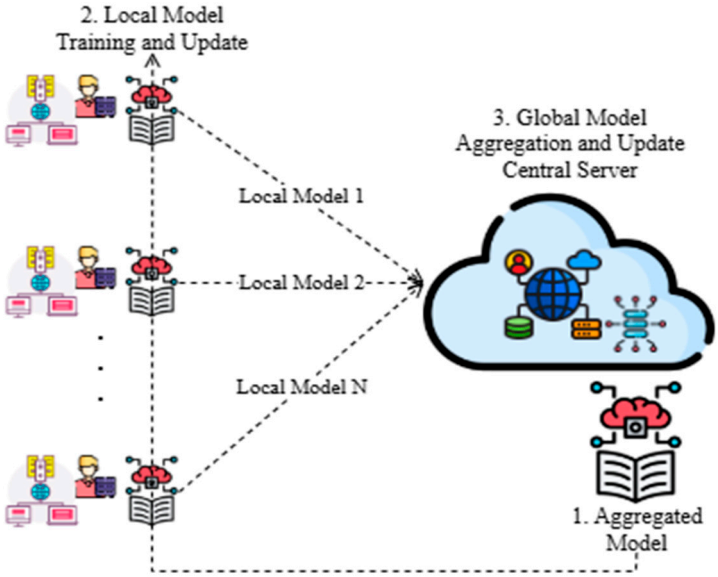

Figure 2. FL architecture for privacy-preserving fraud detection (Sabuhi et al., 2024).

Figure 2 also shows the FL architecture and the training process and as stated above, by using decentralized learning the detection of fraud is boosted while at the same time the privacy of data is maintained. The process starts with the global model being generated in a central server and sending parameters to users (Cho et al., 2020; Nishio & Yonetani, 2019; Wang et al., 2020). Each client can retrain the model on the client's data and then update the parameters whilst preserving the data security of the financial information. These new local models are then combined at the central server through processes like Federated Averaging (FedAvg) to develop a better global model. This process continues until the error of the model is minimized to an acceptable level or until it wants the best accurate results. In this way, FL guarantees compliance with regulations while also improving the financial institutions' ability to fight fraud and protect sensitive data. However, even with all these advantages, there is what is commonly known as the black-box problem when implementing the traditional FL.

To address the dual challenges of data privacy and model interpretability, this research introduces an innovative XFL model for financial fraud detection. Unlike traditional fraud detection systems that either compromise privacy (centralized ML models) or lack transparency (black-box AI models), XFL combines privacy-preserving FL with XAI techniques such as SHAP and LIME. This integration ensures both high accuracy and regulatory compliance, making it a novel contribution to financial fraud detection. The XFL model offers a comprehensive solution that balances security, interpretability, and efficiency, providing financial institutions with a reliable and privacy-preserving fraud detection mechanism. This research work is structured as follows: Section 2 presents a literature review; Section 3 discusses the limitations of previous works; Section 4 highlights the contributions of the proposed approach; Section 5 details the proposed methodology; Section 6 provides simulation results and analysis; and Section 7 concludes the study.

2. Literature Review

Financial fraud is another problem that has continued to plague the banking and financial industries, due to the adverse effects it has had on the economy and security. Traditional systems like rule-based and heuristic have been used but are not very efficient as they cannot adapt to the new trends in fraud. As the use of AI and ML gained prominence, scholars investigated how these technologies can be utilized to improve the specificity of fraud detection without compromising data protection and legal requisites.

In the present study by Talukder et al. (2024), it was indicated that there were high risks of fraud transactions in the financial institutions, hence the need for improved approaches for detection. The research also showed that the data were skewed, with few incidents of fraud, although they were significantly damaging. Traditional fraud models faced this problem and, as a result, they had many false positives. As a solution, the authors presented the IMEML model, which is a combination of an EIC, EBC, and EMC to boost fraud detection. Moreover, advanced techniques such as IHT-EMC/IHT+EMC, CC, and RUS were employed during model learning to enhance the model learning process. The findings of this study tend to support the notion that ensemble methods enhance the approach in addressing the issues related to data imbalance and the likelihood of effective fraud detection in financial transactions.

In the study by Baghdadi et al. (2024), the authors examined credit card fraud a problem that has remained to be focal to institutions and customers because of potential financial threats. Although numerous ML and DL models had been realized, the existing solutions were limited by a lack of balance between the online response time and the predictive capability, together with the lack of willingness of the financial institutions to share their fraud datasets. To address these issues, the researchers presented a novel architecture considering both an Energy-based Restricted Boltzmann Machine (EB-RBM) and Extended Long Short-Term Memory (xLSTM) to increase the efficiency of fraud detection with reasonable computational costs. This was achieved by incorporating evaluation tests, including AUC-ROC, AUC-PR, precision, recall, and F1-score to demonstrate the system's reliability. The proposed model has been further tested and validated against real-world European cardholder data to find that the herein model outperforms existing solutions in terms of accuracy and speed as well as the reliability of fraud detection which is a significant improvement in financial safety.

In Puh and Brkić (2019), the researchers focused on the application and evaluation of different ML techniques for credit card fraud detection. The study focused on the increased cases of credit card fraud and the importance of proper fraud identification models in banks. The authors evaluated several of them: Decision Trees, support vector machines, and Neural Networks and compared them in terms of their accuracy in identifying fraudulent transactions. It also observed that feature extraction and data cleansing, which are critical to fraud detection, toned down the ratio of false positives and negatives. The results also helped to show how ML approaches can be used to improve financial security and confirm the effectiveness of data-driven approaches to combating financial risk.

In the research of Randhawa et al. (2018), the authors discussed the application of the ML approach in the context of improving fraud detection in financial transactions. The paper focused on the increase in credit card fraud and the difficulties in constructing efficient models for its identification, mainly due to the lack of large, genuine fraud datasets. AdaBoost and majority voting were used in combination with twelve ML algorithms that were implemented to help increase classification accuracy. Finally, the potential of the proposed ensemble approach was examined using well-known benchmarking datasets, as well as credit card datasets of real-life scenarios. Also, the study established that using data noise to test the models showed that a majority voting method was accurate regardless of

the level of noise. This study helped to discuss the possibilities of large hybrid ML systems in eliminating false positives and increasing the efficiency of fraud detection to improve financial transaction monitoring.

The study by Sharma et al. (2022) addressed the challenges of detecting fraudulent transactions and the limitations of traditional ML models in adapting to evolving fraud patterns. To enhance fraud detection accuracy, the researchers proposed an unsupervised deep learning approach using autoencoders (AE), implemented with TensorFlow. By analyzing datasets from European, Australian, and Asian credit card transactions, the study examined demographic variations in fraud patterns. The findings demonstrated that autoencoders effectively identified fraudulent transactions by learning intricate transaction behaviors, making them well-suited for dynamic fraud detection. This research emphasized the importance of deep learning (DL) in financial security, showcasing the adaptability of unsupervised models in detecting sophisticated fraud attempts and improving real-time transaction monitoring across financial institutions.

In the study by Bharati et al. (2022), the authors discussed various achievements and challenges of FL in retaining and separating privacy-preserving decentralized ML. Aimed at solving such issues, the study proposed an FL framework that primarily focused on addressing privacy issues while facilitating joint model training across devices, without raw data sharing. Some of the specific fields of interest include healthcare, wireless communication, and service recommendation, and FL aims to improve data protection and develop personalized AI-based services. Therefore, it identified major challenges like high communication costs, system heterogeneity, problems with model uploading, and statistical heterogeneity that affect model convergence and security. The study also discussed privacy-preserving methods, such as homomorphic encryption and secure multiparty computation, to increase FL security. This research was sufficient in presenting an updated understanding of FL's prospects and losses to guide more enhancements in privacy-conscious AI techniques.

The study by Yang et al. (2019), explores the limitations of traditional fraud detection systems and proposes a FL-based framework to enhance privacy-preserving credit card fraud detection. The authors highlight major challenges, including data privacy concerns, skewed class distribution, and real-time detection constraints, which hinder the effectiveness of existing fraud detection systems. To address these issues, the researchers introduce a federated fraud detection (FFD) framework, allowing multiple banks to collaboratively train a shared fraud detection model while preserving data confidentiality. The framework employs an oversampling approach to balance highly skewed datasets, ensuring improved fraud classification. Experimental evaluations using real-world European credit card transaction data demonstrate that the proposed FL-based model outperforms traditional fraud detection systems, achieving a higher AUC and F1-score. This research underscores the potential of FL in financial security, offering a scalable and privacy-aware approach to combating evolving fraud patterns in digital transactions.

The study by Doshi-Velez and Kim (2017), examined the increasing significance of interpretability in ML, particularly in high-stakes domains such as autonomous systems, healthcare, and financial decision-making. The authors highlighted that while ML models achieved high predictive accuracy, their black-box nature restricted trust, accountability, and regulatory compliance. The paper outlined key interpretability evaluation approaches, including application-grounded, human-grounded, and functionally grounded methods, each serving different levels of human involvement. It also discussed various issues within the method for interpreting interpretability, including the lack of a formal definition that consisted of the use of subjective measures, or at most simple heuristics such as sparsity. The authors also discussed the issue of the trade-off between model complexity and

interpretability of the given model and focused on the importance of future unified metrics for comparison of model interpretability. This research established a good foundation for developing interpretable ML and reaffirmed the need to integrate it into the safety and fairness of AI decision-making systems.

The authors of this research (Damanik & Liu, 2025) were able to consider major issues such as class imbalance and overfitting. To address these challenges, they proposed K-meensmoteenn, a combination of oversampling and SMOTE with re-sampling, which focused on synthesizing new instances of both fraudulent and legitimate transactions to overcome noisy synthetic data. Also, they used the stacking ensemble technique that involved combining many ML classifiers to create a better model. This result was better than other single models such as XGBoost as well as Decision Trees with an F1–score of 0.92 and the ROC-AUC of 1.00. The results were quite satisfactory with precision at 0.95, recall at 0.88, as well as AUPRC at 0.96, which would guarantee high accuracy of fraud detection. In addition, the researchers applied XAI through LIME to enhance the interpretability of key decisions and help the financial institutions involved. The research was useful in responding to data imbalance and ignorability assumptions to improve M2M fraud detection for higher predictive accuracy of financial security systems.

Table 1 defines the various fraud detection models in terms of methodology, goals, pre-processing, scalability, privacy preservation, and explainability. Several current strategies for fraud detection use traditional ML, ensemble models (IMEML, AdaBoost, and XGBoost), but none of them meet privacy concerns or offer interpretability. Existing approaches such as autoencoders have the potential to enhance the category of frauds by using DL methods, but they are computationally intensive and are black box. FL methods resolve privacy issues by allowing data processing at the edge but have communication costs and system dissimilarity. To overcome these limitations, the future proposed model, XFL, combines FL with XAI (SHAP & LIME); therefore, it provides privacy, transparency in the model, real-time fraud detection, and higher accuracy in the result.

Table 1 provides a comparative analysis of various fraud detection models, highlighting their strengths and limitations. However, despite the advancements in existing approaches, several challenges remain, as discussed in the following section.

Table 1. Comparative analysis of various fraud detection models.

Ref.	Model Used	Objective	Preprocessing Techniques	Predictive Model	Privacy-Preserving (FL)	Interpretability (XAI)	Scalability	Regulatory Compliance	Real-Time Fraud Detection	Strengths	Limitations
Talukder et al. (2024)	IMEML (EIC, EBC, EMC)	Handling data imbalance	IHT+EMC, cluster centroids, RUS	Ensemble Learning	×	×	Moderate	×	×	Reduces false positives and balances data.	High computational cost; may not generalize well.
Baghdadi et al. (2024) Fraud detection while balancing speed and accuracy	Not explicitly mentioned	Ensemble Learning (RBM + LSTM)	×	×	High	×	☑	×	×	High fraud detection accuracy; real-time processing	Dataset sharing limitations; high model complexity
Puh and Brkić (2019)	Decision Trees, SVM, Neural Networks	Identify fraudulent transactions using ML techniques.	Feature selection; data preprocessing	ML (Supervised)	×	×	Moderate	×	×	Reduces false positives and false negatives; enhances financial security	Performance varies across models and potential data imbalance issues
Randhawa et al. (2018)	AdaBoost, Majority Voting, 12 ML algorithms	Enhance fraud detection using ensemble learning.	Not explicitly mentioned	Ensemble Learning	×	×	High	×	×	Robust against data noise; reduces false positives	Requires diverse datasets; performance depends on dataset quality
Sharma et al. (2022)	Autoencoder (AE)	Improve fraud detection using unsupervised DL	Not explicitly mentioned	DL (Unsupervised)	×	×	High	×	☑	Adapts to evolving fraud patterns; effective anomaly detection	High computational cost; lacks explainability
Bharati et al. (2022)	FL	Address privacy concerns in decentralized ML	Not explicitly mentioned	FL	☑	×	High	☑	×	Ensures data privacy; enables collaborative learning	High communication costs; system heterogeneity
Yang et al. (2019)	Federated Fraud Detection (FFD)	Enhance privacy-preserving fraud detection	Oversampling for class balancing	FL	☑	×	High	☑	☑	Improves fraud classification; preserves data privacy	High computational cost; potential network delays
Doshi-Velez and Kim (2017)	Interpretable ML	Improve ML transparency and trust	Not explicitly mentioned	XAI	×	☑	Moderate	☑	×	Enhances fairness, and accountability	Reduced accuracy; lack of standard evaluation metrics

Table 1. *Cont.*

Ref.	Model Used	Objective	Preprocessing Techniques	Predictive Model	Privacy-Preserving (FL)	Interpretability (XAI)	Scalability	Regulatory Compliance	Real-Time Fraud Detection	Strengths	Limitations
Damanik and Liu (2025)	K-means SMOTEEN, Stacking Ensemble (XGBoost, Decision Trees)	Improve fraud detection accuracy by handling class imbalance	K-means SMOTEENN (resampling for data balancing)	Ensemble Learning	×	☑	High	×	×	High fraud detection accuracy; enhanced interpretability	Computationally expensive; dependency on feature quality
Proposed Model	XFL with SHAP and LIME	Enhance fraud detection with privacy-preserving and XAI	Not explicitly mentioned	FL + XAI	☑	☑	High	☑	☑	Ensures privacy; high fraud detection accuracy; improves interpretability	Higher computational cost; dependency on data quality; potential latency in federated updates

Legend: Symbols used—☑ indicates presence or applicability; × indicates absence or non-applicability.

3. Limitations of Previous Research Works

Despite advancements in AI-driven fraud detection, existing models have several shortcomings that limit their effectiveness. The key challenges identified in the literature are as follows:

3.1. Lack of Privacy-Preserving Mechanisms

Most fraud detection models rely on centralized AI, posing significant risks of data breaches, privacy violations, and regulatory non-compliance. Many studies (Randhawa et al., 2018; Puh & Brkić, 2019; Sharma et al., 2022; Baghdadi et al., 2024; Talukder et al., 2024) applied ML/DL models without privacy-preserving frameworks, exposing sensitive financial data to potential threats.

3.2. Lack of Model Transparency and Explainability

While several fraud detection models (Puh & Brkić, 2019; Randhawa et al., 2018; Sharma et al., 2022) achieved high classification accuracy, they lacked explainability, making them difficult to interpret and trust. Black-box AI models, including DL-based fraud detection systems, fail to provide clear reasoning behind decisions, limiting their acceptance in financial systems. Even FL-based research (Yang et al., 2019; Bharati et al., 2022) failed to integrate XAI, restricting regulatory compliance and decision validation.

4. Contribution of the Proposed Work

To address these limitations, this research introduces an XFL model that enhances both privacy protection and interpretability in fraud detection. The major contributions are as follows:

4.1. Privacy-Preserving AI with FL

It is crucial to have secure and compliant effective fraud detection for the safety of the financial institution as well as the customer. In contrast to the conventional models, the XFL model incorporates FL which enables learning across multiple banks while excluding raw data thus mitigating privacy issues that would lead to compliance with financial laws (Randhawa et al., 2018; Yang et al., 2019; Puh & Brkić, 2019; Bharati et al., 2022; Sharma et al., 2022; Talukder et al., 2024; Baghdadi et al., 2024).

4.2. Enhancing Transparency with XAI

To address the black-box nature of previous models of fraud detection, this study uses SHAP and LIME-based XAI methods to explain the decision-making clearly and understandably. This helps make AI-based fraud classifications reliable for financial analysts and tackles the issues of transparency in prior works (Puh & Brkić, 2019; Randhawa et al., 2018; Sharma et al., 2022) and even FL-based works (Yang et al., 2019; Bharati et al., 2022) where explainability was an issue.

While numerous AI advancements have been made for fraud detection purposes, ML and FL remain problematic in some ways. Almost all the FL models are black-box models, thus, providing FL analysts with limited means of interpreting the reasons a certain fraud has been classified in a particular way. Although XAI methodologies improve interpretability, it is still uncommon to be incorporated with FL-based fraud detection solutions. To fill these gaps, this research put forward an XFL model that enables both high detection accuracy and interpretability with data privacy measures enacted.

5. Proposed Methodology

Financial fraud is one of the pervasive problems that affect organizations around the world; rule-based and centralized ML models fail to counter the emergence of new fraud schemes due to high rates of false positives, lack of flexibility, and data privacy violations. Although FL overcomes the issue of sharing datasets through model learning without data sharing, it has drawbacks such as interpretability, communication overhead, and slow convergence time. To overcome these limitations, this research proposes an XFL model, integrating FL with SHAP and LIME-based XAI techniques for transparent, high-accuracy fraud detection (Rehman & Hashim, 2020; Abbas et al., 2024; Saleem et al., 2024; Shahzad et al., 2024). The XFL model ensures privacy, enhances trust, improves real-time fraud detection, and strengthens regulatory compliance, offering a robust and scalable AI-driven financial security solution. As depicted in Figures 3 and 4, the proposed XFL model progresses from data acquisition and preprocessing to local model training and validation, and finally to a federated fraud detection framework, ensuring privacy, interpretability, and scalability in combating financial fraud.

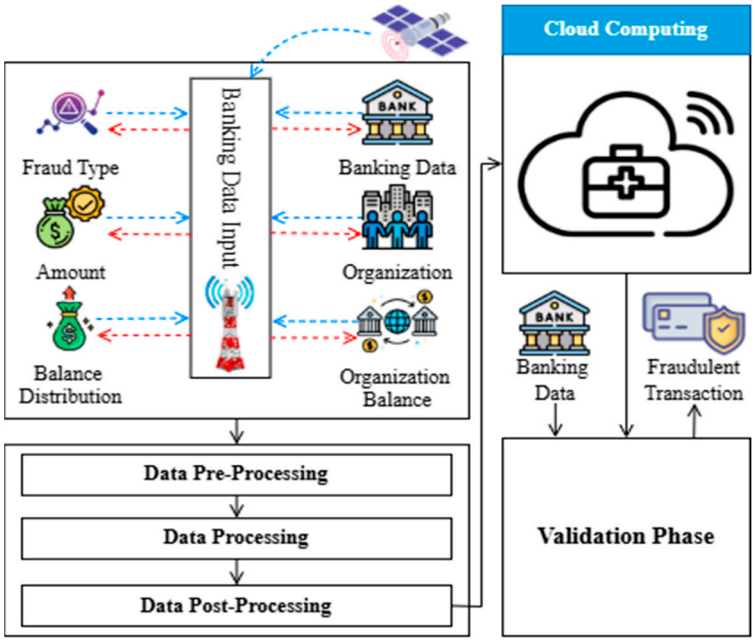

Figure 3. Fraud detection process from data input to validation.

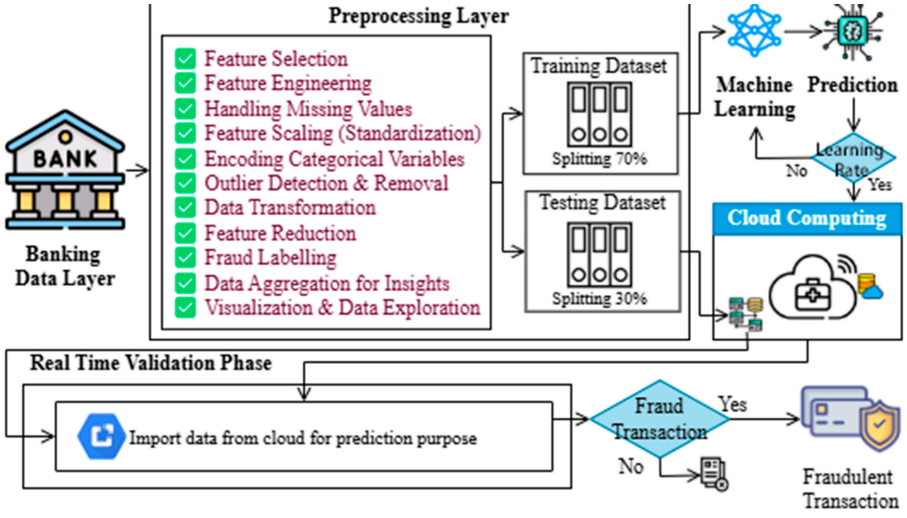

Figure 4. Abstract view of the proposed XFL-based financial fraud detection model (local server).

Figure 3 shows the fraud detection pipeline, beginning with banking data acquisition, where key financial attributes such as fraud type, transaction amount, and balance distribution are collected from various banking institutions and organizations. These data undergo a three-stage processing framework: pre-processing (cleaning and structuring data), processing (applying fraud detection techniques), and post-processing (finalizing structured data for validation). The refined data are then sent to the validation phase, where cloud computing infrastructure analyzes and classifies transactions as fraudulent. This cloud-based validation ensures real-time fraud detection, scalability, and accuracy, forming the foundation for an AI-driven, secure financial fraud prevention system.

Figure 4 illustrates the local model flow for fraud detection in the proposed XFL model (local server). The process begins with the Banking Data Layer, where transaction datasets (Financial Fraud Detection Dataset., n.d.) from financial institutions are collected and forwarded to the Preprocessing Layer. Table 2 shows the features of the dataset.

Table 2. Dataset features description: Financial Fraud Detection Dataset. (n.d.).

Sr. No.	Features	Description
1	step	int64
2	type	object
3	amount	float64
4	nameOrig	object
5	oldbalanceOrg	float64
6	newbalanceOrig	float64
7	nameDest	object
8	oldbalanceDest	float64
9	newbalanceDest	float64
10	isFraud	int64

5.1. Dataset Description

This layer performs essential operations such as feature selection, feature engineering, handling missing values, feature scaling, encoding categorical variables, outlier detection and removal, data transformation, feature reduction, fraud labeling, data aggregation, and visualization to refine the raw financial data.

5.2. Exploratory Data Analysis (EDA)

Figure 5 represents a pie chart distribution of different transaction types, visualizing their relative proportions in the dataset. The chart highlights five transaction categories: CASH_OUT (35.17%) and PAYMENT (33.81%) as the most frequent transaction types, followed by CASH_IN (21.99%), TRANSFER (8.38%), and a minor proportion of DEBIT transactions (0.65%). The exploded sections emphasize the variations in transaction volumes, ensuring clear distinction. This visualization helps in identifying transaction patterns, which is crucial for detecting anomalies and fraudulent activities in financial transactions.

Figure 6 presents a horizontal bar chart displaying the total monetary value (in billions) for each transaction type. The TRANSFER category records the highest transaction volume (USD 485.292 billion), followed by CASH_OUT (USD 394.413 billion) and CASH_IN (USD 236.367 billion). Conversely, PAYMENT transactions account for only USD 28.093 billion, while DEBIT transactions contribute a minimal USD 0.227 billion. This visualization highlights the distribution of financial flow across different transaction types, providing key insights for fraud detection and financial analysis.

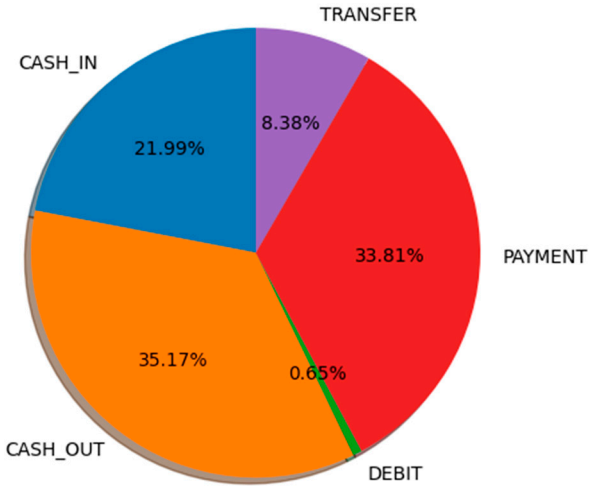

Figure 5. Transaction type distribution with percentages.

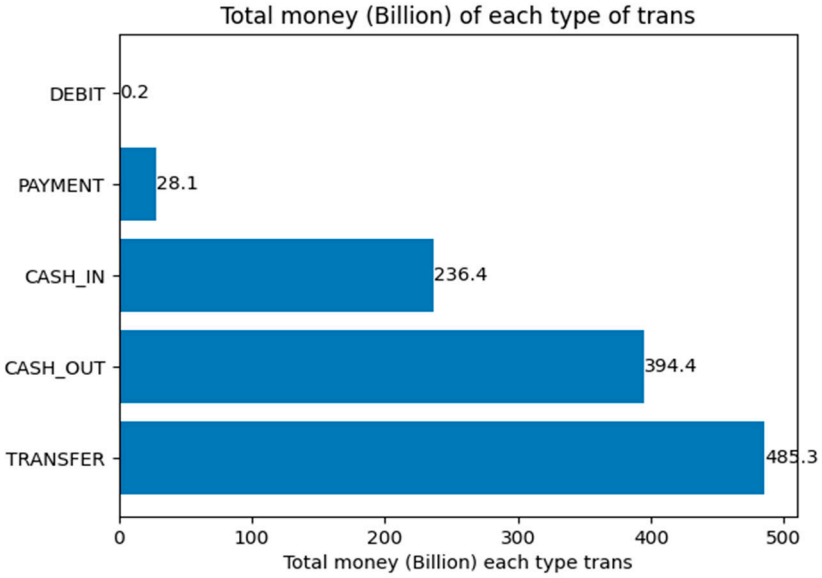

Figure 6. Total monetary value (in billions) across transaction types.

Figure 7 presents a comparative analysis of fraud distribution through two donut charts, illustrating the fraud percentage in total transactions (left) and total monetary value (right). The left chart shows that fraudulent transactions make up only 4.02% of all transactions, while the vast majority (95.98%) are legitimate. However, the right chart highlights that fraud accounts for 11.30% of the total monetary value, revealing that although fraud cases are relatively low in number, they involve significantly higher transaction amounts. Such a disparity gives a vivid picture of how much fraud affects the financial position of an organization thereby promoting the importance of adopting enhanced methods of detecting and preventing fraud.

Figure 8 is a multi-level donut chart, which shows the share of fraudulent and non-fraudulent activities in a total number of transactions and various types of transactions. Concerning the external taxonomy, the outer circle separates the transaction type from its fraud type: CASH_OUT, CASH_IN, TRANSFER, PAYMENT, and DEBIT. In addition, the CASH_OUT and TRANSFER categories have a significantly higher level of fraud compared to the other categories, which include PAYMENT, CASH_IN, and DEBIT, most of which are legitimate transactions. This type of visual representation is helpful in the process of

determining the specific types of transactions that are most likely to be fraudulent to aid in reducing fraud.

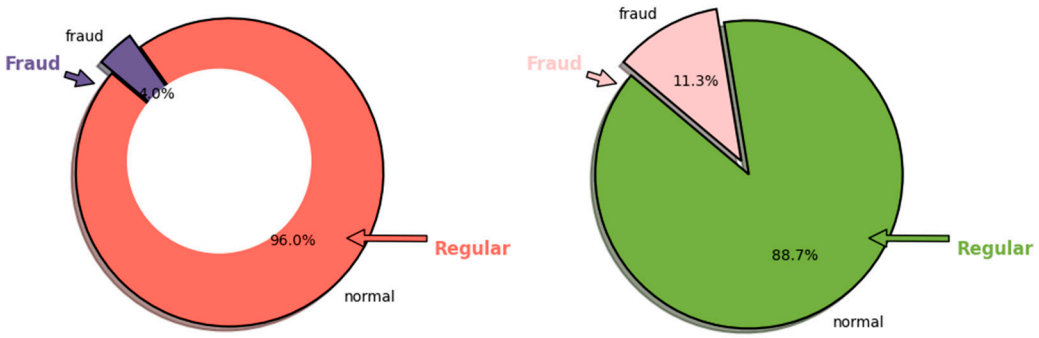

Figure 7. Comparison of fraud percentage in total transactions vs. total monetary value.

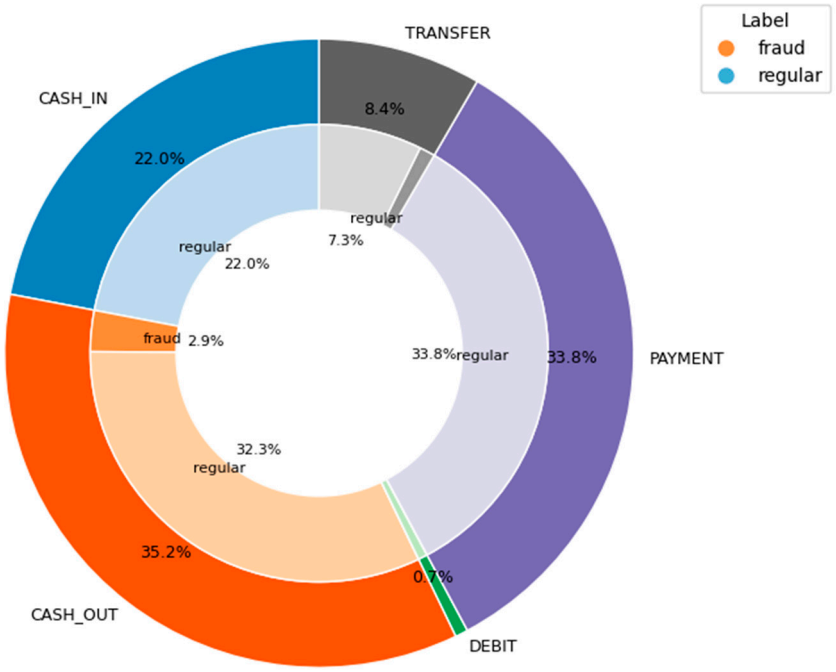

Figure 8. Fraud and non-fraud distribution across transaction types.

Figure 9 shows an established distribution of the numerical transaction features after the removal of the outliers to gain the best framework for fraud detection. The following histograms represent the cleaned-up distributions of the five financial variables: transaction amount, original balance, destination balance, and balance differences. The IQR method was used in cleaning the data by removing the outliers that may affect the classification of fraud hence making the model more reliable. This preprocessing step is a critical step for better quality of data, more accurate identification of fraud incidences, and fewer false alarms of financial transactions.

Figure 10 presents a pair plot visualization illustrating the relationships between key financial attributes (transaction amount, old balance, new balance) and their association with fraud detection. The orange points represent fraudulent transactions, while the blue points indicate legitimate transactions. The plot reveals distinct fraud patterns, particularly in old and new balance correlations, where fraudulent transactions often exhibit outlier-like behavior. This visualization helps in identifying fraud-prone data patterns, assisting ML models in distinguishing between fraudulent and non-fraudulent transactions.

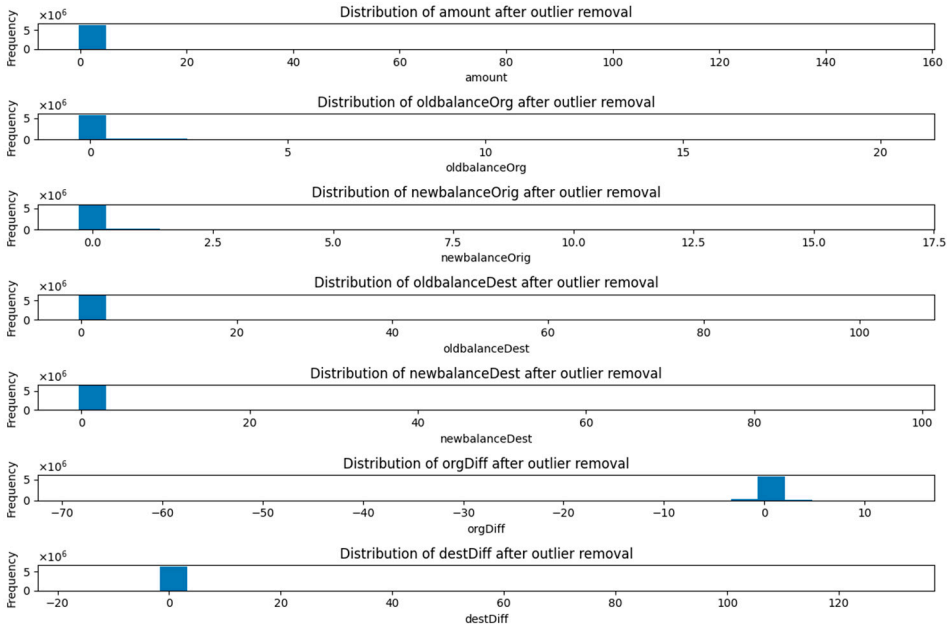

Figure 9. Distribution of key financial features after outlier removal.

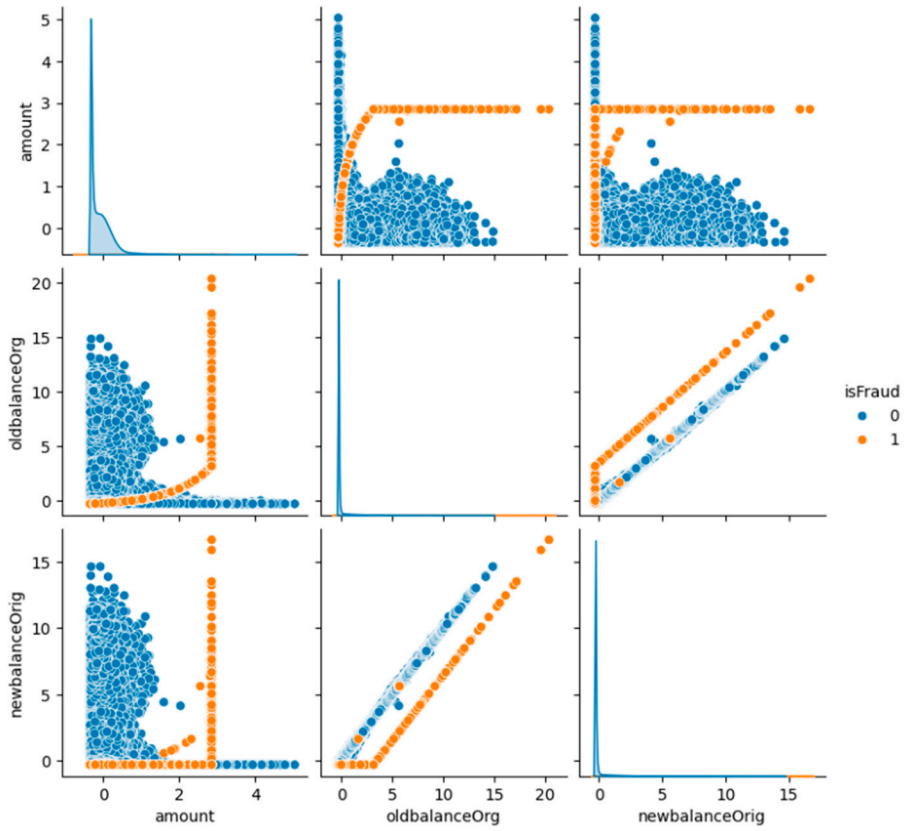

Figure 10. Pair plot analysis of transaction features highlighting fraud patterns.

Figure 11 presents a correlation matrix heatmap, illustrating the relationships between key financial features in the dataset. The color scale ranges from blue (negative correlation) to red (positive correlation), highlighting how different attributes influence each other. Strong correlations are observed between oldbalanceOrg and newbalanceOrig, as well as oldbalanceDest and newbalanceDest, indicating expected financial transitions. The 'isFraud' feature shows weak correlations with most attributes, suggesting that fraud

detection relies on complex patterns rather than direct linear relationships. This heatmap helps in feature selection and understanding dependencies for fraud classification models.

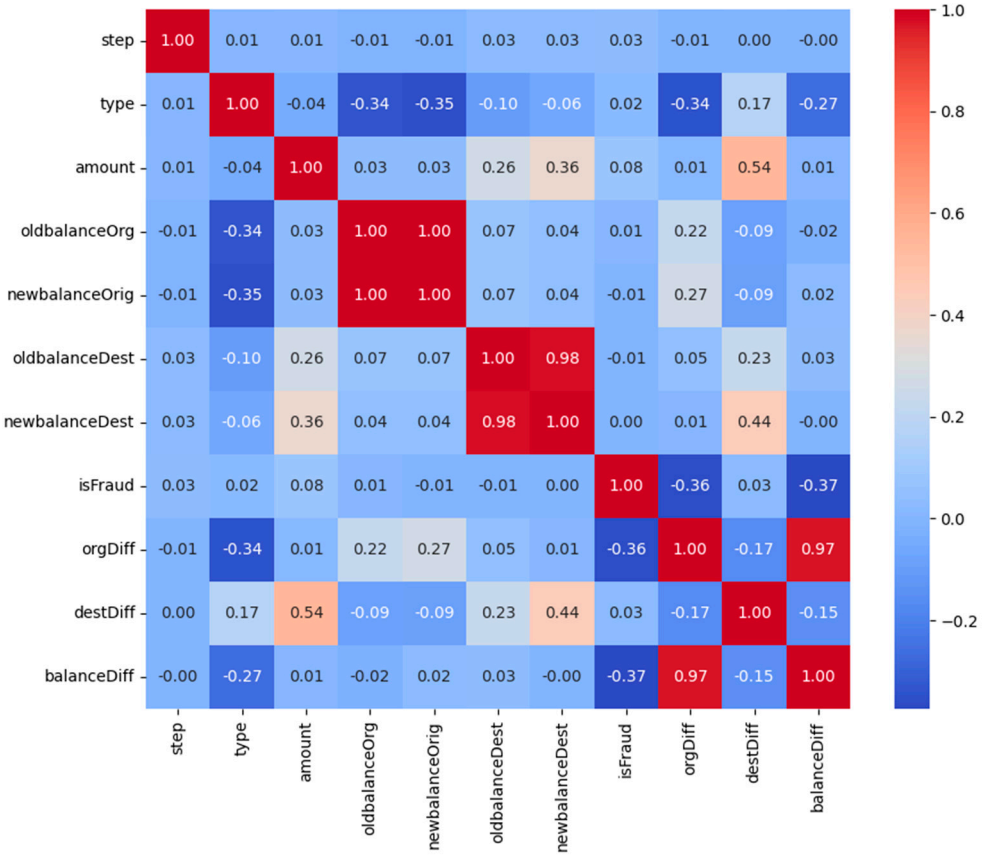

Figure 11. Correlation matrix of financial transaction features for fraud detection insights.

Figure 12 displays a boxplot comparing transaction amounts between legitimate (0) and fraudulent (1) transactions. Fraudulent transactions (1) generally involve higher amounts, whereas legitimate transactions (0) show lower values with some extreme outliers. This distinction in transaction behavior aids in detecting fraud patterns based on transaction value.

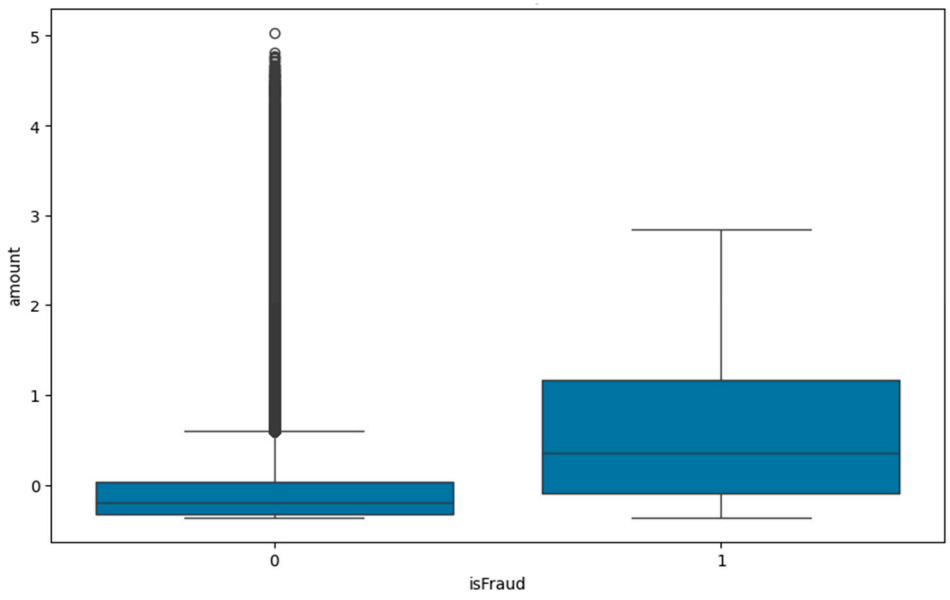

Figure 12. Boxplot of transaction amount distribution for fraudulent and non-fraudulent cases.

Figure 13 illustrates the relationship between the old balance (x-axis) and new balance (y-axis) of the originating account, distinguishing between fraudulent (1) and legitimate (0) transactions using different colors. Fraudulent transactions (1) are clustered in a distinct lower segment, indicating cases where the new balance remains disproportionately low despite significant old balance values, a common fraud indicator. In contrast, legitimate transactions (0) follow a more linear trend, where the new balance is proportionate to the old balance. This pattern helps in identifying fraud by analyzing balance discrepancies.

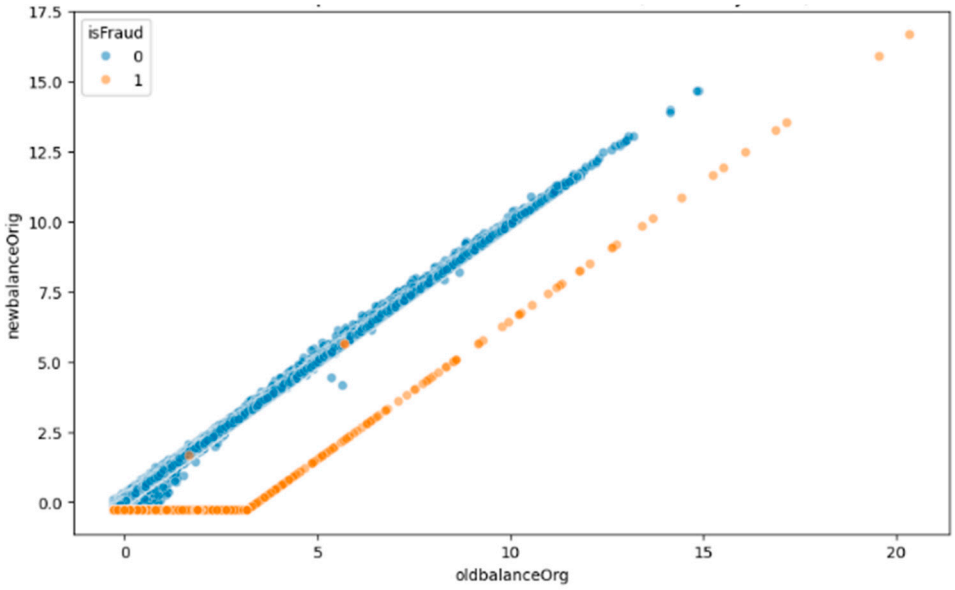

Figure 13. Scatter plot of old and new balances for fraudulent (1) and legitimate (0) transactions.

5.3. Local Model Training and Validation Process

The preprocessed data are split into training (70%) and testing (30%) datasets, where the training set is used to develop ML models at the local banking level to detect fraudulent transactions. These models undergo iterative training, adjusting parameters until an optimal learning rate is achieved. If the required accuracy is not met, the model continues to be refined using localized banking transaction data.

Once the model reaches the desired accuracy, the trained local model is stored on the local server's cloud storage, ensuring secure and scalable processing. The trained models are validated using the testing dataset, allowing for the assessment of fraud detection performance before deployment. The real-time validation phase ensures that the trained local model correctly identifies fraudulent and legitimate transactions, minimizing false positives and false negatives.

While local models efficiently process financial transactions with privacy preservation, their effectiveness is further enhanced by global model aggregation, where insights from multiple local banking institutions are synchronized for improved fraud detection accuracy. Figures 3 and 14 depict the local server-client interaction, where each bank processes transactions independently before integrating into the federated fraud detection model. Table 3 outlines the pseudocode detailing the step-by-step process, from data collection and preprocessing to fraud detection, validation, and synchronization with the global model.

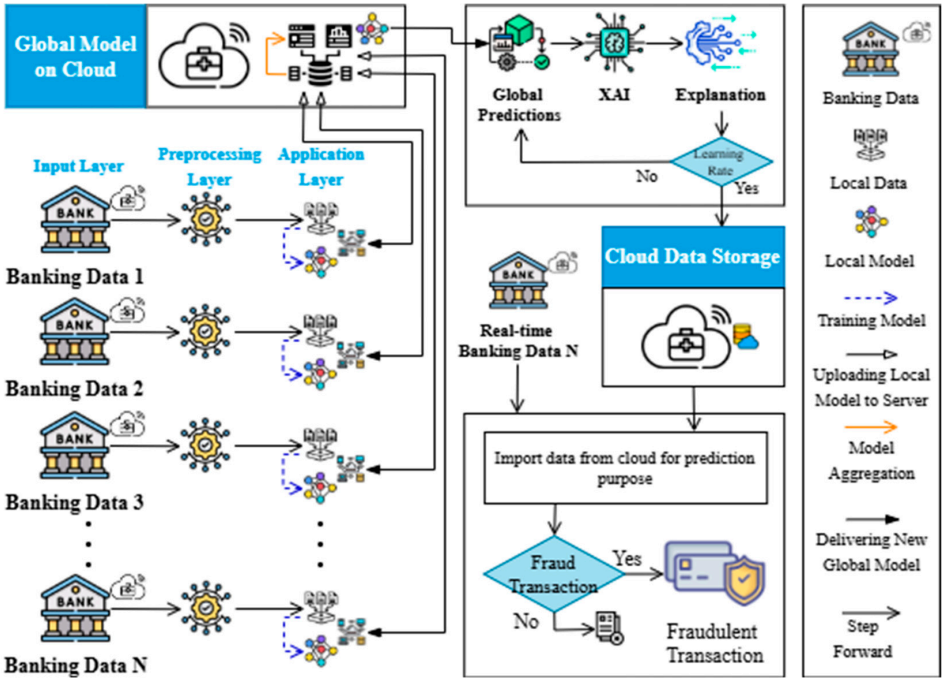

Figure 14. Proposed XFL-based financial fraud detection model (global server).

Table 3. Pseudocode of proposed local server client model for financial fraud detection.

Step	Process
1	Start
2	Data Collection : Gather real-time banking transactions $D = \{d_1, d_2, ..., d_n\}$ (e.g., transaction type, amount, sender and receiver balance).
3	Preprocessing: ☑ Feature Selection ☑ Feature Engineering ☑ Handling Missing Values ☑ Feature Scaling (Standardization) ☑ Encoding Categorical Variables ☑ Outlier Detection and Removal ☑ Data Transformation ☑ Feature Reduction ☑ Fraud Labeling ☑ Data Aggregation for Insights ☑ Visualization and Data Exploration
4	Split Data : Partition dataset into Training $T = 70\%$ and Testing $V = 30\%$.
5	Model Training : Initialize ML models $M = \{M_1, M_2, \ldots, M_k\}$ for fraud classification.
6	Iterative Training: Adjust model parameters $W = W - \eta \nabla W$, optimize learning rate, and retrain models on T until convergence $(E\_train < threshold)$.
7	Validation : Evaluate model $M(V)$ with accuracy $A = f(M, V)$. If $A < threshold$, retrain the model.
8	Store Trained Model : Save optimal parameters M^* to the local banking server.
9	Global Model Sync (if required) : Send M^* to the global system for aggregation.
10	Return Predictions $P = M^*(X)$ to banking professionals/clients.
11	Stop

5.4. Federated Learning and Global Model Aggregation

Figure 14 illustrates the Proposed XFL workflow, where financial transactions from multiple banks are processed independently before contributing to a global fraud detection model. Each local bank (C_i) trains its model using its dataset (D_i) in a privacy-preserving manner, applying preprocessing, feature extraction, and fraud detection model training. The local model updates W_i^t are computed as:

$$W_i^{t+1} = W_i^t - \eta \nabla L(W_i^t, D_i) \tag{1}$$

where W_i^t represents the model weights at iteration t, η is the learning rate, and $\nabla L(W_i^t, D_i)$ is the gradient of the loss function L concerning the local dataset D_i.

Each trained local model sends only its weight updates to the global model, ensuring that no raw transaction data are shared between institutions. The global model is updated by selecting the local model that achieves the highest evaluation score:

$$W^* = \arg\max_{W_i} A(W_i, V_i) \tag{2}$$

where W^* represented the aggregated global model chosen based on performance $A(W_i, V_i)$ of local model W_i on its validation dataset V_i. If no model meets the performance threshold τ, local clients retrain their models with adjusted hyperparameters to improve global accuracy.

If $A(W^*, V^*) < \tau$, then W^* is retrained using additional rounds. This approach ensures that only the best-performing models are aggregated, improving fraud detection across all banks while preserving privacy.

To ensure that the model is trusted and understandable, it incorporates XAI methodologies like SHAP and LIME. SHAP explains each feature by estimating its importance to the model and the degree to which it contributes to the detection of fraud:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{j\}) - f(S)] \tag{3}$$

where ϕ_j is the SHAP value for feature j, F is the feature set, $f(S)$ is the model output when using only feature subset S. To explain fraud predictions for an individual, LIME utilizes a simplified model, which is an approximation of the global model:

$$\hat{f}(x) = arg \min_{g \in G} L(f, g, \pi_x) + \Omega(g) \tag{4}$$

where $f(x)$ is the original model's prediction, g is the interpretable surrogate model, G represents the set of interpretable models from which g is selected, $L(f, g, \pi_x)$ measures how well g approximates f locally, and $\Omega(g)$ is a regularization term that controls model complexity, preventing overfitting and ensuring interpretability by penalizing overly complex functions.

Once optimized, the global model is deployed for real-time fraud detection across institutions. A new transaction X' is classified as fraudulent or legitimate based on the global model:

$$P\left(\text{fraud} \mid X'\right) = f\left(W^{t+1}, X'\right) \tag{5}$$

where $P\left(\text{fraud} \mid X'\right)$ is the probability that X' is fraudulent, and X', $f\left(W^{t+1}, X'\right)$ is the prediction using the global model weights W^{t+1}. If $P(fraud \mid X') > \theta$ (where θ is the decision threshold), then the transaction is flagged as fraudulent, else it is discarded.

This federated framework enhances privacy, security, and interpretability, ensuring that AI-driven fraud detection remains compliant with financial regulations. Table 4 presents the pseudocode of the global server-client model, detailing the key processes from local model aggregation to fraud classification.

Table 4. Pseudocode of the proposed global server-client model for financial fraud detection.

Step	Process
1	Start
2	Initialize global model W^* and set performance threshold τ.
3	Receive trained models W_i from local clientsC_i, where each W_i was trained on the dataset D_i.
4	Evaluate each local model W_i using validation dataset V_i and compute accuracy: $\mathrm{A}(W_i, V_i)$.
5	Select the best-performing model : $W^* = \arg\max_{W_i} A(W_i, V_i)$
6	Check convergence : If $A(W^*, V^*) \geq \tau$, proceed to model deployment.
7	If not converged: Request additional training from local clients with adjusted hyperparameters.
8	Apply XAI techniques (SHAP and LIME) for interpretability
9	SHAP calculation : Compute ϕ_j values using the feature impact formula.
10	LIME Interpretation : Train surrogate model g to approximate f locally.
11	Deploy final global fraud detection model W^*.
12	For a new transaction $X^{'}$, predict fraud probability : $\mathrm{P}\left(\text{fraud} \mid \mathrm{X}^{'}\right)$
13	Decision threshold : If $P\left(fraud \mid X^{'}\right) > \theta$
14	Securely store validated predictions in cloud storage.
15	Stop

6. Simulation Results

With the rise in sophisticated financial fraud schemes, traditional fraud detection systems struggle to balance accuracy, privacy, and interpretability. To address these challenges, the proposed XFL model is simulated on Google Colab in a decentralized banking environment, where multiple financial institutions serve as local clients, training models independently without sharing raw transaction data. A real-world financial fraud dataset (Financial Fraud Detection Dataset., n.d.) is used, with 70% allocated for training and 30% for testing, ensuring a comprehensive evaluation. The model's performance is assessed using key fraud detection metrics such as Accuracy, Sensitivity (TPR), Specificity (TNR), Miss-rate (FNR), Positive Predictive Value (PPV), and Negative Predictive Value (NPV), as described in Equations (6) through (11).

$$\text{Accuracy} = \frac{\sum \text{True Positive} + \sum \text{True Negative}}{\sum \text{Total Population}} \tag{6}$$

$$\text{Sensitivity} = \frac{\sum \text{True Positive}}{\sum \text{Condition Positive}} \tag{7}$$

$$\text{Specificity} = \frac{\sum \text{True Negative}}{\sum \text{Condition Negative}} \tag{8}$$

$$\text{Miss} - \text{Rate} = 1 - \text{Accuracy} \tag{9}$$

$$\text{Positive Predictive Value} = \frac{\sum \text{True Positive}}{\sum \text{Predicted Condition Positive}} \tag{10}$$

$$\text{Negative Predictive Value} = \frac{\sum \text{True Negative}}{\sum \text{Predicted Condition Negative}} \quad (11)$$

Table 5 summarizes the confusion matrix results for the proposed XFL-based fraud detection model, comparing GBM, SVM, and LR. All models effectively detect fraudulent transactions, with GBM achieving the best balance, having high True Positives (1,270,835) and the lowest False Negatives (647), ensuring minimal undetected fraud cases. SVM exhibits zero False Positives, making it highly precise, but at the cost of a higher False Negative rate (1497). Logistic Regression (LR) shows moderate performance, with more misclassifications compared to GBM and SVM. Overall, GBM emerges as the most reliable model, offering optimal fraud detection with minimal errors.

Table 5. Confusion matrix of proposed XFL-based financial fraud detection model.

Confusion Matrix			
	GBM	**SVM**	**LR**
True Positive (TP)	1,270,835	1,270,853	1,270,745
True Negative (TN)	997	147	719
False Positive (FP)	018	000	108
False Negative (FN)	647	1497	925

Table 6 presents the performance metrics of the proposed XFL-based fraud detection model using GBM, SVM, and LR. GBM achieves the highest accuracy (99.95%) and sensitivity (99.95%), ensuring effective fraud detection with minimal false negatives. SVM excels in specificity (100%) and PPV (100%) but struggles with low NPV (8.94%), making it less reliable for identifying non-fraudulent cases. LR offers a balanced performance but falls behind in specificity (86.94%) and NPV (43.73%). Overall, GBM emerges as the most robust model for accurate and reliable fraud detection.

Table 6. Performance matrices of the proposed XFL-based financial fraud detection model.

Performance Matrices	**Algorithms**		
	GBM	**SVM**	**LR**
Accuracy	99.95	99.88	99.92
Sensitivity (TPR)	99.95	99.88	99.93
Specificity (TNR)	98.23	1	86.94
Miss-rate (FNR)	0.05	0.12	0.08
Positive Predictive Value (PPV)	1	1	99.99
Negative Predictive Value (NPV)	60.64	8.94	43.73

After applying FL, GBM emerged as the best-performing model, demonstrating superior accuracy (99.95%) and sensitivity (99.95%) in fraud detection. Due to its high reliability and minimal false negatives, it was selected as the global model for federated aggregation. This ensures that fraud detection benefits from collaborative learning while maintaining data privacy across institutions.

Figure 15 illustrates the explainability of the global model using LIME (Local Interpretable Model-agnostic Explanations) in financial fraud detection. The prediction probabilities indicate that the transaction is classified as "Not Fraud" with a probability of 93% and "Fraud" with only 7% confidence, highlighting the model's decision boundary. The middle section presents a Decision Tree-like breakdown, where key features such as oldbalanceOrg, orgDiff, transaction type, step, and amount influence the classification. The rightmost table ranks these features by their contribution, showing oldbalanceOrg and transaction type as the most influential factors in determining whether a transaction is fraudulent. This LIME-generated explanation provides transparency into the model's decision-making, ensuring interpretability and trust in fraud detection.

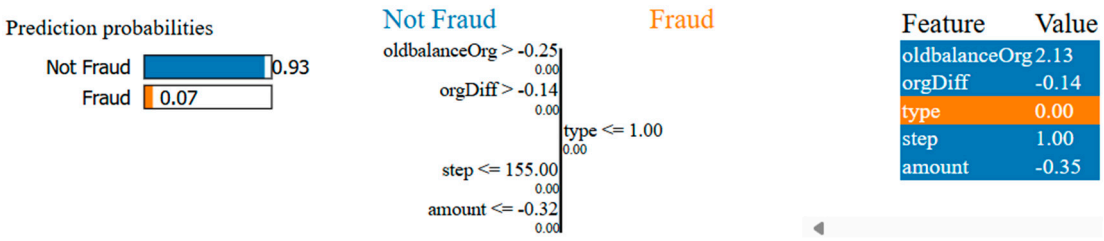

Figure 15. LIME explanation for proposed XFL-based financial fraud detection model.

Figure 16 is the SHAP summary plot that illustrates the impact of key features on the global fraud detection model's predictions. The x-axis represents SHAP values, showing each feature's contribution to fraud or non-fraud classification, while the y-axis ranks features by importance. OldbalanceOrg, orgDiff, and newbalanceOrig are the most influential factors. The color gradient (blue to red) indicates feature values, with red representing high values and blue having low values. This analysis enhances model transparency, helping identify the most critical fraud indicators.

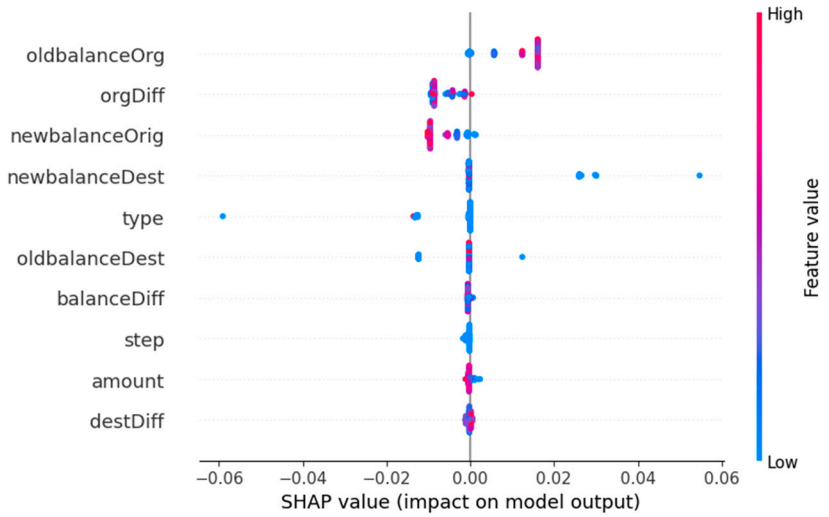

Figure 16. SHAP explanation for proposed XFL-based financial fraud detection model.

Table 7 compares various fraud detection models based on accuracy and miss rate, highlighting the effectiveness of different ML techniques. Traditional models like AE (81.6%), VAE (93.8%), and FL+ID3 (89%) show moderate accuracy with relatively higher miss rates. In contrast, the proposed XFL model (XAI + FL) achieves the highest accuracy (99.95%) with the lowest miss rate (0.05%), demonstrating its superior fraud detection capability, enhanced privacy preservation, and improved interpretability over existing approaches (Cecchini et al., 2010; Glancy & Yadav, 2011; Humpherys et al., 2011; Behera & Panigrahi, 2015; Kazemi & Zarrabi, 2017; Askari & Hussain, 2017; Kirkos et al., 2007; Sweers et al., 2018).

Table 7. Comparison of the proposed XFL-based financial fraud detection model.

References	Model	Accuracy (%)	Miss-Rate (%)
Behera and Panigrahi (2015)	FCM-MLP	94	6
Kazemi and Zarrabi (2017)	AE	81.6	18.4
Sweers et al. (2018)	VAE	93.8	6.2
Kirkos et al. (2007)	DT, NN, BNN	DT = 73.6, NN = 80, BNN = 90.3	DT = 26.4, NN = 20, BNN = 9.7
Cecchini et al. (2010)	Ontology + WN	83.87	16.3
Humpherys et al. (2011)	LR, NB, SVM, C4.5, LWL	LR = 63.4, NB = 67.3, SVM = 65.8, C4.5 = 67.3, LWL = 60.4	LR = 36.6, NB = 32.7, SVM = 34.2, C4.5 = 32.7, LWL = 39.6
Glancy and Yadav (2011)	CFDM	90.9	9.1
Askari and Hussain (2017)	FL+ID3	89	11
Proposed XFL-based financial fraud detection model	XAI+FL	99.95	0.05

7. Conclusions

Financial fraud detection faces significant challenges, including data privacy risks, high false positive rates, lack of model interpretability, and evolving fraud tactics that traditional rule-based and centralized AI models struggle to address. Traditional fraud detection techniques and models are semi-effective, but they are rigid, highly regulated, and lack transparency in their decision-making, thus proving to be unfit for use in complex financial environments.

To address these issues, this research has proposed the use of FL to enable distributed and private fraud detection, as well as utilizing XAI methods such as SHAP and LIME to facilitate the interpretation of the model. It was trained across several financial institutions but with the assurance that raw data would not be exchanged between institutions. According to the comparative analysis of the developed predictive models, the global model, namely, GBM outperformed other ML technologies, in terms of specificity, sensitivity, and accuracy. The integration with XAI helped to increase the level of trust in the fraud detection decisions and compliance with the regulation. The results approve the proposed XFL has better conveys fraud detection with benchmark accuracy of 99.95% over the previous works (Cecchini et al., 2010; Glancy & Yadav, 2011; Kirkos et al., 2007; Behera & Panigrahi, 2015; Humpherys et al., 2011; Kazemi & Zarrabi, 2017; Askari & Hussain, 2017; Sweers et al., 2018) and incorporates the credibility and confidentiality in the big financial institutions, which allow XFL as a radical solution for financial security. This proposed model can be seamlessly integrated into financial institutions via a cloud-based infrastructure, ensuring scalability, real-time fraud detection, and regulatory compliance.

8. Limitations and Future Considerations

The proposed XFL model ensures privacy-preserving and interpretable fraud detection, but it has certain practical challenges. Federated Learning (FL) introduces computational overhead, requiring efficient resource allocation. Data heterogeneity across financial institutions may impact model generalization, while latency in model aggregation could delay real-time fraud detection. XAI techniques (SHAP and LIME) improve transparency, but their interpretations may still require refinement for non-technical users. Additionally, cross-border regulatory compliance remains a challenge due to varying data privacy laws.

Future research should focus on enhancing FL efficiency through gradient compression and selective client updates to reduce computational costs. Adaptive learning mechanisms can improve fraud detection by dynamically adjusting to evolving fraud tactics. Further advancements in XAI techniques could provide more intuitive fraud explanations. Ad-

ditionally, developing customized FL frameworks aligned with regional regulations will facilitate broader adoption across financial institutions.

Author Contributions: S.K.A. and S.J.A. have collected data from different resources and contributed to writing—original draft preparation. S.K.A., H.D. and M.A.K. performed formal analysis and simulation. H.D. and M.A.K. performed writing—review and editing. M.A.K. performed supervision. S.K.A. drafted pictures and tables. S.J.A. and H.D. performed revision and improved the quality of the draft. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The simulation files/data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Abbas, S., Qaisar, A., Farooq, M. S., Saleem, M., Ahmad, M., & Khan, M. A. (2024). Smart vision transparency: Efficient ocular disease prediction model using explainable artificial intelligence. *Sensors*, *24*(20), 6618. [CrossRef] [PubMed]

Abolarin, J. (2025). *Banking law and financial regulations: The imperatives for managing stability in the banking sector*. Available online: https://books.google.com/books?hl=en&lr=&id=Ces9EQAAQBAJ&oi=fnd&pg=PP11&dq=Banking+Law+and+Financial+Regulations:+The+Imperatives+for+Managing+&ots=RYehI4bpju&sig=s0iGlQW9gEJHdyxk0yP5dScStCc (accessed on 2 November 2024).

Al-dahasi, E. M., Alsheikh, R. K., Khan, F. A., & Jeon, G. (2025). Optimizing fraud detection in financial transactions with machine learning and imbalance mitigation. *Expert Systems*, *42*(2), e13682. [CrossRef]

Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Computer Science Review*, *40*, 100402. [CrossRef]

Askari, S. M. S., & Hussain, M. A. (2017, May 5–6). *Credit card fraud detection using fuzzy ID3*. IEEE International Conference on Computing, Communication and Automation, ICCCA 2017 (pp. 446–452), Greater Noida, India. [CrossRef]

Baghdadi, P., Korukoglu, S., Bilici, M. A., & Onan, A. (2024). Ensemble learning approach using energy-based RBM and xLSTM for predictive analytics in credit card fraud detection. *Authorea preprints*. [CrossRef]

Barker, L. (2024). *Managing a company in an environment with significant corruption*. Available online: https://www.proquest.com/openview/964c25d7eac52c3156f88664a095096f/1?cbl=18750&diss=y&pq-origsite=gscholar (accessed on 4 March 2025).

Barnes, J. (2020). Fraud detection: Forensic accounting education and CFE designation impact on auditor's confidence levels. *Journal of Accounting and Finance*, *20*(4), 62–75. Available online: http://www.na-businesspress.com/JAF/JAF20-4/5_BarnesFinal.pdf (accessed on 4 March 2025).

Behera, T. K., & Panigrahi, S. (2015, May 1–2). *Credit card fraud detection: A hybrid approach using fuzzy clustering & neural network*. 2015 2nd IEEE International Conference on Advances in Computing and Communication Engineering, ICACCE 2015 (pp. 494–499), Dehradun, India. [CrossRef]

Bharati, S., Mondal, M. R. H., Podder, P., & Prasath, V. B. S. (2022). Federated learning: Applications, challenges and future directions. *International Journal of Hybrid Intelligent Systems*, *18*(1–2), 19–35. [CrossRef]

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Making words work: Using financial text as a predictor of financial events. *Decision Support Systems*, *50*(1), 164–175. [CrossRef]

Cho, Y. J., Wang, J., & Joshi, G. (2020). Client selection in federated learning: Convergence analysis and power-of-choice selection strategies. *arXiv*, arXiv:2010.01243.

Choi, D., & Lee, K. (2018). An artificial intelligence approach to financial fraud detection under IoT environment: A survey and implementation. *Security and Communication Networks*, *2018*(1), 5483472. [CrossRef]

Damanik, N., & Liu, C. M. (2025). Advanced fraud detection: Leveraging K-SMOTEENN and stacking ensemble to tackle data imbalance and extract insights. *IEEE Access*, *13*, 10356–10370. [CrossRef]

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*, arXiv:1702.08608.

Farooq, M. S., Muhammad, M. H. G., Ali, O., Zeeshan, Z., Saleem, M., Ahmad, M., Abbas, S., Khan, M. A., & Ghazal, T. M. (2024). Developing a transparent anaemia prediction model empowered with explainable artificial intelligence. *IEEE Access*, *13*, 1307–1318. [CrossRef]

Financial Fraud Detection Dataset. (n.d.). Available online: https://www.kaggle.com/datasets/sriharshaeedala/financial-fraud-detection-dataset (accessed on 5 March 2025).

Gandomi, A. H., Abualigah, L., Saleh Alfaiz, N., & Fati, S. M. (2022). Enhanced credit card fraud detection model using machine learning. *Electronics*, *11*(4), 662. [CrossRef]

Ghazal, T. M., Iqbal Janjua, J., Abbas, S., Fatima, A., Saleem, M., Khan, M. A., & Alqarafi, A. (2024, May 3). *Fuzzy-based weighted federated machine learning approach for sustainable energy management with IoE integration*. 2024 Systems and Information Engineering Design Symposium, SIEDS 2024 (pp. 112–117), Charlottesville, VA, USA. [CrossRef]

Glancy, F. H., & Yadav, S. B. (2011). A computational model for financial reporting fraud detection. *Decision Support Systems*, *50*(3), 595–601. [CrossRef]

Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, *50*, 585–594. Available online: https://www.sciencedirect.com/science/article/pii/S0167923610001338 (accessed on 4 March 2025). [CrossRef]

Ikemefuna, C. D., Okusi, O., Iwuh, A. C., & Yusuf, S. (2024). Adaptive fraud detection systems: Using ML to identify and respond to evolving financial threats. *International Research Journal of Modernization in Engineering*, *6*, 2077–2092. Available online: https://www.researchgate.net/profile/Oluwatobiloba-Okusi/publication/384319231_Adaptive_Fraud_Detection_SystemsUsing_Machine_Learning_To_Identify_and_Respond_To_Evolving_Financial_Threat/links/66f3db50869f1104c6b488e2/Adaptive-Fraud-Detection-SystemsUsing-Machine-Learning-To-Identify-and-Respond-To-Evolving-Financial-Threat.pdf (accessed on 4 March 2025).

Javadpour, A., Ja'fari, F., Taleb, T., Shojafar, M., & Benzaïd, C. (2024). A comprehensive survey on cyber deception techniques to improve honeypot performance. *Computers & Security*, *140*, 103792. [CrossRef]

Jessica, A., Raj, F. V., & Sankaran, J. (2023, October 29–31). *Credit card fraud detection using machine learning techniques*. ViTECoN 2023—2nd IEEE International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies, Proceedings, Lagos, Nigeria. [CrossRef]

Johnson, D. L. (2022). Demystifying the elusive quest for cyber insurance protection: The need for new contract language. *Cardozo Law Review*, *44*, 2361. Available online: https://heinonline.org/HOL/Page?handle=hein.journals/cdozo44&id=2451&div=&collection= (accessed on 4 March 2025).

Kazemi, Z., & Zarrabi, H. (2017, December 22). *Using deep networks for fraud detection in credit card transactions*. 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation, KBEI 2017 (pp. 630–633), Tehran, Iran. [CrossRef]

Khan, M. A., Farooq, M. S., Saleem, M., Shahzad, T., Ahmad, M., Abbas, S., & Abu-Mahfouz, A. M. (2025). Smart buildings: Federated learning-driven secure, transparent, and smart energy management system using XAI. *Energy Reports*, *13*, 2066–2081. [CrossRef]

Khan, M. A., Sabahat, Z., Farooq, M. S., Saleem, M., Abbas, S., Ahmad, M., Mazhar, T., Shahzad, T., & Saeed, M. M. (2024). Optimizing smart home energy management for sustainability using machine learning techniques. *Discover Sustainability*, *5*(1), 430. [CrossRef]

Khetani, V., Gandhi, Y., Bhattacharya, S., Ajani, S. N., & Limkar, S. (2023). Cross-domain analysis of ML and DL: Evaluating their impact in diverse domains. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(7s), 253–262. Available online: https://www.ijisae.org/index.php/IJISAE/article/view/2951 (accessed on 4 March 2025).

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, *32*(4), 995–1003. [CrossRef]

Kose, I., Gokturk, M., & Kilic, K. (2015). An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. *Applied Soft Computing*, *36*, 283–299. [CrossRef]

Lim, W. Y. B., Luong, N. C., Hoang, D. T., Jiao, Y., Liang, Y. C., Yang, Q., Niyato, D., & Miao, C. (2020). Federated learning in mobile edge networks: A comprehensive survey. *IEEE Communications Surveys and Tutorials*, *22*(3), 2031–2063. [CrossRef]

Mahboubi, A., Luong, K., Aboutorab, H., Bui, H. T., Jarrad, G., Bahutair, M., Camtepe, S., Pogrebna, G., Ahmed, E., Barry, B., & Gately, H. (2024). Evolving techniques in cyber threat hunting: A systematic review. *Journal of Network and Computer Applications*, *232*, 104004. [CrossRef]

Modi, K., & Dayma, R. (2018, June 23–24). *Review on fraud detection methods in credit card transactions*. 2017 International Conference on Intelligent Computing and Control, I2C2 2017 (pp. 1–5), Coimbatore, India. [CrossRef]

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. [CrossRef]

Nishio, T., & Yonetani, R. (2019, May 20–24). *Client selection for federated learning with heterogeneous resources in mobile edge*. IEEE International Conference on Communications, Shanghai, China. [CrossRef]

Puh, M., & Brkić, L. (2019, May 20–24). *Detecting credit card fraud using selected machine learning algorithms*. 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia. Available online: https://ieeexplore.ieee.org/abstract/document/8757212/ (accessed on 4 March 2025).

Ramachandran, K., Kayathwal, K., Wadhwa, H., & Dhama, G. (2023, June 18–23). *FraudAmmo: Large scale synthetic transactional dataset for payment fraud detection*. International Joint Conference on Neural Networks, Gold Coast, Australia. [CrossRef]

Randhawa, K., Loo, C. K., Seera, M., Lim, C. P., & Nandi, A. K. (2018). Credit card fraud detection using AdaBoost and majority voting. *IEEE Access*, *6*, 14277–14284. [CrossRef]

Rehman, A., & Hashim, F. (2020). Impact of Fraud Risk Assessment on Good Corporate Governance: Case of Public Listed Companies in Oman. *Business Systems Research: International Journal of the Society for Advancing Innovation and Research in Economy*, *11*(1), 16–30. [CrossRef]

Ruposky, T. J. (2022). The exponential rise of cybercrime. *University of Central Florida Department of Legal Studies Law Journal*, *5*. Available online: https://heinonline.org/HOL/Page?handle=hein.journals/ucflaegs5&id=138&div=&collection= (accessed on 4 March 2025).

Sabuhi, M., Musilek, P., & Bezemer, C. P. (2024). Micro-FL: A fault-tolerant scalable microservice-based platform for federated learning. *Future Internet*, *16*(3), 70. [CrossRef]

Saleem, M., Farooq, M. S., Shahzad, T., Hassan, A., Abbas, S., Ali, T., Aggoune, E. H. M., & Khan, M. A. (2024). Secure and transparent mobility in smart cities: Revolutionizing AVNs to predict traffic congestion using MapReduce, Private Blockchain and XAI. *IEEE Access*, *12*, 131541–131555. [CrossRef]

Shahzad, T., Saleem, M., Farooq, M. S., Abbas, S., Khan, M. A., & Ouahada, K. (2024). Developing a transparent diagnosis model for diabetic retinopathy using explainable AI. *IEEE Access*, *12*, 149700–149709. [CrossRef]

Sharma, M. A., Raj, B. R. G., Ramamurthy, B., & Bhaskar, R. H. (2022). Credit card fraud detection using deep learning based on auto-encoder. *ITM Web of Conferences*, *50*, 01001. [CrossRef]

Sweers, T., Heskes, T., & Krijthe, J. (2018). *Autoencoding credit card fraud* [Bachelor Thesis, Radboud University]. Available online: https://www.cs.ru.nl/bachelorscripties/2018/Tom_Sweers___4584325___Autoencoding_credit_card_fraude.pdf (accessed on 4 March 2025).

Talukder, M. A., Khalid, M., & Uddin, M. A. (2024). An integrated multistage ensemble machine learning model for fraudulent transaction detection. *Journal of Big Data*, *11*(1), 1–25. [CrossRef]

Wang, J., Liu, Q., Liang, H., Joshi, G., & Poor, H. V. (2020). Tackling the objective inconsistency problem in heterogeneous federated optimization. *Advances in Neural Information Processing Systems*, *33*, 7611–7623.

Yan, C., Li, Y., Liu, W., Li, M., Chen, J., & Wang, L. (2020). An artificial bee colony-based kernel ridge regression for automobile insurance fraud identification. *Neurocomputing*, *393*, 115–125. [CrossRef]

Yang, W., Zhang, Y., Ye, K., Li, L., & Xu, C. Z. (2019). FFD: A federated learning-based method for credit card fraud detection. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), 11514 LNCS* (pp. 18–32). Springer. [CrossRef]

Zhu, X., Ao, X., Qin, Z., Chang, Y., Liu, Y., He, Q., & Li, J. (2021). Intelligent financial fraud detection practices in the post-pandemic era. *Innovation*, *2*(4), 100176.