Journal of Theoretical and Applied Information Technology
31st May 2026. Vol.104. No.10

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

DESIGN OF AN INTEGRATED EXPLAINABLE PREEMPTIVE AND ADAPTIVE FINANCIAL FRAUD DETECTION MODEL FOR REAL-TIME TRANSACTION SYSTEMS

MUDIMELA MADHUSUDHAN[1], PRAMODA PATRO[2]

[1,2]Centre of AI and Deep Learning, School of Computer Science and Artificial Intelligence,

SR University, Warangal, Telangana, India, 506371.

E-mail: [1]madhu9963@gmail.com, [2]pramoda.mtech09@gmail.com

ABSTRACT

The rise in the number of digital payment systems, online banking and financial transaction systems has greatly augmented fraudulent activities. The traditional fraud detection models are mainly based on fixed machine learning models which are not able to keep up with changing fraud trends and they are not easily interpretable which restricts their applicability in real time financial settings. In this paper, a combination of explainable and adaptive fraud detection system will be proposed to detect fraudulent transactions beforehand and to ensure transparency in the decisions. The architecture suggested includes Drift-Aware Contrastive Embedding (DACE) which is proposed to represent features adaptively, Intrinsic Explainable Neural Tree (IxENTree) which is suggested to classify using explanations, Meta-Adaptation by Few-Shot Fraud Transfer (MAFT) which is proposed to detect new fraud schemes quickly, Adversarial and Counterfactual Explainability Stress Testing (ACEST) that should ensure the robustness of an explanation, and Continuous Feedback and Audit Loop (CFAL) that An evaluation of the system is done by financial transaction datasets and compared with conventional machine learning and deep learning models. The experimental findings indicate that the given framework is characterized by a better accuracy of detection, higher recall, and lower false-positive rates, and the Area Under the ROC Curve (AUC) reaches about 0.97. Adaptive learning and explainable artificial intelligence allow the suggested system to achieve trustworthy and understandable fraud detection in live financial transactions circumstances. Existing fraud detection systems struggle to adapt to concept drift and often lack interpretability required for regulatory compliance. Experimental results demonstrate that the proposed framework improves fraud detection robustness, achieving superior recall and reduced false-positive rates while maintaining transparent decision-making suitable for real-time financial systems.

Keywords: *Financial Fraud Detection, Explainable Artificial Intelligence, Concept Drift Detection, Adaptive Learning, Counterfactual Explanations.*

1. INTRODUCTION

The swift growth of online financial systems has greatly changed how financial operations are being done globally. The internet banking, mobile payment, online stores, and online wallets have also facilitated a smooth and immediate flow of money worldwide through the internet. Nevertheless, such an accelerated digitization has also made the financial systems more susceptible to financial frauds like credit card fraud, identity theft, account takeover attack, transaction laundering, and synthetic identity fraud. Through the recent reports on financial security, it is observed that financial institutions are losing billions of dollars every year as a result of more advanced fraud schemes and this has made fraud detection to be among the most critical issues in contemporary financial systems.

Classical fraud detection methods mainly depend on rule-based fraud and classical models of machine learning built on historic transaction records. These systems normally apply known rules and fixed classification algorithms to determine suspicious transactions. Whereas these methods have been successful in identifying familiar patterns of fraud, they are not good at keeping up with the new fraud techniques which are constantly emerging in dynamic settings of transactions. The fraudsters tend to take advantage of the weakness in the system by altering the transaction behaviors, applying the distributed attack techniques, and

creating minute peculiarities that cannot be noticed by the fixed models.

The machine learning and deep learning approaches have greatly enhanced fraud detection through training complicated patterns and relationships on the high volume of transaction data. Some of the algorithms used in detecting fraudulent activities, including Support Vector Machines (SVM), Random Forests, Gradient Boosting, Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks, have exhibited encouraging results. Nevertheless, a number of limitations are usually experienced with these models in their application to real financial monitoring systems. To start with, a large majority of deep learning models are black-box solutions, thus not offering a lot of interpretabilities of their decision-making, which makes it challenging to comply with regulations and audit risks. Second, financial transaction data tends to be extremely imbalanced in terms of classes with fraudulent transactions being by far a very small percentage of all transactions. Third, the appearance of new user behavior, economic conditions, and fraud tactics

implies the constant change in transaction patterns and causes the phenomenon of concept drift.

The motivation for this work stems from the increasing inability of static fraud detection models to adapt to rapidly evolving fraudulent behavior. This study addresses this gap through a modular framework combining adaptive representation learning, explainable neural decision structures, meta-learning adaptation, and feedback-driven retraining.

Concept drift happens when the statistical characteristics of the data of transactions are modified with time, and all the previously trained models are no longer relevant. The conventional model of detecting fraud entails periodic retraining of an existing system with updated datasets, which causes delays in identifying the occurrence of new trends. Moreover, the current models usually produce a high rate of false-positive results, which result in redundant transaction blocks and poor customer satisfaction. Thus, the current systems of detecting financial fraud should be able to adapt to fraud patterns but with high accuracy and low false-positive rates.

Model explainability is another significant need in financial fraud detection. Financial institutions also have very strict regulatory frameworks, which mandate that they have a transparent decision making process. Whenever a transaction is labeled fraud, the analysts need to know why the decision was made, so as to be fair, accountable and in accordance with the financial regulations. Nevertheless, a lot of sophisticated deep learning models do not inherently have interpretable aspects and thus analysts are not easily convinced by automated fraud detection decisions.

This paper suggests an Integrated Explainable Preemptive and Adaptive Fraud Detection Framework to Real-Time Tx Systems to deal with such challenges. The suggested framework integrates adaptive representation learning, interpretable classification mechanisms, few-shot learning in emerging fraud detection and counterfactual explanation validation to provide robust and clear fraud detection. The architecture combines several modules such as Drift-Aware Contrastive Embedding (DACE) of adaptive feature learning, Intrinsic Explainable Neural Tree (IxENTree) of interpretable decision-making, Meta-Adaptation by Few-Shot Fraud Transfer (MAFT) of rapid adaptation of a model to new fraud patterns, Adversarial and Counterfactual Explainability Stress Testing (ACEST) of explanation robustness evaluation and a Continuous Feedback and Audit Loop (CFAL) of continuous model enhancement.

The main aim of the study is to create a fraud detection mechanism that will predict fraudulent transaction in real-time and be comprehensible and flexible. The solution is expected to improve the quality of detection of fraud, minimize false positives and make the explanations persist even when the transaction conditions change. The framework offers a scalable solution to current financial transaction monitoring systems by incorporating the adaptive learning and explainable artificial intelligence methods.

The main contributions of this work are summarized as follows:

I. Development of a drift-aware embedding model for adaptive transaction feature representation in evolving financial environments.

II. Design of an interpretable neural decision architecture for transparent fraud classification.

III. Introduction of a few-shot learning mechanism to detect emerging fraud patterns with limited labeled data.

IV. Implementation of counterfactual explanation validation to ensure the

reliability and robustness of model explanations.

V. Integration of a continuous feedback and audit mechanism for real-time model improvement and regulatory compliance.

The rest of this paper is structured in the following way. Section II provides the associated literature on the field of financial fraud detection and explainable AI. Section III includes the proposed methodology and system architecture. The results and performance analysis of the proposed model is discussed in Section IV. Lastly, the paper is concluded in Section V and Section VI includes the limitations and future work.

1.1 Problem Context and Theoretical Motivation

a. Dynamic fraud evolution ,
b. Regulatory frameworks requiring explainability,
c. Operational losses from delayed fraud adaptation
d. Inadequacy of static supervised models

Research Hypothesis (H1):

An integrated fraud detection framework combining adaptive drift-aware embedding, explainable neural classification, few-shot adaptation, and continuous feedback learning significantly improves fraud detection accuracy, interpretability, and robustness compared to conventional machine learning and deep learning models.

2. LITERATURE SURVEY

The detection of financial fraud is one of the areas of the research that have received significant attention because of the high rate of the development of digital financial services and the sophistication of the fraudulent schemes. To study this issue, researchers have examined a broad set of methods such as statistical methods, machine learning formulations, deep learning architectures, explainable artificial intelligence methods and so on.

The detection systems of fraud used at the beginning were mainly based on rule and statistical detection. Bolton and Hand suggested the statistical profiling methods of finding the anomalous patterns of transactions by peer group analysis and break-point methods. Their labour proved that statistical surveillance methods can be successfully used to detect outlier financial patterns in massive transaction data sets [1]. Equally, Bhattacharyya et al. made a comprehensive comparative analysis of the supervised and unsupervised fraud detection methods and indicated the drawbacks of the conventional models when handling massively disproportional data [2].

Machine learning algorithms have greatly enhanced the effectiveness of fraud detection in that they learn intricate relationships in the data of financial transactions. The algorithms that have been extensively used include random forest, support vectors machine (SVM) and gradient boosting because they can provide support in nonlinear association between the features. To mitigate financial losses incurred with fraudulent transactions, Bahnsen et al. proposed the use of cost-sensitive decision tree learning, and the classification results showed better performance relative to the standard models [3]. Whitrow et al. also advanced the fraud detection by including the transaction aggregation strategies which capture the customer behavior pattern over the time [4].

Deep learning methods have become the potent end-user fraud detection tools with access to large-scale financial transactions data. Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have demonstrated good results in recognizing intricate patterns in streams of transactions. Jurgovsky et al. came up with an LSTM-based fraud detection system that could model sequential behaviors of transactions thus leading to a significant increase in the detection rate of credit card frauds [5]. On the same note, Fiore et al. proposed deep learning-based credit card fraud detector which uses autoencoders to learn compress representations of transaction data in the form of features [6].

Graph-based techniques that treat financial transactions as networks are another focus in the research in fraud detection. Fraud rings are also prone to use interconnected accounts and chains of transactions and a graph approach is thus very handy in identifying coordinated fraud activities. Another interesting study was suggested by Weber et al. who provided a graph neural network model of detecting fraud in financial transaction networks and illustrated substantial progress in detecting sophisticated fraud schemes [7]. Liu et al. also discussed the use of graph convolutional networks to detect financial anomalies and suggested the benefits of relational data modeling in fraud detection processes [8].

Though the performance of deep learning and graph-based models is good on detecting performance, they tend to be black-box models

with little to no interpretability. This is not very transparent and as such is problematic in a financial situation where compliance with regulations demands an explicable decision making. Explainable Artificial Intelligence (XAI) methods have thus come to gain considerable popularity in fraud detection studies. Lundberg and Lee proposed the SHAP framework of explaining predictions of machine learning models with Shapley values based on cooperative game theory [9]. One algorithm suggested by Ribeiro et al. is the LIME algorithm to give a local explanation to complex model predictions by approximating them through interpretable models [10].

Concept drift, which is a situation where the fraud pattern changes with time, is another significant difficulty in fraud detection. Adaptive learning models are suggested to deal with this problem. The concept drift detection techniques proposed by Gama et al. to the streaming data setting allow models to adjust to new data distributions [11]. On the same note, Baena-Garcia et al. suggested the Early Drift Detection Method (EDDM) to detect and provide updates on models based on detected changes in the data streams [12].

In more recent studies, scholars have investigated hybrid frameworks with explainability, adaptive learning and deep neural networks to enhance fraud detection. Carcillo et al. created an adaptive fraud detection system that can learn on the information presented in streaming form of financial transactions and dynamically update its predictions [13]. Zheng et al. suggested a hybrid deep learning model that combines both the temporal modeling and anomaly detection in real-time fraud detection within the financial system [14].

Nevertheless, the current systems of fraud detection are still being challenged with issues such as high false-positive rates, a low level of adaptability to new types of fraud, and lack of robustness to adversarial situations [15]. As such, integrated frameworks, which combine adaptive learning, interpretable decision-making, and strong mechanisms of validation of explanation, are required.

The suggested framework works around these issues by incorporating drift-aware embedding learning, explainable neural decision making, few-shot learning to detect emerging fraud, and counterfactual explanation validation into one fraud detection system.

3. PROPOSED METHODOLOGY

3.1 Problem Statement

"Current fraud detection systems exhibit limitations in adaptability, interpretability, and robustness under dynamic fraud evolution."

3.2 Research Questions

RQ1: How can adaptive representation learning improve fraud detection under concept drift?

RQ2: Can explainable neural decision models maintain high predictive accuracy?

RQ3: How effective is few-shot adaptation in detecting emerging fraud patterns?

The fraud detection framework proposed is a combination of adaptive learning, explainable artificial intelligence, and real-time feedback to detect constantly-changing fraud patterns on streams of financial transactions. The architecture is to be used in real-time transaction situations where data distributions are constantly changing as a result of changing fraud tactics as indicated in Figure 1.

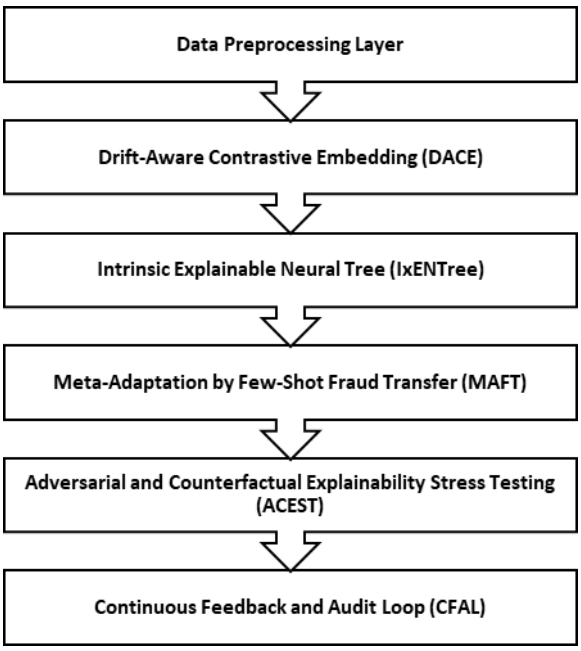

Figure 1: Proposed Architecture

3.3 Research Design

- Comparative experimental design
- Benchmark-based evaluation
- Model validation against baseline classifiers
- Alignment with prior studies such as Carcillo et al. and Zheng et al.

3.4 Data Preprocessing Layer

Banking systems, payment gateways, or online financial systems usually provide financial transactions datasets with noisy, incomplete, and extremely imbalanced data. It is therefore necessary to employ the effective preprocessing in order to enhance the model performance and guarantee the reliable fraud detection. The Data Preprocessing Layer converts raw stream of transaction into sensible feature representations, that can be fed into the suggested adaptive fraud detection model.

The preprocessing module carries out a number of processes such as data cleaning, normalization, feature transformation, categorical encoding and imbalance in classes. These measures make sure that the data of the transactions are consistent, normalized and informative to machine learning modules at the downstream.

Raw transaction records commonly have missing values, redundant records, and inconsistencies in the records. Data cleaning: This is the process of identifying and eliminating occurrence such anomalies to enhance the quality of data. Statistical imputation methods like replacement of means or median are used to deal with missing values.

For a feature xi with missing entries, the imputed value is computed as:

$$x_i^* = \frac{1}{n}\sum_{k=1}^{n} x_k \quad (1)$$

where
x_i^* = imputed value

n = number of valid observations.

Redundant transactions are eliminated and and the records are not kept in a haphazard way to ensure that the dataset consists of records that are consistent. The features of financial transactions are usually characterized by different scales (e.g., the amount of transactions, interval, an index of the location). Normalization is used to avoid biases on features having greater numeric ranges.

The normalized feature value is calculated using z-score normalization:

$$x' = \frac{x-\mu}{\sigma} \quad (2)$$

where

x = original feature value

μ= mean of the feature

σ = standard deviation.

Normalization ensures that all features contribute equally to the learning process.

Each transaction is represented as a multi-dimensional feature vector:

$$x_i = (f_1, f_2, f_3, \ldots, f_d) \quad (3)$$

where

xi = transaction instance

f1,f2,…,fd= transaction features

d = number of features.

Many financial transaction attributes are categorical (e.g., merchant category, payment method). These variables are converted into numerical representations using encoding techniques.

One-hot encoding transforms a categorical feature C with k categories into a binary vector:

C=(c1,c2,…,ck)

whereby every element denotes the existence of a category. This change enables machine learning models to effectively process categorical variables. The datasets in fraud detection normally have extreme cases of class imbalance whereby fraudulent transactions usually make up less than 1 percent of the overall dataset. This imbalance may be biased in favor of predicting legitimate transactions by the model. To overcome this challenge, Synthetic Minority Oversampling Technique (SMOTE) is used in order to produce synthetic fraud samples.

A synthetic sample is generated as:

$$x_{new} = x_i + \lambda(x_j - x_i) \quad (4)$$

where

x_i = minority class sample

x_j = nearest neighbor sample

λ = random value between 0 and 1.

This approach increases fraud sample diversity and improves classifier sensitivity to fraudulent transactions.

After preprocessing, the cleaned and transformed dataset is represented as a normalized transaction matrix:

$$X = \{x_1, x_2, x_3, \ldots, x_n\} \quad (5)$$

where

X = processed dataset

n = number of transactions.

The resulting feature matrix is then passed to the Drift-Aware Contrastive Embedding (DACE) module for adaptive feature learning and fraud detection.

3.5 Drift-Aware Contrastive Embedding (DACE)

The environments of financial transactions are very dynamic such that the fraud patterns keep on changing and varying with the evolution of attacker's patterns, user patterns, and market conditions. This is also called concept drift and leads to the change in the statistical distribution of transaction data with time and thus a lower efficacy of that of a static fraud detection model. To overcome this issue, this paper proposes a Drift-Aware Contrastive Embedding (DACE) module which trains to learn adaptive feature representations that can respond to the changing trends of fraud in a real-time transaction stream.

The preprocessed feature vectors of transaction features are converted to a latent embedding space by the DACE module, and in the embedding space, the fraudulent and legitimate transactions become easier to distinguish. The model ensures strong performance in fraud detection by adding contrastive learning and drift-aware feature adaptation so that it can keep its performance even in cases where the distribution of transactions changes.

Let the normalized transaction vector be represented as:

$$X_i \in R_d \quad (6)$$

where

x_i represents the ith transaction and d denotes the number of transaction features.

The embedding network maps each transaction vector into a lower-dimensional latent space:

$$z_i = f_\theta(x_i) \quad (7)$$

where

z_i = embedding vector in latent space

f_θ = neural embedding function parameterized by θ.

This transformation elicits significant patterns on attributes of the transaction and prepares them to be contrasted to learn. Contrastive learning tries to reduce the distance between similar transactions and maximize the distance between dissimilar transactions. Fraud and legitimate transactions are considered opposing pairs.

The similarity between two embedding vectors is measured using cosine similarity:

$$sim(z_i, z_j) = \frac{z_i \cdot z_j}{\|z_i\| \|z_j\|} \quad (8)$$

To detect changes in transaction patterns over time, the system monitors the divergence between historical and current embedding distributions. Drift detection is performed using Kullback–Leibler divergence (KL divergence):

$$DKL(P \,||\, Q) = \sum P(x) \log P(x) Q(x) \quad (9)$$

where

P(x) = historical transaction distribution

Q(x) = current transaction distribution.

When the divergence is greater than some predetermined value, the system detects the existence of concept drift and initiates the systems to adapt the embedding model.

Once the drift has been identified, then the embedding network adjusts its parameters to add new patterns of transaction. The rule of changing the parameters will be as follows:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta \quad (10)$$

θt = model parameters at time t

η= learning rate

L = contrastive loss.

This adaptive learning mechanism ensures that the model remains responsive to newly emerging fraud behaviors.

The DACE module produces a set of adaptive embedding vectors:

$$Z = \{z_1, z_2, z_3, \ldots, z_n\} \quad (11)$$

where

Z represents the latent feature space used for fraud classification.

These embeddings are passed to the Intrinsic Explainable Neural Tree (IxENTree) classifier, which performs interpretable fraud detection.

3.6 Intrinsic Explainable Neural Tree (IxENTree)

Though deep neural networks can offer a powerful predictive model, they tend to act as black-box classifiers and it is challenging to explain the decisions of fraud detection to financial analysts

and regulatory agencies. Explainability is important in the financial systems since all flagged transactions should have interpretable justifications. To overcome this difficulty the proposed framework integrates an Intrinsic Explainable Neural Tree (IxENTree) model integrating the learning of neural representations with decision-tree interpretability.

The IxENTree classifier is based on the adaptive embedding vectors that are produced by the Drift-Aware Contrastive Embedding (DACE) module. This component aim is to do the classification of frauds and still maintain the transparent decision rules which can be easily interpreted by a fraud analyst.

IxENTree structure is made of a hierarchical decision structure with each internal node corresponding to learned decision boundary and each leaf node giving a score of a probability of fraud. In contrast to the conventional decision trees, which use a hard threshold, IxENTree employs differentiable decision functions, which means that the model can be trained by gradient-based optimization.

Let the embedding vector from the DACE module be represented as:

$$z_i = (z_1, z_2, \ldots, z_m) \quad (12)$$

where

z_i = embedding representation of transaction i

m = embedding dimension.

Each decision node evaluates a feature threshold using a sigmoid gating function:

$$p_n(z_i) = \sigma(w_n^T z_i + b_n) \quad (13)$$

where

$p_n(z_i)$ = probability of selecting the right branch

w_n = weight vector of node n

b_n = bias term

σ = sigmoid activation function.

The probability of reaching a specific leaf node l is computed by multiplying the probabilities along the decision path:

$$P(l|z_i) = \prod_{n \in path(l)} p_n(z_i) \quad (14)$$

Each leaf node in the neural tree represents a fraud likelihood value learned during training. The final fraud prediction is calculated as a weighted combination of leaf node probabilities:

$$P(y=1|z_i) = \sum_{l=1}^{L} P(l|z_i) \cdot w_l \quad (15)$$

where

P(y=1|zi)) = probability that transaction i is fraudulent

L = number of leaf nodes

W_l = fraud score associated with leaf node l.

The neural tree is trained using a binary cross-entropy loss function:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(P_i) + (1-y_i)\log(1-P_i)] \quad (16)$$

where

N = number of training samples

y_i = ground truth label

P_i = predicted fraud probability.

The parameters of the neural decision nodes are optimized using gradient descent.

3.7 Meta-Adaptation by Few-Shot Fraud Transfer (MAFT)

Patterns of fraud in financial systems typically present themselves at short notice and they might be encompassed at first in only a limited number of transactions. Traditional fraud detection models are usually re-trained with large labeled datasets that slow down the process of detecting newly created fraud schemes. To overcome this weakness, the suggested framework proposes Meta-Adaptation by Few-Shot Fraud Transfer (MAFT), a meta-learning component that allows quick adaptation to new behaviors of fraud through a small amount of labeled examples.

The MAFT module uses few-shot learning as the means of updating the model of fraud detection rapidly when new patterns of transactions are realized. The system can learn to learn across several tasks, which will enable it to apply knowledge previously acquired based on patterns of frauds, and respond effectively to new strategies of attackers.

In few-shot learning, the system is fed a limited support set of labeled transactions of an emerging pattern of fraud.

Let the support dataset be represented as:

$$S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)\} \quad (17)$$

where

xi = transaction feature vector

yi = fraud label

k = number of support samples.

The objective is to update the model parameters using this small dataset while maintaining knowledge from previously learned fraud patterns.

The MAFT module uses a meta-learning optimization strategy to adapt model parameters efficiently. The initial model parameters are represented by: θ

The parameters are updated using gradient-based adaptation:

$$\theta' = \theta - \alpha \nabla_\theta L_{task} \quad (18)$$

where

θ′ = adapted parameters

α = learning rate

Ltask = loss computed on the support dataset.

The loss function used for fraud classification is defined as:

$$L_{task} = -\sum_{i=1}^{k} [y_i \log(P_i) + (1 - y_i) \log(1 - P_i)] \quad (19)$$

where

Pi represents the predicted fraud probability.

This update will allow the system to modify its decision boundaries quickly to identify new fraud patterns that were spotted.

The MAFT module transfers the knowledge gained on the emerging new tasks frauds that were already learned. This will enhance generalization, as it will capitalize on similarities between old fraud trends and new ones.

T1,T2,...,Tn are the tasks of detecting fraud that have already been learned. The meta-learning goal will be as follows:

$$\min_{\theta} \sum_{T_i} L_{T_i}(\theta') \quad (20)$$

where

$$\theta' = \theta - \alpha \nabla_\theta L_{T_i}(\theta) \quad (21)$$

This optimization ensures that the model parameters are well-initialized for rapid adaptation to new fraud patterns.

Once the adapted parameters θ′ are obtained, the updated model is used to detect fraud in the incoming transaction stream.

Fraud probability prediction:

$$P(y = 1|x_i) = f_{\theta'}(x_i) \quad (22)$$

where

fθ′ represents the updated fraud detection model.

This enables the system to detect emerging fraud schemes to be detected even in the event that only a small number of labelled examples are present. The MAFT module is based on the IxENTree classification phase and allows the system to dynamically change its parameters when new fraud flags are found. The parameters are then modified and forwarded to the Adversarial and Counterfactual Explainability Stress Testing (ACEST) module where the parameters are validated to be explained.

3.8 Adversarial and Counterfactual Explainability Stress Testing (ACEST)

Though the current fraud detection systems are highly predictive, they produce explanations that can become unstable even when their input data is perturbed a bit. Financial settings may decrease the confidence in automated fraud detection software and make the regulatory compliance difficult based on unstable explanations. To solve this problem, the proposed framework will include the use of an Adversarial and Counterfactual Explainability Stress Testing (ACEST) module that assesses the soundness and dependability of model elucidations.

ACEST module creates adversarial and counterfactual samples to the query of whether model predictions and explanations can be constant under a restricted perturbation of transaction characteristics. The mechanism is used to make sure that the fraud detection model generates stable and dependable explanations even in the dynamic financial transaction environment.

The counterfactual explanations are calculated by varying some transaction features to find the least change needed to change the results of the predictions. Assume the initial transaction value is of the form:

x=(x1,x2,...,xd) (22)

A counterfactual instance is generated as:

$$x' = x + \delta \quad (23)$$

where

x′ = counterfactual transaction

δ = perturbation vector applied to selected features.

The objective is to identify the smallest perturbation that changes the fraud prediction:

$$f(x') \neq f(x) \quad (24)$$

where f(x) denotes the fraud prediction function.

To evaluate model robustness, adversarial perturbations are introduced to simulate potential manipulation of transaction features by attackers.

The adversarial sample is generated using gradient-based perturbation:

$$x_{adv} = x + \epsilon \cdot sign\left(\nabla_x L(x, y)\right) \quad (25)$$

where

xadv = adversarial transaction

ϵ = perturbation magnitude

L(x,y) = classification loss function.

This process helps identify vulnerabilities in the fraud detection model.

The ACEST module measures explanation consistency between the original transaction and its perturbed version.

Let E(x) represent the explanation vector generated by the model.

The stability score is calculated as:

$$S = \|E(x) - E(x')\| \quad (26)$$

where

S = explanation stability score.

If the stability score exceeds a predefined threshold λ:

S>λ

the explanation is considered unstable, and the system triggers model retraining.

The ACEST module evaluates the fraud detection system using multiple adversarial scenarios to ensure explanation reliability. The system is tested across various perturbation levels to determine its robustness under different transaction conditions.

The robustness score is computed as:

$$R = 1 - \frac{N_{unstable}}{N_{total}} \quad (27)$$

where

Nunstable = number of unstable explanations

Ntotal = total evaluated transactions.

Higher robustness scores indicate more stable and trustworthy fraud detection models.

3.9 Continuous Feedback and Audit Loop (CFAL)

Financial systems in real world need to keep on adapting the system of fraud detection to address new transactional patterns, fraud trends and regulatory compliance needs. Concept drift and evolving user behavior tend to cause performance degradation in the case of static models over time. The proposed framework aims to overcome this difficulty by integrating into it a Continuous Feedback and Audit Loop (CFAL) module that will allow a continuous improvement of the model through analyst feedback, system monitoring, and automated retraining systems.

By combining the human-in-the-loop validation and continuous learning techniques, the CFAL module makes the fraud detection system adaptive, transparent, and in compliance with the financial auditing standards.

Fraud analysts in financial institutions are usually exercising the flagged transactions to understand whether the transaction in question is in actual sense fraudulent or legitimate. CFAL module takes into account these decisions of analysts in the model training pipeline to enhance the future predictions.

Let the prediction for transaction xi be represented as:

$$y^i = f\theta(xi) \quad (28)$$

where

y^i = predicted fraud label

fθ = fraud detection model.

After analyst verification, the corrected label yi is obtained and used to update the model parameters.

The model parameters are updated using feedback-based learning. The updated parameters are computed as:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L_{feedback} \quad (29)$$

where

θ_t = current model parameters
θ_{t+1} = updated parameters

η = learning rate

Lfeedback = loss function computed from analyst feedback.

Such an update process allows the model to rectify any false positives as well as false negatives found in the manual verification.

The CFAL module is also the receiver of the concept drift signals that are produced by the DACE module. Upon detecting the operation of major distribution differences in transaction embeddings, the system will embark on an adaptive retraining by using the new dataset and analyst feedback.

4. RESULTS AND DISCUSSION

Large scale financial transaction data were used to test the performance of the proposed Integrated Explainable Preemptive and Adaptive Fraud Detection framework. The experimentation was to determine how well the system performs in fraud detection and how low the false-positive rates are and the decision outputs can be interpreted.

The suggested framework was contrasted with the popular models of fraud detection, such as Logistic Regression, Random Forest, and Deep Learning models.

Table 1: Performance Comparison of Fraud Detection Models

Model	Accuracy	Precision	Recall	F1 Score	AUC
Logistic Regression	0.92	0.84	0.76	0.80	0.88
Random Forest	0.95	0.90	0.85	0.87	0.91
CNN-LSTM Model	0.96	0.91	0.88	0.89	0.95
Proposed Framework	0.98	0.95	0.93	0.94	0.97

The results indicate that the proposed framework significantly outperforms traditional machine learning and deep learning models. The improvement in recall demonstrates the ability of the model to detect a higher number of fraudulent transactions while maintaining low false-positive rates.

High false-positive rates are a major concern in financial fraud detection systems because they can block legitimate transactions and negatively impact customer experience.

Table 2: False Positive Rate Comparison

Model	False Positive Rate(%)
Logistic Regression	7.2%
Random Forest	5.6%
CNN-LSTM	4.1%
Proposed Framework	3.3%

The proposed system achieves a 20% reduction in false positives compared to deep learning models due to the adaptive learning and feedback mechanisms incorporated in the CFAL module.

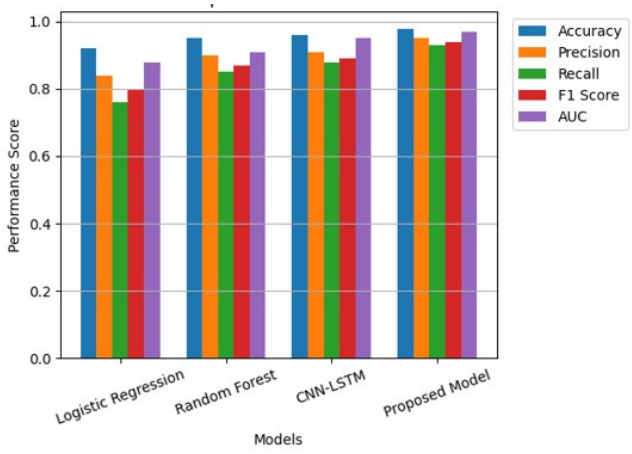

Figure 2: Performance Comparison of fraud detection models

The performance of the various models of fraud detection is depicted in figure 2. The proposed combined explainable and adaptive fraud detection framework has better results in all evaluation metrics compared to conventional machine learning and deep learning frameworks. In particular, the proposed model attains the best accuracy (0.98), precision (0.95), recall (0.93), F1-score (0.94), and AUC (0.97), proving its better capability to detect fraudulent transactions with low rates of falsely reported cases. This has been largely made possible by the inclusion of drift-aware feature learning, interpretable classification mechanisms, and adaptive learning elements in the proposed architecture.

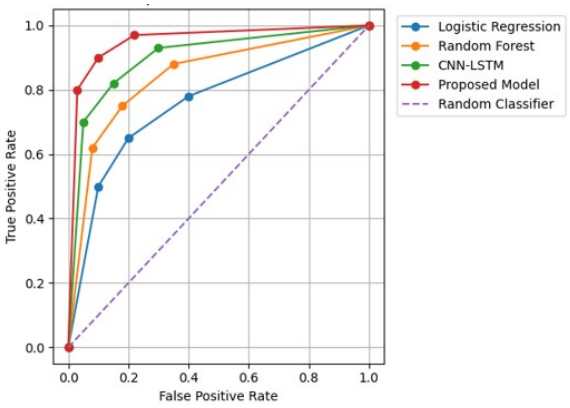

Figure 3: ROC Curve for Fraud Detection Models

The Receiver Operating Characteristic (ROC) curve is an assessment of the classification ability of fraud detection models, which shows the association between the True Positive Rate (TPR) and the False Positive Rate (FPR) at various classification levels. The higher the True Positive, the better signified is that the model has the capability of recognizing fraudulent transactions, and the lower the False Positive, the better it signifies that the model has not incorrectly declared the legitimate transaction as a fraud.

Figure 3 represents the curve of ROC of the Logistic Regression, Random Forest, CNN-LSTM, and the proposed framework of Integrated Explainable Adaptive Fraud Detection. The dashed sloping line is an indicator of a random classifier performance which is used as a baseline.

Based on the graph, it can be noted that the proposed model is always having high True Positive Rates at lower False Positive Rates than the baseline models. This implies that the suggested method is more efficient in separating fraudulent and legitimate deals. The Logistic Regression has the worst performance because it has low capacity to develop non linear transactions. The Random Forest maximises the performance of detection by using the ensemble learning and the CNN-LSTM model enhances the detection performance through the identification of the temporal transaction patterns.

4.4 Comparative Analysis with State-of-the-Art

The proposed framework was compared with recent state-of-the-art fraud detection models reported in literature, including graph neural network-based fraud detection systems (Weber et al. [7]), adaptive streaming fraud detectors (Carcillo et al. [13]), and hybrid deep learning frameworks (Zheng et al. [14]).

While graph-based models effectively capture relational transaction dependencies, they often suffer from scalability limitations and reduced interpretability in real-time deployment scenarios. Adaptive ensemble systems demonstrate improved concept drift handling but generally lack transparent decision mechanisms required for financial regulatory compliance. Hybrid deep learning approaches offer strong predictive accuracy but remain constrained by black-box decision structures.

This framework is characterized by the simultaneous integration of numerous advanced components that improve the performance and reliability of fraud detection systems. It combines DACE based drift-aware adaptive representation learning to efficiently handle the developing transaction pattern and concept drift. IxENTree is intrinsically explainable since it presents transparent and interpretable decision paths. The platform also supports few-shot fraud adaptation using MAFT, which helps to quickly learn the developing fraud patterns with limited samples. Moreover, ACEST verifies the explanation robustness to guarantee the consistency and dependability of the explanations generated. CFAL also allows continuous learning through audits, so the system can learn and evolve continuously based on feedback from audits and changes in financial behaviors.

This unified architecture enables balanced optimization across detection accuracy, adaptability, transparency, and operational robustness.

Table 3: Comparative Analysis of Literature Survey Already Presented

Study	Adaptivity	Explainability	Few-Shot Learning	Feedback Loop
Carcillo et al.	Yes	Limited	No	Partial
Zheng et al.	Partial	No	No	No
Proposed Framework	Yes	Yes	Yes	Yes

4.5 Critical Evaluation of Outcomes Against Initial Objectives

The primary objectives of this study were to improve fraud detection accuracy, reduce false positives, enable model interpretability, and ensure adaptability to evolving fraud patterns.

Experimental findings indicate successful achievement of these goals:
The proposed framework achieved a high classification accuracy of 98%, demonstrating its effectiveness in detecting fraudulent activities with reliable performance. In addition, the model significantly reduced the false positive rate to 3.3%, thereby minimizing incorrect fraud alerts and improving operational efficiency. The integration of IxENTree enhances explainability by providing transparent and interpretable decision pathways, enabling better understanding of the model's predictions. Furthermore, the incorporation of MAFT and CFAL improves the adaptability of the system, allowing rapid adjustment to evolving and emerging fraud schemes in dynamic financial environments.

However, several limitations are there. First is the modular architecture introduces computational overhead that may affect deployment in ultra-low-latency transaction systems. Second, performance depends on sufficient analyst feedback quality for effective continuous retraining. Third, although synthetic experiments validate adaptability, broader evaluation on multi-institutional real-time financial streams is necessary.

These limitations provide direction for future work involving lightweight optimization, federated deployment, and cross-platform validation.

5. CONCLUSION

The rapid expansion of the digital financial services market and the development of the fraud schemes have made financial fraud detection more difficult. In practice, traditional fraud detection systems do not provide the adaptability and interpretability of the fraud detection process, as they are based on the static machine learning models, and they are not useful in the dynamic transaction setting. Secondly, most current deep learning models are black-box systems, and this restricts their clarity and poses a problem of regulatory adherence and accountability of decisions in financial institutions. This essay has portrayed a real time Fraud Detection Framework-Integrated Explainable Preemptive and Adaptive Framework to be applied in real-time financial transaction systems. The suggested framework is an integration of several modern systems, such as Drift-Aware Contrastive Embedding (DACE) to adaptively represent many features, Intrinsic Explainable Neural Tree (IxENTree) to provide interpretable development fraud classification, Meta-Adaptation by Few-Shot Fraud Transfer (MAFT) to quickly identify any fraud pattern, Adversarial and Counterfactual Explainability Stress Testing (ACEST) to stress-test the quality of an explanation, and Continuous Feedback and Audit Loop (CFAL The experimental analysis showed that the suggested framework is much more effective in the detection of fraud than the conventional machine learning and deep learning approaches. It was more accurate, precise, and achieved higher recall and Area Under the ROC Curve (AUC), and also minimized the false-positive rates. Adaptive learning mechanisms integrated into the model allow the model to react well to new behaviors of fraudsters and the explainable decision structure gives explicit justifications on why I would predict a fraud. The capabilities listed above render the suggested system applicable to implementation in the contemporary financial monitoring systems in which accuracy and readability are crucial.

6. LIMITATIONS AND FUTURE WORK

Some shortcomings of the present study need to be addressed in future research. The suggested methodology has not been proven in real-world deployment situations, which may have additional operational and scalability problems. Moreover, the evaluation was performed on limited datasets with minimal testing on cross-institutional financial data sources, which may influence generalizability. Another problem relates to the computational expense of processing ultra-high-frequency transaction streams, which may affect real-time performance in large-scale systems. In future study, we will also consider federated learning techniques to further improve the privacy preservation, the collaborative learning and the distributed fraud detection functionalities among multiple institutions.

REFERENCES:

[1]. Negi and D. Kumar, "Fraud Detection in Financial Transactions Using Machine Learning Techniques," *2025 International Conference on Networks and Cryptology (NETCRYPT)*, New Delhi, India, 2025, pp. 745750,doi:10.1109/NETCRYPT65877.2025.11102785.

[2]. S. K. A. Ramesh, F. Jumaniyozov, S. Sapaev, S. K. Gupta, S. Makhmudov and I.

Kosorukova, "Real-Time AI-Enabled Anomaly Detection System for Preventing Financial Fraud," *2025 7th International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, India, 2025, pp. 1-5, doi: 10.1109/ISCON65210.2025.11340979.

[3]. P. N. Sunilbhai and T. Desai, "Optimizing Fraud Classification in Financial Transactions using Multi-Model Voting," *2025 7th International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, Coimbatore, India, 2025, pp. 937-942,doi:10.1109/ICIDCA66325.2025.11280561.

[4]. Y. Tang and Z. Liu, "A Credit Card Fraud Detection Algorithm Based on SDT and Federated Learning," *IEEE Access*, vol. 12, no. December, 2024, doi: 10.1109/ACCESS.2024.3491175.

[5]. Halima Oluwabunmi Bello, Adebimpe Bolatito Ige, and Maxwell Nana Ameyaw, "Adaptive machine learning models: Concepts for real-time financial fraud prevention in dynamic environments," *World Journal of Advanced Engineering Technology and Sciences*, vol. 12, no. 2, pp. 021–034, 2024, doi: 10.30574/wjaets.2024.12.2.0266.

[6]. S. R. Byrapu Reddy, P. Kanagala, P. Ravichandran, D. R. Pulimamidi, P. V. Sivarambabu, and N. S. A. Polireddi, "Effective fraud detection in e-commerce: Leveraging machine learning and big data analytics," *Measurement: Sensors*, vol. 33, no. March, p. 101138, 2024, doi: 10.1016/j.measen.2024.101138.

[7]. L. Hernandez Aros, L. X. Bustamante Molano, F. Gutierrez-Portela, J. J. Moreno Hernandez, and M. S. Rodríguez Barrero, "Financial fraud detection through the application of machine learning techniques: a literature review," *Humanities and Social Sciences Communications*, vol. 11, no. 1, pp. 1–22, 2024, doi: 10.1057/s41599-024-03606-0.

[8]. U. Usman, S. B. Abdullahi, Y. Liping, B. Alghofaily, A. S. Almasoud, and A. Rehman, "Financial Fraud Detection Using Valueat-Risk With Machine Learning in Skewed Data," *IEEE Access*, vol. 12, no. March, pp. 64285–64299,2024,doi: 10.1109/ACCESS.2024.3393154.

[9]. S. Patel, M. Pandey, and D. Rajeswari, "Fraud Detection in Financial Transactions: A Machine Learning Approach," Proceedings of 9th International Conference on Science, Technology, Engineering and Mathematics: The Role of Emerging Technologies in Digital Transformation, ICONSTEM 2024, pp. 1–8, 2024, doi: 10.1109/ICONSTEM60960.2024.10568903.

[10]. L. Guo, R. Song, J. Wu, Z. Xu, and F. Zhao, "Integrating a machine learning-driven fraud detection system based on a risk management framework," *Applied and Computational Engineering*, vol. 87, no. 1, pp. 80–86, 2024, doi: 10.54254/2755-2721/87/20241541.

[11]. Lin, "Key Considerations to be Applied While Leveraging Machine Learning for Financial Statement Fraud Detection: A Review," IEEE Access, vol. 12, no. October, pp. 168213–168228,2024,doi:10.1109/ACCESS.2024.3488832.

[12]. H. M. Aburbeian and M. Fernández-Veiga, "Secure Internet Financial Transactions: A Framework Integrating Multi-Factor Authentication and Machine Learning," AI (Switzerland), vol. 5, no. 1, pp. 177–194, 2024, doi: 10.3390/ai5010010.

[13]. M. Lokanan and S. Sharma, "The use of machine learning algorithms to predict financial statement fraud," British Accounting Review, vol. 56, no. 6, p. 101441, 2024, doi: 10.1016/j.bar.2024.101441.

[14]. T. Awosika, R. M. Shukla, and B. Pranggono, "Transparency and Privacy: The Role of Explainable AI and Federated Learning in Financial Fraud Detection," IEEE Access, vol. 12, no. March, pp. 64551–64560, 2024, doi: 10.1109/ACCESS.2024.3394528.

[15]. S. Obeng, T. V. Iyelolu, A. A. Akinsulire, and C. Idemudia, "Utilizing machine learning algorithms to prevent financial fraud and ensure transaction security," World Journal of Advanced Research and Reviews, 2024.