Journal of Theoretical and Applied Information Technology
30th April 2025. Vol.103. No.8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AI-DRIVEN FRAUD DETECTION AND SECURITY SOLUTIONS: ENHANCING ACCURACY IN FINANCIAL SYSTEMS

JANJHYAM VENKATA NAGA RAMESH [1 A,B], DR SUDHANSU SEKHAR NANDA [2], VAMSI KRISHNA CHIDIPOTHU [3], VEMULA JASMINE SOWMYA [4], AMIT VERMA [5], TAOUFIK SAIDANI [6*]

[1 a] Adjunct Professor, Department of CSE, Graphic Era Hill University, Dehradun,248002, India.
[1 b] Adjunct Professor , Department of CSE, Graphic Era Deemed To Be University , Dehradun, Uttarakhand, India.
2 Associate Professor, Sri Sri University, Cuttack, Odisha, India
[3] Research Student, Department of Information Technology, University of the cumberlands, college Station drive. Williams burg, Kentucky.
[4] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India.
[5] University Centre for Research and Development, Chandigarh University, Gharuan Mohali, Punjab, India.
[6] Center for Scientific Research and Entrepreneurship, Northern Border University, 73213, Arar, Saudi Arabia.
E-mail: [1] jvnramesh@gmail.com, [2] nandasudhansusekhar.87@gmail.com, [3] krishynavamsi@gmail.com, [4] vemulajasmine@gmail.com, [5] amit.e9679@cumail.in, [6] . Taoufik.Saidan@nbu.edu.sa

ABSTRACT

Financial institutions worldwide are facing a major challenge in identifying unauthorized transactions. Financial organizations need sophisticated fraud detection systems to safeguard their financial stability and maintain customer confidence; nevertheless, the design of such systems may be hindered by many complexities. The infrequency of illegal transactions and the disproportion in several transaction datasets, namely the few occurrences of fraudulent transactions relative to legitimate ones, exemplify these characteristics. The dataset disparity may undermine the accuracy and effectiveness of a fraudulent activity detection program. Disseminating customer information to provide a more effective centralized approach is impractical, since each bank is required to comply with data protection standards. To ensure the user experience remains unaltered, the fraud detection solution must be both accessible and easily observable. As a result, this study presents an innovative approach in tackling these issues by combining Federated Learning (FL) with Explainable AI (XAI). FL financial institutions work together to train a model that can identify fraudulent activities without sharing client information directly. This protects the privacy and security of client information. The addition of XAI makes it definite that human specialists can understand and analyse the model's results, which gives the system more reliability and transparency. The FL-based fraudulent detection system regularly shows high-performance measures, as shown by experiments employing real transaction datasets. This study demonstrates that federated learning might serve as an effective and confidential instrument in combating fraudulent activity.

Keywords: *Federated Learning, Explainable AI, Fraud Detection Systems, Data Privacy, Client Information*

1. INTRODUCTION

Financial transaction integrity is crucial in digital banking [1]. Digital platforms provide exceptional convenience, but they also expose customers to rising cyber dangers, notably bank account theft. Fraudulent actions threaten the global economy, financial security, and financial institution confidence [2]. Fraud, particularly in electronic banking and credit card transactions, causes yearly losses that need powerful detection techniques [3, 4]. Financial organizations are researching new

fraud tactics [5]. The dynamic nature of unlawful operations, especially bank account fraud, which varies from credit card theft in its nuanced forms such as unauthorized transfers and account takeovers, provides distinct issues [6]. The psychological and economic effects on victims need detailed investigations using varied datasets. ML is widely used in fraud detection systems. However, data privacy problems and class disparities in financial datasets, which have a disproportionately low number of fraudulent transactions, impede model construction [7]. While centralized ML techniques can handle vast amounts of data, they fail to adapt to new fraud trends across institutions [8]. This study uses Federated Learning (FL), a privacy-preserving distributed learning framework, to train models collaboratively without data sharing [9]. FL allows organizations to train models locally and aggregate updates, enabling knowledge transfer and data confidentiality [10]. Explainable AI (XAI) approaches improve model transparency and confidence in AI-driven fraud detection, which is 'black box' [11].

This research integrates FL and XAI to improve bank account fraud detection. The study has certain goals:

- Implement FL-based fraud detection systems with local data storage and centralized model update aggregation to improve data privacy. Show that the recommended strategy is feasible using a web app.
- Create a reliable deep neural network (DNN) model to identify fraudulent transactions in federated datasets. Improve model transparency and trust by using XAI approaches to interpret choices.
- Impact on Practice: The FL-XAI system has several practical advantages. Financial institutions may work together to prevent fraud without sacrificing customer privacy, building confidence and security. The openness of XAI helps comprehend and validate model predictions, improving regulatory compliance and operational efficiency.

The modular solution is straightforward to implement and integrate into current financial systems.[12] This study advances safe and interpretable AI-driven fraud detection, providing a scalable banking industry solution. Protecting the honesty and authenticity of monetary transactions is of the highest priority in this stage of digital banking. Despite the incredible ease this transition brings, it also leaves clients at risk from cyber challenges, with bank account hacking being a major concern.[13] There is an immense risk to the worldwide economy, people's financial security, and the credibility of financial institutions from financial scams, especially in the field of electronic banking with transactions made with credit cards. Fraud causes billions of dollars in deficits every year, which shows that there is a need for stronger techniques for identifying it.[14] As a result, financial institutions have initiated and supported research to detect and prevent fraudulent activities of any sort. However, because of its varied behaviours and constantly changing strategies, illicit activity is still challenging to grasp. The focus of this research is bank-related fraudulent activity, which is a common area that has been the focus of several studies.[15] The techniques, effects, and difficulties in detecting bank account fraud are different from those of other financial frauds. Bank account fraud might appear in more subdued forms, such as illegal transfers of cash, account acquisitions, or even fraud that results in the opening of fresh accounts, in contrast with credit card theft, where illegal activities can be quickly identified because of odd spending patterns. The victim may have over time psychological and economic consequences. Extensive study supported by large and varied datasets is necessary to comprehend and prevent these obstacles. Machine Learning (ML) is often used to train systems to make exact guesses based on data inputs, which is useful for making systems that can spot financial accounts' fraudulent activities. Selecting a machine learning method depends on the type of information and the fraud that the intended victim is trying to find. The records of bank account activities not only contain private information, but they also show an imbalance, with fewer fake transactions than real ones. These features make it harder to create a strong method for finding fraudulent transactions.[16] As a reflection of a centralized ML technique, bankers use their information to train several ML models to detect possibly illicit actions. This centralized approach is now extensively used in the financial services industry, mostly because of its capacity to handle massive volumes of data and detect underlying trends. However, there is a downside to the centralized approach: different institutions deal with different types of fraud, and it can be harder for them to spot new forms of abuse. At this moment, federated learning becomes relevant.

New Federated Learning (FL) method for autonomous machine learning that protects privacy. It offers an alternate solution to this problem by

letting models be trained in local gadgets and only sharing changes that have been authorized by everyone. Collaboration and data security are the main areas where the unified method is different. The first one only works in one financial institution, but the second one involves numerous financial institutions working together [16]. More than a new technology approach, FL stands out in the fraud detection environment as an indication of a secret and collaborative countermeasure to illicit activities [17-19]. The reason is that it can integrate data from several sources without requiring direct data interchange, which is becoming an increasingly important feature [20,21]. Furthermore, FL prioritizes the communication of model changes over heavy data, making the process quicker, and more efficient, and protecting client data from any compromise [22].

AI-based methods for finding scams are also unclear and function in a "black box." Artificial intelligence must be efficient and reliable for important tasks like finding financial criminal activity [23-25]. Researchers solve the issue by adding XAI techniques to the FL-based financial issues scam detection described above. Not only does the proposed approach protect user privacy, but it also offers a way for people to work together to train artificial intelligence models, and it can be trusted. As a result, this study suggests a method for detecting fraud that makes use of FL & XAI together. There are a lot of positive features to this approach, but two of the most important are the openness and protection of client confidentiality. A synopsis of the study's findings is provided below.

- By using an FL method for an improved fraud detection system, the privacy of financial datasets may be improved by ensuring that individual information is stored locally and only model updates are centralized. This approach naturally protects financial dataset privacy, which is a significant benefit in this era of data sensitivity.

- By adding the suggested FL-based system to a web-based app, to see if the suggested method is feasible. Design a highly accurate model for a deep neural network, that can detect anomalies linked to forged transactions across multiple databases.

- By incorporating the XAI technology, the model's choices become transparent, guaranteeing a new approach for identifying fraudulent activity systems.

The final parts of this paper are shown below. In the second section, researcher discusses about the problem statement. In Section Three, methodology is provided, in Section 4, the researcher goes in-depth about how the plan is implemented. The findings are provided in section five. In the sixth part, the study's conclusion is presented. Section 3 discusses the technique, Section 4 implements it, Section 5 offers the data and analysis, and Section 6 summarizes the research.

1.1 PROBLEM STATEMENT

Financial institutions have an increasing problem in efficiently identifying and mitigating bank account fraud, attributed to the advancing complexity of cyber threats and the intrinsic constraints of conventional, centralized machine learning methodologies. The limitations include the potential exposure of sensitive client information, challenges in reacting to changing fraud tendencies across varied datasets, and the absence of transparency in AI-driven decision-making. There is an urgent need for a strong, privacy-preserving, and interpretable fraud detection system that can use collaborative learning while maintaining data security and ensuring clarity in model predictions.

2. METHODOLOGY:

The researcher described the suggested system design and the research methods in this section.

2.1 Architectural Design for the Proposed System

The researcher provided the suggested system architecture diagram that describes the framework and process of the FL technique for detecting bank fraud in Figure 1. A server that coordinates the training and aggregation of models across several financial institutions is at the basis of this system. Data remains inside each institution, protecting it against unauthorized access, as each node houses local DNN models trained on private data. These local models share their opinions, rather than the data itself, with a global model regularly. This helps the global model get better. This design not only uses the combined knowledge of all the organizations involved, but it also recognizes the necessity it is to keep financial data safe.

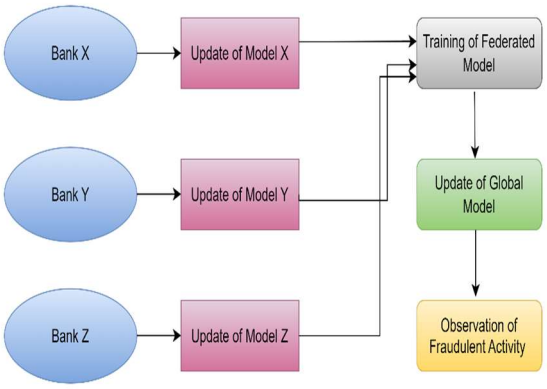


Figure 1 Block Diagram of the Suggested Federated Learning

Individualistic roles and actions are performed by both the client & the server. The global model has been started by the server, usually from a pre-trained model or with random values, and then sent to all participating banks to undergo local training. The server is in charge of combining the model updates that clients give back after finishing their local training to optimize the global model. More complex aggregation methods can also be used instead of summing weights. The server may use a held-out validation dataset to make sure the newly updated global model meets the speed standards after aggregation. Assuring that clients are using the most up-to-date edition of the global model; the server manages the linking of model changes. Additionally, it handles any contact that needs to happen between clients, even though the direct discussion between clients is usually very limited.
Additionally, the client must train the obtained global model using their local information set. This requires running the model through many training epochs to fine-tune it using their unique data. When clients finish training locally, they update the model by sending any changes back to the main server. Protecting sensitive information is not an issue since no raw data is sent. Once it comes to raw data, clients must guarantee that it stays in the local setting. For privacy reasons and to make sure that privacy laws are followed, all information preparation, cleaning, and training are carried out in-house. Clients get the new global model for the next round of training after the server collects and improves it. To find fraud in new deals in real-time, clients can use the locally learned model and the FL system's group intelligence, all without putting data security at risk. The convergence and consistency in globally shared models are both guaranteed by this method. To identify patterns linked to fraudulent actions across federated datasets, a customized DNN framework was developed rather than relying on generic models. To better detect and target illicit activity, the model's design makes use of the massive computing power of DNNs.
In addition, a unique method of fraud detection is presented by the suggested system model, which integrates FL and XAI. Since FL guarantees effective model training across many platforms without affecting the confidentiality of information, XAI provides model conclusions that are clear and understandable. Particularly in fields where comprehending the reasoning behind a model's choice is just as important as the choice itself, this combination is crucial. There are several advantages to FL. Firstly, there is a marked improvement in the security and confidentiality of data. This is because, unlike when transferred to centralized servers, raw data stays at its source, reducing the likelihood of breaches. This is an extremely important precaution to take since sensitive financial data is susceptible to online attacks.
An improved and up-to-date fraud detection system is the result of FL's use of effective data usage, which involves training models on diverse data collected in real time from different sources.

2.2. Suggested FL-based Model:

The use of FL in the detection of fraud not only supports increased cooperation among institutions to address fraud in a dynamic environment but also utilizes the strength of collective information without affecting individual data privacy.
Here is an example of a standard centralized ML model update using Stochastic Gradient Descent (SGD):

$$P_i + 1 = P_i - R\hat{a}\hat{} \ddagger L(P_i)$$

Here P stands for the model parameters, and ∇L(Pi) is the change in the loss function L concerning the model's parameters at iteration i.

The latest update is not administered centrally in FL. Rather, the global model is updated by aggregating the updates computed by each client. Updating the normal SGD to calculate multiple updates in every client and then average them on the server is the fundamental principle underlying the Federated averaging technique.

Based on its data, each client C figures out its update for the specified global model G:

$$P_i^C + 1 = P_i - R\hat{a}^{\wedge} \ddagger L_C(P_i)$$

Here L_C stands for the loss in client C. The server puts together all of the local changes that have been made by each client to make the global model update. Federated average usually does this by adding up all the local changes and giving them weight.

$$P_i + 1 = \hat{a}\frac{S_C}{S} P_i^C + 1$$

Assuming that S is the sum of all customers' data points and S_C is the number of points for client C.

Several rounds of this method are repeated until convergence is achieved. One important benefit is that data privacy is preserved since only model changes are shared. Federated averaging, which essentially provides a middle ground between local and centrally managed learning, enables models to make use of a variety of local information sources without affecting user privacy.

2.3. Integration of Explainable AI:

Interpretability is essential in the banking sector. Evaluating the choices & predictions generated by models is crucial because they have significant real-world consequences. To fill this gap, among the need for openness and the opacity of particular models, XAI has developed. When included in the FL model, these XAI approaches provide stakeholders insights that increase trust in the model's predictions. Troubleshooting and refining models, addressing any prejudices or malware, is made easier with an awareness of feature significance.

Shapley Additive Explanations (SHAP) is a well-known artificial intelligence method. Emerging from game theory, SHAP values give a single way to measure the significance of a feature by calculating the difference between what the model says will happen and what the average predicts will happen for each feature. An instance's SHAP value is that feature's average addition to all the possible combos of features for that instance.

Feature's SHAP value is determined by:

$$= \hat{a}_{B\hat{a}\check{S}\dagger N[k]} \frac{\hat{a}^{\wedge}\ldots_k(f)}{\frac{|B|!\,(|N|-|B|-1)!}{|N|!}} [f(B\hat{a}^{\wedge a}(k)] - f(B)]$$

All features are included in the set N. To remove feature K from N, B is taken as a subset of N. For each set B of input characteristics, f(B) represents the model's prediction.

3 SPECIFICATIONS OF THE IMPLEMENTATION

The steps for implementing the suggested method are illustrated in Figure 2. Examinations are done first, data is processed, FL is built, and XAI interaction methods are used.

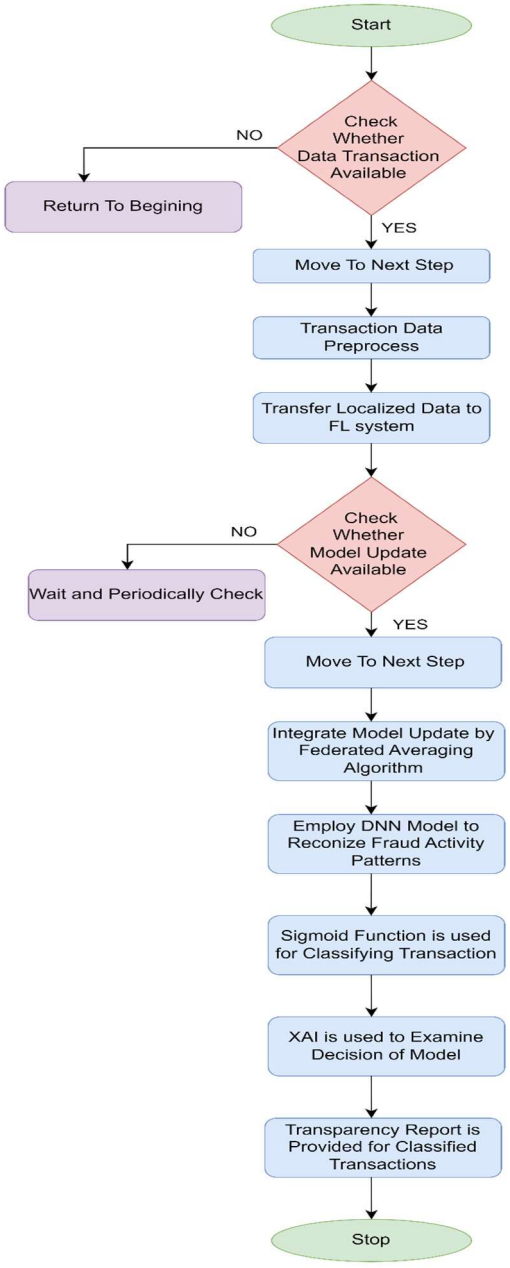


Figure 2 Flowchart of the Suggested System

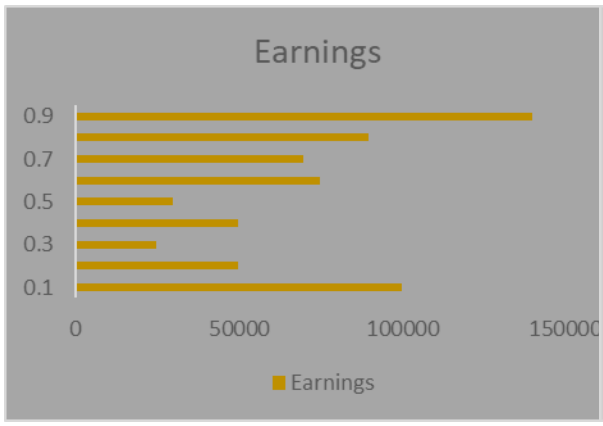
3.1. Dataset

Based on a real-world sample from the present day, the study includes realistic data for finding fraudulent activities. With over 33 different features, the sample has 29,043 items. Different types of data are in this large collection. There are 18 columns of numeric data, 11 columns of floating-point data, and 6 columns of category or string data. The following are the attributes found in the dataset:

i. **Earnings:** This is the user's cash, and it is a float-type constant number.
ii. **client_age:** The customer's years of age.
iii. **planned_balcony_amount**: The amount deposited in the balcony.
iv. **credit_risk_rate:** An indicator of the possibility of a threat of extending credit to the person or business.
v. **phone_house_authentic, phone_cell_authentic:** Home and mobile phone number validity indicators based on points or binaries.
vi. **email_costless:** Some email providers, like Gmail and Yahoo, are free. This is shown by a binary number.
vii. **month:** Most often, this indicates the month in which the financial transaction or information was recorded.
viii. **job_status, transaction_mode, residential_status, device_os, transaction_mode:** The client's job condition, residential situation, data source, and operating system are all indicated by these category qualities, together with the mode of transaction.
ix. **request_lasting_days:** A variable that always shows the days that have passed since a certain request, like a request for an account transaction statement.
x. **present_residence_count_month, former_residence_count_month:** Measures of a period, measured in months, spent at the present and former addresses.
xi. **uniformity_name_email:** A constantly changing variable that stores the score for how close an individual's name and email address are.
xii. **fraud bool:** This variable serves as a target for predictive models and is characterized by being binary. It may be used to determine whether the record was fake (1) or not (0). Figures (3A – 3I) show the unique features of the dataset's distribution.

Journal of Theoretical and Applied Information Technology
30th April 2025. Vol.103. No.8

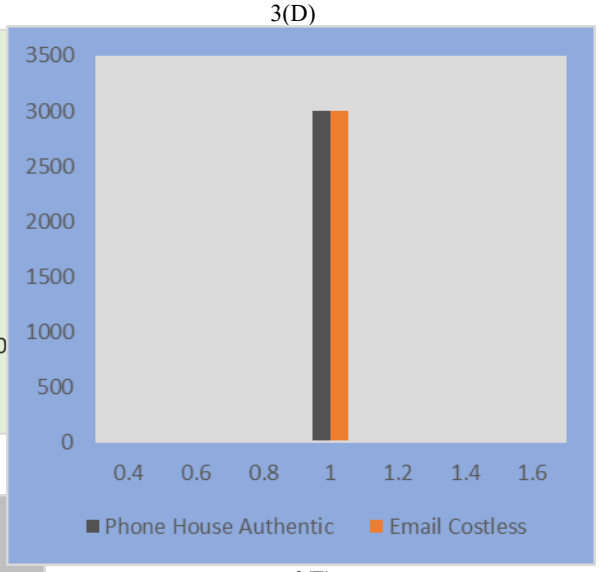
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



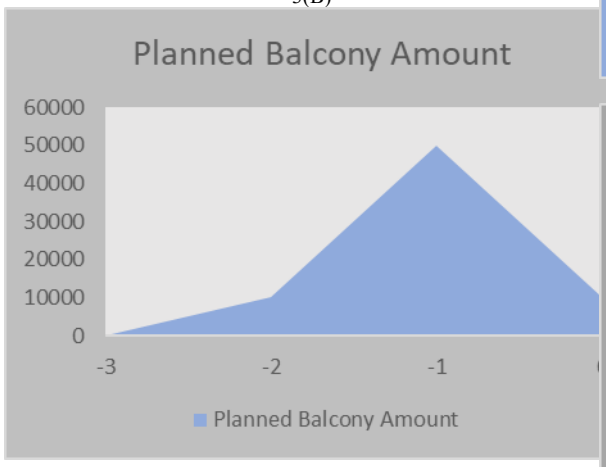
3(A)

3(D)

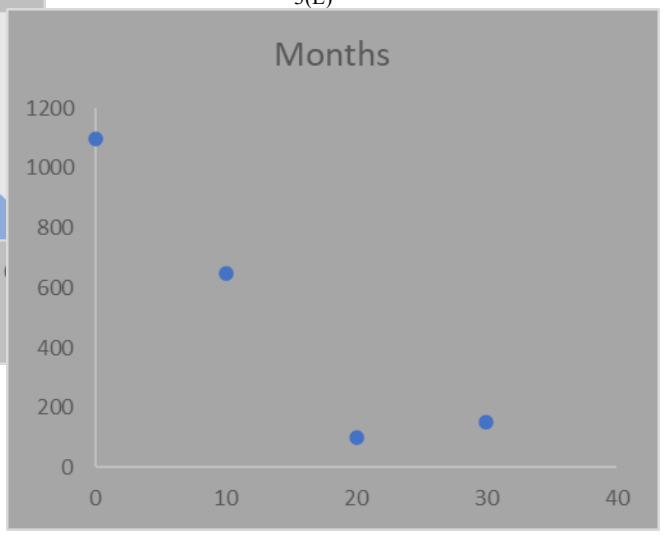
3(B)



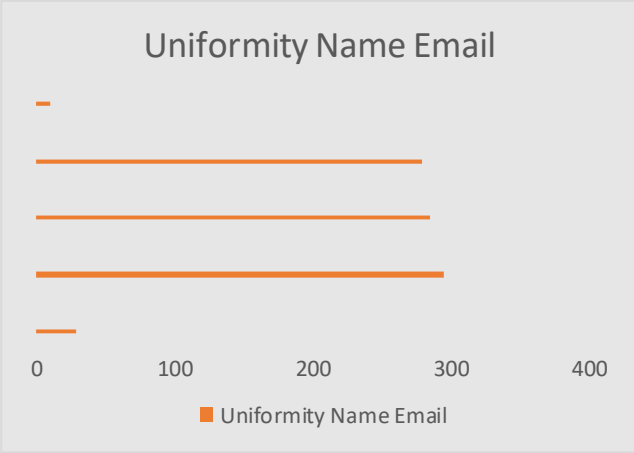
3(E)



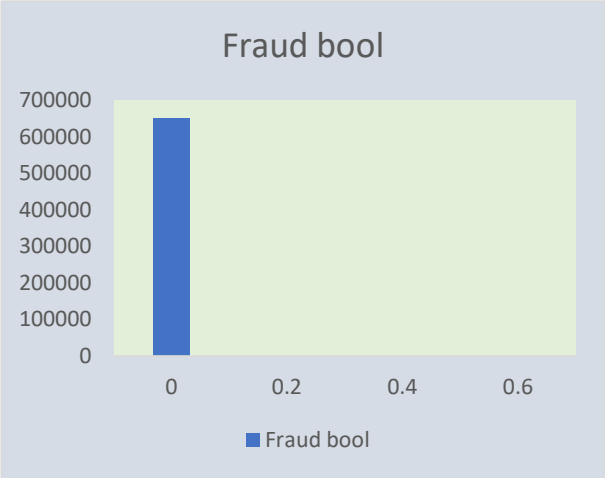
3(C)



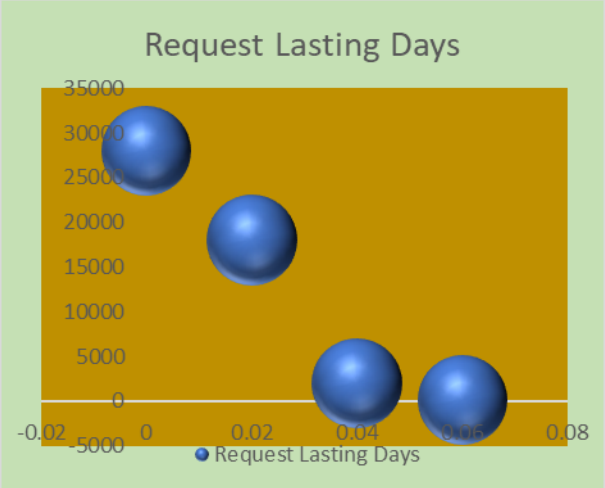
3(F)



3(G)



3(H)



3(I)

Figures (3A-3I) Dataset Characteristics

3.2. Equalizing Data

The Synthetic Minority Oversampling Technology (SMOTE) was used to balance the data because it was very unbalanced, as shown in Figure 4. A common way to fix class mismatch is to use the SMOTE method, which creates fake samples in the feature space. It is used to make sure that there are similar numbers of samples for both the majority and minority groups. This evens out any imbalances in the training data.

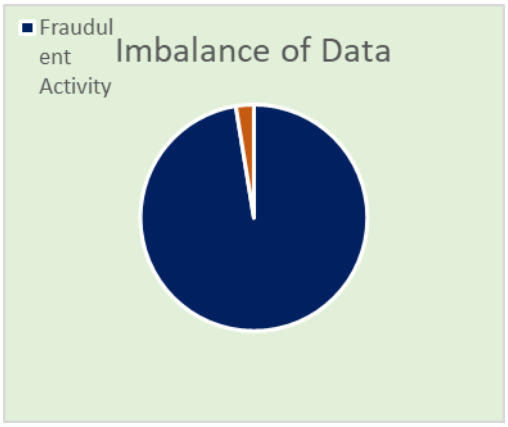


Figure 4 Unbalanced Dispersion of Suggested Dataset

3.3. Pre-Processing of Dataset:

Two separate approaches were used in the study to handle the dataset's missing data. To fill in the details for quantitative characteristics, researchers utilized the average value of the corresponding column. The most common value is used for inputting with categorical characteristics, on the other hand.

For columns with floating-point values, outlier elimination was done during the data preparation step. Any data values that exceeded 1.5 times the IQR of the initial (Q1)/third quartile (Q3) were classified as outliers and then removed from the dataset using that IQR approach to increase the data's resilience.

Following this, as shown in Figure 5, a correlation matrix was generated to investigate the connections and possible correlations among the numerical characteristics. Then, a heatmap was created from this matrix, which showed the linear correlations between variables in a colour-coded form. The significant correlations or possible multicollinearity situations can be easily seen using the colours in this heatmap, which vary from shades signifying

ideal negative correlation to shades indicating ideal positive correlation.

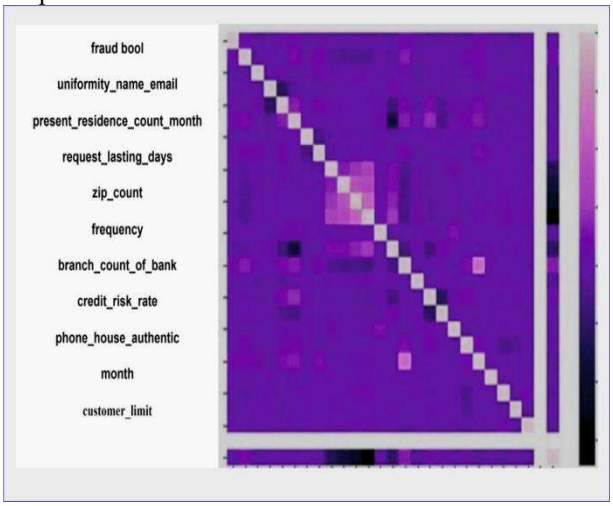


Figure 5 Correlation Matrix for Fraudulent Dataset

3.4. Figuring out Attributes

Feature engineering was used to make the data easier for the model to understand so it could learn from it:

- **Data Binning:** To construct the binned income column, the continuous income columns were split into intervals. One method for dividing continuous data into discrete sets is binning. Instead of using individual income values, the model may be able to identify patterns or trends more readily across income ranges by splitting the income into 10 discrete bins and indicating them with numbers.
- **One-Hot Encoding:** Data columns including housing status, source, payment type, employment status, and device os were encoded using one-hot. To help machine learning algorithms improve their prediction abilities, one-hot encoding may be used to transform categorical data into a more suitable format. A new binary column is generated for every distinct value in the primary category column. For models like neural networks, which perform more effectively with numerical input, this translation is crucial.

3.5. DL Model:

A 3-layer dense neural network that makes up the deep learning (DL) model was used for this investigation. The Rectified Linear Unit (ReLU) activation function is used by the first layer, which has sixty-four neurons and receives the input shape from the characteristic dimension of the validation datasets. There are thirty-two neurons in the next layer, which is also triggered by ReLU. The terminal layer is made up of a single neuron that is triggered by a sigmoid function and is intended for binary classification. The function of binary cross-entropy (BCE) assesses prediction losses, and the Adam algorithm for optimization is used to modify the model's weights.

3.6. Evaluation and Training:

Data splitting was done in this study after the data had undergone pre-processing. This study used the commonly used technique of dividing the data in 81% for training & 19% for testing to ensure impartial and random division. The modified training dataset was divided into three sections, designated Y train1, Y train2, and Y train3, in order to integrate these datasets into FL.

3.7. Indicators of Performance:

This is essential to choose a suitable measure that represents the ML model's precision and dependability to evaluate the model's efficiency. One popular and easy-to-understand statistic is the confusion matrix, which is beneficial for binary classification tasks like differentiating between the two results. Four separate prediction results are provided by this confusion matrix:

i. If a financial transaction is correctly identified as fraud, it is called a "Correct Positive."
ii. If a transaction is wrongly classified as fraud, it is called an "Incorrect Positive."
iii. If a transaction is erroneously identified as legitimate, it is called an "Incorrect Negative."
iv. If a transaction is precisely identified as legitimate, it is called a "Correct Negative."

Employing the validation dataset, a multifaceted metric approach was used to determine model performance:

- Precision measures the degree to which suitable statements end up coming true.
- Recall Indicates the number of correctly sorted hits.
- F1-Score strikes an appropriate balance between accuracy and memory, which is especially important when there are class differences.

- The number of correct guesses made is shown by accuracy.

3.8. Explicable AI Integration

To demonstrate the relevance of features, researchers use the SHAP approach, which provides an assessment of the way each characteristic corresponds to a predefined baseline.

3.8.1. Configuration for Simulations:

The simulation setup offers a comprehensive view of the way individual customer-side models might aid in the learning of a global model by using current DL frameworks and including FL. Without exchanging raw data, the FL architecture seeks to decentralize model training across many clients. Researchers clarify model choices using SHAP after training to improve their interpretability. A set of software tools was used to create the simulation to support FL and further assessments. For server-client communication, a lightweight web application framework was created using FastAPI. DL procedures, including the creation, training, and assessment of the neural network model, were made easier using the PyTorch library. AJAX cross-origin capability was ensured by using the FastAPI-CORS extension to handle Cross-Origin Resource Sharing (CORS). Port 5000 was used to reach the server, which was set up to run locally.

4. ANALYSIS AND OUTCOMES:

The success measures of the shared approach were well evaluated in this part. The researcher also showed off the web-based system that was made for this study. The strength of XAI was also used to figure out the way the model made its choices and find key traits that are essential for finding scams. The goal of this research is to confirm that this method works as well as offer new ideas that might change the method by which fraudulent activity is found.

4.1. Web-based Architecture:

The FL configuration is shown in Figures 6A and 6B. Each client's progress is shown on the HTML page as it uses the global model that was retrieved at the central server to train its local dataset before updating the server's database with its model weights, not actual data. The clients have three status updates: idle, training, and updating. While the client waits for the server to provide the updated global model, its status becomes idle. The number "33" on the Gradient Updates tab indicates how many times the federated learning paradigm has been trained on client-side systems as the server and client communicate in real-time. The website is updated based on recall, F1-score, accuracy, and precision, for every iteration.

Customer	Status
Customer 1	Updating
Customer 2	Training
Customer 3	Updating

Figure 6A Status of Customer

Update	Metrics
1	["precision" (0.8557932376861573),"loss" (0.336732387542724 7)]
2	["precision" (0.8550292253494264),"loss" (0.335765540599824)]
3	["precision" (0.85592120885850),"loss" (0.33408019180297 86)]
4	["precision" (0.8858200907707215),"loss" (0.276101887226104 75)]
5	["precision" (0.8855137228965760),"loss" (0.27654233574867 26)]

Figure 6B Training Metrics

9	["precision" (0.900946855545045),"loss" (0.244583696126378 8)]
10	["precision" (0.900576770305633 3),"loss" (0.24451807415218353)]
11	["precision" (0.9095793842147478),"loss" (0.22610188722610475)]
12	["precision" (0.9023875124775611),"loss" (0.22786278525587447)]
13	["precision" (0.9124733581217387),"loss" (0.211357655405998241)]
14	["precision" (0.9159212088585011),"loss" (0.21080191802978617)]
15	["precision" (0.9185820090770721 5),"loss" (0.21610188722610475)]
16	["precision" (0.9151372289657601 8),"loss" (0.21276542335748672)]
17	["precision" (0.9125029225349426),"loss" (0.21335765540599824)]
18	["precision" (0.9155921208855011 4),"loss" (0.21334080191802978)]
19	["precision" (0.9285820090777215),"loss" (0.20276101887226104)]
20	["precision" (0.9288551372289676 0),"loss" (0.20765423357486726)]
21	["precision" (0.9288582009070721 5),"loss" (0.19761018872261047)]
22	["precision" (0.9288551372289656 0),"loss" (0.19654233574867261)]
23	["precision" (0.9288582009077072 1),"loss" (0.19761018872261047)]
24	["precision" (0.9288551372289657 6),"loss" (0.1965423357486726)]
25	["precision" (0.9288582009077072 1),"loss" (0.19610188722610475)]
26	["precision" (0.9288551372289657 6),"loss" (0.19654233574867261)]

Figure 6C Training Dashboard of FL

A graph showing the accuracy of updates in real-time is shown in Figure 6D. As the precision of the computations grows and the Gradient Update rises, this graph likewise updates itself. From the results, it is observed that the FL model converges in accuracy as a result of the time required to collect information from all clients. There is an apparent 94% accuracy score, and the accuracy improves dramatically.

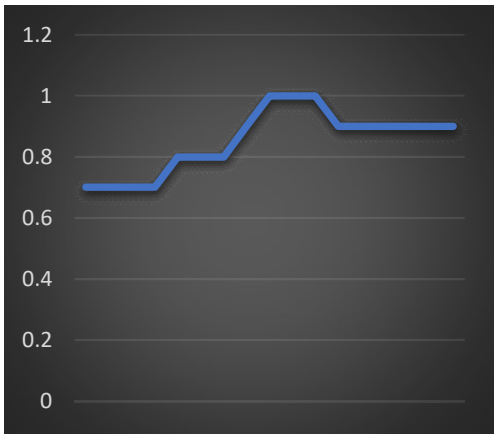


Figure 6D Performance Metrics of FL

4.2. Performance Metrics for the Model:

Figure 7A shows the accuracy and F1 Score of the FL model. This suggests that the model convergence occurs because it takes time to group information from all clients. There is a clear precision rating of 94%, and the degree of precision goes up exponentially.

As the number of epochs grows, the FL model's accuracy rises linearly until around 40 epochs, as shown in Figure 7B. The accuracy begins to oscillate and does not rise as much after around forty epochs. This might mean that the model is gaining information from the training data to the extent that it can. The FL model's recall goes up as the total number of epochs grows as shown in Figure 7C. The model can learn more from the training dataset and become better at finding successful instances over time. There are more epochs in Figure 7A, which shows that the F1-score for the FL model rises high. This shows that the model can become more effective in predicting future outcomes by absorbing additional information from the training set. On the other hand, as the total number of epochs grows, the graph reveals that the improvement rate falls. The reason is that at some point, the model will have learned all it can from the training data, and there will be little opportunity for progress after that.

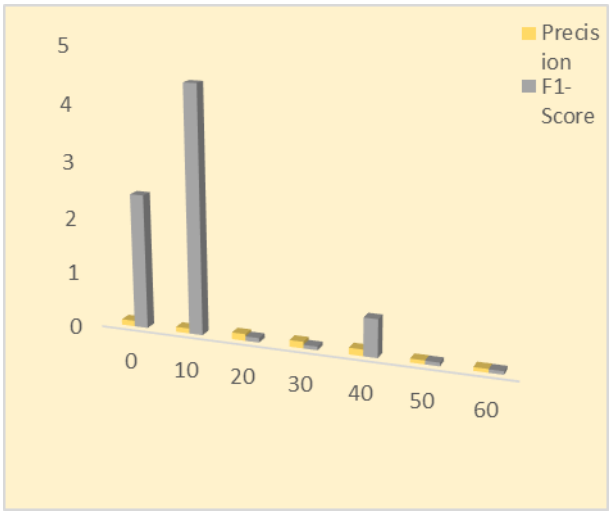


Figure 7A Accuracy and F1 Score of FL Model

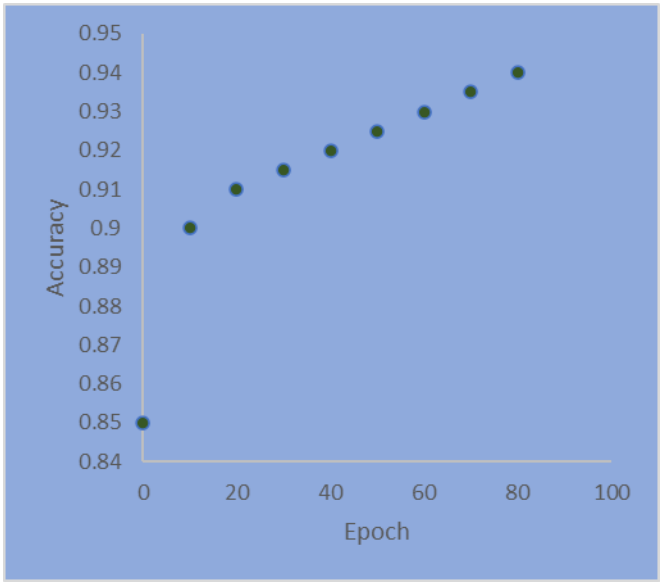


Figure 7B FL Model's Accuracy

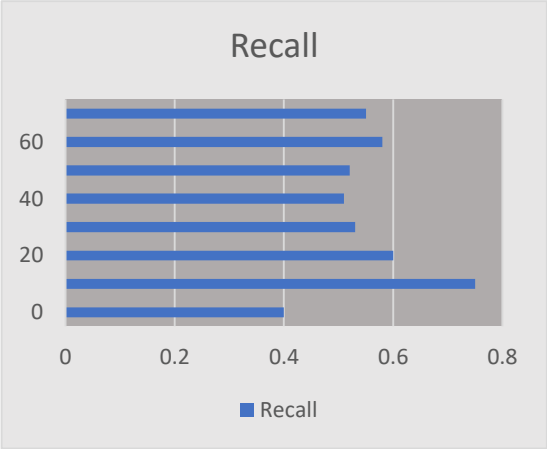


Figure 7C FL Model's Recall

The objective of the study was to find out where information privacy benefits come from using an FL-based method for advanced malicious detection technologies. When FL was contrasted to other centralized machine learning models, it was clear that the real data was not shared. Instead, model weights were transmitted to a central computer to be combined and a global model was developed. The idea behind the shared learning model came from this iterative method.

4.3. Observations from Explainable Artificial Intelligence:

The model's explication has been assessed using the SHAP plots. Both the positive and negative impacts of attribute values on output prediction, as compared to the baseline value, are shown by the colour labelling (Purple and Green) in the SHAP plots. All of the model's output values averaged out across the dataset make up the baseline value, which is utilized as the point of reference. The researcher observed that in the SHAP plots, the purple area represents an attribute value that improved the prediction value. Whereas the green colour indicates that the estimated amount was reduced due to a certain attribute value. In the purple side of each of the customer plots, feature 2.087 shows that it positively affects the FL model's prediction. The degree to which a characteristic influences the SHAP values may be determined by looking at their magnitude, or by observing how far they deviate from the baseline. Larger magnitudes indicate that the attribute has a significant influence on the estimation, whether they are positive or negative. For instance, in the subsequent SHAP plot (Figure 8B), characteristics 2.16, 0.419, as well as 0.518 carried a greater influence on the model's prediction than features 1.16 and -0.9.

Figure 8A SHAP Plot for Customer 1

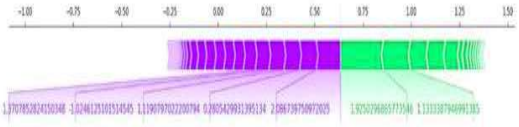


Figure 8B SHAP Plot for Customer 2

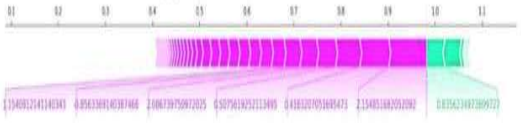


Figure 8C SHAP Plot for Customer 3

5. Conclusion

This study offered a novel approach to identifying bank account fraud by combining Explainable AI (XAI) with Federated Learning (FL) in a synergistic way. The FL-based technology allowed for collaborative model training across financial institutions without jeopardizing confidential customer data, which is a departure from conventional centralized machine learning techniques. This tackles a significant issue in the financial industry, where privacy issues often obstruct efficient data exchange and teamwork. The system's originality was further increased with the addition of XAI, more especially Shapley Additive Explanations (SHAP), which made the model's decision-making process transparent and interpretable. This promotes operational effectiveness and regulatory compliance in addition to increasing stakeholder confidence. We showed how crucial financial security judgments may be both accurate and explicable by illuminating the "black box" of AI. This study has important practical implications. The viability of implementing such a system in actual financial situations was shown by the web-based implementation. Our strategy gives institutions the ability to work together to fight fraud while protecting data privacy by facilitating safe, cooperative learning. Financial institutions'

approaches to fraud detection might be revolutionized by this, shifting toward AI-driven solutions that are more transparent and safer.

Furthermore, the flexibility and efficacy of our approach are shown by the adaptation of a deep neural network (DNN) architecture for federated datasets. The strong performance metrics attained during testing highlight how this strategy has the potential to greatly lower financial losses and improve security. Essentially, this study offers a novel approach for tackling the twin problems of model interpretability and data privacy in fraud detection. The foundation for future developments in safe and open AI-driven financial systems is laid by this study, which shows the effectiveness of FL and XAI in a real-world application. This study has an effect by offering a scalable and privacy-preserving technique that boosts confidence in AI-driven fraud detection systems and facilitates improved financial institution cooperation.

Acknowledgement:

The authors extend their appreciation to Northern Border University, Saudi Arabia, for supporting this work through project number (NBU-CRP-2025-2225)

REFERENCES:

[1]. Javaid, Haider Ali. "How Artificial Intelligence is Revolutionizing Fraud Detection in Financial Services." Innovative Engineering Sciences Journal 4.1 (2024).

[2]. Islam, Md Zahidul, Sanjib Kumar Shil, and Md Rashed Buiya. "AI-Driven Fraud Detection in the US Financial Sector: Enhancing Security and Trust." International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence 14.1 (2023): 775-797.

[3]. Chirra, Bharadwaja Reddy. "AI-Driven Fraud Detection: Safeguarding Financial Data in Real-Time." Revista de Inteligencia Artificial en Medicina 11.1 (2020): 328-347.

[4]. Zanke, Pankaj. "AI-driven fraud detection systems: a comparative study across banking, insurance, and healthcare." Advances in Deep Learning Techniques 3.2 (2023): 1-22.

[5]. Potla, Ravi Teja. "AI in fraud detection: leveraging real-time machine learning for financial security." Journal of Artificial Intelligence Research and Applications 3.2 (2023): 534-549.

[6]. Inampudi, Rama Krishna, Thirunavukkarasu Pichaimani, and Yeswanth Surampudi. "AI-Enhanced Fraud Detection in Real-Time Payment Systems: Leveraging Machine Learning and Anomaly Detection to Secure Digital Transactions." Australian Journal of Machine Learning Research & Applications 2.1 (2022): 483-523.

[7]. Mewada, Shivlal, et al. "Smart diagnostic expert system for defect in forging process by using machine learning process." Journal of Nanomaterials 2022.1 (2022): 2567194..

[8]. Johora, Fatema Tuz, et al. "AI Advances: Enhancing Banking Security with Fraud Detection." 2024 First International Conference on Technological Innovations and Advance Computing (TIACOMP). IEEE, 2024.

[9]. Mewada, Shivlal, Anil Saroliya, N. Chandramouli, T. Rajasanthosh Kumar, M. Lakshmi, S. Suma Christal Mary, and Mani Jayakumar. "Smart Diagnostic Expert System for Defect in Forging Process by Using Machine Learning Process." Journal of Nanomaterials 2022 (2022).

[10]. Odeyemi, Olubusola, et al. "Integrating AI with blockchain for enhanced financial services security." Finance & Accounting Research Journal 6.3 (2024): 271-287.

[11]. Gautam, Ayush. "Evaluating the impact of artificial intelligence on risk management and fraud detection in the banking sector." AI, IoT and the Fourth Industrial Revolution Review 13.11 (2023): 9-18.

[12]. Bello, Oluwabusayo Adijat, et al. "AI-Driven Approaches for Real-Time Fraud Detection in US Financial Transactions: Challenges and Opportunities." European Journal of Computer Science and Information Technology 11.6 (2023): 84-102.

[13]. Odeyemi, Olubusola, et al. "Reviewing the role of AI in fraud detection and prevention in financial services." International Journal of Science and Research Archive 11.1 (2024): 2101-2110.

[14]. Bello, Oluwabusayo Adijat, and Komolafe Olufemi. "Artificial intelligence in fraud prevention: Exploring techniques and applications, challenges and opportunities." Computer Science & IT Research Journal 5.6 (2024): 1505-1520.

[15]. Mujtaba, Numan, and Alan Yuille. "AI-Powered Financial Services: Enhancing Fraud Detection and Risk Assessment with Predictive Analytics."

[16]. Shoetan, Philip Olaseni, and Babajide Tolulope Familoni. "Transforming fintech fraud detection with advanced artificial

intelligence algorithms." Finance & Accounting Research Journal 6.4 (2024): 602-625.
[17]. Shabir, Ghulam, and Najif Khalid. "AI-Powered Fraud Detection and Risk Assessment: The Future of Financial Services."
[18]. Narsina, Deekshith, et al. "AI-Driven Database Systems in FinTech: Enhancing Fraud Detection and Transaction Efficiency." Asian Accounting and Auditing Advancement 10.1 (2019): 81-92.
[19]. Ismaeil, Mohamed Kamal Aldin. "Harnessing AI for Next-Generation Financial Fraud Detection: A Data-Driven Revolution." Journal of Ecohumanism 3.7 (2024): 811-821.
[20]. Kotagiri, Anudeep. "Mastering Fraudulent Schemes: A Unified Framework for AI-Driven US Banking Fraud Detection and Prevention." International Transactions in Artificial Intelligence 7.7 (2023): 1-19.
[21]. Dayalan, Padmalosani, and Balaji Sundaramurthy. "Exploring the Implementation and Challenges of AI-Based Fraud Detection Systems in Financial Institutions: A Review." Creating AI Synergy Through Business Technology Transformation (2025): 25-38.
[22]. Adhikari, Prabin, Prashamsa Hamal, and Francis Baidoo Jnr. "Artificial Intelligence in fraud detection: Revolutionizing financial security." (2024).
[23]. Thilagavathi, M., et al. "AI-driven fraud detection in financial transactions with graph neural networks and anomaly detection." 2024 International Conference on Science Technology Engineering and Management (ICSTEM). IEEE, 2024.
[24]. Gayam, Swaroop Reddy. "AI-Driven Fraud Detection in E-Commerce: Advanced Techniques for Anomaly Detection, Transaction Monitoring, and Risk Mitigation." Distributed Learning and Broad Applications in Scientific Research 6 (2020): 124-151.
[25]. Inampudi, Rama Krishna, Yeswanth Surampudi, and Dharmeesh Kondaveeti. "AI-Driven Real-Time Risk Assessment for Financial Transactions: Leveraging Deep Learning Models to Minimize Fraud and Improve Payment Compliance." Journal of Artificial Intelligence Research and Applications 3.1 (2023): 716-758.