Journal of Engineering Managment and Soft Computing (JEMSC)

Proposing a Data Governance Model for Fraud Detection in Executive Agencies Based on Federated Learning in a Cloud Computing Environment

Masoomeh Mojtabaee[1], Seyed Javad Iranbanfard[2]✉, Sara Najafzadeh[3] and Mostafa Kolahdoozi[4]

1. Ph.D. Student Department of Information Technology Management, Ki.C., Islamic Azad University , Kish, Iran.
2. Associate Prof., Department of Management, Shi.C., Islamic Azad University, Shiraz, Iran.
3. Assistant Prof., Department of computer, YI.C., Islamic Azad University, Tehran, Iran.
4. Associate Prof., Department of Information Technology Management, SR.C. Islamic Azad University, Tehran, Iran.

Article Info

Article type:
Research Article

Article history:
Received 4 October 2025
Received in revised form 17 February 2026
Accepted 1 March 2026
Published online 1 April 2026

Keywords:
Data Governance, Fraud Detection, Executive Agencies, Federated Learning, Cloud Computing.

ABSTRACT

With the growing volume of data interactions and increasing complexity of oversight in executive agencies, the need for innovative approaches to fraud detection and data governance has become more pressing than ever. Given that data is stored in a distributed manner, cloud computing emerges as an effective solution. However, security concerns hinder direct data exchange between organizations. To address this challenge, the present study proposes a decentralized, data-driven governance model based on federated learning and cloud infrastructure. In this model, each organization preprocesses its data at the edge and extracts fraud-related features. These results are then transmitted to a central server, where deep learning techniques are used to predict new inter-organizational fraud patterns. This approach preserves confidentiality and enables collaborative analysis without requiring data aggregation. Experimental results show that the proposed method reduces computational complexity by 60% and achieves a fraud detection accuracy of 97.6%, demonstrating its high effectiveness in multi-organizational and distributed environments.

Cite this article: Iranbanfard, S,J. & et al, (2026)., Proposing a Data Governance Model for Fraud Detection in Executive Agencies Based on Federated Learning in a Cloud Computing Environment. *Engineering Management and Soft Computing,* 12 (2). 109-140.

DOI: https://doi.org/10.22091/jemsc.2026.14081.1309

DOI: https://doi.org/10.22091/jemsc.2026.14081.1309

Publisher: University of Qom

1) Introduction

Today, the world is witnessing a profound change. One is a change driven by ongoing innovations and evolving technologies and will have a profound impact on people's lives, organizational structures and mindsets, and even on interactions among nations. Understanding this change and keeping pace with it in a timely way seems essential for all organizations regardless of their size. The emergence of potent new and powerful digital technologies, digital platforms, and digital infrastructures has dramatically transformed innovation and entrepreneurship (Mirfallah Lialestani et al., 2021). As With the digitization of processes, a vast huge amount trove of sensitive data has been generated in executive government agencies organizations (Ahmad et al, 2024). However, the lack of integration and deficiencies weakness in data governance have made it difficult to effectively utilize these resources (Canton, 2021). The increasing volume and complexity of information have paved the way for emerging new fraud schemes types of fraud and illegal transactions, that traditional fraud detection systems, based on predefined rules, are woefully ineffective against (Alazzabi et al, 2021; Moghadam et al., 2023). These systems are only able to identify known fraud patterns, and their the inability to detect new ones leads to a flood of many false positives alarms, diminished reduced efficiency, and escalated operational costs (O. A. Bello et al., 2023; Ha et al., 2025). In addition, these methods are unable to analyze behavioral and cross-channel data and are ill-suited to cloud infrastructures (Islam et al., 2024). These limitations are especially pronounced in high-stakes systems, such as banking and e-commerce, that require real-time processing of massive transaction volumes of data (Adejumo et al., 2025).

In response to these challenges, innovative solutions grounded in artificial intelligence and cloud computing have emerged. AI Artificial intelligence algorithms, such as machine learning and deep learning, are able to identify intricate complex fraud patterns and evolve to adapt to new fraud tactics by continuously learning from new data (He, 2021; Mohanty et al., 2023; Shi et al., 2023). Supervised learning can distinguish fraudulent transactions from legitimate ones, while deep learning uncovers hidden data interdependencies to significant detection accuracy by discovering complex dependencies in data (Ashtiani et al,. 2021; H. O. Bello et al,. 2023). Meanwhile, unsupervised learning enables the identification of unknown types of frauds without relying on the need for labeled datasets (Ali et al., 2022). Complementing these advances, In addition, cloud computing, with its scalability and real-time processing capabilities, facilitates the rapid analysis of large -scale volumes of transactions and secure data sharing across institutions (Stojanović et al,. 2025; Upreti et al., 2022). Modern cloud platforms, by integrating multi-factor authentication and federated learning, strike a delicate balance between security and performance efficiency (Kanamori et al., 2022; Myalil et al,. 2023; Zhang et al., 2022). However, challenges such as safeguarding privacy, ensuring algorithm transparency, and navigating legal requirements remain—hurdles requiring more robust governance frameworks to harmonize operational efficiency with ethical considerations (Rehan, 2021).

This paper proposes a novel data governance framework for cross-organizational executive agencies fraud detection using federated learning, and cloud, edge computing. The main problem addressed in this research is the security and privacy challenges faced when analyzing sensitive and distributed data across organizations. Traditional systems aimed to solve this problem usually collect data in a central location, which is both costly and heightens the risk of privacy violations.

The main issue is how to detect complex fraud patterns in distributed data without aggregating raw data. This paper addresses this challenge through a federated learning approach. In this model, each organization acts as an independent "participant" federator and does not transmit its data to central servers. Instead, each participant federator performs preprocessing in the edge environment, i.e., on its local servers. This preprocessing involves using methods, such as Pearson correlation, to select features that are highly correlated with fraud in the local data. In other words, each organization only sends the "knowledge" extracted from its data, and not the raw data itself, to a central federated entity in the cloud.

The central federated entity does not receive raw data; rather, it receives a collection of this processed knowledge from different federated entities and combines it. In this step, convolutional neural networks (CNNs) and ensemble crowd-learning methods are used to analyze these extracted features,

to identify complex patterns of inter-organizational fraud, and build a central predictive model. This model can then be used to detect fraud in new data. This innovative approach, by simultaneously leveraging edge computing for privacy and cloud computing for extensive analytics, offers a secure and scalable solution that overcomes the limitations of traditional methods while boosting efficiency and security.

The main innovation of this research centers on the development of a unified framework for data governance and cross-organizational fraud detection, that simultaneously addresses three core fundamental challenges in operational environments, namely data confidentiality, challenges in the impossibility of aggregating sensitive data, and the need for cross-organizational intelligent analytics. Unlike conventional approaches that have either focused on fraud detection or focused solely on the technical aspects of federated learning, this paper proposes federated learning as the core of a data governance model. In this framework, the entire process of feature selection, local learning, knowledge aggregation, and the training of advanced machine learning models is structured and designed in a cloud computing environment. The application of correlation-based feature selection methods at the federated nodes level and the deployment of deep learning models in the central layer enable the extraction of complex fraud patterns without violating data privacy. This purposeful integration of data governance, federated learning, and intelligent analytics stands as the main distinguishing feature of the present research compared to previous studies.

In summary, this paper contributes to important advances in distributed data analytics by presenting several key innovations. Among its most important contributions are:

- A new data governance model: Introducing a comprehensive, federated learning-based framework designed to solve the problem of cross-organizational fraud detection.
- Privacy preservation through edge computing: By intelligently using each organization's local servers for data preprocessing, which eliminating effectively the need to transfer raw data.
- Intelligent feature extraction: Applying Pearson correlations in each participant federated entity to identify factors contributing to fraud, thereby transforming raw data into processed knowledge.
- Central deep learning: Designing a deep learning architecture based on convolutional neural networks (CNNs) and ensemble crowd learning in the cloud environment to analyze complex, inter-organizational, and common fraud patterns.
- Scalability and security: Providing a trusted infrastructure that allows organizations to analyze data collaboratively without the risk of worrying about privacy breaches.
- This hybrid approach strikes an optimal balance between system efficiency, data security, and privacy, thereby providing a practical and reliable solution for fraud detection in intelligent surveillance systems.

The rest of the paper is organized as follows. In second Section, related works are reviewed. In Section three, the details of the proposed method are presented. In the fourth Section, the experiments and their analysis are presented. Finally, in the fifth Section, the conclusions are presented.

2) Literature Review

Given the critical importance of cross-organizational fraud detection, this section examines some key recent approaches in this field. Traditional fraud detection systems in the financial sector face serious challenges in real-time processing of high-volume traffic data, which often results in delayed responses and incomplete fraud detection.

In Hasan et al.'s (2025) study, a systematic review of previous research reveals that federated learning, as an organizational framework, enables secure and scalable exploitation of distributed data while simultaneously addressing meeting the privacy, security, and governance requirements of AI in various domains. Oyekunle et al. (2025) and Rivandi (2025) on studying real banking data demonstrated that combining behavioral biometrics with data governance frameworks effectively enhances fraud detection systems and significantly improves fraud detection accuracy and reduces fraud rates while

maintaining ethical and privacy considerations. In Khan (2025), a novel approach to deploying data governance in an automated and integrated in-built manner in ETL processes is presented, which simultaneously ensures operational agility and compliance requirements in modern data systems by providing real-time monitoring, data lineage, and proactive regulatory compliance. Alademehin (2025) highlighted that applying AI-based methods to automate metadata management, anomaly detection, and policy optimization, data governance can be transformed from a reactive approach to a predictive, secure, and compliant system in various domains. In Alamu (2025), the focus is on cognitive data governance, an AI-based framework that makes data governance smarter, more adaptable, and more efficient in various industrial domains by automating classification, real-time monitoring, and policy enforcement. In Kumar (2025), an AI-based governance framework is presented that enables real-time fraud detection, improves transparency, and strengthens regulatory oversight in the financial system by intelligently analyzing transactional data and automating compliance processes.

In Rahmati (2025), a real-time framework for financial fraud detection was presented, which, by simultaneously utilizing adaptive neural networks, federated learning, and explainable artificial intelligence methods, enabled the discovery of emerging fraud patterns without violating data privacy and significantly increased the detection accuracy. Aljunaid et al. (2025) proposed an explainable federated learning model for bank fraud detection, which significantly improved the detection accuracy and minimized false positive errors by maintaining data confidentiality and enhancing the transparency of decisions. Manwani (2025), examining federated learning, demonstrated that banks can jointly train fraud detection models without directly exchanging sensitive data, thereby improving financial security and regulatory compliance in fraud detection. In Gimah (2025), by implementing federated learning based on local training of deep learning models and aggregation of updates, credit card fraud was effectively detected, while the raw data of each institution remained on-site in the same place without violating privacy. Claus et al. (2025) showed that, using federated learning, fraud detection models can be trained collaboratively across financial institutions while preserving data privacy and improving detection accuracy.

In Rehan (2021), the role of artificial intelligence and cloud computing in the evolution of fraud detection is examined, and the benefits, architectures, implementation challenges, and prospects for future directions of this approach are also analyzed. Hassan et al. (2025) employ strong corporate governance practices to detect possible fraud within the organization. It further argues that to design and implement information technologies tailored to the specific needs of each organization is necessary. In addition, the authors advocate that the government should increase awareness about the provision of data by relevant institutions, organizations, and individuals. Andayani et al. (2023) argue that social responsibility, good corporate governance, and fraud detection in financial statements can be effective in reducing financial fraud and should be strengthened through ethical values and organizational culture. In Pamisetty et al. (2022), extensive government investments in FinTech technologies are reviewed, which are used to improve tax compliance, combat fraud in public financing, and increase the efficiency of public resources, by using tools such as blockchain, machine learning, artificial intelligence, data mining, and electronic financial systems. In Salmanov (2025), a hybrid model for fraud detection using statistical techniques and machine learning algorithms is proposed, which improves the detection rate and promotes transparency in the governance structures of organizations. In Pamisetty (2023), the main and basic elements of cloud computing are reviewed and its application in improving e-government is analyzed; the results of the analysis of six key areas confirm the viability of this method and introduce the main effective players in the development and growth of cloud-based e-governance systems. Favour (2022) highlights the challenges of regulatory compliance in AI-based fraud detection systems in cloud computing environments, especially in the areas of privacy, data security, and transparency, and the paper provides solutions for designing architectures that comply with regulations. In Das et al. (2023), a combined approach of human expertise and advanced AI techniques is proposed under the name of "collaborative intelligence," which enables accurate and large-scale fraud detection by combining diverse data sources and utilizing machine learning, graph analysis, and natural language processing. Moreover, the alerts generated by AI are supplemented by human review, and compliance with

governance principles in privacy, security, and legal requirements is likewise ensured. In Samuel (2023), the integration of AI with cloud-based big data analytics is reviewed to improve the performance of financial fraud detection systems and a scalable and adaptable framework for real-time fraud detection is presented. Furthermore, while analyzing the security challenges of cloud computing in detail, a balanced framework between analytical accuracy and security resilience is proposed. In Katari et al. (2022), data governance challenges in multi-cloud environments for financial services companies are examined and solutions such as the use of unified frameworks, advanced cryptography, AI-based real-time monitoring, based on artificial intelligence, and process automation are proposed to enhance security, efficiency, and regulatory compliance. Halbouni et al. (2016) examine the role of corporate governance and IT in fraud prevention and detection in the UAE and indicate that both factors play a moderate role; therefore, there is a need for a better understanding of the importance of monitoring by senior management, strengthening the culture of integrity, and greater use of technology to combat fraud. Yandrapalli (2024) present an automated framework for data quality assurance in banking using statistical methods and machine learning, in which the performance of machine learning models is improved by identifying and removing outliers. Table 1 summarizes the advantages and disadvantages of the previous approaches.

Table 1 - Advantages and Disadvantages of Previous Methods

Author and Year	Method Used	Advantages	Limitations
Hasan, 2025	Federated Learning	Scalable, privacy-preserving	Management complexity
Oyekunle et al., 2025	Behavioral Biometrics + Data Governance	Improved fraud detection accuracy, privacy compliance	Requires complex infrastructure
Khan, 2025	Automated Data Governance in ETL	Real-time monitoring, proactive compliance	High implementation cost
Alademehin, 2025	AI-based Automated Metadata Management	Predictive and secure governance	Dependent on data quality
Alamu	AI-driven Cognitive Data Management	More intelligent and efficient	Technological complexity
Kumar, 2025	Transaction Analytics + AI	Real-time fraud detection, transparency	Risk of AI errors
Rahmati, 2025	Graph Neural Networks + FL + XAI	Detection of emerging patterns	Model complexity
Aljunaid et al., 2025	Explainable Federated Learning	Higher accuracy, reduced errors	Computationally intensive
Manwani	Interbank Federated Learning	Privacy preservation, financial security	Dependence on trusted networks
Gimah, 2025	Federated Learning with Local Training	Effective detection, data remains local	High bandwidth requirements
Claus et al., 2025	Cross-institutional Federated Learning	Improved accuracy, privacy preservation	Operational complexity
Rehan, 2021	AI + Cloud Computing	Transformation of fraud detection	Security and privacy challenges
Hassan et al., 2025	Corporate Governance + IT	Detection of internal organizational fraud	Requires tailored technology design
Andayani & Wuryantoro, 2023	Social Responsibility + Governance	Fraud reduction, ethical reinforcement	Dependent on organizational culture
Pamisetty et al., 2022	FinTech + Blockchain + AI	Improved tax discipline, fraud prevention	Requires significant investment
Salmanov, 2025	Hybrid Statistical + ML Model	Enhanced detection and transparency	Model complexity
Pamisetty, 2023	Cloud Computing & E-Government	Improved data governance	Security challenges

Favour, 2022	Cloud-based AI & Regulation	Privacy and security preservation	Operational complexity
Das et al., 2023	Collaborative Intelligence (Human + AI)	High accuracy, governance compliance	Requires human–machine coordination
Samuel, 2023	AI + Cloud Big Data	Scalable framework, real-time processing	Security and analytical complexity
Katari & Ankam, 2022	Data Governance in Multi-cloud	Security, automation	Requires advanced infrastructure
Halbouni et al., 2016	Corporate Governance + IT	Fraud prevention and detection	Moderate impact, management training required
Yandrapalli, 2024	Data Quality Assurance + ML	Outlier removal, model improvement	Dependent on high-quality data

According to the methods reviewed, it can be seen that the method proposed in this paper has significant advantages over previous methods; first, by utilizing federated learning and data processing in the edge computing environment, without the need to aggregate raw data, the privacy of organizations is preserved and the security risks arising from data sharing are prevented. Unlike conventional methods that require transferring complete information to the central server, in this approach only the knowledge extracted from local data (important features identified with high correlation) is transferred to the central server, which not only increases processing efficiency but also improves the scalability of the system. Additionally, combining the results of different federates in the cloud environment and utilizing convolutional neural networks and ensemble classification algorithms increases the accuracy and comprehensiveness in identifying inter-organizational fraud patterns; an issue that has been less addressed in traditional methods in a distributed and privacy-oriented manner.

A review of previous studies reveals that most of the research has either focused on fraud detection using centralized machine learning methods or has examined federated learning as a technical solution to protect data privacy. However, there is still a gap in a comprehensive model that applies federated learning within a data governance framework that is tailored to the structure of the executive agencies. The present study fills this gap by presenting an inter-organizational model in which data processing and feature selection are performed at the federate level and only the extracted knowledge is transferred to the cloud layer. Furthermore, combining statistical feature selection with deep learning models and comparing them with ensemble learning methods provides a different approach to existing work. Therefore, the innovation of this paper can be explained not only in the use of federated learning, but also in how it is applied as a practical data governance model for fraud detection in multi-organizational executive environments.

3) Research Methodology

In this paper, we introduce a new model for data governance and inter-organizational fraud detection that is designed based on federated learning and cloud computing infrastructure. The main goal of this model is to accurately and timely detect fraud among multiple organizations, without the need to directly share raw and sensitive data. In this model, by using the federated approach, each organization independently and in its own dedicated environment (edge or fog computing), performs the initial data processing, and only the processed results and extracted knowledge are shared. In addition to preserving data privacy, this approach also greatly increases information security. The proposed solution method in this research is designed based on a phased and distributed process that continues from the moment the data enters the executive organizations to the final decision-making in the central layer. In the first step, each executive agency is considered an independent federate that owns its local data and is responsible for storing, cleaning, and initial data preparation. This data includes non-fraudulent and fraudulent samples and is processed without leaving the organizational environment. Next, to reduce the dimensionality of the data and eliminate low-impact features, a feature selection process based on the

Pearson correlation coefficient is performed in each federate; in such a way that only the features that have the highest statistical relationship with the fraud label are retained. This step plays an important role in reducing the computational complexity and increasing the stability of the learning models.

After feature selection, each federate uses its local data to train the initial fraud detection model and, instead of sending raw data, only sends the extracted results, including the set of selected features and learning parameters, to the central server. The central server, which is located on a cloud computing platform, is responsible for aggregating this distributed knowledge. In this layer, selected features are shared among federates, forming a common feature space that represents cross-organizational fraud patterns. Then, based on these aggregated features, advanced machine learning models, especially convolutional neural networks, are trained to detect complex and non-linear fraud patterns.

In the final stage, the central trained model is used to evaluate and detect fraud on new data. This detection can be performed online or near online, and in the event of new data, the model is modified without the need to aggregate raw data and only by updating the federated parameters. In this way, the proposed solution method, while making the data flow and learning process transparent, provides a practical, secure, and scalable mechanism for detecting fraud in multi-organizational environments that is compatible with data governance requirements and privacy constraints. The conceptual model of the proposed method is presented in Figure 1.

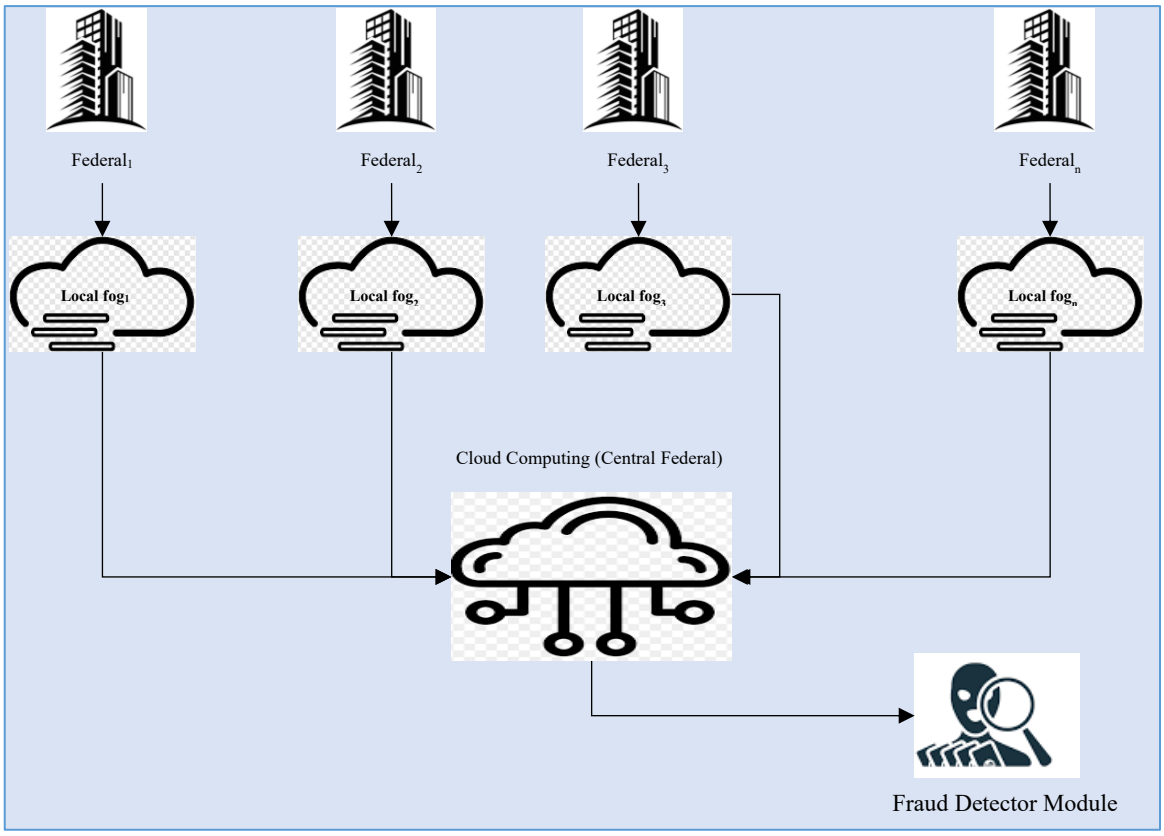

Figure 1. Overview of the Proposed Method

As can be seen in Figure 1, in the first step, each organization or federate stores and preprocesses local data related to fraudulent and non-fraudulent reports on its servers. The preprocessing stage involves selecting the features that have the highest correlation with the occurrence of fraud. For this purpose, the feature selection method based on the Pearson correlation coefficient is used, which is one of the well-known statistical methods for analyzing the relationship between variables. In this step, without the need to transfer raw data to other federates or to the central server, each organization only works on its own data and extracts the features that play an important role in fraud detection. The

knowledge extracted from the data of each federate, which is actually a combination of statistical features and fraud detection factors, is transferred to the central server. The central server is located in the cloud computing infrastructure and is responsible for combining the information obtained from different federates. This information does not include raw data, but rather processed results that allow for analysis and learning of predictive models without revealing the original content of the information. In this way, data security is maintained and common problems, such as privacy violations, are avoided.

To make privacy compliance in the proposed model more transparent, it is necessary to mention the details of the sensitive data processing. In the presented model, each organization or federate independently stores and processes its data in its own dedicated environment, and no raw data or sensitive information is transferred to other federates or to the central server. Local processing includes preprocessing, cleaning and extraction of features related to fraud detection, and only the statistical results and extracted knowledge, which do not contain personal or sensitive information, are sent to the central server. This knowledge includes weights, statistical indices, and fraud detection factors, and no raw data that could reveal the identity or confidential details of individuals or organizations is transferred.

The central server also trains deep learning models using only this processed and combined data, and the training process is designed in such a way that it is not possible to reconstruct the original data. In addition, the models are updated in an incremental and secure manner, and at all stages, privacy policies are respected, and access to confidential information is limited to the same local organization. In this way, in addition to maintaining data security, this structure prevents privacy violations and allows for accurate analysis and training of models without exposing sensitive data.

This approach provides several layers of privacy:

- Local data processing: Each organization preprocesses its data in its own dedicated environment and extracts important features, without sharing raw data.
- Sharing processed results: What is transferred to the central server is a combination of statistical analysis results and profiles extracted from the data, and there is no raw or sensitive information in it.
- No disclosure of the original content of the information: The federated learning architecture ensures that even during the training phase of deep learning models, the raw data never leaves the organizational environment and only the weights and model updates are exchanged.
- Continuous update without transmitting new data: The model can be continuously updated without the need to send new data, which in addition to maintaining security, is also legally and regulatory compliant with privacy standards.

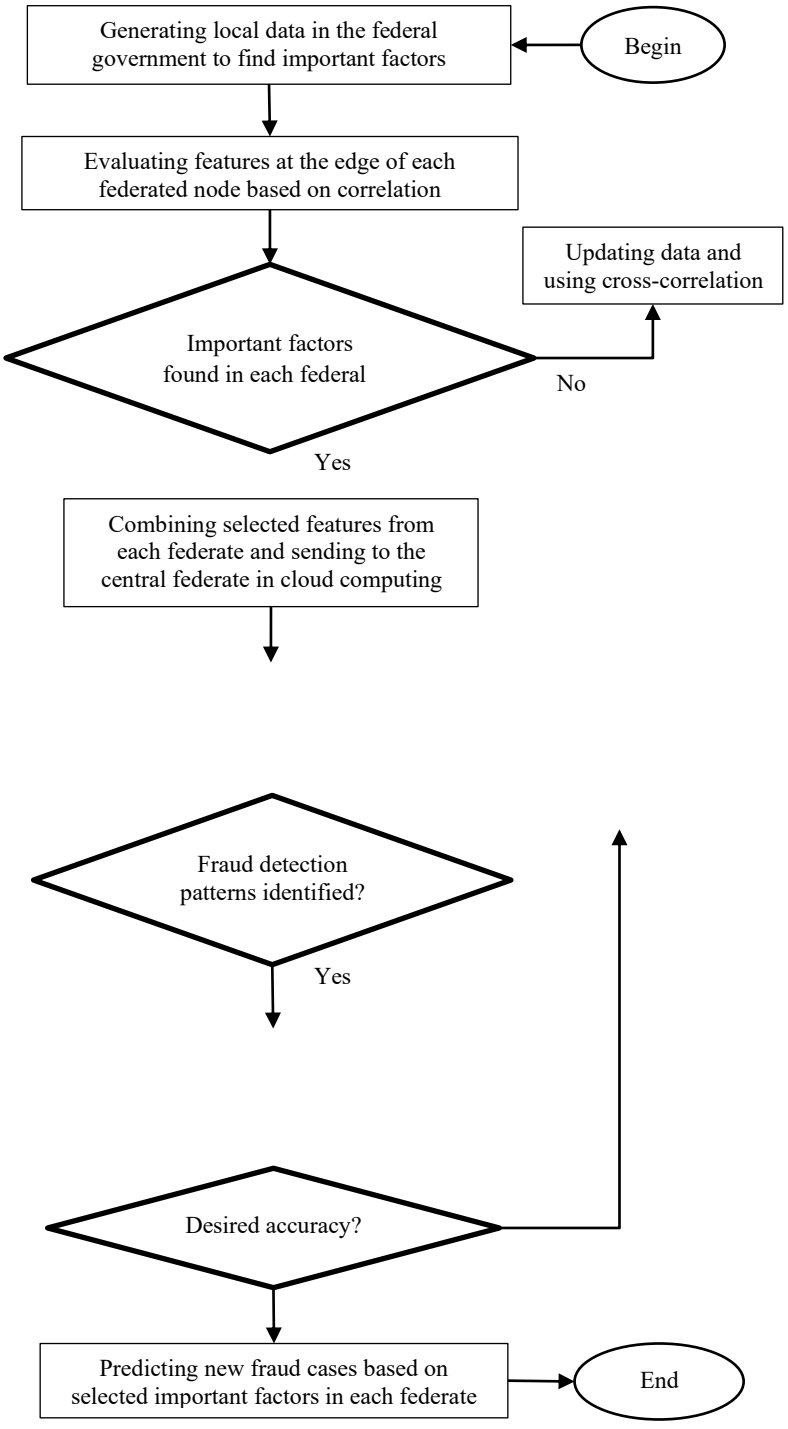

Figure 2. Flowchart of the Proposed Method

In the central part, the factors received from the federates are combined and the final features are extracted to train deep learning models. In this process, convolutional neural networks (CNN) are used to identify complex patterns in inter-organizational data. Additionally, ensemble methods are employed to compare the accuracy of the final model in fraud detection based on deep learning and ensemble learning.

Finally, the patterns discovered by the machine learning models are applied to new samples to enable real-time and high-accuracy fraud detection. Since these patterns are formed based on information from multiple federates, they are more capable of detecting hidden patterns of inter-

organizational fraud. Moreover, due to the structure based on federated learning, the model can be continuously updated without the need to transfer new data. As a result, this model provides a practical and efficient solution for distributed data governance and fraud detection in multi-organizational environments, which is reliable in terms of both data security and performance accuracy. Figure 2 illustrates the flowchart of the proposed method.

3.1 Correlation-Based Feature Selection

The proposed method in this paper has a significant advantage over previous methods by combining federated learning and the correlation-based feature selection (CFS) algorithm. In this method, each organization independently processes its internal data in the edge computing environment and selects the features that have the highest correlation with the class label (i.e., fraud occurrence or non-occurrence). The CFS algorithm uses the merit function (Equation (1)) to evaluate the quality of a subset of features:

Equation (1)

$$Merit_s = \frac{kr_{cf}}{\sqrt{k + (k+1)r_{ff}}}$$

Where:

k is the number of features selected in the S subset.

rcf is the average correlation between the features and the class label.

rff is the average correlation between the features (Yildirim, 2015).

This function seeks to maximize the predictive power of the class (numerator) and simultaneously minimize the redundancy between the features (denominator). Using this algorithm, only those features that are of high importance in fraud detection and have little information overlap with other features are selected. This is very efficient in distributed federated environments, as there is no possibility of sharing raw data and only the extracted knowledge (selected features) is sent to the central federate.

The CFS algorithm uses the Pearson correlation coefficient to measure the degree of linear relationship between features and class labels. The Pearson correlation coefficient between two random variables X and Y is defined as Equation (2):

Equation (2)

$$r_{xy} = \frac{\sum(x_i - \bar{x})\sum(y_i - \bar{y})}{\sqrt{(\sum(x_i - \bar{x}))^2}\sqrt{(\sum(y_i - \bar{y}))^2}}$$

Where

- $\sqrt{(\sum(x_i - \bar{x}))^2}$ is the mean of variable X
- $\sqrt{(\sum(y_i - \bar{y}))^2}$ is the mean of variable Y

This measure takes a value between -1 and 1, indicating the intensity and type of linear relationship between the two variables. In this method, each federate uses this coefficient to select the features that have the highest predictive power for fraud detection and sends only these features (not raw data) to the central federate in the cloud environment. Finally, the central federate extracts inter-organizational fraud patterns by integrating the selected features with deep learning methods such as convolutional neural networks. This process not only increases the accuracy of the final model but also ensures privacy, data security, and computational efficiency (Zhou et al., 2016).

3.2 Deep Neural Network

Deep neural networks (DNN) consist of several types of layers. These are input layers, convolutional layers, pooling layers, and fully connected layers. When these layers are stacked, a deep neural network architecture is formed. A simplified architecture of deep neural networks for classification is presented in Figure 3.

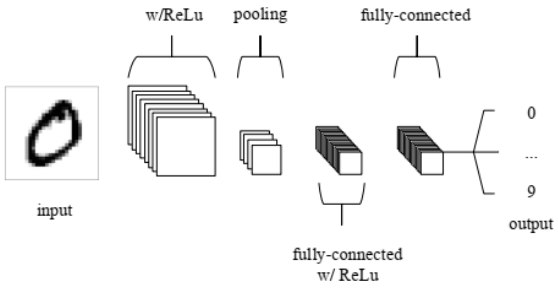

Figure 3. Simple Architecture of a Deep Neural Network Consisting of Five Layers (O'Shea & Nash, 2015)

According to Figure 3, the layers of deep neural networks are presented as follows.

- Input layer

The input layer in neural networks plays an important role in processing and optimizing the data. This layer first centers the data by reducing the mean of the data to zero and creates favorable conditions for training the network. It then normalizes the data to the interval [0,1] to avoid large fluctuations and speed up the convergence of learning. In addition, unnecessary dependencies between features are reduced by using data whitening techniques, and principal component analysis (PCA) reduces the dimensionality of the data and increases the focus on key factors. These steps make the data available to the learning algorithms in a compact and concentrated form, improving the performance of the neural network in learning (Du, 2018).

- Convolutional layer (CONV)

The convolutional layer plays a key role in extracting local features in deep neural networks, especially for data processing. This layer extracts specific features of each part of the data using convolutional kernels (filters) that slide on the inputs and produces convolutional results. The weights of the convolutional kernel are applied equally to all data points, which is called "common weighting," and allows the network to identify similar features at different points in the data. Adjusting parameters such as kernel size, depth, step size, and other filter settings helps the convolutional layer to recognize patterns effectively and perform better in data analysis. The algorithm for calculating the output size is defined based on Equation (3) (Du, 2018).

Equation (3)

$$\mathrm{H_{out}} = 1 + \frac{\mathrm{H_{in}} + (2 * \mathrm{pad}) - \mathrm{K_{height}}}{\mathrm{S}};\ \mathrm{W_{out}} = 1 + \frac{\mathrm{W_{in}} + (2 * \mathrm{pad}) - \mathrm{K_{width}}}{\mathrm{S}}$$

- Activation Layer

Nonlinearization of the output of the convolutional layer in neural networks is performed through activation layers, which is essential to solve the vanishing gradient problem and improve the learning of models. Various activation functions, including Sigmoid, Tanh, ReLU, Leaky ReLU, ELU, and Maxout, are used for this purpose. In particular, Leaky ReLU has been recognized as a popular choice in recent years due to its high convergence speed and lack of dead zones. The appropriate choice of activation function can have a great impact on the performance and training speed of deep neural networks (Du, 2018).

- Pooling layer

The pooling layer in deep neural networks is used to reduce both the dimensionality of extracted features and computational complexity, and helps prevent overfitting. This layer includes three main types: non-overlapping pooling, which involves max and average pooling with a stride equal to the window size, overlap pooling, which preserves more information with overlapping regions, and spatial pyramid pooling, which allows the network to convert the features of data of different input sizes into uniform dimensions, without damaging the information structure and preventing data loss. Spatial

pyramidal pooling has become a key component in the design of deep neural networks due to its ability to preserve information and improve performance when dealing with diverse data (Du, 2018).

- Fully connected layer

Fully connected layers are used at the end of deep neural networks to transfer and analyze the information extracted from previous layers to the final output. In these layers, each neuron is connected to all neurons in the previous layer, which helps to use all the extracted features simultaneously and produce more accurate outputs. After several convolutional layers have extracted different features from the data, the fully connected layers analyze this information comprehensively and help to create a more robust representation of the data. These layers help to simplify the calculations and increase the processing speed, and are therefore of great importance in the design and final performance of deep neural networks (Wu, 2017).

3.3 Evaluation Criteria

After implementing the proposed method, we evaluate its performance using new transactions as test data. Fraud detection is very important and several criteria have been developed to evaluate the improvement and effectiveness of fraud prediction methods. As mentioned earlier, the dataset used in the proposed approach is divided into training and testing sets. During the evaluation phase, we compare the predicted labels generated by the proposed model with the actual labels by keeping the labels of the test samples hidden. The accuracy of the proposed model is evaluated based on its ability to correctly identify fraudulent transactions, which is quantified by comparing the predicted outputs with the actual class labels of the test set. To achieve this, a confusion matrix is generated as output that highlights the number of samples that are correctly classified versus those that are incorrectly classified during the testing phase. This matrix provides a detailed insight into the performance of the model and reflects its effectiveness in distinguishing between fraudulent and non-fraudulent samples. This matrix consists of four elements: true positive (TP), false positive (FP), true negative (TN), and false negative (FN), which are as follows (Dorostkar Navaei et al., 2025).

- TN: A test transaction that is detected as non-fraudulent and its true label is non-fraudulent.
- TP: A test transaction that is detected as fraudulent and its true label is fraudulent.
- FP: A test transaction that is detected as fraudulent but its true label is non-fraudulent.
- FN: A test transaction that is detected as non-fraudulent but its true label is fraudulent.

By extracting the parameters of the confusion matrix, evaluation criteria can be calculated based on it to compare the results of the proposed model on the test set. The most common criteria in this context are accuracy, recall, precision, and F-measure. These criteria are defined based on Equations (3) to (6).

(Equation 3)

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP+TN+FP+FN}}$$

(Equation 4)

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

(Equation 5)

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F (Equation 6)

$$\text{F} - \text{measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

The evaluation criteria introduced are used as a tool to measure the quality of the proposed method and compare it with other existing methods.

3.4 Time Complexity Analysis

To analyze the time complexity of the proposed model, it is necessary to examine each stage of data processing and model training. In the local preprocessing stage, each federate works only on its own data, and the calculations include feature selection and the calculation of Pearson correlation coefficients. The time complexity of this stage depends on the number of local samples n and the number of features d, and is approximately equal to O(n.d), since for each feature, the correlation with the target variable is calculated.

After feature extraction, the processed data and fraud detection indicators are sent to the central server. The synthesis of this data at the center involves statistical aggregation and input preparation for the convolutional neural networks. The time complexity of this step mainly depends on the number of federates F and the number of selected features d′ and can be considered O(F.d′), since the central server receives only summarized and processed data and no raw data is transmitted. In the training part of the deep learning model, convolutional neural networks are trained on the combined features. The time complexity of training CNN networks varies depending on the number of samples N, the number of layers L, and the size of each layer and filters k, and is approximately estimated to be $O(N.L.k^2)$. Since the data is updated through federated learning, each training session involves collecting updates from federates and applying them to the central model, which also has a time complexity of O(F.U), where U represents the time to aggregate updates from each federate. Overall, the federated learning-based architecture reduces the computational burden on each organization, as the raw data remains in place and only the extracted knowledge is transferred. This approach, while maintaining data privacy and security, allows for higher scalability and parallel processing, and the time complexity of the entire system is significantly reduced by distributing the computations among federates. Therefore, in addition to being efficient in fraud detection, the proposed model is also practical and time-optimized for implementation in multi-organizational environments.

4) Findings

As mentioned, in this paper, a novel model for data governance using federated learning and based on cloud computing infrastructure enables cross-organizational fraud detection without the need for data aggregation. In this paper, the financial fraud dataset (Lopez-Rojas et al., 2016) is used. Real data in the field of financial services, especially mobile transactions, is not publicly available, which makes research in the field of fraud detection difficult. To solve this problem, the PaySim synthetic dataset is generated by simulating normal and fraudulent transaction behavior to evaluate the performance of fraud detection methods.

All experiments in this research were conducted in the MATLAB R2021b environment using the Classification Learner toolbox and standard MATLAB statistical analysis functions. The data was first preprocessed and important features were extracted using Pearson correlation coefficient and then divided into several federates; therefore, each federate was trained independently and only the processed results were transferred to the central server. To identify complex patterns, machine learning models, including KNN, Random Forest, Bagging, XGBoost, and Convolutional Neural Networks (CNN), were used. The accuracy, recall, precision, and F-measure metrics were also calculated to evaluate the performance of the models. All calculations and training of the models were performed in the same MATLAB environment, and the feature correlation matrix for each federate was displayed as a heatmap to ensure the transparency of the proposed model performance. The simulation environment was run on an ASUS N56J laptop system with an Intel Core i7 processor and 8 GB of RAM and the Windows 10 (64-bit) operating system, which allows for an accurate reproduction and evaluation of the models' performance. Table 2 represents the characteristics of the dataset.

Table 2. Features of the PaySim Dataset (Lopez-Rojas et al., 2016)

Number	Feature name	Number	Feature name
1	step	7	nameDest
2	type	8	oldbalanceDest
3	amount	9	newbalanceDest
4	nameOrig	10	isFlaggedFraud
5	oldbalanceOrg	11	isFraud
6	newbalanceOrig		

As shown in Table 2, there are 10 features in the dataset, only a part of which are associated with the class label. Therefore, selecting the features that have the highest correlation with the class label (such as the fraudulent or non-fraudulent status of the transaction) plays a very important role in improving the performance of machine learning models. By focusing on relevant features and eliminating redundant or unimportant information, the model can more accurately identify hidden patterns in financial data and inter-organizational transactions. This not only increases the accuracy of fraud prediction but also reduces the computational complexity and processing time of the model, ultimately helping to develop smarter and more efficient systems to detect suspicious behaviors and prevent financial losses.

As mentioned, in the proposed method, data is divided among federates so that each federate performs preprocessing operations on the existing data locally and in its own edge environment. This preprocessing involves selecting the features that have the highest correlation with the class label. The selected features are sent as knowledge extracted from the data to the central server in the cloud computing platform to prevent the direct transmission of raw data, maintain information privacy, and prevent the transmission of large volumes of data. This process saves computational resources and network bandwidth and increases the efficiency of the distributed learning system. Table 3 shows the correlation matrices for each of the federates.

Table 3. (a-e) Correlation Matrix of Each Federate Relative to the Class Label

The correlation matrix of federal number 1

Features												
	1	1.000	0.011	0.072	0.000	0.000	-0.014	0.000	0.033	0.037	0.016	0.137	1
	2	0.011	1.000	0.096	0.000	-0.322	-0.346	0.000	-0.125	-0.082	0.011	0.092	0.8
	3	0.072	0.096	1.000	0.000	0.104	0.001	0.000	0.165	0.338	0.017	0.312	
	4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.6
	5	0.000	-0.322	0.104	0.000	1.000	0.987	0.000	0.066	0.051	0.014	0.047	
	6	-0.014	-0.346	0.001	0.000	0.987	1.000	0.000	0.071	0.043	0.014	-0.037	0.4
	7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	8	0.033	-0.125	0.165	0.000	0.066	0.071	0.000	1.000	0.973	-0.002	-0.025	0.2
	9	0.037	-0.082	0.338	0.000	0.051	0.043	0.000	0.973	1.000	-0.002	0.007	

10	0.016	0.011	0.017	0.000	0.014	0.014	0.000	-0.002	-0.002	1.000	0.041	0
11	0.137	0.092	0.312	0.000	0.047	-0.037	0.000	-0.025	0.007	0.041	1.000	
	1	2	3	4	5	6	7	8	9	10	11	-0.2

(a)

The Correlation Matrix of Federal Number 2

Features												
1	1.000	0.023	0.056	0.000	-0.002	-0.014	0.000	0.030	0.031	0.000	0.141	
2	0.023	1.000	0.083	0.000	-0.322	-0.347	0.000	-0.108	-0.073	0.000	0.087	0.8
3	0.056	0.083	1.000	0.000	0.102	0.006	0.000	0.217	0.371	0.000	0.275	
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.6
5	-0.002	-0.322	0.102	0.000	1.000	0.986	0.000	0.050	0.037	0.000	0.049	
6	-0.014	-0.347	0.006	0.000	0.986	1.000	0.000	0.054	0.030	0.000	-0.034	0.4
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
8	0.030	-0.108	0.217	0.000	0.050	0.054	0.000	1.000	0.979	0.000	-0.014	0.2
9	0.031	-0.073	0.371	0.000	0.037	0.030	0.000	0.979	1.000	0.000	0.013	
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0
11	0.141	0.087	0.275	0.000	0.049	-0.034	0.000	-0.014	0.013	0.000	1.000	
	1	2	3	4	5	6	7	8	9	10	11	-0.2

(b)

The Correlation Matrix of Federal Number 3

Features												
1	1.000	0.018	0.057	0.000	-0.002	-0.013	0.000	0.014	0.016	0.013	0.121	
2	0.018	1.000	0.098	0.000	-0.318	-0.343	0.000	-0.121	-0.073	0.011	0.090	0.8
3	0.057	0.098	1.000	0.000	0.096	0.002	0.000	0.216	0.394	0.057	0.294	
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.6
5	-0.002	-0.318	0.096	0.000	1.000	0.988	0.000	0.079	0.060	0.024	0.045	
6	-0.013	-0.343	0.002	0.000	0.988	1.000	0.000	0.085	0.053	0.024	-0.036	0.4

7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
8	0.014	-0.121	0.216	0.000	0.079	0.085	0.000	1.000	0.971	-0.002	-0.035	0.2
9	0.016	-0.073	0.394	0.000	0.060	0.053	0.000	0.971	1.000	-0.003	-0.001	
10	0.013	0.011	0.057	0.000	0.024	0.024	0.000	-0.002	-0.003	1.000	0.041	0
11	0.121	0.090	0.294	0.000	0.045	-0.036	0.000	-0.035	-0.001	0.041	1.000	
	1	2	3	4	5	6	7	8	9	10	11	-0.2

(c)

The Correlation Matrix of Federal Number 4

Features												
1	1.000	0.018	0.053	0.000	-0.001	-0.012	0.000	0.022	0.025	0.019	0.137	1
2	0.018	1.000	0.096	0.000	-0.326	-0.348	0.000	-0.094	-0.051	0.017	0.085	0.8
3	0.053	0.096	1.000	0.000	0.073	-0.001	0.000	0.267	0.442	0.050	0.244	
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.6
5	-0.001	-0.326	0.073	0.000	1.000	0.989	0.000	0.054	0.037	0.018	0.040	
6	-0.012	-0.348	-0.001	0.000	0.989	1.000	0.000	0.058	0.031	0.018	-0.036	0.4
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
8	0.022	-0.094	0.267	0.000	0.054	0.058	0.000	1.000	0.972	-0.003	-0.029	0.2
9	0.025	-0.051	0.442	0.000	0.037	0.031	0.000	0.972	1.000	-0.003	-0.001	
10	0.019	0.017	0.050	0.000	0.018	0.018	0.000	-0.003	-0.003	1.000	0.066	0
11	0.137	0.085	0.244	0.000	0.040	-0.036	0.000	-0.029	-0.001	0.066	1.000	
	1	2	3	4	5	6	7	8	9	10	11	-0.2

(d)

The Correlation Matrix of Federal Number 5

Features												
1	1.000	0.026	0.064	0.000	-0.002	-0.014	0.000	0.026	0.028	0.023	0.135	1
2	0.026	1.000	0.097	0.000	-0.313	-0.339	0.000	-0.123	-0.076	0.016	0.094	0.8

3	0.064	0.097	1.000	0.000	0.106	0.017	0.000	0.231	0.406	0.083	0.279	
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.6
5	-0.002	-0.313	0.106	0.000	1.000	0.988	0.000	0.067	0.047	0.026	0.047	
6	-0.014	-0.339	0.017	0.000	0.988	1.000	0.000	0.073	0.043	0.026	-0.031	0.4
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
8	0.026	-0.123	0.231	0.000	0.067	0.073	0.000	1.000	0.961	-0.003	-0.029	0.2
9	0.028	-0.076	0.406	0.000	0.047	0.043	0.000	0.961	1.000	-0.003	-0.001	
10	0.023	0.016	0.083	0.000	0.026	0.026	0.000	-0.003	-0.003	1.000	0.057	0
11	0.135	0.094	0.279	0.000	0.047	-0.031	0.000	-0.029	-0.001	0.057	1.000	
	1	2	3	4	5	6	7	8	9	10	11	-0.2

(e)

As can be seen in Tables 3(a-e), examining the correlation of features in the dataset shows that only a limited number of these features have significant correlation with the class label, while many features either have weak correlation with the target variable or produce redundant and useless information. This highlights the importance of the feature selection process as a key step in improving the performance of classification algorithms, especially in the federated learning framework. This is because selecting relevant features and eliminating redundant features not only reduces computational complexity but can also significantly increase the accuracy and efficiency of the model in each federation. Accordingly, in this study, using the correlation criterion, the features that were most related to the class label were selected as inputs to the classification models, and the results of this selection are presented in Table 4, indicating that each federation is able to have a favorable performance in detection or prediction by using an optimal subset of features.

Table 4. Selected Features in Each Federation

Features Indexes	Federal number
{1,2,3}	1
{1,2,3}	2
{1,2,3}	3
{1,2,3,10}	4
{1,2,3,10}	5

As can be seen in Table 4, similar features have been selected in most federations, which indicates the importance of the main factors in determining intra-organizational fraud in the financial dataset. In fact, in the proposed method, the original dataset, consisting of 10 features, is reduced to 4 features, and in the central federation, the computational complexity is reduced by 60% for final processing and the prediction of new samples.

4-1- Evaluation of Selected Features

In addition, in this study, a lightweight and fast classification algorithm, such as the k-nearest neighbor (KNN) method, was used to accurately evaluate the performance of the selected features in each federation and to examine the impact of these features on the quality of classification. This algorithm is considered a suitable option for examining the efficiency of the selected subsets of features in each

federation due to its simple structure and high execution speed. In this regard, KNN models are first trained on the training data using the features selected separately in each federation to assess the model's ability to learn patterns from real data at the local level.

These trained models are then tested using the testing data to calculate the exact values of the performance evaluation criteria, including precision, sensitivity, overall accuracy, and F-measure, in each federation. This approach allows for qualitative and quantitative comparisons of each federation performance with respect to the selected features and helps to better understand the strengths and weaknesses of feature selection in the federated learning process. The results of these evaluations are presented as comparative graphs in the results section to clearly demonstrate the impact of feature subsets on the quality of prediction in real distributed data conditions.

Figure 4 shows a comparison of the prediction accuracy of test samples in different federations. Figure 5 illustrates a comparison of the prediction sensitivity of test samples in different federations, as well as the average values of the evaluation criteria in each federation. Figure 6 presents the comparison of the precision of test samples in different federations. Figure 7 shows the comparison of the F-measure of the prediction of test samples in different federations.

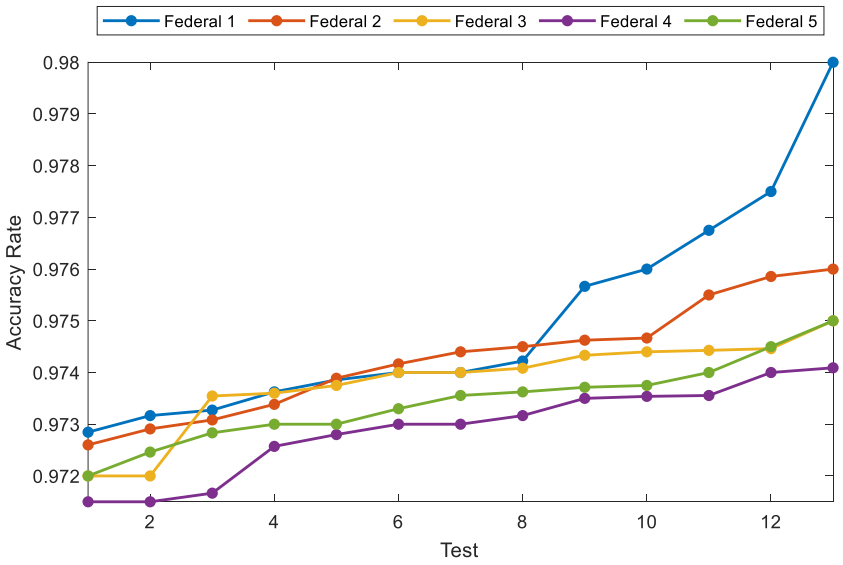

Figure 4. Comparison of the Accuracy of Test Samples in Different Federates Based on Selected Features

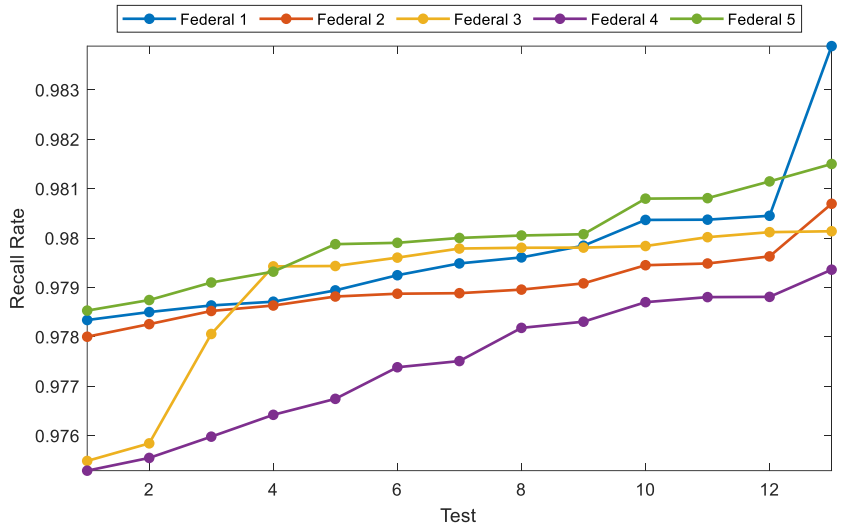

Figure 5. Comparison of the Recall of Test Samples in Different Federates Based on Selected Features

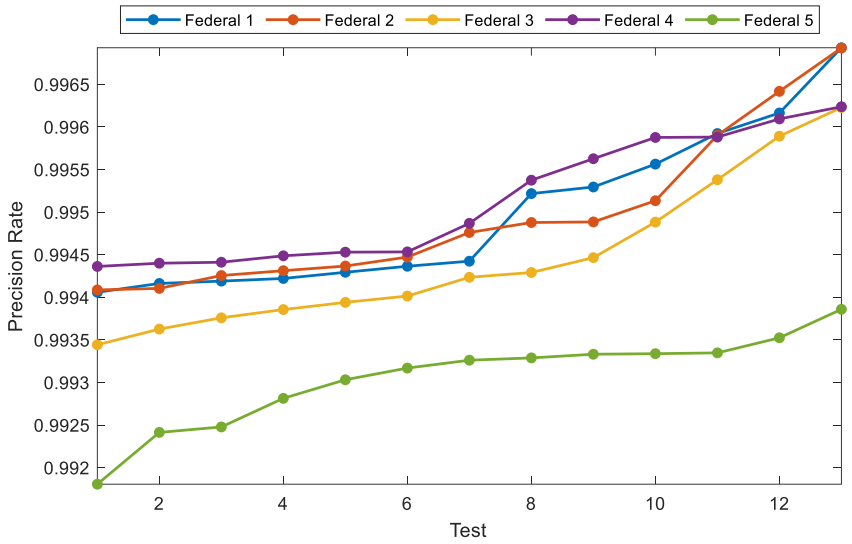

Figure 6. Comparison of the Precision of Test Samples in Different Federates Based on Selected Features

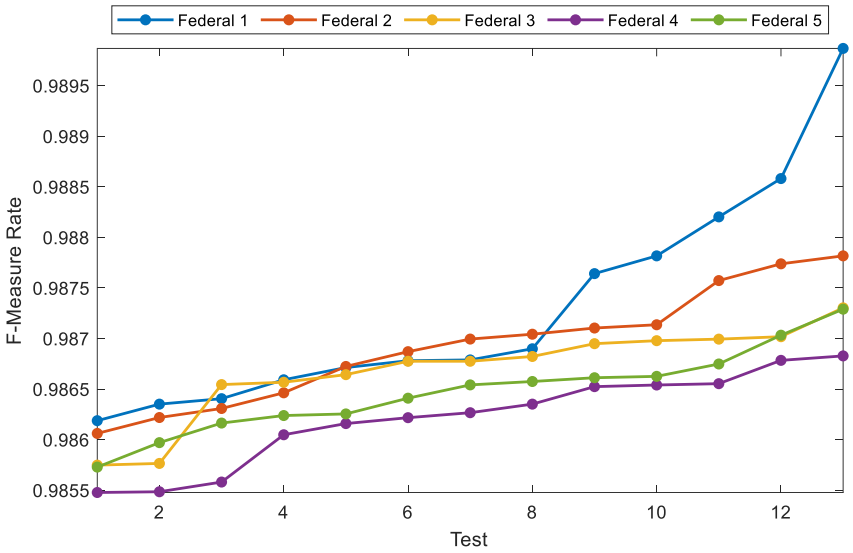

Figure 7. Comparison of the F1-Score of Test Samples in Different Federates Based on Selected Features

Table 5. Average Values of Evaluation Criteria in Different Federates

	Accuracy	Recall	Precision	F-Measure
Federal 1	**0.9775**	**0.9813**	**0.9959**	**0.9886**
Federal 2	**0.9693**	**0.9752**	**0.9935**	**0.9843**
Federal 3	**0.9730**	**0.9798**	**0.9927**	**0.9862**
Federal 4	**0.9720**	**0.9780**	**0.9936**	**0.9857**
Federal 5	**0.9713**	**0.9798**	**0.9909**	**0.9853**

As shown in Figures 4 to 7 and Table 5, the use of correlation-based feature selection enables identifying important factors in determining intra-organizational fraud in financial report data well. Hence, the performance of the proposed classification model in federations has yielded good results for each of the evaluation criteria, including accuracy, sensitivity, precision, and F-measure.

4-2- Central Federation Training

As mentioned, the feature selection results in each federation are sent as output to the central federation in cloud computing. In the central federation, the features selected in the federations are merged to ultimately train the convolutional neural network model based on these features. Convolutional neural network models, with deep training on the features selected in the federations, try to extract accurate intra-organizational fraud detection patterns based on financial data and reports. Figure 8 shows the process of training convolutional neural networks in the central federation in cloud computing.

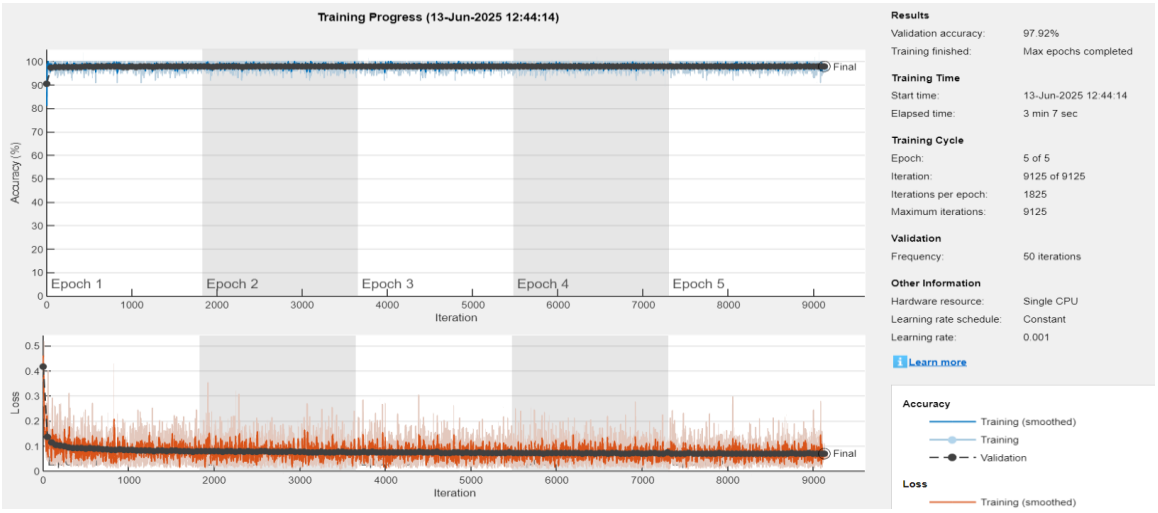

Figure 8. Training Process of Convolutional Neural Networks in Central Federated Cloud Computing

As can be seen in Figure 8, neural networks in the early stages of training usually have a relatively high error rate due to the lack of processed training data and the incompleteness of the learning process. At this stage, the weights and internal parameters of the network are not yet optimally adjusted, and the model is identifying initial patterns in the input data. This behavior is normal in many deep learning networks and is part of the gradual convergence process of the model towards a stable and optimal state.

As the iterations progress and the number of training cycles increases, the network gradually learns the appropriate structure of the relationships and dependencies in the data, and the weights of the different layers are adjusted in a targeted manner. As a result, the error rate decreases and the model tends to converge to an acceptable and stable error value. This gradual reduction in error indicates the efficiency of the neural network's learning and optimization process and its ability to correctly generalize the learned knowledge to new and unknown data. The error downward trend in Figure 9 clearly demonstrates this step-by-step improvement and confirms the correctness of the network structure design and the quality of training data.

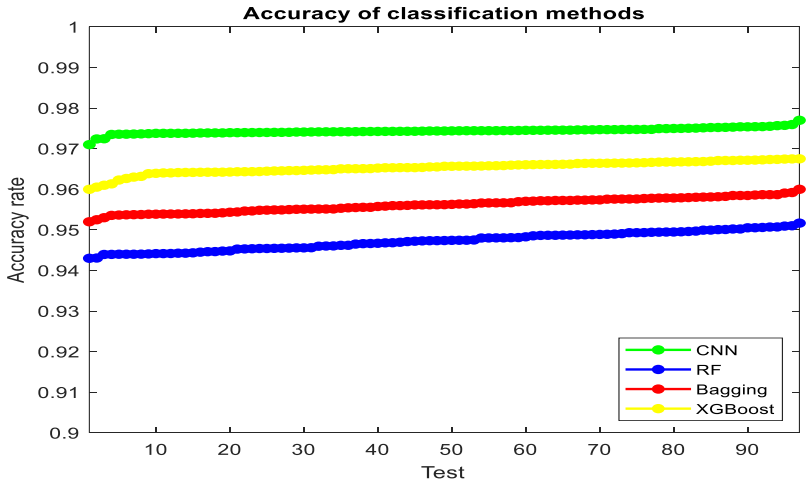

Figure 9. Comparison of the Accuracy of Different Classification Methods on Test Data in the Central Federated System

4.3. Evaluation of the Proposed Method

In the proposed method, various classification algorithms, such as Random Forest (RF), Bagging, and Extreme Gradient Boosting (XGBoost), are employed to classify fraudulent instances and predict new samples. These algorithms, by leveraging their learning capabilities, extract patterns related to fraud from the features selected within the federated systems. These features constitute the basis of the classification process in the central federated system deployed on the cloud computing server. Ultimately, the extracted patterns are utilized to determine the status of new samples.

In this process, 30% of the original data were randomly selected as the test dataset, whose true labels were already known. These labels served as the reference for evaluating the performance of the classification models, and the predictions generated by the models were compared against them. The outcomes of this comparison are presented in the form of a confusion matrix, which provides a standard basis for assessing classification performance using key evaluation metrics.

Figure 9 illustrates the comparison of prediction accuracy on the test samples in the central federated cloud environment across different classification methods. Figure 10 presents the comparison of prediction sensitivity (recall). Figure 11 shows the comparison of prediction precision, while Figure 12 depicts the comparison of the F1-score obtained by the different classifiers.

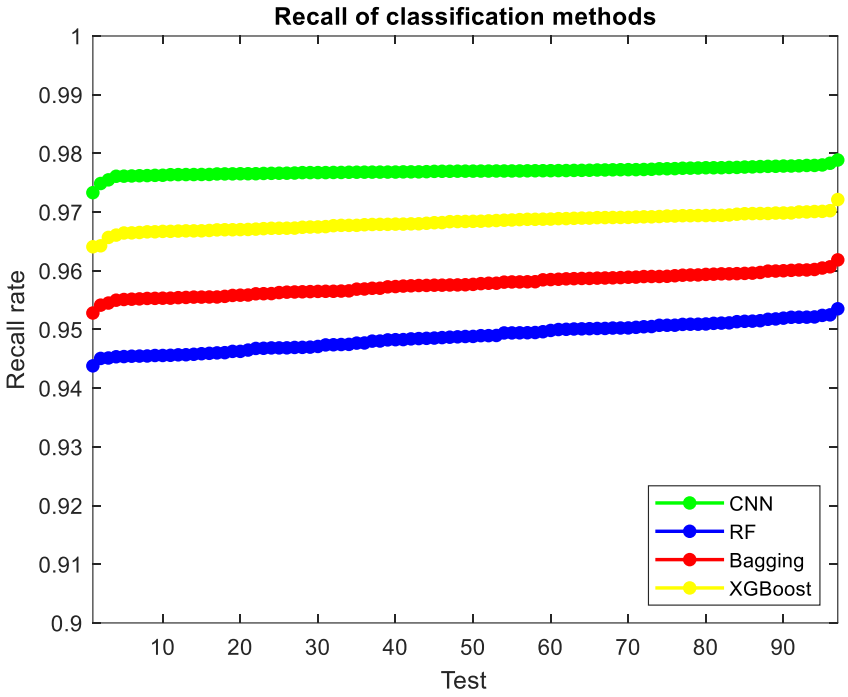

Figure 10. Comparison of Recall of Different Classification Methods on Test Data in the Central Federated System

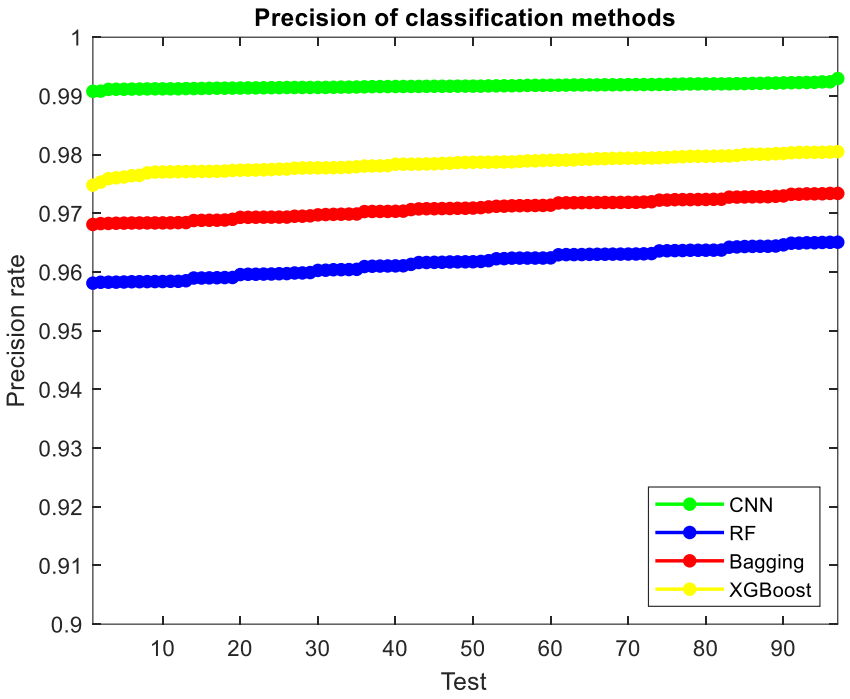

Figure 11. Comparison of Precision of Different Classification Methods on Test Data in the Central Federated System

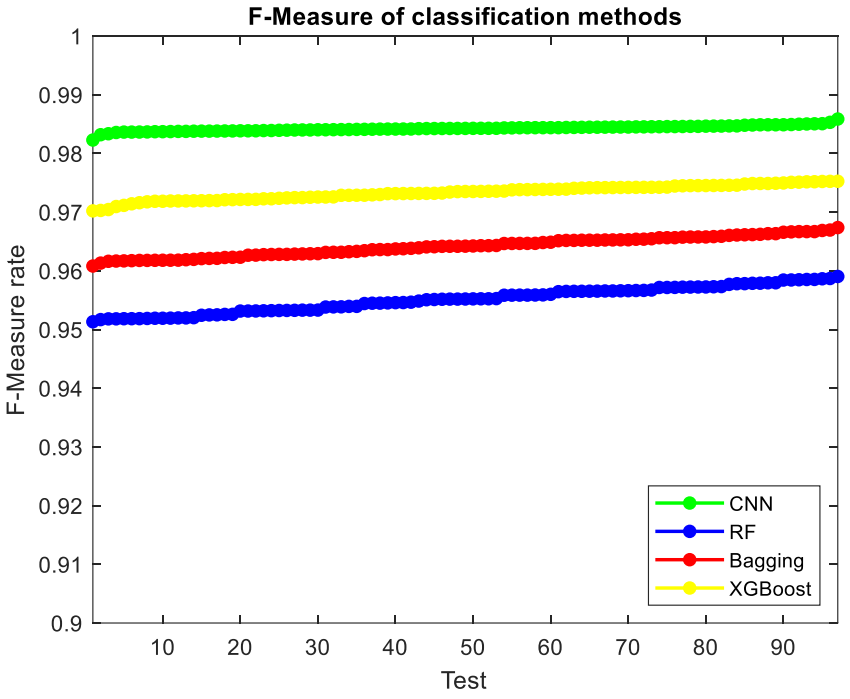

Figure 12. Comparison of F1-Score of Different Classification Methods on Test Data in the Central Federated System

As observed in Figures 9 to 12, the integration of correlation-based feature selection within the federated framework, along with the utilization of the selected features at the central cloud server, has resulted in an effective approach for detecting and predicting fraud-related patterns in inter-

organizational reports. Among the evaluated classifiers, the Convolutional Neural Network (CNN) demonstrates superior performance in terms of overall accuracy compared to other methods.

Figure 13 displays the comparison of average accuracy at the central server based on a box plot. Figure 14 illustrates the comparison of average sensitivity. Figure 15 presents the comparison of average precision, and Figure 16 shows the comparison of the average F1-score, all measured within the central cloud environment.

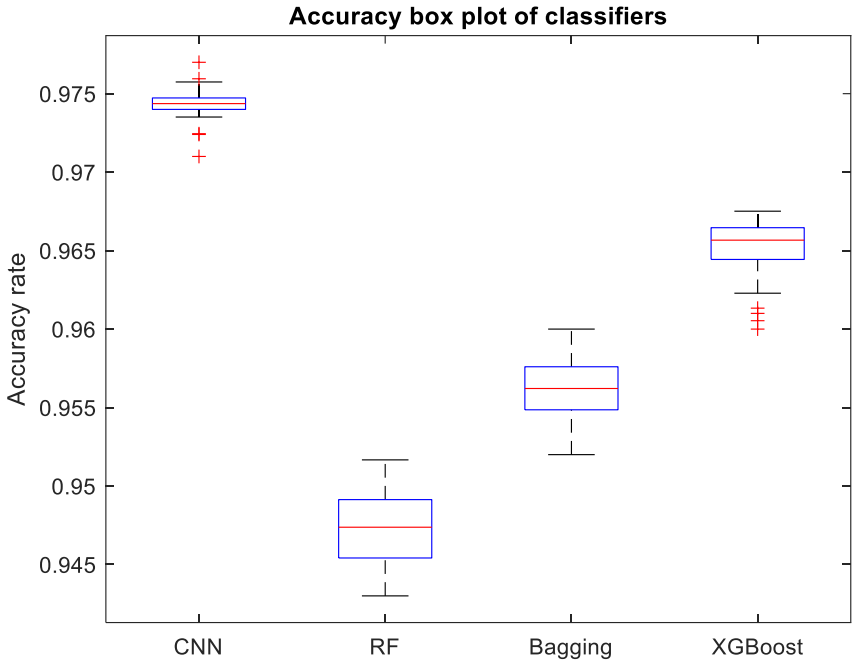

Figure 13. Box Plot Comparing the Accuracy of Different Classification Methods

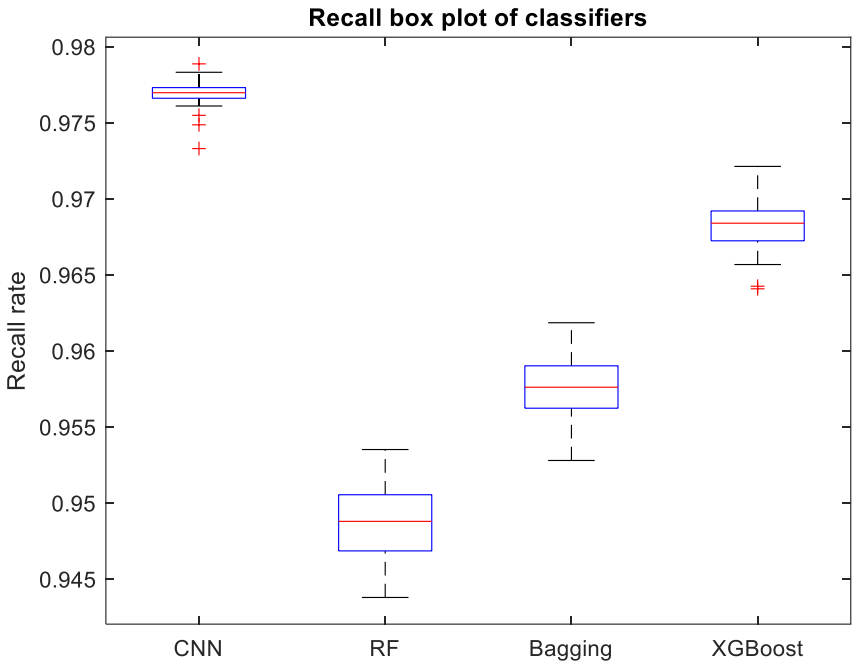

Figure 14. Box Plot Comparing the Recall of Different Classification Methods

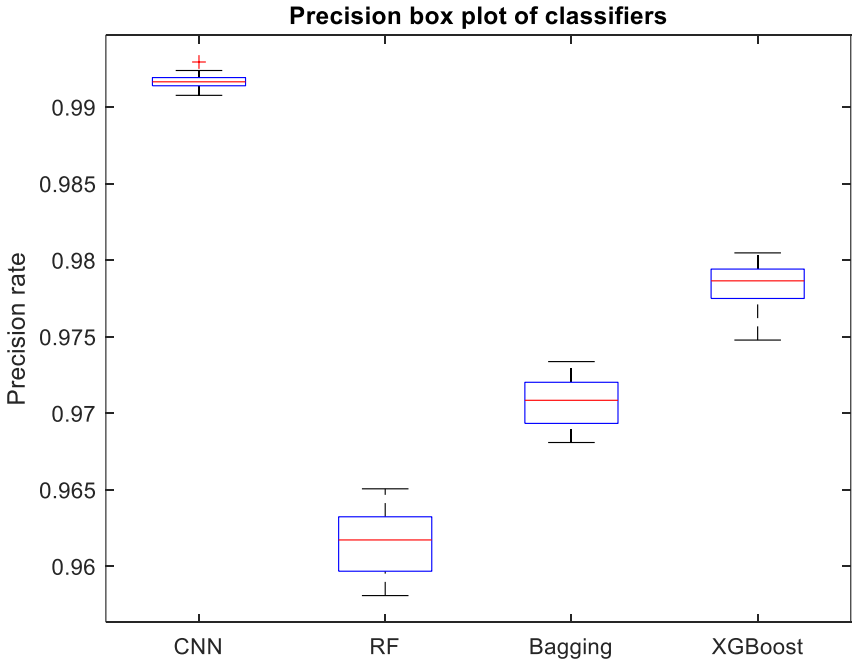

Figure 15. Box Plot Comparing the Precision of Different Classification Methods

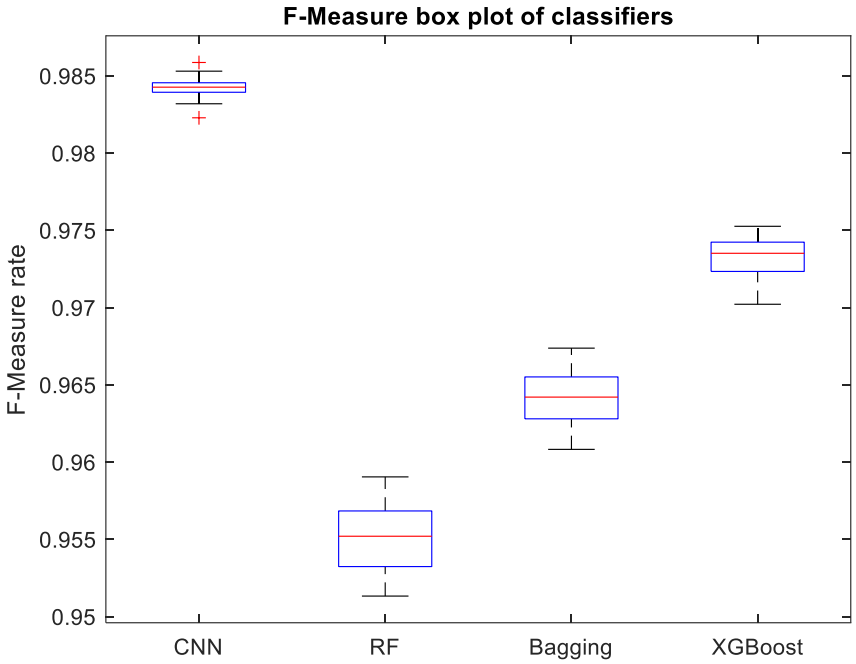

Figure 16. Box Plot Comparing the F1-Score of Different Classification Methods

According to Figures 13 to 16, CNNs, owing to their deep layered architecture, demonstrate a stronger capability to extract more discriminative fraud patterns. Consequently, when these learned patterns were applied to the test dataset, the model produced superior evaluation metric values relative to other classification methods.

4.4. Comparison of the Proposed Method with Previous Studies

Following the implementation and evaluation of the proposed method, its effectiveness was further assessed through comparison with prior studies (Almazroi et al., 2023; Fu, 2022; Mubalaike et al., 2018; Paulraj, 2024) under comparable experimental settings. Given the significance of the accuracy metric in detecting fraudulent samples within financial reporting contexts, Figure 17 presents a comparative analysis of the proposed method against the referenced approaches based on classification accuracy.

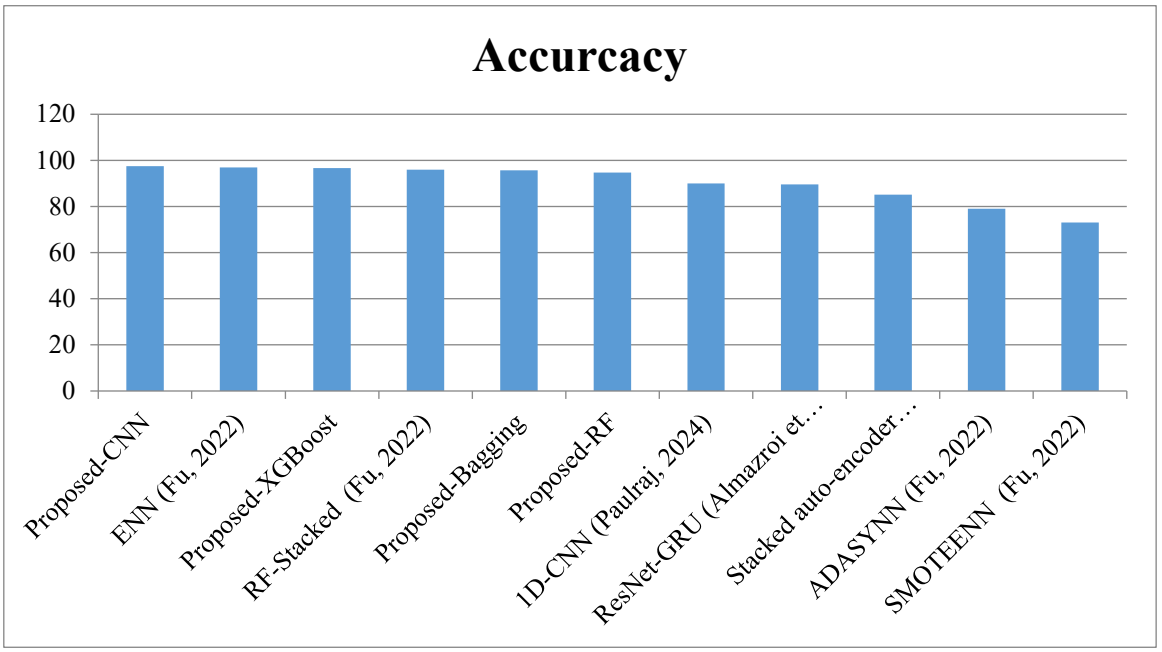

Figure 17. Comparison of the Proposed Method with Previous Methods Based on the Accuracy Metric

According to Figure 17, the proposed method, by employing local feature selection in the fog/edge environment of each federated node (organization), successfully identifies intra-organizational fraud-related factors. These factors are then used as inputs to the Convolutional Neural Network (CNN) model at the central federated layer in the cloud, enabling the prediction of fraud in new financial reports. The utilization of these factors results in improved classification accuracy and more reliable predictions. Under the same experimental conditions, the proposed method demonstrates superior performance compared to previous approaches in predicting fraud within organizational financial reports.

4.5. Analysis of Findings

The proposed model facilitates accurate fraud detection at both intra-organizational and inter-organizational levels. Local data processing and distributed feature selection contribute to privacy preservation while also reducing computational overhead. Furthermore, centralized learning using CNN and ensemble learning methods enhances the system's capability to capture complex and evolving fraud patterns. The federated–cloud architecture exhibits desirable properties in terms of scalability and security, making it suitable for deployment in intelligent monitoring systems across governmental and large-scale organizational contexts. The findings are summarized as follows:

Step 1: Local processing in each federated node

Each organization preprocesses its data within the edge/fog layer as an independent federated entity. The results indicate that the local models achieve high performance, with accuracy values ranging from 0.9693 to 0.9775, and F1-scores ranging from 0.9843 and 0.9886. This stage confirms that privacy-preserving local learning, without transferring raw data, is sufficient for extracting meaningful fraud-related patterns.

Step 2: Feature selection using Pearson correlation

Within each federated node, features associated with fraud occurrence are selected using the Pearson correlation coefficient. This process reduces data dimensionality, accelerates processing, and emphasizes the most informative variables. The distributed nature of feature selection provides an effective basis for improving central model performance.

Step 3: Central learning with deep learning architecture

The selected features from federated nodes are transmitted to the central cloud layer. The CNN-based central model captures complex inter-organizational fraud patterns, achieving an accuracy of

0.9744 and an F1-score of 0.9843. Other ensemble methods, including XGBoost, Bagging, and Random Forest (RF), also yield competitive results with accuracy values exceeding 0.94. These findings highlight the effectiveness of integrating federated learning with centralized deep learning.

Step 4: Inter-organizational fraud detection

By applying the central model, fraud detection is performed without exposing sensitive organizational data. The overall accuracy of inter-organizational fraud detection reaches 97.6%, while computational complexity is reduced by approximately 60%, underscoring the efficiency of the distributed architecture.

The comparative evaluation across different learning scenarios demonstrates that the proposed model is well aligned with real-world operational conditions. In practical environments, organizational data are inherently distributed, heterogeneous, and privacy-sensitive, rendering full centralization impractical. The first-stage results show that each federated entity, analogous to a real organization, can achieve high fraud detection accuracy using only its local data—an essential requirement in privacy-constrained systems.

During the feature selection phase, the reduction of data volume and focus on key attributes mirror the constraints of real monitoring systems, where computational resources are limited and data are often noisy. This step enables the central model to learn from compact yet informative representations, leading to faster response times and more stable decisions.

A comparison of CNN with conventional classifiers, such as RF, Bagging, and XGBoost, reveals that although all methods provide acceptable performance, deep learning exhibits superior capability in modeling nonlinear and sparse fraud patterns, particularly at the inter-organizational level.

Finally, the inter-organizational detection results indicate that the proposed federated–cloud architecture satisfies not only accuracy requirements but also efficiency and scalability considerations. The substantial reduction in computational burden, coupled with the elimination of raw data transfer, supports the practical applicability of the model in large-scale and dynamic organizational ecosystems. Overall, these observations suggest that the proposed framework extends beyond a purely experimental setting and can be realistically deployed in executive and governmental environments.

5) Conclusion and Suggestions

With the expansion of information systems and the digitalization of processes in executive agencies, a vast volume of sensitive and heterogeneous data is generated in a distributed manner. Despite its significant analytical value, such data are often underutilized due to data fragmentation, security constraints, and the absence of real-time analytical infrastructures. This condition not only complicates transaction monitoring but also contributes to the growth of fraud and corruption, particularly in governmental institutions that still rely on traditional rule-based approaches for fraud detection. While these approaches may be effective for recognizing predefined patterns, they are typically ineffective against emerging and adaptive fraud schemes, often resulting in excessive false alarms and increased investigation costs.

Under these circumstances, adopting modern solutions grounded in artificial intelligence, machine learning, and cloud computing becomes essential. Machine and deep learning techniques can substantially enhance fraud detection accuracy by identifying complex and nonlinear behavioral patterns in near real time, while cloud computing provides scalable computational resources and supports secure inter-organizational collaboration. Nevertheless, the deployment of such technologies requires robust data governance and AI governance frameworks to ensure an appropriate balance among efficiency, transparency, and privacy.

In this study, a novel data governance model was proposed based on a decentralized, data-centric architecture, integrating federated learning with cloud computing infrastructure to enable cross-organizational fraud detection without centralized data aggregation. Within this framework, each organization operates as an independent federated entity, performing data preprocessing at the edge/fog layer using local resources. This preprocessing stage involves selecting features exhibiting the strongest

association with fraud occurrence, implemented through Pearson correlation-based feature selection, thereby identifying the most informative factors for subsequent analysis.

The experimental findings demonstrate the effectiveness of the proposed model. As indicated by the federated evaluation results, each federated entity achieved high predictive performance using only local data, with accuracy values ranging from 0.9693 to 0.9775. The corresponding F1-score, precision, and recall values fall within 0.9843–0.9886, 0.9909–0.9959, and 0.9752–0.9813, respectively. These results confirm the capability of the system to correctly identify fraudulent samples while minimizing false positives. Moreover, they underscore the importance of correlation-driven feature selection in reducing redundant data processing and enhancing model efficiency. A key advantage of this mechanism is that sensitive organizational data remain local, ensuring strict privacy preservation, which is critical in inter-organizational environments.

At the centralized level, both deep learning and ensemble learning models exhibited strong performance. The Convolutional Neural Network (CNN) achieved the best overall results, with an accuracy of 0.9744 and an F1-score of 0.9843, while XGBoost, Bagging, and Random Forest (RF) also delivered competitive accuracy values exceeding 0.94. These observations indicate that aggregating knowledge from distributed federated entities within a central deep learning framework significantly strengthens the detection of complex inter-organizational fraud patterns. The combined use of CNN and ensemble strategies contributes to improved generalization and reduced classification errors, enabling the detection of previously unseen fraud behaviors.

Comparative analysis further validates the superiority of the proposed framework. The achieved accuracy of 97.46% surpasses that of several benchmark and hybrid models, including ENN, RF-Stacked, ResNet–GRU, and Stacked Auto-Encoder. This outcome highlights the effectiveness of integrating federated learning, intelligent feature selection, and centralized deep learning, which collectively address limitations inherent in purely rule-based or historically trained systems.

The principal practical contributions of the proposed model can be summarized as follows:

- Data privacy and security: Data processing remains local to each federated entity, with only abstracted knowledge or selected features transmitted, thereby preventing sensitive data exposure.
- Scalability and real-time capability: The integration of cloud and edge/fog computing enables efficient handling of high-volume transactional data and supports near real-time fraud detection.
- Cross-organizational generalization: The central model benefits from knowledge extracted across multiple organizations, improving robustness in heterogeneous operational settings.
- Reduction of false alarms: Effective feature selection and deep learning mechanisms contribute to lower false positive rates and optimized resource utilization.
- Alignment with governance requirements: The architecture is inherently compatible with privacy regulations and data governance principles, promoting secure collaboration.

Despite these advantages, certain challenges warrant consideration. Effective deployment requires standardized communication protocols, interoperable security mechanisms, and reliable strategies for knowledge aggregation among federated entities. Additionally, federated learning environments characterized by data imbalance or feature heterogeneity may necessitate advanced weighting, normalization, and adaptive optimization techniques to maintain model stability and fairness. Addressing these factors represents a valuable direction for future research and practical refinement.

Overall, the findings indicate that the proposed model is not only capable of accurately detecting intra-organizational and inter-organizational fraud but also offers a practical and sustainable solution for government information systems and large organizations through privacy preservation, cost reduction, and enhanced scalability. This study emphasizes the critical role of data governance, the design of federated–cloud architectures, and the integration of federated learning with deep learning in achieving an appropriate balance among security, detection accuracy, and operational efficiency.

The extracted features and learned patterns, representing the factors associated with fraudulent behavior, are transmitted to the central cloud layer, where deep learning-based aggregation and fraud prediction processes are performed at the inter-organizational level. Experimental results further demonstrate that the proposed method, with an approximate 60% reduction in computational complexity, achieves a detection accuracy of 97.6% for newly observed fraud cases across organizations. This outcome reflects the effectiveness of the decentralized architecture and distributed feature selection strategy compared to conventional centralized approaches.

Considering these results, it is recommended that managers and policymakers responsible for data governance and monitoring in executive agencies reassess fully centralized data architectures and progressively adopt distributed federated structures. The application of federated learning enables organizations to benefit from shared analytical knowledge without violating privacy regulations or transferring sensitive raw data. Consequently, the development of macro-level policies centered on exchanging "knowledge rather than data" can substantially strengthen inter-organizational trust and collaboration.

It is also advisable that data governance frameworks clearly define the responsibilities of the central federated entity as a knowledge aggregator and supervisory component. Such responsibilities should remain transparent, restricted, and auditable in order to mitigate risks associated with excessive centralization and potential security vulnerabilities. Furthermore, investment in secure cloud infrastructure and the training of specialized human resources in federated learning and distributed AI systems represent essential managerial prerequisites for the successful operationalization of this model.

From an applied perspective, the results suggest that the proposed framework can serve as a foundational component of intelligent fraud detection systems within governmental and regulatory environments. In practical deployments, each executive agency should be modeled as an independent federated node, with preprocessing and feature selection procedures adapted to the specific characteristics of its local data to preserve model accuracy. A phased implementation strategy—initially through decision support systems followed by real-time detection mechanisms—may further reduce operational risks.

The adoption of this model in domains such as financial supervision, public procurement, subsidy allocation, and audit systems can facilitate earlier identification of concealed fraud patterns. Additionally, incorporating mechanisms for continuous performance monitoring and periodic model updates, aligned with evolving fraud behaviors, is strongly recommended to ensure long-term system effectiveness.

Resources

Adejumo, A., & Ogburie, C. (2025). Forensic accounting in financial fraud detection: Trends and challenges. *International Journal of Science and Research Archive, 14,* 1219-1232. https://doi.org/10.30574/ijsra.2025.14.3.1209.

Ahmad, M., Ahmed, Z., Alvarado, R., Hussain, N., & Khan, S. A. (2024). Financial development, resource richness, eco-innovation, and sustainable development: Does geopolitical risk matter? *Journal of Environmental Management, 351*, 119824. https://doi.org/10.1016/j.jenvman.2023.119824.

Alazzabi, W. Y. E., Mustafa, H., & Karage, A. I. (2023). Risk management, top management support, internal audit activities and fraud mitigation. *Journal of Financial Crime, 30*(2), 569-582. https://doi.org/10.1108/JFC-11-2019-0147

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., . . . Saif, A. (2022). Financial fraud detection based on machine learning: A systematic literature review. *Applied Sciences, 12*(19), 9637. https://doi.org/10.3390/app12199637

Almazroi, A. A., & Ayub, N. (2023). Online payment fraud detection model using machine learning techniques. *IEEE Access, 11*, 137188-137203. https://doi.org/10.1109/ACCESS.2023.10341223

Andayani, W., & Wuryantoro, M. (2023). Good corporate governance, corporate social responsibility and fraud detection of financial statements. *International Journal of Professional Business Review: Int. J. Prof. Bus. Rev., 8*(5), 9. https://doi.org/10.26668/businessreview/2023.v8i5.1051

Ashtiani, M. N., & Raahemi, B. (2021). Intelligent fraud detection in financial statements using machine learning and data mining: A systematic literature review. *IEEE Access, 10*, 72504-72525 https://doi.org/10.1109/ACCESS.2021.3080243

Bello, H. O., Ige, A. B., & Ameyaw, M. N. (2024). Deep learning in high-frequency trading: Conceptual challenges and solutions for real-time fraud detection. *World Journal of Advanced Engineering Technology and Sciences, 12*(02), 035-046. https://doi.org/10.5281/zenodo.10812284

Bello, O. A., Folorunso, A., Ejiofor, O. E., Budale, F. Z., Adebayo, K., & Babatunde, O. A. (2023). Machine learning approaches for enhancing fraud prevention in financial transactions. *International Journal of Management Technology, 10*(1), 85-108.

Canton, H. (2021). Organisation for economic co-operation and development—OECD. In *The Europa Directory of International Organizations 2021* (pp. 677-687). Routledge.

Das, R., Sirazy, M., Khan, R. S., & Rahman, S. (2023). A collaborative intelligence (ci) framework for fraud detection in us federal relief programs. *Applied Research in Artificial Intelligence and Cloud Computing, 6*(9), 47-59. https://doi.org/10.5281/zenodo.10812282

Dorostkar Navaei, Y., Rezvani, M. H., & Eftekhari Moghadam, A. M. (2025). Identifying key influencers in social networks: A new neighborhood-based method. *Journal of Soft Computing and Information Technology, 14*(1), 11-26. https://doi.org/10.22034/jscit.2025.219059

Du, J. (2018). Understanding of object detection based on CNN family and YOLO. *Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/1004/1/012029

Favour, A. A. (2022, September 2025). *Regulatory compliance challenges in cloud-based AI fraud detection systems*. ReseachGate. www.researchgate.net/publication/391454540

Fu, Z. (2022). Check for updates stacking model for financial fraud detection with synthetic data. In *The Proceedings of the 2022 International Conference on Bigdata Blockchain and Economy Management (ICBBEM 2022)*. https://doi.org/10.2991/978-94-6463-030-5_8

Ha, W., Gang, S., Navaei, Y. D., Gezawa, A. S., & Nanehkaran, Y. A. (2025). Ordered clustering-based semantic music recommender system using deep learning selection. *Computers, Materials & Continua, 83*(2), https://doi.org/10.32604/cmc.2025.061343.

Halbouni, S. S., Obeid, N., & Garbou, A. (2016). Corporate governance and information technology in fraud prevention and detection: Evidence from the UAE. *Managerial Auditing Journal, 31*(6/7), 589-628. https://doi.org/10.1108/MAJ-12-2014-1118

Hassan, S. W. U., Kiran, S., Gul, S., Khatatbeh, I. N., & Zainab, B. (2025). The perception of accountants/auditors on the role of corporate governance and information technology in fraud detection and prevention. *Journal of Financial Reporting and Accounting, 23*(1), 5-29. https://doi.org/10.1108/JFRA-05-2023-0235

He, Fen, et al. (2021). Applications of deep learning techniques for pedestrian detection in smart environments: A comprehensive study. *Journal of Advanced Transportation, 2021*(1), 5549111. https://doi.org/10.1155/2021/5549111

Islam, S., Haque, M. M., & Karim, A. N. M. R. (2024). A rule-based machine learning model for financial fraud detection. *International Journal of Electrical & Computer Engineering (2088-8708), 14*(1). https://doi.org/10.11591/ijece.v14i1.pp1-10

Kanamori, S., Abe, T., Ito, T., Emura, K., Wang, L., Yamamoto, S., . . . Nojima, R. (2022). Privacy-preserving federated learning for detecting fraudulent financial transactions in japanese banks. *Journal of Information Processing, 30*, 789-795. https://doi.org/10.2197/ipsjjip.30.789

Katari, A., & Ankam, M. (2022). Data governance in multi-cloud environments for financial services: Challenges and solutions. *Educational Research (IJMCER), 4*(1), 339-353. https://doi.org/10.6731/TPCC_proceedings.003b-001-R5-00009062

Mirfallah Lialestani, M., Khamseh, A., Radfar, R., (2021). Digital transformation model, based on grounded theory. *Journal of Information Systems and Telecommunication, 9*(4). https://doi.org/10.58278/2345-2775-9-4-9

Moghadam, S. S., Ghorbani, A., & Forouzesh, R. (2021). Review and analysis of INTOSAI standards for good governance. *International Journal of Political Science, 11*(3), 63-76. https://doi.org/10.22034/ijps.2021.270134.1049

Mohanty, B., & Mishra, S. (2023). Role of artificial intelligence in financial fraud detection. *Academy of Marketing Studies Journal, 27*(S4). https://doi.org/10.54615/AMSJ.2023.S4.001

Mubalaike, A. M., & Adali, E. (2018). Deep learning approach for intelligent financial fraud detection system. In *The 2018 3rd International Conference on Computer Science and Engineering (UBMK)*. https://doi.org/10.1109/UBMK.2018.8566574

Myalil, D., Rajan, M., Apte, M., & Lodha, S. (2021). Robust collaborative fraudulent transaction detection using federated learning. In *The 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*. https://doi.org/10.1109/ICMLA52953.2021.00110

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv*, 1511.08458. https://doi.org/10.48550/arXiv.1511.08458

Pamisetty, V. (2023, December 15). Leveraging AI, big data, and cloud computing for enhanced tax compliance, fraud detection, and fiscal impact analysis in government financial management. *Fraud Detection, and Fiscal Impact Analysis in Government Financial Management*. https://doi.org/10.21275/SR23122164932

Pamisetty, V., Pandiri, L., Annapareddy, V. N., & Sriram, H. K. (2022, June 15). Leveraging AI, machine learning, and big data for enhancing tax compliance, fraud detection, and predictive analytics in government financial management. *Machine Learning, And Big Data For Enhancing Tax Compliance, Fraud Detection, And Predictive Analytics In Government Financial Management*. https://doi.org/10.21275/SR22615175938

Paulraj, B. (2024). Machine learning approaches for credit card fraud detection: A comparative analysis and the promise of 1D convolutional neural networks. In *The 2024 7th International Conference on Information and Computer Technologies (ICICT)*. https://doi.org/10.1109/ICICT59662.2024.00035

Rehan, H. (2021). Leveraging AI and cloud computing for Real-Time fraud detection in financial systems. *Journal of Science & Technology, 2*(5), 127. https://doi.org/10.52451/jst.v2i5.127

Rivandi, E., & Jamili Oskouei, R. (2025). A novel approach for developing intrusion detection systems in mobile social networks. *Journal of Soft Computing and Decision Analytics, 3*(1), 158-170. https://doi.org/ 10.31181/jscda31202535

Salmanov, T. (2025). Enhancing corporate governance via machine learning and statistical tools for fraud detection. *Advances in Corporate Governance, 2*(1), 6. https://doi.org/10.54616/acg.2025.02106

Samuel, A. (2023). Enhancing financial fraud detection with AI and cloud-based big data analytics: Security implications. *SSRN 5273292*. https://doi.org/10.2139/ssrn.5273292

Shi, F., & Zhao, C. (2023). Enhancing financial fraud detection with hierarchical graph attention networks: A study on integrating local and extensive structural information. *Finance Research Letters, 58*, 104458. https://doi.org/10.1016/j.frl.2023.104458

Stojanović, B., & Božić, J. (2022). Robust financial fraud alerting system based in the cloud environment. *Sensors, 22*(23), 9461. https://doi.org/10.3390/s22239461

Upreti, K., Vats, P., Srinivasan, A., Daya Sagar, K., Mahaveerakannan, R., & Charles Babu, G. (2025). Detection of banking financial frauds using hyper-parameter tuning of DL in cloud computing environment. *International Journal of Cooperative Information Systems, 34*(01), 2350024. https://doi.org/ 10.1142/S0218843023500248

Wu, J. (2017). Introduction to convolutional neural networks. National key lab for novel software technology. *Nanjing University. China, 5*(23), 495. https://doi.org/10.13140/RG.2.2.36353.22883

Yandrapalli, V. (2024). AI-powered data governance: A cutting-edge method for ensuring data quality for machine learning applications. In *The 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE).* https://doi.org/10.1109/ICETITE57412.2024.00023

Yildirim, P. (2015). Filter based feature selection methods for prediction of risks in hepatitis disease. *International Journal of Machine Learning and Computing, 5*(4), 258. https://doi.org/10.7763/IJMLC.2015.V5.525

Zhang, R., Shu, H., & Navaei, Y. D. (2022). Load balancing in edge computing using integer linear programming based genetic algorithm and multilevel control approach. *Wireless Communications and Mobile Computing, 2022*(1), 6125246. https://doi.org/10.1155/2022/612524

Zhou, H., Deng, Z., Xia, Y., & Fu, M. (2016). A new sampling method in particle filter based on Pearson correlation coefficient. *Neurocomputing, 216*, 208-215. https://doi.org/10.1016/j.neucom.2016.07.036