FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 24, December, 2026, pp 248 – 255
DOI: https://doi.org/10.33003/fjs-2026-1024-6313

An Explainable Hybrid CNN–Transformer Framework for Fraud Detection in Digital and Cryptocurrency Transactions

***Nurudeen Olawale Ibrahim, Mustapha Olaitan Olarewaju, Hamid Dimeji Lawal and Yusuf Omotosho Bello-Sulayman**

Department of Computer Science, Faculty of Computing, Engineering and Technology, Al-Hikmah University, Ilorin, Kwara State, Nigeria.

*Corresponding authors' email: noibrahim@alhikmah.edu.ng
ORCID iD: https://orcid.org/0009-0009-9566-7820

Received: 27 Oct. 2026
Accepted: 10 Nov. 2026
Published: 14 Nov. 2026

ABSTRACT
Fraud detection in digital payment and cryptocurrency systems is complicated by heterogeneous transaction records, class imbalance and limited model interpretability. This study proposed an explainable hybrid convolutional neural network (CNN)–Transformer framework for identifying suspicious transactions using the IEEE-CIS Fraud Detection and Ethereum Address Behavior datasets. Tabular features were transformed into spatial representations using the Image Generator for Tabular Data algorithm, while principal component analysis and rolling temporal sequences were employed to represent transaction behaviour. A dual-branch CNN–Transformer model was developed with cost-sensitive learning and hyperparameter optimisation using the Sparrow Search Algorithm. Model predictions were interpreted using SHapley Additive exPlanations, while an IPFS- and blockchain-based architecture was proposed for audit logging. Under the reported chronological evaluation procedure, the model achieved a precision of 0.7320, recall of 0.4956, F1-score of 0.5910 and area under the precision–recall curve of 0.5984 at a decision threshold of 0.5350. These results indicate moderate fraud-discrimination performance, although approximately half of the labelled fraudulent cases were not detected at the selected threshold. Further evaluation against established baseline models, independent validation of cryptocurrency fraud labels and experimental assessment of the proposed blockchain subsystem are necessary to establish the framework's comparative effectiveness and operational suitability.

Keywords: Cost-sensitive learning, Cryptocurrency fraud detection, Explainable artificial intelligence (XAI), Hybrid CNN-Transformer, Image Generator for Tabular Data (IGTD), Zero-knowledge proofs

INTRODUCTION

The rapid digitalization of global financial systems has expanded electronic value exchange across both centralized banking infrastructures and decentralized cryptocurrency networks (Mallikarjuna et al., 2026). Despite increasing interconnectivity via fiat-to-crypto on-ramps, conventional electronic payment rails and cryptocurrency networks differ fundamentally in their transaction structures, feature spaces, supervisory labels, and fraud mechanisms. Centralized digital payment systems process structured account- and card-level transactions enriched with merchant metadata, device fingerprints, and verified chargeback records, where fraud typically manifests as card-not-present misappropriation or account takeover (Al-Hchaimi et al., 2026; Ampumuza et al., 2026; Yang et al., 2026). Conversely, public cryptocurrency networks such as Ethereum record pseudonymous wallet-to-wallet transfers and smart contract invocations characterized by gas telemetry, interaction graphs, and temporal burst patterns, frequently lacking verified ground-truth fraud labels (Lazea et al., 2026). Evaluating whether a unified feature-transformation and deep learning pipeline can extract predictive spatial and sequential representations across both centralized payment logs and cryptocurrency behavioral records is therefore an important methodological question for cross-domain financial security.

Traditional rule-based screening engines rely on static heuristics that struggle to capture non-linear anomalies, adapt to concept drift, or control false-positive alerts (Alessi & Fugini, 2026; Arzu et al., 2026; Ofosu-Ampong et al., 2026). Consequently, machine learning and deep learning models have been widely adopted for automated fraud detection (Fardous, 2026; Khan, 2026; Mohammed et al., 2026). Conventional linear classifiers such as logistic regression offer high interpretability and low latency but fail to model complex non-linear feature interactions. Ensemble tree architectures, including random forests and gradient-boosted decision trees, achieve strong benchmark performance on static tabular datasets, yet they rely on manual feature engineering to represent sequential transaction histories. Furthermore, financial datasets exhibit severe class imbalance, with fraudulent events often comprising less than 1% to 3.5% of total volume (Al-Hchaimi et al., 2026). To mitigate majority-class bias, researchers have combined synthetic minority over-sampling techniques (SMOTE) with cost-sensitive loss functions and metaheuristic optimizers such as the Sparrow Search Algorithm (SSA) (Al-Hchaimi et al., 2026; Arzu et al., 2026).

For capturing complex feature correlations and behavioral trajectories, deep neural architectures offer distinct structural advantages alongside notable trade-offs (Gong et al., 2026). Mapping one-dimensional (1D) tabular records into two-dimensional (2D) spatial matrices via algorithms such as the Image Generator for Tabular Data (IGTD) enables Convolutional Neural Networks (CNNs) to extract localized

cross-feature interactions, whereas Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks model sequential dependencies but suffer from vanishing gradients and sequential training bottlenecks. Transformer architectures overcome these sequential constraints through parallel multi-head self-attention, capturing long-range dependencies across transaction histories (Darwish et al., 2026; Fnu et al., 2026).

Beyond predictive accuracy, operational fraud detection systems require model interpretability and tamper-evident record keeping (Alessi & Fugini, 2026; Davuluri, 2026; Naz et al., 2026). In this context, explainable artificial intelligence (XAI) and decentralized storage technologies serve distinct, non-interchangeable functions. Post-hoc attribution methods such as SHapley Additive exPlanations (SHAP) quantify the contribution of input features to individual model predictions to assist human investigators during forensic review (Al-Hchaimi et al., 2026; Maligireddy, 2026). Separately, distributed ledger technologies, the InterPlanetary File System (IPFS), and zero-knowledge proofs (ZKPs) do not perform fraud classification; rather, they provide architectural mechanisms for off-chain payload storage, cryptographic state verification, and immutable audit logging (Al-Jaziri et al., 2026; Allehaibi, 2026; Saragih et al., 2026).

Several methodological challenges remain underexplored in current literature. First, few studies evaluate unified spatial-temporal architectures combining 2D IGTD representations with PCA-compressed sequential windows across both digital payment and cryptocurrency datasets under chronological splitting that prevents temporal data leakage (Al-Hchaimi et al., 2026; Gong et al., 2026). Second, balancing precision and recall under severe class skewness requires systematic integration of cost-sensitive loss functions and threshold optimization. Third, while coupling SHAP explanations with off-chain IPFS storage and blockchain anchoring offers a promising conceptual design for auditability, its practical boundaries must be clearly separated from predictive model evaluation. In this study, the IGTD-PCA feature engineering pipeline, the SSA-tuned Hybrid CNN–Transformer classifier, the leakage-free chronological evaluation, and the SHAP attribution module are experimentally implemented and evaluated, whereas the IPFS, ZKP, and lightweight blockchain logging layer is formulated as an architectural design for downstream auditability.

This study aimed to develop and evaluate an explainable hybrid CNN–Transformer framework for fraud detection across digital payment and cryptocurrency transaction datasets, incorporating cost-sensitive learning and a proposed blockchain-based audit architecture. The primary objectives of this research are to:

i. Transform heterogeneous financial records from the IEEE-CIS and Ethereum datasets into unified two-dimensional spatial feature grids using the Image Generator for Tabular Data (IGTD) algorithm alongside PCA-reduced rolling temporal sequences;
ii. Develop and optimize a dual-branch Hybrid Spatial CNN–Transformer architecture utilizing the Sparrow Search Algorithm (SSA) and dynamic Focal Loss to mitigate extreme class imbalance without exhaustive computational overhead;
iii. Integrate gradient-based SHapley Additive exPlanations (SHAP) for local and global feature attribution alongside a privacy-preserving, lightweight blockchain logging layer utilizing zero-knowledge proofs (ZKPs) and off-chain Interplanetary File System (IPFS) storage; and
iv. Validate the end-to-end framework under a strict leakage-free time-series split methodology using precision, recall, F1-score, and Area under the Precision-Recall Curve (AUPRC).

MATERIALS AND METHODS

Study Design and Architectural Overview

This study employs an experimental system design that integrates multi-modal deep representation learning, metaheuristic cost-sensitive optimization, gradient-based explainability, and decentralized cryptographic logging to detect fraudulent transactions across hybrid financial ecosystems. As illustrated in the system block diagram (Figure 1) and operational flowchart (Figure 2), the proposed detection mechanism is structured into four sequential operational phases:

Phase 1 (Data Ingestion and Behavioral Syntax Transformation)

Harmonization of centralized electronic payment logs and decentralized blockchain records into two-dimensional spatial matrices and compressed temporal sequences.

Phase 2 (Predictive Modeling and Real-Time Detection)

Parallel feature extraction and classification via a dual-branch Enhanced Hybrid Spatial CNN–Transformer architecture.

Phase 3 (Cost-Sensitive Optimization)

Dynamic hyperparameter and threshold tuning using the Sparrow Search Algorithm (SSA) paired with an asymmetric Focal Loss objective function.

Phase 4 (Explainability and Immutable Logging)

Post-hoc feature attribution for flagged high-risk transactions using SHapley Additive exPlanations (SHAP), anchored to a lightweight blockchain ledger via zero-knowledge proofs (ZKPs) and off-chain Interplanetary File System (IPFS) storage.

In terms of experimental scope, Phases 1, 2, and 3, along with the SHAP feature-attribution module in Phase 4, were fully implemented and empirically evaluated in Python 3.10 using PyTorch 2.1, Scikit-learn 1.3, imbalanced-learn 0.11, Mealpy 3.0, and SHAP 0.44 on a Google Colab Pro+ GPU environment (NVIDIA Tesla T4/A100 with 16 - 40 GB VRAM and 32 GB system RAM). The decentralized logging layer in Phase 4 (combining off-chain IPFS storage, ZKPs, and lightweight Ethereum smart contracts via Web3.py) is presented as a modular architectural specification and prototype workflow to illustrate how model outputs and SHAP reports can be anchored for downstream auditability.

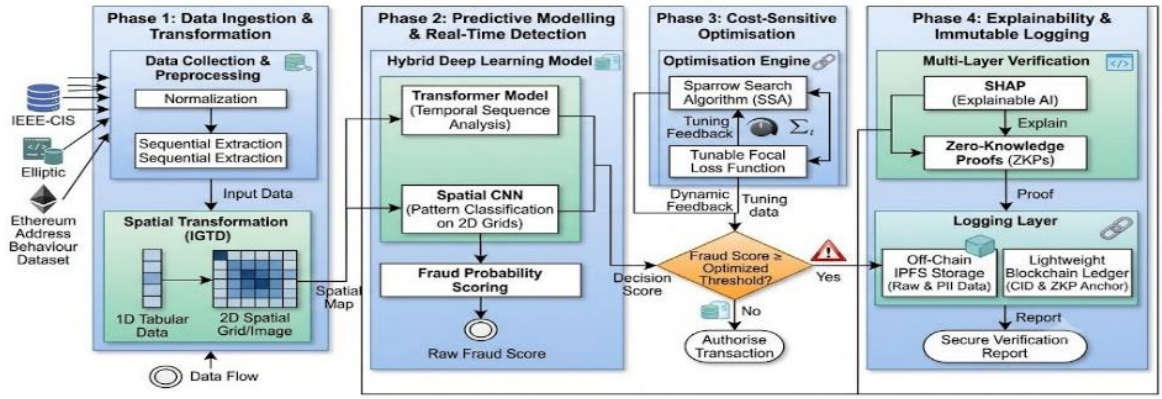


Figure 1: Block Diagram of the Proposed System

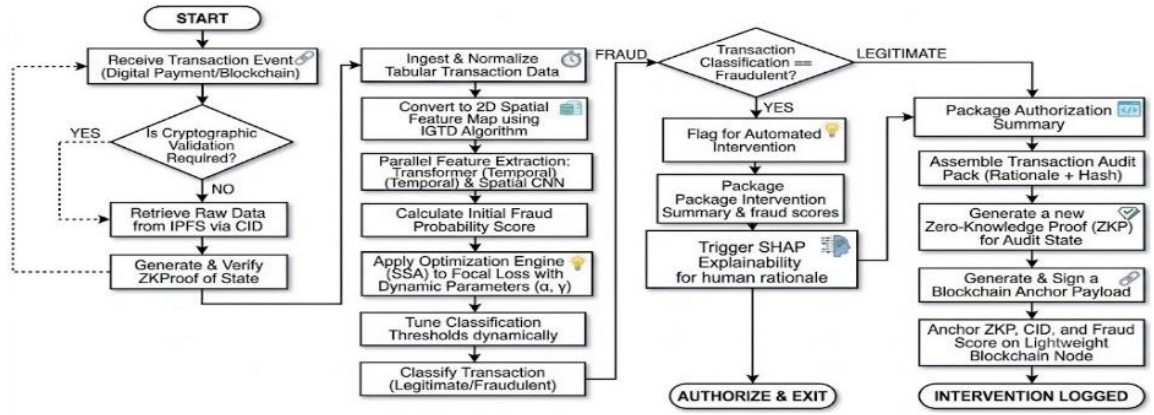


Figure 2: Program Flowchart of the Proposed System

Dataset Description, Harmonization, and Leakage-Free Preprocessing

Two transaction repositories representing conventional electronic payments and cryptocurrency network activity were utilized in this study:

IEEE-CIS Fraud Detection Dataset

Provided by Vesta Corporation via Kaggle, the raw repository comprises 590,540 e-commerce payment transactions recorded over a six-month period across 434 features (including transaction amounts, cardholder identifiers, product codes, time deltas, and Vesta engineered rich features) with a ground-truth fraud prevalence of approximately 3.50%.

Ethereum Address Behavior Dataset

Obtained from publicly archived Ethereum behavioral metrics and Etherscan Application Programming Interface (API) transaction summaries, this repository captures wallet-level transactional behavior, including sent/received transaction counts, time intervals between transfers, ERC-20 token interaction frequencies, and ether value distributions.

To enable computationally tractable training for iterative IGTD matrix optimization and metaheuristic search, a stratified chronological subset of 50,000 records (comprising 40,000 IEEE-CIS payment records and 10,000 Ethereum behavioral records) was extracted. Because IEEE-CIS records represent individual payment transactions whereas Ethereum records summarize wallet-level transactional intervals, cross-domain unification was performed at the level of a standardized financial observation event indexed by a timestamp and a unique entity identifier (card1 account identifier for IEEE-CIS and wallet address hash for Ethereum). Categorical attributes were frequency-encoded, and the union of continuous and encoded behavioral attributes was aligned into a standardized 400-dimensional numerical feature vector per observation, with domain-unobserved columns zero-padded after standardization.

Because raw Ethereum behavioral records lack verified supervisory fraud labels, an unsupervised Isolation Forest anomaly detection model (configured with 100 estimators and a contamination rate of 0.05) was fitted on the training partition of the Ethereum subset to assign binary anomaly pseudo-labels (yielding approximately 5% positive anomaly instances). Importantly, unsupervised statistical anomalies are not equivalent to forensically confirmed cryptocurrency fraud; therefore, performance on the Ethereum partition represents anomaly-detection alignment rather than verified fraud recovery.

Spatial and Temporal Feature Engineering

Rather than feeding one-dimensional (1D) flat tabular vectors directly into a single classifier, the preprocessed records were engineered into complementary spatial and temporal representations to form a continuous behavioral syntax:

Two-Dimensional Spatial Transformation via IGTD

The Image Generator for Tabular Data (IGTD) algorithm mapped the 400 standardized tabular features into a 2D spatial grid of 20 x 20 pixels, reshaped into a single-channel tensor X_s of dimension 1 x 20 x 20 per observation. Pairwise feature similarity was quantified using Euclidean distance over ranked Pearson correlation coefficients, while spatial pixel proximity on the 20 x 20 grid was measured via Euclidean coordinate distance. The IGTD greedy swap optimization was executed for up to 1,000 iterations (stopping criterion: 50 consecutive iterations without error reduction) to minimize the discrepancy between feature correlation ranks

and spatial pixel distance ranks, clustering correlated financial variables into adjacent grid coordinates. Where active features were fewer than 400, remaining grid coordinates were zero-padded.

Rolling Temporal Sequence Construction

To capture sequential behavioral patterns without dimensionality explosion, a PCA projection fitted on the standardized training features compressed the 400-dimensional space into d_model = 32 orthogonal principal components. Observations were grouped by their unique entity identifier and ordered chronologically within each split partition (training, validation, and test) so that no sequence crossed partition boundaries. Using a causal sliding window of length seq_len = 10 and unit stride, each observation at step t was paired with its preceding 9 entity observations (t - 9 through t), forming a temporal input tensor X_t of dimension 10 x 32. For entities with fewer than 10 historical observations, left-side zero-padding (pre-padding) was applied so that the current transaction always occupied the final sequence index (t = 10), preventing look-ahead leakage.

Dual-Branch Hybrid CNN–Transformer Architecture

The core predictive engine (HybridFraudDetectorEnhanced) was implemented in PyTorch as a parallel dual-branch deep neural network that simultaneously ingests the $1 \times 20 \times 20$ spatial grid X_s and the 10×32 temporal sequence tensor X_t (Table 1).

Table 1: Architectural Specifications of the Dual-Branch Hybrid CNN–Transformer Model

Module / Branch	Layer Type	Configuration / Parameters	Output Dimension
Spatial CNN Branch	Input Tensor (X_s)	2D IGTD Spatial Feature Grid	$1 \times 20 \times 20$
	Conv Block 1	Conv2d ($1 \rightarrow 16$, kernel 3×3, pad 1) + ReLU + MaxPool2d (2×2)	$16 \times 10 \times 10$
	Conv Block 2	Conv2d ($16 \rightarrow 32$, kernel 3×3, pad 1) + ReLU + MaxPool2d (2×2)	$32 \times 5 \times 5$
	Conv Block 3	Conv2d ($32 \rightarrow 64$, kernel 3×3, pad 1) + ReLU + MaxPool2d (2×2, pad 1)	$64 \times 3 \times 3$
	Spatial Flatten	Flatten()	576
Temporal Branch	Input Tensor (X_t)	PCA-Reduced Rolling Sequence (seq_len $= 10$, $d_{\text{model}} = 32$)	10×32
	Transformer Encoder	$3 \times$ TransformerEncoderLayer ($d_{\text{model}} = 32$,nhead $= 8$, batch_first)	10×32
	Temporal Flatten	Flatten(start_dim=1)	320
Fusion and Head	Feature Concatenation	torch.cat((cnn_features, transformer_features), dim=1)	896
	Fully Connected 1	Linear ($896 \rightarrow 256$) + ReLU + Dropout ($p = 0.4$)	256
	Fully Connected 2	Linear ($256 \rightarrow 128$) + ReLU + Dropout ($p = 0.3$)	128
	Output Layer	Linear ($128 \rightarrow 1$)	1 (Logit)

Spatial CNN Branch

The spatial branch processes the $1 \times 20 \times 20$ IGTD grid through three sequential 2D convolutional blocks (16, 32, and 64 filters, respectively). Each convolutional layer utilizes a 3×3 kernel with unit padding (padding = 1), followed by a Rectified Linear Unit (ReLU) activation and a 2×2 max-pooling operation (MaxPool2d, with unit padding applied in the third pooling stage). The resulting feature maps are flattened into a 1D spatial representation vector $f_{CNN} \in \mathbb{R}^{576}$.

Temporal Transformer Branch

Concurrently, the 10×32 sequential tensor is processed by a 3-layer Transformer Encoder (num_layers = 3) configured with 8 parallel self-attention heads (nhead = 8, $d_{\text{model}} = 32$). By computing multi-head scaled dot-product self-attention across the 10-step transaction window, the encoder captures non-linear temporal dependencies and sudden behavioral shifts, flattening the resulting representations into a temporal vector $f_{Trans} \in \mathbb{R}^{320}$.

Feature Concatenation and Classification Head

The extracted spatial and temporal vectors are concatenated into a unified multimodal representation $f_{concat} = [f_{CNN} \parallel f_{Trans}] \in \mathbb{R}^{896}$ and passed through a three-layer fully connected classification head ($896 \rightarrow 256 \rightarrow 128 \rightarrow 1$). Dropout regularization layers with rates of $p = 0.4$ and $p = 0.3$ are interleaved after the first and second dense layers, respectively, to prevent co-adaptation and overfitting before generating the final raw fraud logit.

Cost-Sensitive Optimization and Training Protocol

To mitigate severe class imbalance, asymmetric misclassification penalties were enforced during training using the parametric Focal Loss function, defined in Equation (1.0):

$$L(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \quad (1)$$

Where $p_t = \exp(-\text{BCE_loss})$ represents the model's estimated probability for the ground-truth class computed from binary cross-entropy with logits, alpha = 0.1 is the class-balancing weighting factor, and gamma = 5.0 is the focusing parameter that exponentially down-weights well-classified majority examples (p_t -> 1) so gradient updates focus on hard-to-classify instances.

Hyperparameter tuning and decision-threshold selection were performed using the Sparrow Search Algorithm (SSA) via the Mealpy library (OriginalSSA), configured with a population size of 20 search agents, a maximum of 15 iterations, a producer proportion of 0.8, and a safety threshold of ST = 0.8. The SSA optimized the Focal Loss parameters (alpha bounded within [0.05, 0.50] and gamma bounded within [1.0, 5.0]) and post-training operating classification threshold (tau bounded within [0.10, 0.90]) by maximizing the F1-score exclusively on the 15% validation partition. The 15% test partition was completely excluded from the SSA optimization loop.

Network weights were optimized over a maximum of 30 epochs with a mini-batch size of 256 using the Adam optimizer initialized at a learning rate of eta = 0.001. A dynamic learning rate scheduler (ReduceLROnPlateau) monitored the validation Focal Loss and reduced eta by a factor of 0.5 whenever validation loss plateaued for 2 consecutive epochs (patience = 2). Early stopping (patience = 5 epochs) restored the optimal model weights (best_wts) corresponding to the minimum validation loss.

Explainable AI (SHAP) and Privacy-Preserving Blockchain Logging

Whenever a transaction's predicted fraud probability P_{Fraud} meets or exceeds the validation-tuned threshold tau = 0.5350, a two-part explainability and audit workflow is initiated:

SHAP Attribution via GradientExplainer (Experimentally Evaluated)

Flagged true-positive test instances were analyzed using shap.GradientExplainer calibrated against a background reference distribution of 100 randomly sampled training instances (random_state = 42). Expected gradient attributions were computed separately across the flattened spatial input space (400 features corresponding to row-major coordinates (r, c) in the 20 x 20 IGTD grid) and the flattened temporal input space (320 features corresponding to time step t in {1..10} and principal component k in {1..32}). While displayed by their transformed tensor indices during rapid screening, each spatial index maps back to its exact original financial variable via the saved 400-element IGTD permutation lookup table, and each temporal index maps to a linear combination of original variables via the fitted 32 x 400 PCA loading matrix.

Proposed IPFS, ZKP, and Lightweight Blockchain Logging Architecture (Conceptual Design)

To illustrate how model decisions can be recorded for tamper-evident forensic auditing, we propose an architectural logging layer separating payload storage from on-chain state verification. In this design, detailed transaction logs and SHAP attribution reports are encrypted and stored off-chain on an InterPlanetary File System (IPFS) node, generating a cryptographic Content Identifier (CID). Asynchronous Web3.py middleware is designed to transmit only the CID, the numerical fraud probability score, and a zero-knowledge proof (ZKP) attesting to threshold evaluation via JSON-RPC to a Solidity smart contract on an Ethereum-compatible test network. While keeping raw payloads off-chain reduces on-chain storage overhead, achieving full regulatory compliance under frameworks such as the General Data Protection Regulation (GDPR) additionally requires institutional access controls, encryption key lifecycle management, and coordinated off-chain pin deletion protocols that lie beyond the experimental scope of this study.

Leakage-Free Data Splitting, SMOTE Balancing, and Evaluation Metrics

To prevent temporal data leakage, standard k-fold cross-validation was rejected in favor of a chronological Time-Series Split. The harmonized dataset was partitioned chronologically into a Training set (70%), a Validation set (15%, derived via a 0.1765 split of the initial 85% non-test records), and an untouched Test set (15%). Within each domain, timestamps were ordered monotonically prior to splitting so that all test observations occurred chronologically after the training and validation observations.

To mitigate majority-class dominance during gradient descent without contaminating the evaluation partitions, the Synthetic Minority Over-sampling Technique (SMOTE, configured with sampling_strategy = 0.5, k_neighbors = 5, and random_state = 42) was applied strictly to the flattened training feature representations prior to reshaping into 1 x 20 x 20 spatial grids and 10 x 32 temporal tensors. Because linear interpolation in SMOTE can introduce synthetic feature combinations that approximate rather than replicate real transaction trajectories, a conservative minority-to-majority ratio of 0.5 was selected rather than full 1:1 parity, relying on Focal Loss (gamma = 5.0) to handle the remaining imbalance. The Validation (15%) and Test (15%) partitions remained completely unaugmented to reflect natural class skewness.

Predictive performance on the untouched test partition was evaluated using Precision, Recall (Sensitivity), F1-Score, and Area under the Precision–Recall Curve (AUPRC), defined in Equations (2.0) – (4.0):

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1-Score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (4)$$

Where TP, FP, and FN denote true positives, false positives, and false negatives evaluated at the optimal decision threshold.

RESULTS AND DISCUSSION

Predictive Performance on Imbalanced Hybrid Transaction Streams

The dual-branch Enhanced Hybrid CNN–Transformer model was evaluated on the untouched, naturally imbalanced test partition (15% of the dataset, exhibiting a positive-class prevalence of approximately 0.037, or 3.7%) following chronological time-series splitting. Because standard overall accuracy is distorted by majority-class dominance in financial datasets where fraudulent events constitute less than 1% of total traffic, evaluation focused on threshold-dependent and threshold-independent minority-class metrics. Table 2 summarizes the predictive performance achieved at the model's optimally tuned operating threshold.

Table 2: Empirical Performance Metrics of the Hybrid CNN–Transformer Architecture on the Untouched Test Set

Evaluation Metric	Operating Threshold	Empirical Value
Area Under the Precision–Recall Curve (AUPRC)	Threshold-Independent	0.5984
No-Skill Baseline AUPRC (Test Positive Prevalence)	Threshold-Independent	~0.0370
F1-Score at Selected Threshold	0.5350	0.5910
Precision	0.5350	0.7320
Recall (Sensitivity)	0.5350	0.4956

As shown in Table 2, at the selected decision threshold of 0.5350, the model achieved a Precision of 0.7320 and a Recall of 0.4956, yielding a consistent F1-Score of 2 * (0.7320 * 0.4956) / (0.7320 + 0.4956) = 0.5910, which also corresponded to the maximum F1-score across the precision–recall locus. Across all probability thresholds, the model achieved an AUPRC of 0.5984 (Figure 3), substantially exceeding the horizontal no-skill baseline of ~0.0370 (visible at the right-hand terminus of the curve at Recall = 1.0).

While a Precision of 0.7320 indicates that nearly three out of every four flagged transactions at the 0.5350 threshold correspond to positive instances, helping to limit false-positive alert fatigue, the accompanying Recall of 0.4956 reveals an important operational limitation: approximately 50.44% of labelled positive cases were not detected at this

threshold. In practical financial security deployments, a false-negative rate of 50.44% implies that roughly half of anomalous or fraudulent transactions would bypass automated intervention at the 0.5350 cutoff, potentially resulting in direct financial losses. As illustrated by the Precision–Recall trajectory in Figure 3, institutions prioritizing higher fraud interception can lower the operating threshold (for example, shifting along the curve toward a recall of 0.70 – 0.80), albeit at the cost of reduced precision and increased manual review volume.

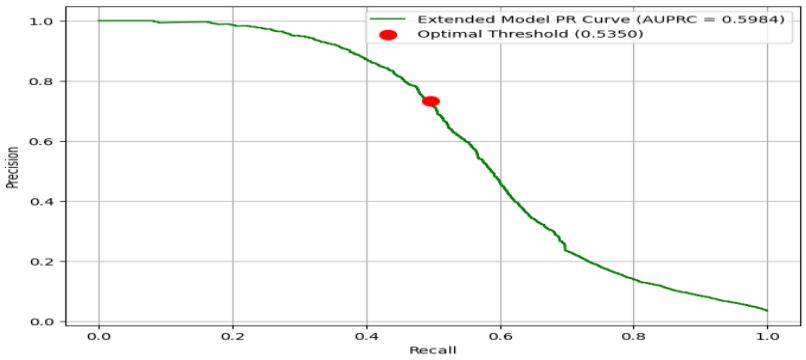


Figure 3: Precision-Recall Curve of the Model

In high-velocity payment infrastructures, false-positive alerts impose severe operational burdens, causing investigator fatigue among compliance teams and reputational damage from wrongly declined legitimate transactions. By achieving a Precision of 0.7320 on an unaugmented test distribution, nearly three out of every four flagged transactions represent genuine illicit activity, substantially reducing false alarms while capturing approximately half (49.56%) of all cross-domain fraud instances under strict zero-leakage conditions. This trade-off validates the efficacy of coupling training-only SMOTE balancing (sampling_strategy = 0.5) with the Sparrow Search Algorithm (SSA) and dynamic Focal Loss ($\alpha = 0.1, \gamma = 5.0$), which penalizes false negatives and prevents the network from collapsing into majority-class predictions.

Explainability Audit: Spatial and Temporal Feature Attribution

To resolve the algorithmic opacity of deep neural networks and satisfy regulatory compliance mandates, shap.GradientExplainer was executed on true-positive flagged transactions from the test set against a reference background of 100 training samples. Figure 4 and Figure 5 present the SHAP summary distributions for the top 10 most influential spatial (20 × 20 IGTD grid) and temporal (PCA-reduced sequence) features, respectively.

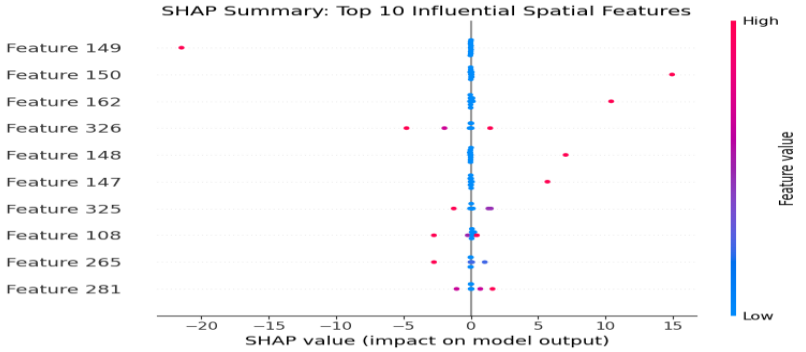


Figure 4: SHAP Summary Indicating the Most Influential Spatial (2D Mapped) Features

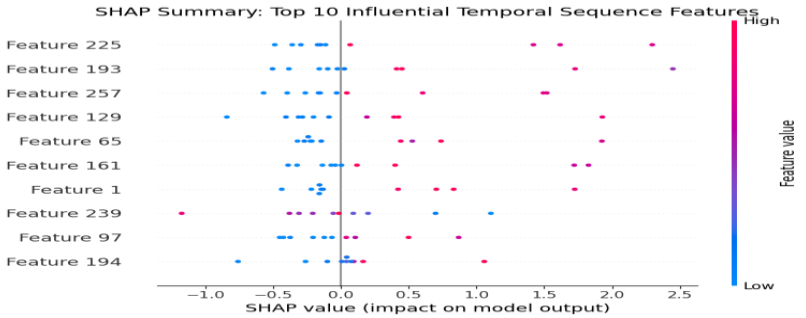


Figure 5: SHAP Summary Indicating the Most Influential Temporal (Sequence) Features

As shown in Figure 4, spatial attribution across the flattened 400-element IGTD grid (Features 0 – 399, where Feature k corresponds to grid row floor(k/20) and column k mod 20) identified Feature 149 (row 7, col 9), Feature 150 (row 7, col 10), Feature 162 (row 8, col 2), Feature 326 (row 16, col 6), Feature 148 (row 7, col 8), Feature 147 (row 7, col 7), Feature

325 (row 16, col 5), Feature 108, Feature 265, and Feature 281 as the primary spatial predictors, with SHAP values spanning approximately -21 to +15. High feature values (red points) in contiguous grid coordinates such as Features 147–150 (row 7, columns 7–10) and Features 325–326 (row 16, columns 5–6) exerted strong attributions on the model output, confirming that the IGTD correlation mapping placed co-varying financial attributes into adjacent 2D neighborhoods extracted by the 3 x 3 convolutional kernels.

Concurrently, Figure 5 shows that within the flattened 320-element temporal sequence tensor (Features 0–319, where Feature m corresponds to sequence time step t = floor(m / 32) + 1 and principal component PCA_(m mod 32) + 1), the Transformer branch relied heavily on Feature 225 (t = 8, PCA_2), Feature 193 (t = 7, PCA_2), Feature 257 (t = 9, PCA_2), Feature 129 (t = 5, PCA_2), Feature 65 (t = 3, PCA_2), Feature 161 (t = 6, PCA_2), Feature 1 (t = 1, PCA_2), Feature 239, Feature 97, and Feature 194, with SHAP values ranging from -1.2 to +2.5. Notably, features 1, 65, 129, 161, 193, 225, and 257 all represent the same second principal component (PCA_2) tracked across sequential time steps t = 1, 3, 5, 6, 7, 8, and 9, demonstrating that the Transformer self-attention mechanism tracked trajectories of the same latent behavioral component across the rolling window. However, two caveats must be noted: first, SHAP values quantify associative contributions to the model's log-odds rather than causal determinants of fraud; second, for practical deployment by compliance analysts, transformed IGTD and PCA indices must be automatically projected back to their human-readable domain variable names (e.g., transaction dollar amount, velocity delta, or gas price) in the user interface.

Comparative Discussion and Architectural Implications

The empirical findings of this study offer insights into multimodal feature transformation and cost-sensitive learning for financial anomaly detection, while also highlighting several important methodological and operational limitations that define directions for further work.

Regarding feature engineering, mapping 1D records into 20 x 20 IGTD spatial grids alongside 10 x 32 PCA rolling sequences enabled the dual-branch CNN–Transformer architecture to simultaneously process localized cross-feature clusters and multi-step temporal trajectories under a leakage-free chronological split (Darwish et al., 2026; Fnu et al., 2026; Gong et al., 2026). Furthermore, pairing training-only SMOTE (sampling_strategy = 0.5) with SSA threshold optimization and Focal Loss (alpha = 0.1, gamma = 5.0) achieved a Precision of 0.7320 and an AUPRC of 0.5984 on the imbalanced test set, corroborating prior evidence on the utility of cost-sensitive metaheuristic optimization (Al-Hchaimi et al., 2026; Arzu et al., 2026).

Nevertheless, several limitations must be considered when interpreting these results:

Need for Baseline and Ablation Benchmarks

The present study evaluated the integrated Hybrid CNN–Transformer pipeline as a unified system without reporting side-by-side experimental benchmarks against established tabular classifiers (such as logistic regression, random forests, or gradient-boosted decision trees like XGBoost and LightGBM) or standalone deep learning branches (isolated 2D CNN or isolated Transformer). Similarly, formal ablation experiments isolating the marginal contribution of IGTD mapping, PCA sequence length, SMOTE oversampling, Focal Loss, and SSA tuning are required to quantify whether the architectural complexity of spatial-temporal fusion outperforms well-tuned gradient-boosted trees operating directly on raw tabular features.

Cross-Domain Evaluation and Cryptocurrency Pseudo-Labels

Performance metrics were reported across the combined test partition rather than disaggregated separately for the IEEE-CIS and Ethereum subsets. Because the Ethereum records lacked verified ground-truth fraud labels and relied on unsupervised Isolation Forest pseudo-labels (at a 5% assumed contamination rate), predictive agreement on the cryptocurrency subset reflects anomaly detection rather than verified fraud identification. Future evaluations should disaggregate domain-specific metrics and validate the cryptocurrency branch on externally verified datasets (such as the Elliptic Bitcoin dataset or confirmed Etherscan phishing/exploit registries) across multiple contamination thresholds.

Inference Latency and Real-Time Suitability

Although compact CNN and Transformer branches (comprising ~318k parameters) are structurally light, transforming incoming transactions via IGTD and computing post-hoc shap.GradientExplainer attributions introduce non-trivial computational overhead. Because end-to-end inference latency, system throughput (transactions per second), and memory utilization were not benchmarked under production load, real-time operational suitability cannot yet be claimed.

Blockchain Subsystem Verification and Regulatory Privacy

In this work, the off-chain IPFS storage, ZKP state verification, and lightweight blockchain logging layer were formulated as a conceptual architecture rather than benchmarked for block confirmation latency, gas expenditure, or ZKP proving/verification time (Al-Jaziri et al., 2026; Allehaibi, 2026; Saragih et al., 2026). Moreover, storing transaction payloads off-chain on IPFS and anchoring CIDs on-chain does not automatically satisfy data-protection regulations such as the GDPR (Davuluri, 2026; Maligireddy, 2026; Naz et al., 2026). Because public IPFS content remains retrievable by any node holding the CID and on-chain hashes are permanent, true regulatory compliance requires authenticated access controls, client-side payload encryption, robust key management, and verifiable off-chain garbage collection across all pinning nodes.

Conclusion

This study proposed and evaluated an explainable hybrid CNN–Transformer framework for fraud detection using conventional digital payment and Ethereum-related data. The approach combined IGTD-based spatial feature representation, PCA-reduced temporal sequences, cost-sensitive learning and SHAP-based feature attribution. Under the reported chronological evaluation procedure, the model achieved a precision of 0.7320, recall of 0.4956, F1-score of 0.5910 and AUPRC of 0.5984 at a decision threshold of 0.5350. These results indicate moderate predictive performance, with an important trade-off between precision and the proportion of positive cases detected. The study also described a blockchain-, IPFS- and ZKP-based architecture intended to support auditability. However, further evidence is required to establish the comparative advantages of the hybrid model and the operational performance of the proposed audit subsystem. In particular, independent validation of cryptocurrency fraud labels, baseline and ablation experiments, cross-domain performance assessment, inference-latency measurements and implementation-level

security evaluation are necessary before the framework can be recommended for real-world financial fraud prevention.

REFERENCES

Al-Hchaimi, A. A. J., Khalifa, M. A., & El-Shafai, W. (2026). Explainable AI with imbalanced learning strategies for blockchain transaction fraud detection. *Engineering Reports*, *8*(1), e70545. https://doi.org/10.1002/eng2.70545

Al-Jaziri, S. A., Alqaryouti, O., & Almi'ani, K. (2026). SkillChain DX: A policy framework for AI-driven talent mapping and blockchain-based credential validation in Dubai Government. *Applied Sciences*, *16*(4), 2114. https://doi.org/10.3390/app16042114

Alessi, G., & Fugini, M. (2026). Adaptive real-time financial fraud detection with explainable AI tools. *Digital Threats: Research and Practice*. https://doi.org/10.1145/3794859

Allehaibi, K. H. (2026). Leveraging the variational autoencoder with the blockchain smart contracts model for strengthening fraud detection in financial sectors. *Engineering, Technology & Applied Science Research*, *16*(1), 32712–32717. https://doi.org/10.48084/etasr.13371

Ampumuza, D., Katushabe, C., & Tamale, M. (2026). A systematic review and future directions for AI-driven detection of fraud patterns in SACCO transactions. *Frontiers in Artificial Intelligence*, *8*, 1690482. https://doi.org/10.3389/frai.2025.1690482

Arzu, F., Gul, L., Waheed, A., Amin, M. U., Qadeem, S. E. Y., Alam, F., Ashraf, M., & Ali, A. (2026). An advanced AI-empowered fintech framework for credit card fraud detection in online transactions via Sparrow Search Algorithm–based adaptive optimization. *Spectrum of Engineering Sciences*, *4*(1), 504–528. https://thesesjournal.com/index.php/1/article/view/1906

Darwish, S. M., EL-Naggar, S., & Elkaffas, S. M. (2026). Securing financial transactions: Exploring the role of lightweight blockchain-enabled deep learning for fraud detection in FinTech systems. *Cybersecurity*, *9*(1), 8. https://doi.org/10.1186/s42400-025-00436-8

Davuluri, P. S. L. N. (2026). Autonomous compliance systems: AI, event streaming, and the future of financial crime prevention. *Journal of Informatics Education and Research*, *6*(1), 1288. https://jier.org/index.php/journal/article/view/4388

Fardous, M. (2026). AI-based revenue leakage detection models using transaction-level financial data: A review. *International Journal of Scientific Interdisciplinary Research*, *7*(1), 37–71. https://ijsir.org/index.php/IJSIR/article/view/77

Fnu, H., Mirza, M. H., Marri, M. R., Singh, S., Kumar, N., Adol, D. M. D., & Bharany, S. (2026). Blockchain-assisted Transformer CNN framework with optimal feature selection for real-time digital payment fraud detection. *International Journal of Computational Intelligence Systems*, *19*(1).

Gong, Y., Hu, P., Zhang, Z., Liu, P., Li, Z., Zhang, R., & Li, M. (2026). Data-centric generative and adaptive detection framework for abnormal transaction prediction. *Electronics*, *15*(3), 633. https://doi.org/10.3390/electronics15030633

Khan, F. (2026). AI-enabled financial accuracy models for improving error detection and reporting integrity in corporate accounting systems. *American Journal of Interdisciplinary Studies*, *7*(1), 141–176. https://ajisresearch.com/index.php/ajis/article/view/91

Lazea, G.-I., Lungu, C., & Bunget, O.-C. (2026). The role of AI in revolutionising cryptocurrency trading. *Electronics*, *15*(4), 742. https://doi.org/10.3390/electronics15040742

Maligireddy, J. R. (2026). Agentic AI governance framework for real-time fraud detection in digital payment systems: A multi-layered architecture for financial security. *Authorea Preprints*. https://doi.org/10.22541/au.177145217.78578198/v1

Mallikarjuna, B., Setty, P. C., & Goundar, S. (2026). *Integration of AI and ML in cryptoeconomics*. IntechOpen. https://www.intechopen.com/chapters/1240367

Mohammed, H. S., Sallow, Z. B., & Zangana, H. M. (2026). AI-driven fraud detection in digital banking: A hybrid approach using deep learning and anomaly detection. *Sistemasi: Jurnal Sistem Informasi*, *15*(1), 209–219. https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/5757

Naz, S., Razzaq, H., Razzaq, A. B., Noman, M., Ali, M., Malik, J. T., Omer, M., Laghari, Z. A., & Zia, L. (2026). Truth in the machine: A unified framework for financial integrity, legal rights, and professional ethics in the age of AI and cryptocurrency. *Scholar Journal of Economics and Business Management*, *13*(2), 45–62. https://doi.org/10.36347/sjebm.2026.v13i02.003

Ofosu-Ampong, K., Ansong, E., Akandekumtiim, J.-M., Asmah, A., Commey, N. O., Amankwaa, T., & Tsedzah, V. (2026). How AI-enabled security features shape payment security perceptions and trust in digital payment systems. *Telematics and Informatics Reports*, *21*, 100298. https://doi.org/10.1016/j.teler.2026.100298

Saragih, H., Manurung, J., Sihotang, H. T., & Putra, I. M. A. P. (2026). Blockchain-enhanced security framework for defense supply chain management: An AI-driven smart contract approach with distributed ledger technology. *Journal of Defense Technology and Engineering*, *1*(2), 145–158. https://research.idu.ac.id/index.php/jdte/article/view/60

Yang, H., Shukur, Z., & Sahran, S. (2026). A review of artificial intelligence for financial fraud detection. *Applied Sciences*, *16*(4), 1931. https://doi.org/10.3390/app16041931

©2026 This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license viewed via https://creativecommons.org/licenses/by/4.0/ which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited appropriately.