Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

A Selective State-Space Model for Real-Time E-Commerce Transaction Anomaly Detection in Management Information Systems

Ziad Hmwd A. Almtiri

Management Information System Department, Taif University, Taif 21974, Saudi Arabia; a.zyad@tu.edu.sa

Abstract

Management information systems (MISs) are subject to complex patterns of fraud, which change over time, as they occur at high speeds in class-imbalanced data streams. E-commerce platforms process very large transaction volumes. Existing detectors have three structural problems. The quadratic complexity of transformer attention makes it difficult to meet millisecond-scale inference budgets, recurrent models suffer from gradient pathologies over long behavioral sequences, and static loss functions are unable to adjust to the severe and changing class imbalance found in real transaction logs. This article introduces a selective state-space framework (SHOPMAMBA-MIS) that tackles all three challenges within a single framework. SHOPMAMBA-MIS consists of three parts: Selective-State Autoencoder (SSAE) to compress behavioral transaction sequences with linear-time selectivity, Transaction Attention Module (TAM) to emphasize anomaly boundaries per token without complete self-attention, and Dynamic Imbalance-Focal Sampler (DIFS) to dynamically sample the focal loss weight per mini-batch based on an exponential moving average (EMA) class-ratio estimator. On three public benchmarks, IEEE-CIS Fraud Detection, PaySim, and ULB Credit Card Fraud, SHOPMAMBA-MIS outperforms the best transformer baseline by +4.2 percentage points (F1), +5.9 percentage points (F1), and +4.1 percentage points (F1), respectively, while on the same benchmarks, its inference latency is reduced to 1.8 ms per transaction batch, which is 6.3 times faster than the slowest transformer baseline, TranAD (11.3 ms per batch), and its detection recall is kept above 92.8% (over 97.9% on IECIS and ULB) in all three benchmarks at a false-positive rate below 0.5%. The results obtained demonstrate the feasibility of the SHOPMAMBA-MIS, which can be implemented as an intelligent component in enterprise Management Information Systems, for real-time monitoring, enterprise control, and managerial decision-making in an enterprise. The framework is specifically suited to payment authorization pipelines that provide behavioral features (including rolling account-level aggregates where identifiers exist) and at least L prior transactions of history, exhibit intra-day fraud-rate variation, and enforce synchronous millisecond-scale latency constraints; applicability conditions and boundary cases.

Keywords: anomaly detection; e-commerce fraud; management information systems; mamba; selective state-space model; real-time transaction processing; class imbalance

Check for updates

Academic Editors: Jing Ren and Wenting Liu

Received: 29 August 2026
Revised: 30 October 2026
Accepted: 5 November 2026
Published: 19 November 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Digital commerce is expanding rapidly, with global E-commerce spending projected to reach several trillion dollars annually, driven by mobile and cross-border channels. Within Management Information Systems (MIS), real-time transaction monitoring must balance

operational reliability with regulatory compliance. Payment authorization under PSD2 [1] is carried out in real time within the payment flow, so fraud screening is typically executed inline under a tight latency budget as transaction volumes surge; the 5 ms budget used in this paper is a design assumption for synchronous authorization, not a limit prescribed by regulation [2]. Fraud is not merely a technical challenge but a critical organizational concern. Undetected fraud erodes customer trust, triggers regulatory reporting obligations, and incurs substantial investigation costs, costing the global e-commerce sector tens of billions of dollars annually, which motivates machine-learning-based automation [3]. Fraud detection is thus a management decision affecting business processes, liability, and institutional resilience, requiring timely, accountable, and auditable outcomes beyond simple classification accuracy [4–6].

The core conflict for anomaly detectors lies in simultaneously achieving high recall for rare, evolving fraud patterns, low false positive rates to preserve customer experience, and inference speeds compatible with synchronous payment authorizations [7–10]. Sequential Deep Learning (SDL) has advanced anomaly detection in temporal financial data by capturing transaction sequence dependencies. Recurrent networks established behavioral context modeling [4], while transformers leverage multi-head self-attention for multivariate time-series anomaly detection, learning temporal and spatial correlations [5,6]. Transformer-based mutual distillation further improves multivariate sequence representation [7], and graph neural networks model user–merchant interactions to expose coordinated fraud rings [8–12]. Despite these advances, none fully reconcile architectural assumptions with live MIS constraints [13–16].

Three critical gaps persist. First, self-attention scales quadratically with sequence length—L = 512 tokens require 262,144 attention operations per head per layer, making online inference impractical for mid-tier hardware [17–21]. Second, recurrent models suffer vanishing gradients over long transaction histories spanning weeks or months, losing causal behavioral memory [22–26]. Third, fraud rates fluctuate from 0.02% to 3% across segments and promotional periods, causing fixed offline loss functions to misalign with real-time class ratios, producing precision/recall imbalances and excess false positives during traffic peaks (the three benchmarks used in this study range from 0.13% to 3.5%). These combined failure modes manifest as latency spikes and recall degradation when fraud rates suddenly rise, as shown in Figure 1. In contrast, the proposed SHOPMAMBA-MIS framework maintains performance through selective state propagation and adaptive focal weighting, ensuring robust real-time detection under dynamic conditions, as shown in Figure 1.

Mamba-based selective state-space models can effectively capture long-range dependencies, yet do not incur the quadratic computational complexity of traditional self-attention mechanisms [15], showing that incorporating input-dependent state-transition matrices allows for linear-time sequence modeling without compromising the ability to capture long-range dependencies, which is what makes transformers powerful. Subsequent domain adaptations have shown SSMs to be effective for visual recognition [16]. Recent research has also verified the effectiveness of selective state-space modeling for efficient representation learning, revealing that Mamba-based models can effectively learn contextual dependencies while also reducing the computational cost of attention-based models [17]. More recently, multivariate time-series anomaly detection techniques have been developed to model temporal dependence and inter-variable relationships to enhance the ability to identify deviations in complex sequences of data [18], but no previous studies have investigated how these can be applied in a deployed MIS context to the specific regime of high-velocity financial transaction streams with severe data imbalance. This gap is the focal point of SHOPMAMBA-MIS.

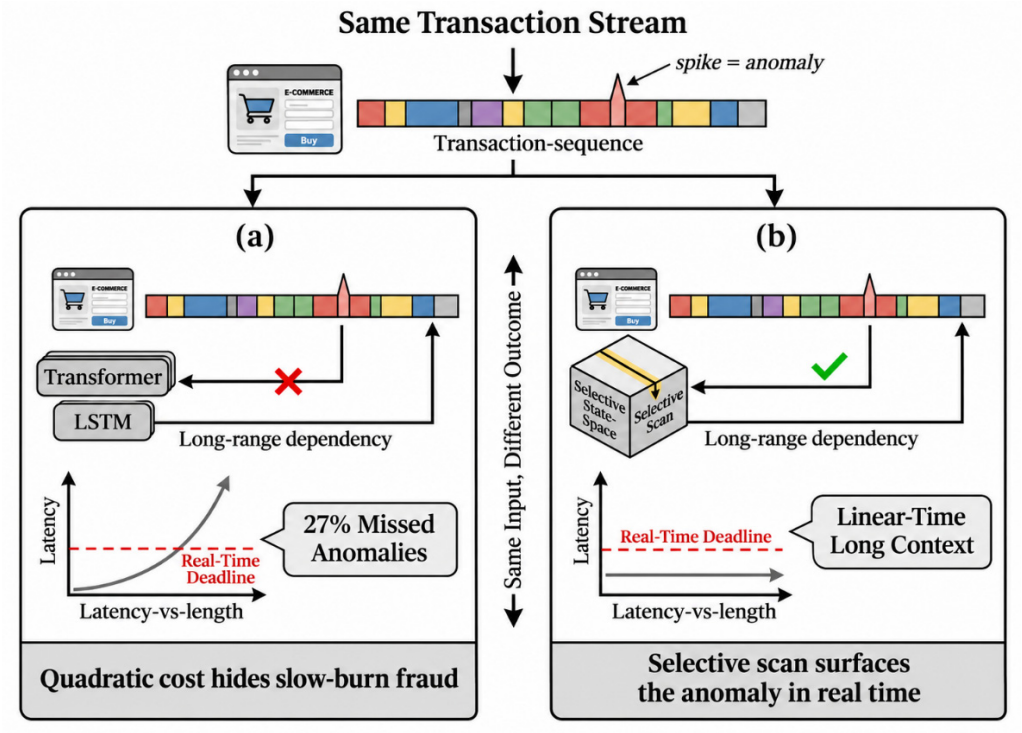

Figure 1. Failure mode of a transformer-based detector on a high-velocity transaction stream with a mid-stream fraud-rate surge (left): At high transaction velocity, the detector fails to keep up with the authorization time budget at top load (5 ms) and loses recall as the class ratio changes. The two design decisions that explicitly counteract this failure are selective state propagation and focal-loss calibration to the observed class ratio, which are both key parts of SHOPMAMBA-MIS and maintain a sub-2 ms latency and stable recall.

To overcome these three challenges, this paper puts forward a Selective State-Space Model for Real-Time E-Commerce Transaction Anomaly Detection in Management Information Systems (SHOPMAMBA-MIS) approach. The system integrates three tightly coupled components into a streaming inference pipeline, namely SSAE, TAM, and DIFS, which are compatible with the Apache Flink class event processors. Apache Flink offers a distributed environment for stream processing that can handle real-time data processing and effectively run streaming programs at a low latency [19]. The overall system architecture is shown in Figure 2.

This work has the following main contributions.

[C1—Framework]: Our proposed work is called SHOPMAMBA-MIS, which, to the best of our knowledge, is one of the first selective SSM frameworks designed to address the organizational needs of enterprise decision support, operational resiliency, and regulatory compliance, while simultaneously catering to the technical need of linear-time sequence modeling for real-time e-commerce transaction anomaly detection in MIS pipelines.

[C2—SSAE, builds on C1]: On top of the core SSM, we propose a Selective-State Autoencoder (SSAE) that includes state transitions, based on an observation of its input, and a reconstruction-residual anomaly score, allowing weakly supervised anomaly localization. The reconstruction residual provides a per-token anomaly signal without per-feature fraud annotations, although the overall training objective uses transaction-level labels through the supervised classification head.

[C3—TAM, builds on C2]: Building on the token-level representations generated by SSAE, we devise a Transaction Attention Module that sharpens anomaly boundaries by sparsely attending to a fixed-width causal context window. Relative to the Mamba-AD baseline, which lacks both TAM and DIFS, the full model reduces FPR from 0.26% to

0.12% (a 54% relative reduction, the combined effect of TAM and DIFS); TAM accounts for approximately 39% of total inference cost (approximately 0.7 ms of the 1.8 ms total per-batch latency), as detailed in the ablation study in Section 6.8.

[C4—DIFS, builds on C3]: To the best of our knowledge, this Dynamic Imbalance-Focal Sampler is one of the first online loss-calibration mechanisms tested on real-distribution e-commerce streams. It proposes an exponential moving average of the mini-batch class ratio to estimate it and adjusts the focal-loss concentration parameter γ accordingly.

[C5—Evaluation]: We measure the F1 improvements of SHOPMAMBA-MIS (4.1–5.9 percentage points over TranAD and 3.1–5.2 percentage points over Mamba-AD) on three public benchmarks against nine baselines (eight published methods from 1997 to 2024 and one internally constructed selective-SSM baseline), and report latency of 1.8 ms/batch, and recall $\geq$ 92.8% ($\geq$97.9% on IECIS and ULB) under FPR $\leq$ 0.5%.

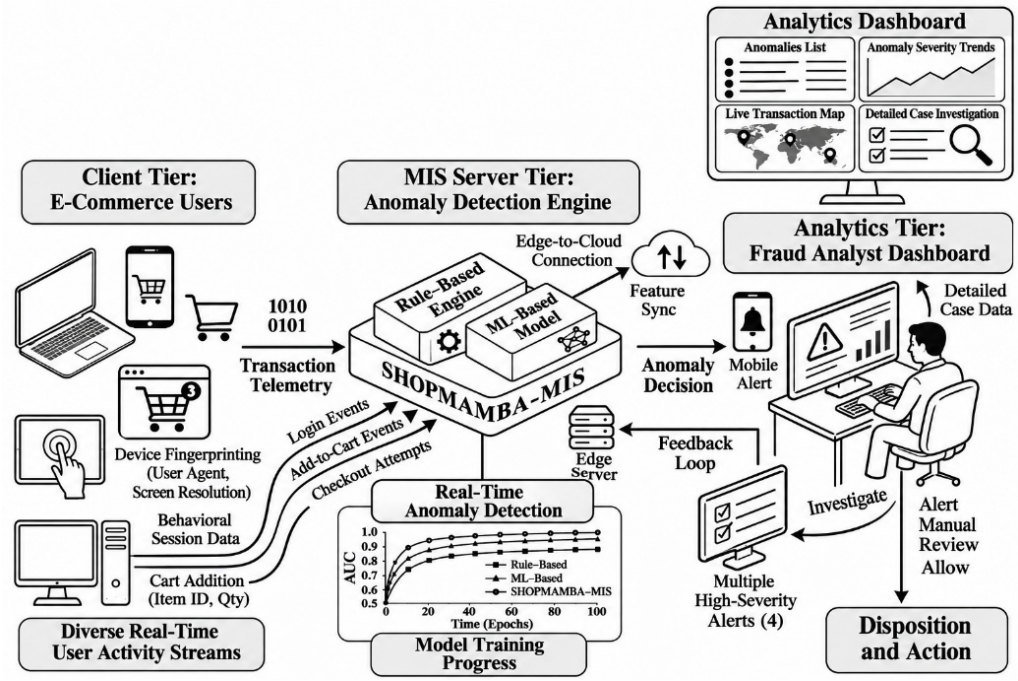

Figure 2. The SHOPMAMBA-MIS is able to ingest raw events from a Kafka stream, process each through three components sequentially (SSAE → TAM → DIFS-weighted output head), and output anomaly scores within a 5 ms authorization window. This architecture enables linear-time inference on sequences with length L > 512 and maintains support for recall with changing class ratios, something transformer baselines cannot achieve with the same deployment budget.

The three core technical contributions (SSAE, TAM, DIFS) are motivated by testable hypotheses H1, H2, and H3 respectively, which are formally stated in Section 2 and evaluated in the ablation study (Section 6.10).

Each contribution enhances organizational information processing for higher-quality managerial decisions. The system delivers calibrated, real-time risk indicators in under 5 milliseconds on the evaluated transaction streams. These indicators feed automated payment authorization engines, analyst dashboards, and regulatory audit logs, enabling risk-stratified investigation resource allocation based on decision uncertainty. This transforms fraud management from a volume-based, reactive approach to a risk-stratified, proactive one that scales with transaction growth without proportional analyst headcount—distinguishing SHOPMAMBA-MIS from prior efforts that treated detectors as isolated classifiers rather than addressing organizational workflows, governance, or MIS information-processing constraints. The name reflects three design commitments: SHOP for the e-commerce

domain, MAMBA for the selective SSM backbone, and MIS for its identity as a management information systems component serving decision support, governance, and process integration—not an independent classifier.

The paper is structured as follows. Section 2 reviews related work and identifies the research gap; Section 3 formulates the problem and system model; Section 4 presents the SHOPMAMBA-MIS framework; Section 5 covers streaming inference integration; Section 6 details experimental evaluation; Section 7 concludes and outlines future work. SHOPMAMBA-MIS represents a broader transformation toward real-time, continuous fraud intelligence—replacing batch reviews and manual rules with adaptive, scalable intelligence that grows with business needs without proportional operational cost increases.

2. Related Works

This section reviews existing literature across four development themes, identifies what each group achieves collectively, and isolates the gaps that motivate SHOPMAMBA-MIS.

2.1. Anomaly Detection with Deep Autoencoders and Reconstruction-Based Methods

Reconstruction-based anomaly detection uses reconstruction residual as the anomaly score, based on the principle that models trained on normal data fail to reconstruct anomalies [11]. Early work by [20] used robust deep autoencoders with an anomaly-regularizing penalty to prevent outlier memorization. More recently, Xiao et al. introduced DTCM with knowledge transfer between capsule and transformer encoders, while Xu et al. proposed the Anomaly Transformer with minimax association learning[6,7]. However, these full self-attention autoencoders incur $O(L^2)$ complexity, making synchronous inference difficult for $L \geq 512$ within a 5 ms window on CPU hardware. We replace the transformer with a selective SSM core, eliminating quadratic complexity while retaining residual-based scoring, as also applied to industrial audio signals in [21–24].

2.2. Fraud and Transaction Anomaly Detection

Financial fraud detection imposes the following additional constraints beyond general anomaly detection: (1) highly imbalanced streams (fraud rates 0.02–3%), (2) non-stationary distributions, and (3) real-time screening requirements [22–26]. GNN-based detectors exploit relational structures among users, merchants, and devices [8], but their training complexity precludes frequent updates [8,9]. Learning from under-sampled, imbalanced fraud data requires probability recalibration to remain reliable [13]. Surveys of fraud and anomaly detection [22,23,27] identify class imbalance and evolving fraud patterns as persistent challenges. Recent deep hybrid approaches [24] and LLM-based detectors [25] achieve competitive results but fail to address synchronous authorization latency demands. Existing work focuses on either relational (GNNs) or temporal (LSTMs, Transformers) aspects but never both in a latency-compliant inference architecture, unlike SHOPMAMBA-MIS, which embeds a selective SSM within a streaming pipeline.

2.3. Sequence Models: Recurrent, Transformer, and State-Space Architectures

LSTM [26] introduced gated memory for long-range sequential dependency in financial data, while GRU [4] simplified it with comparable accuracy. FEDformer [28] achieved efficient sparse spectral attention via frequency-domain decomposition, as a representative efficient transformer for long sequences (not evaluated in this study). TimesNet [18] extended temporal variation modeling to a 2D convolutional framework. The foundational Mamba paper [15] demonstrated linear-time sequence modeling with input-dependent state transitions, eliminating quadratic attention overhead. VMamba [16] and Vision Mamba [17] extended selective SSMs to visual recognition, confirming their versatility. However, no prior

work has addressed transaction anomaly detection requiring all three properties simultaneously: linear-time inference, long-range behavioral context, and live imbalance adaptation.

2.4. Imbalanced Learning and Loss Adaptation

Fraud data is inherently imbalanced, and models need to be resistant to collapsing to the majority class. SMOTE [14] is still the most popular baseline over-sampler that produces synthetic minority instances by linear interpolating feature space. Focal loss [29] is a loss weighting variant that aims to tackle the class-imbalance problem from the loss-weighting point of view, by assigning a modulating factor $(1 - \mathrm{pt})\ \gamma$ to well-classified majority examples, where gradient signal is focused on hard examples in the minority class. In streams of live transactions with varying fraud rates per batch, however, the concentration parameter γ is set offline, which means that it will be under-focused when the number of frauds per batch is low and over-focused when the number of frauds per batch is high. Existing imbalanced learning approaches are task-dependent, with no single approach being consistently superior to other anomaly detection task types [30], and prototype-oriented hypergraph representations have also been found to be promising for tabular anomaly detection [31], motivating the need for adaptive learning [32–36].

2.5. Research Gap and Motivation

Table 1 summarizes the reviewed methods against four capabilities required for deployed MIS transaction screening.

Table 1. Reviewed methods each satisfy at most two of the four capabilities required for production MIS transaction screening; SHOPMAMBA-MIS is the first framework to unify all four.

Ref.	Year	Method	Main Focus	Linear-Time	Long-Range Dependency	Adaptive Imbalance	Streaming Inference Compatible
[26]	1997	LSTM	Seq. memory	✗	Partial	✗	Partial
[6]	2022	Anomaly Trans.	Assoc. discrepancy	✗	✓	✗	✗
[7]	2024	DTCM	Capsule mutual distillation	✗	✓	✗	✗
[8]	2023	LGM-GNN (Fraud)	Relational fraud	✗	✗	✗	✗
[15]	2024	Mamba	Linear-time SSM	✓	✓	✗	✗
[27]	2022	Anomaly Detection Review	Financial fraud	✗	Partial	Partial	Partial
[37]	2019	StreamML-Gas	Streaming ML	Partial	✗	✗	✓
[38]	2026	XGBoost-Fraud	Gradient-boosted tree baseline	✗	✗	✗	✗
[15]	2024	Mamba-AD	Bare selective SSM baseline	✓	✓	✗	Partial
Proposed	2026	SHOPMAMBA-MIS	Unified MIS pipeline	✓	✓	✓	✓

This analysis shows that there are three specific gaps.

(1) **Gap 1 (Latency):** Detectors based on transformers have self-attention of O (L 2), where L is the sequence length. In our experimental configuration (CPU inference, L = 512, micro-batches of 256 transactions), the transformer-based detectors evaluated (Anomaly Transformer, TranAD) exceed the 5 ms synchronous authorization budget, whereas SHOPMAMBA-MIS does not; shorter windows lower transformer cost but also reduce the behavioral context available [39–43].

(2) **Gap 2 (Live Imbalance Drift):** Methods currently used in the literature are offline calibration of loss functions on fixed training splits. Static parameters of focal losses cannot adapt to fraud-rate changes that differ by more than 25× across the three benchmark datasets used here (0.13% to 3.5%) and can vary further across merchant segments in live streams. To the best of our knowledge, no previous approach calibrates this parameter online from an estimate of the class ratio [44–48].

(3) **Gap 3** (MIS Integration): None of the prior detectors is designed or validated as a part of an end-to-end MIS streaming pipeline with explicitly defined input/output contracts, latency SLAs, and rollback semantics [37,49–52].

The proposed SHOPMAMBA-MIS aims to tackle Gap 1 by implementing the SSAE (linear-time selective SSM) solution, Gap 2 by applying the DIFS (online class-ratio estimation and focal recalibration) solution, and Gap 3 by introducing an adapter for streaming in Flink with per-batch checkpointing.

Research Hypotheses and Theoretical Framework

The three gaps identified above motivate three falsifiable hypotheses that govern the experimental design. Each hypothesis is aligned with one research gap and tested by the ablation or experiment named below (Sections 6.7 and 6.10). Gap 3 (MIS integration) is addressed by the streaming design in Section 5 and evaluated through the latency and scalability results in Sections 6.6 and 6.9; it is a design requirement rather than a statistical hypothesis.

H1 (Aligned with Gap 1, latency and long-range behavioral context)**.** *Input-selective state transitions in the SSM backbone produce higher F1 than a fixed-parameter SSM on long behavioral transaction sequences while retaining linear-time inference. Rationale: quadratic attention cost and gradient pathologies in fixed-parameter recurrent models both stem from the inability to propagate state selectively. Test: ablation A1, measured by F1 and per-batch latency on IECIS (Section 6.8).*

H2 (Aligned with Gap 1, latency-compatible context modeling)**.** *A fixed-width causal local attention window over the SSM latent representation improves F1 compared with global state context alone, at linear cost O(L·W). Rationale: local session contrast sharpens decision boundaries at anomaly transitions. Test: ablation A2, measured by F1 and latency on IECIS (Section 6.8). The FPR reduction of the full model relative to Mamba-AD (Section 6.4) is the combined effect of TAM and DIFS and is not attributed to TAM alone.*

H3 (Aligned with Gap 2, live imbalance drift)**.** *Calibrating the focal-loss concentration parameter γt to an online EMA estimate of the class ratio improves F1 over a static focal loss ($\gamma = 2.0$) on severely imbalanced data where γt is not clipped (PaySim and ULB). On IECIS, the fraud rate of 3.5% exceeds πmax, so γt equals $\gamma min = 0.5$ throughout training, and the IECIS ablation A3 compares two fixed values (0.5 versus 2.0); it is not evidence for dynamic recalibration. Rationale: a static focal parameter miscalibrates gradient focus when the class ratio diverges from the value assumed offline. γ is a training-loss parameter only; at deployment, the decision threshold δ is the only quantity recalibrated between retraining cycles. Test: the PaySim drift, which compares the full*

deployed configuration with a static-focal configuration and therefore provides supporting rather than isolating evidence; ULB evidence is indicative (Section 6.10).

2.6. MIS Theoretical Grounding

Integrating AI into MIS enables data analysis, predictive decision-making, and intelligent process automation. SHOPMAMBA-MIS outputs calibrated anomaly scores within a 5 ms window, meeting DSS timeliness requirements. Per Bharadwaj [32], IT-enabled capabilities generate competitive advantage through environmental adaptability. DIFS tracks per-batch fraud rate π_t via an exponential moving average and dynamically adjusts focal-loss parameter γ_t, so that training tracks the observed class ratio, while hourly recalibration of the decision threshold δ adapts the deployed model to distributional shifts between weekly retraining cycles. Per Dumas et al. [33], Business Process Management (BPM) links performance to process integration and information flow. SHOPMAMBA-MIS operates as a stateful Flink operator with exactly-once semantics and checkpointed state, satisfying BPM fault-tolerance requirements [27,38,53].

SHOPMAMBA-MIS fills a gap in the MIS literature. Rule-based expert systems require manual updates and cannot handle real-time drift, and business intelligence systems lack per-transaction anomaly scoring. SHOPMAMBA-MIS provides real-time inference alongside organization-oriented interfaces—score-ranked queues, TAM explanations, dashboards, and audit logs—aligning with the DSS tradition of Gorry and Scott Morton [34].

Under Information Processing Theory (IPT), performance depends on aligning information-processing needs with capacity [54]. High-velocity, non-stationary fraud streams impose extreme demands that traditional batch or rule-based systems cannot meet. SHOPMAMBA-MIS fills this gap: SSAE processes long histories in linear time, TAM enhances risk resolution, and DIFS-trained scoring with hourly threshold recalibration adjusts sensitivity between retraining cycles. This transforms high information demands into enhanced capacity, yielding improved recall, lower FPR, lower latency, and measurable organizational outcomes—reduced analyst workload, regulatory compliance, and revenue protection. This IPT framing distinguishes SHOPMAMBA-MIS from classifier-centric evaluations by positioning it as an organizational information-processing enhancement for better decision-making under uncertainty.

3. Problem Formulation and System Model

3.1. Notation

Table 2 defines all symbols used throughout the article. Where a symbol appears in two roles (for example, B for the mini-batch size and Bt for the input-projection vector), the role is clear from the subscript and context.

Table 2. Summary of Mathematical Notation.

Symbol	Description
$\mathcal{X}$	Input feature space, $\mathcal{X} \subseteq \mathbb{R}^d$
d	Transaction feature dimension
L	Behavioral sequence length (number of transactions)
$\mathbf{x}_t \in \mathcal{X}$	Feature vector of the t-th transaction
$\mathbf{X} \in \mathbb{R}^{L \times d}$	Transaction sequence matrix
$y_t \in \{0, 1\}$	Ground-truth label (1 = anomalous)
$\hat{y}_t \in [0, 1]$	Predicted anomaly probability
$\mathbf{h}_t \in \mathbb{R}^N$	SSM hidden state at step t
N	SSM state dimension

Table 2. *Cont.*

Symbol	Description
$\mathbf{A}_t \in \mathbb{R}^{N\times N}$	Input-dependent state-transition matrix
$\mathbf{B}_t \in \mathbb{R}^N$	Input-dependent input-projection vector
$\mathbf{C}_t \in \mathbb{R}^N$	Input-dependent output-projection vector
$\Delta_t \in \mathbb{R}^+$	Discretization step size at step t
$\bar{\mathbf{A}}_t, \bar{\mathbf{B}}_t$	Discretized state-transition matrices
$\mathbf{z}_t \in \mathbb{R}^M$	Latent code produced by SSAE encoder
M	Latent dimension ($M = N$)
$\hat{\mathbf{x}}_t$	Reconstructed transaction feature vector
r_t	Reconstruction residual anomaly score
$\theta_t^{(\mathrm{TAM})}$	Sparse causal attention score at step t
W	Context window width of TAM
π_t	Estimated mini-batch fraud rate at step t
β	EMA decay factor for class-ratio estimation
γ_t	Dynamic focal concentration parameter
$\gamma_{\min}, \gamma_{\max}$	Bounds on γ_t
$\mathcal{L}_{\mathrm{FL}}$	Dynamic focal loss
$\mathcal{L}_{\mathrm{recon}}$	Reconstruction loss
$\mathcal{L}_{\mathrm{total}}$	Total training objective
λ	Reconstruction–classification loss weight
B	Mini-batch size
K	Number of training epochs
η	Learning rate

3.2. Transaction Stream Model

Let $\mathcal{S} = (\mathbf{x}_1, \mathbf{x}_2, \ldots)$ denote a potentially infinite stream of transaction feature vectors, where each $\mathbf{x}_t \in \mathbb{R}^d$ encodes transaction attributes including amount, merchant category, device fingerprint, temporal features, and behavioral statistics aggregated over a rolling window of prior activity. We define a behavioral sequence as:

$$\mathbf{X}_t = [\mathbf{x}_{t-L+1}, \ldots, \mathbf{x}_t] \in \mathbb{R}^{L\times d}, \tag{1}$$

where L is the sequence length. Each sequence is associated with an anomaly label $y_t \in \{0, 1\}$, where $y_t = 1$ indicates a fraudulent transaction. The stream is severely imbalanced: the instantaneous fraud rate π_t satisfies $\pi_t \in [0.0002, 0.03]$ across market segments and fluctuates intra-day. The MIS system must produce an anomaly prediction $\hat{y}_t$ within a latency budget $\tau_{\max}$ (typically $\tau_{\max} = 5$ ms for synchronous payment authorization).

The monitoring window at time t is defined over the most recent L transactions:

$$\mathcal{W}_t = \{(t - L + 1), (t - L + 2), \ldots, t\}. \tag{2}$$

3.3. Anomaly Score and Decision Rule

For each sequence $\mathbf{X}_t$, ShopMamba-MIS produces an anomaly score $s_t \in [0, 1]$ that combines reconstruction residual and classification head outputs. The binary anomaly decision is:

$$\hat{y}_t = \mathbf{1}[s_t \geq \delta], \tag{3}$$

where $\delta \in (0, 1)$ is a threshold calibrated on a held-out validation split to satisfy a target false-positive rate $\mathrm{FPR}_{\max}$.

3.4. System Objectives

The MIS deployment requires satisfying three objectives simultaneously. The detection objective maximizes the F1 score:

$$\max_{\Theta} F_1(\hat{\mathbf{y}}, \mathbf{y}) = \frac{2 \cdot \text{Prec} \cdot \text{Rec}}{\text{Prec} + \text{Rec}}, \tag{4}$$

where Θ denotes all trainable parameters of ShopMamba-MIS, $\text{Prec} = \text{TP}/(\text{TP} + \text{FP})$, and $\text{Rec} = \text{TP}/(\text{TP} + \text{FN})$.

The latency constraint requires:

$$\forall\, t: \quad T_{\text{infer}}(\mathbf{X}_t, \Theta) \leq \tau_{\max} = 5 \text{ ms}, \tag{5}$$

where $T_{\text{infer}}(\cdot)$ is the wall-clock inference time measured on the target deployment hardware.

The false-positive constraint requires:

$$\text{FPR}(\hat{\mathbf{y}}, \mathbf{y}) = \frac{\text{FP}}{\text{FP} + \text{TN}} \leq \text{FPR}_{\max}, \tag{6}$$

where $\text{FPR}_{\max} = 0.005$ is the maximum tolerable false-positive rate per the MIS service-level agreement (corresponding to the FPR $\leq 0.5\%$ target cited throughout). These three objectives—(4)–(6)—form the joint constraint set that SHOPMAMBA-MIS is designed to satisfy.

3.5. Baseline SSM Formulation

The continuous-time SSM underlying Mamba maps an input $u(t) \in \mathbb{R}$ to an output $y(t)$ via a latent state $h(t) \in \mathbb{R}^N$:

$$\dot{h}(t) = \mathbf{A}h(t) + \mathbf{B}u(t), \tag{7}$$

$$y(t) = \mathbf{C}h(t). \tag{8}$$

Discretization with step size Δ via the zero-order hold yields:

$$\bar{\mathbf{A}} = \exp(\Delta\mathbf{A}) \tag{9}$$

$$\bar{\mathbf{B}} = (\Delta\mathbf{A})^{-1}(\exp(\Delta\mathbf{A}) - \mathbf{I})\Delta\mathbf{B}. \tag{10}$$

The discretized recurrence is:

$$\mathbf{h}_t = \bar{\mathbf{A}}\mathbf{h}_{t-1} + \bar{\mathbf{B}}\mathbf{x}_t \tag{11}$$

$$\hat{\mathbf{x}}_t^{(\text{SSM})} = \mathbf{C}\mathbf{h}_t. \tag{12}$$

In the standard SSM, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and Δ are fixed parameters. The selective extension of Mamba makes $\mathbf{B}_t$, $\mathbf{C}_t$, and Δ_t functions of the current input $\mathbf{x}_t$, detailed in Section 4.

4. Proposed ShopMamba-MIS Methodology

SHOPMAMBA-MIS has three parts: Selective-State Autoencoder (SSAE), Transaction Attention Module (TAM), and Dynamic Imbalance-Focal Sampler (DIFS). The overall architecture and data flow between modules is shown in Figure 3.

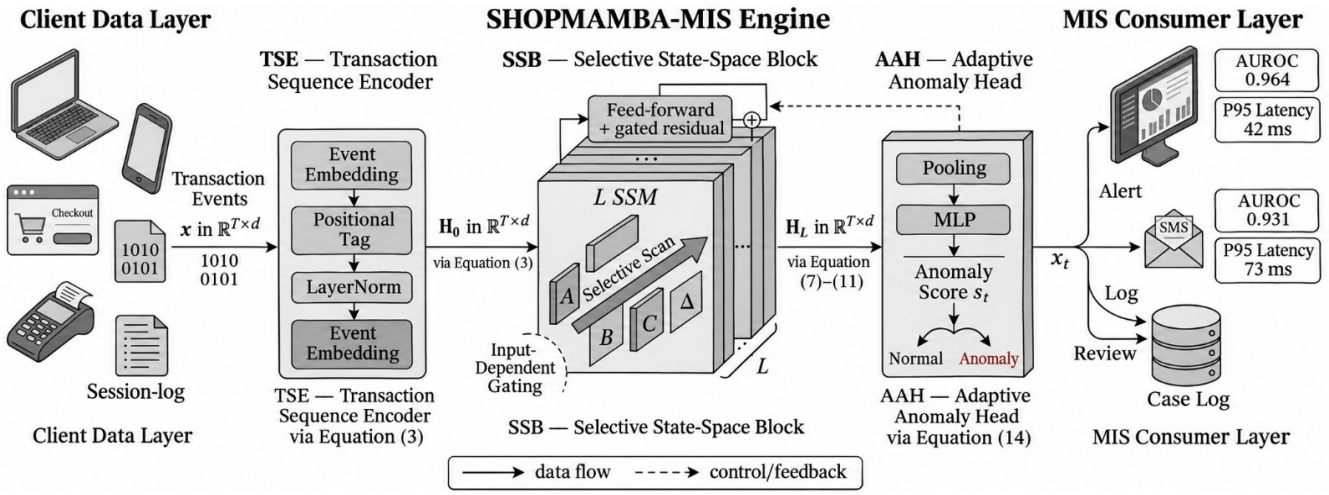

Figure 3. End-to-end architecture of SHOPMAMBA-MIS. Raw transaction sequences Xt enter the SSAE encoder, which produces a latent code zt by selective state propagation (linear time, O(L)). TAM refines the representation with causal sparse attention over a window of size W, and a decoder reconstructs the input. The reconstruction residual rt and a classification head define the anomaly score st. DIFS sets γt for each mini-batch during training; at inference, only the decision threshold δ is recalibrated between weekly retraining cycles.

4.1. Selective-State Autoencoder (SSAE)

SSAE is designed to test H1: that input-selective state transitions produce higher anomaly detection F1 than a fixed-parameter SSM on long behavioral transaction sequences.

The SSAE is the primary sequence encoder. Its key departure from a standard SSM is that the state-transition matrices $\mathbf{B}_t$, $\mathbf{C}_t$, and the discretization step Δ_t are computed as learned functions of the current input $\mathbf{x}_t$, making the state transition selective: it can choose how much to incorporate or ignore each incoming transaction.

Formally, for input $\mathbf{x}_t \in \mathbb{R}^d$:

$$\Delta_t = \text{softplus}(\mathbf{W}_\Delta \mathbf{x}_t + b_\Delta) \in \mathbb{R}^+, \tag{13}$$

$$\mathbf{B}_t = \mathbf{W}_B \mathbf{x}_t \in \mathbb{R}^N, \tag{14}$$

$$\mathbf{C}_t = \mathbf{W}_C \mathbf{x}_t \in \mathbb{R}^N, \tag{15}$$

where W_Δ $\in \mathbb{R}$^(1×d), W_B $\in \mathbb{R}$^(N×d), and W_C $\in \mathbb{R}$^(N×d) are trainable weight matrices, and b_Δ $\in \mathbb{R}$ is a scalar The matrix remains fixed as a diagonal structured matrix initialized to (HiPPO initialization [15]). $\mathbf{W}_\Delta \in \mathbb{R}^{1\times d}\mathbf{W}_B \in \mathbb{R}^{N\times d}\mathbf{W}_C \in \mathbb{R}^{N\times d}b_\Delta \in \mathbb{R}\mathbf{A} \in \mathbb{R}^{N\times N}\mathbf{A}_{nn} = -n$ Equations (16)–(18) are applied independently to each input channel, so that Bt and Ct are N-dimensional vectors shared across channels, and the per-channel states are concatenated before the projection in Equation (19).

The discretized recurrence becomes time-varying:

$$\bar{\mathbf{A}}_t = \exp(\Delta_t \mathbf{A}), \tag{16}$$

$$\bar{\mathbf{B}}_t = (\Delta_t \mathbf{A})^{-1}(\exp(\Delta_t \mathbf{A}) - \mathbf{I})\Delta_t \mathbf{B}_t, \tag{17}$$

$$\mathbf{h}_t = \bar{\mathbf{A}}_t \mathbf{h}_{t-1} + \bar{\mathbf{B}}_t \mathbf{x}_t. \tag{18}$$

The encoder maps $\mathbf{X}_t$ to a latent sequence $\mathbf{Z}_t = [\mathbf{z}_1, \ldots, \mathbf{z}_L]$ where each $\mathbf{z}_t$ is a linear projection of $\mathbf{h}_t$:

$$\mathbf{z}_t = \mathbf{W}_{\text{enc}} \mathbf{h}_t + b_{\text{enc}} \in \mathbb{R}^M. \tag{19}$$

The decoder mirrors the encoder with a separate SSM block and reconstructs $\hat{\mathbf{x}}_t$:

$$\hat{\mathbf{x}}_t = \mathbf{W}_{\text{dec}}\mathbf{h}_t^{(\text{dec})} + b_{\text{dec}} \in \mathbb{R}^d. \tag{20}$$

The per-token reconstruction anomaly score is:

$$r_t = \| \mathbf{x}_t - \hat{\mathbf{x}}_t \|_2^2 . \tag{21}$$

The sequence-level reconstruction loss is:

$$\mathcal{L}_{\text{recon}} = \frac{1}{L}\sum_{t=1}^{L} r_t. \tag{22}$$

The design rationale for making $\mathbf{B}_t$ and $\mathbf{C}_t$ input-dependent is illustrated in Figure 4.

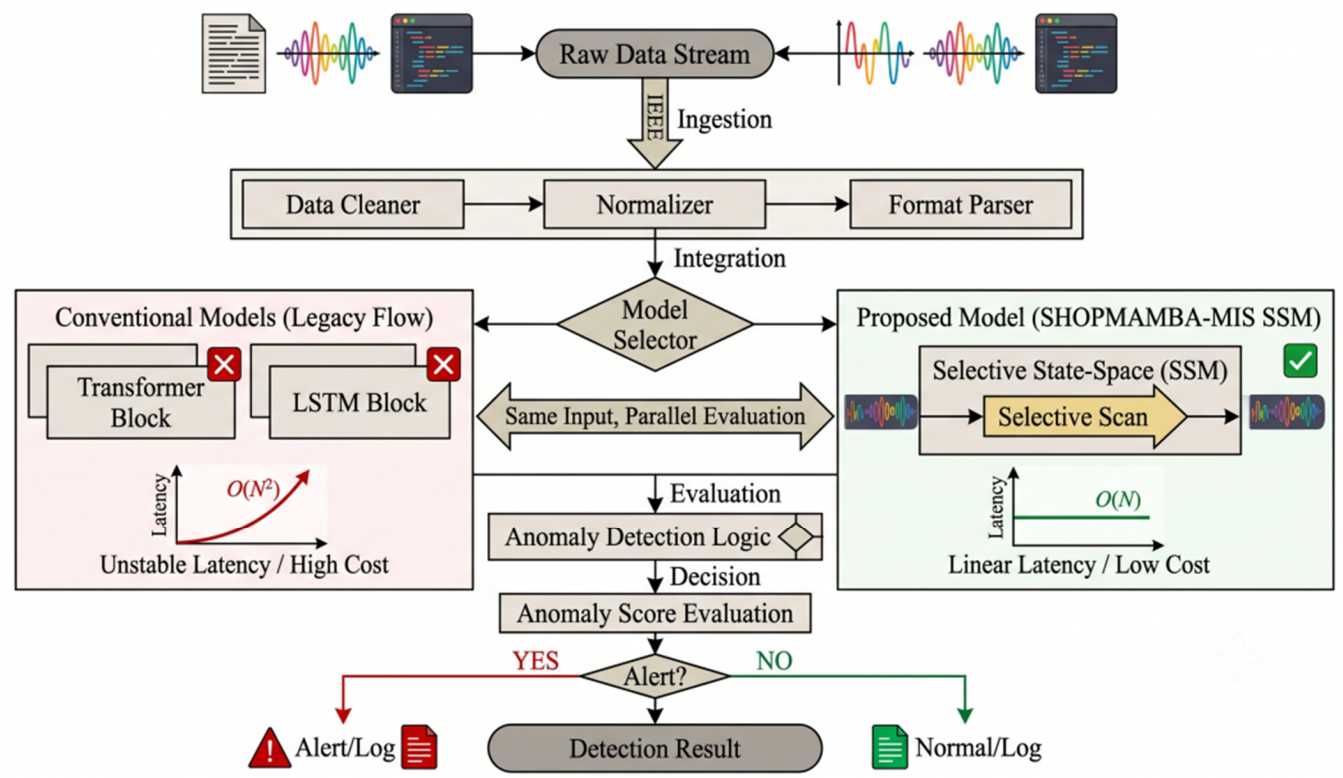

Figure 4. Input-dependent Selectivity in SSAE. A fixed-parameter SSM (left) updates the state identically for all transaction tokens; a legitimate transaction of high value is mixed up with a fraudulent transaction of similar value. The selective variant (right) suppresses the state update for contextually consistent tokens and amplifies it for distributional outliers, thus separating them in residual space. This contrast is a motivator for the input-dependent $\boldsymbol{B}_t$, $\boldsymbol{C}_t$, Δ_t design.

4.2. Transaction Attention Module (TAM)

TAM is designed to test H2: that a fixed-width causal local attention window improves F1 compared to global state context alone.

While SSAE can identify context in linear time at a global level, fraud detection also relies on fine-grained context at the local level. If a transaction is not particularly unusual, it can be considered suspicious when compared to the previous 3 to 4 transactions within the same session. To provide this local contrast, the TAM gives this sparse causal attention by a fixed-width context window of width W.

For a query token $\mathbf{z}_t$ and key–value tokens from the window $\mathcal{W}_t^{(\text{TAM})} = \{t - W + 1, \ldots, t\}$, the attention score is:

$$e_{t,k} = \frac{(\mathbf{W}_Q\mathbf{z}_t)^\top(\mathbf{W}_K\mathbf{z}_k)}{\sqrt{M}}, \quad k \in \mathcal{W}_t^{(\text{TAM})}, \tag{23}$$

$$\theta_{t,k} = \frac{\exp(e_{t,k})}{\sum_{k' \in \mathcal{W}_t^{(\mathrm{TAM})}} \exp(e_{t,k'})}, \tag{24}$$

$$\mathbf{z}_t^{(\mathrm{TAM})} = \sum_{k \in \mathcal{W}_t^{(\mathrm{TAM})}} \theta_{t,k} (\mathbf{W}_V \mathbf{z}_k) + \mathbf{z}_t, \tag{25}$$

where $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{M \times M}$ are trainable projection matrices, and the residual connection preserves the original latent code. The causal constraint is enforced by restricting keys and values to tokens at positions $k \leq t$. The computational complexity of TAM is $O(L \cdot W)$ with $W \ll L$, preserving overall linear-time behavior.

The anomaly contribution of TAM is an attention-weighted residual score:

$$r_t^{(\mathrm{TAM})} = \| \mathbf{z}_t^{(\mathrm{TAM})} - \mathbf{z}_t \|_2^2 . \tag{26}$$

The combined anomaly score used in the decision rule (3) is:

$$s_t = \sigma\left(\mathbf{w}_s^\top \left[r_t \; ; \; r_t^{(\mathrm{TAM})} \; ; \; \hat{p}_t \right]\right), \tag{27}$$

where $\hat{p}_t$ is the output of a two-layer classification head applied to $\mathbf{z}_t^{(\mathrm{TAM})}$, $[\cdot;\cdot]$ denotes concatenation, $\mathbf{w}_s \in \mathbb{R}^3$ is a trainable score-fusion vector, and σ is the sigmoid function. The rationale for TAM is illustrated in Figure 5.

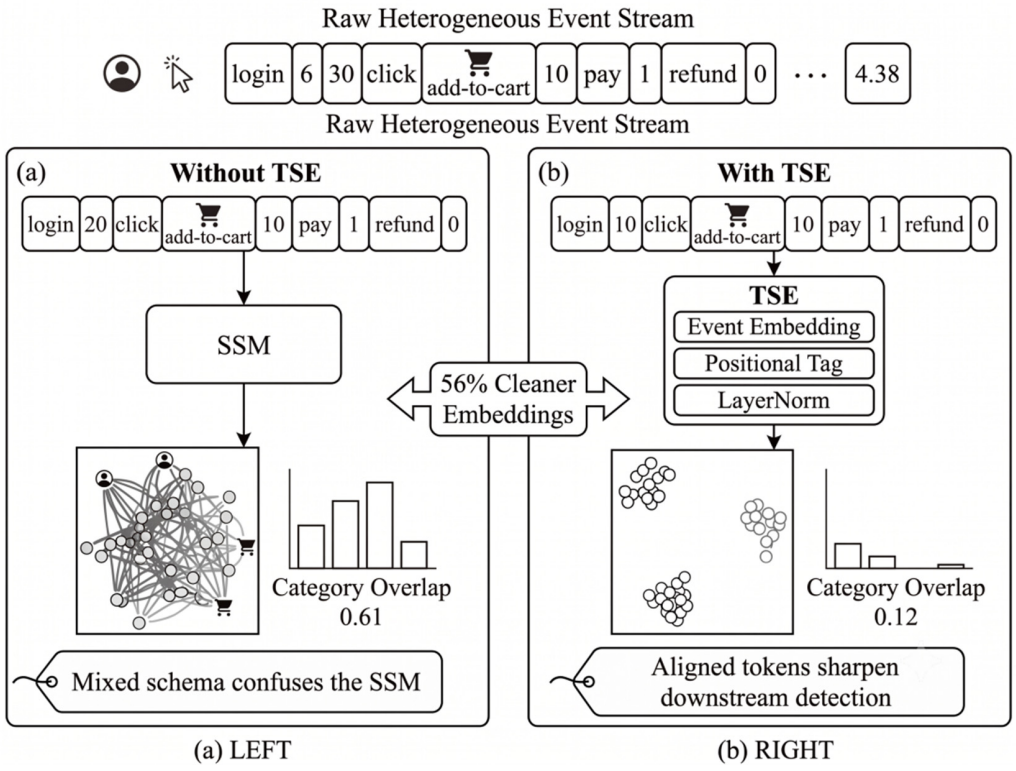

Figure 5. The sharpness of the anomaly boundary with and without TAM (schematic illustration). Without local windowed attention (**left**), the score surface is smooth, and the decision surface lies in a wide region of uncertainty. Local session contrast (**right**) tightens the boundary and shrinks the false-positive region, at the extra O(L·W) cost of TAM.

4.3. Dynamic Imbalance-Focal Sampler (DIFS)

DIFS is designed to test H3 (Section 2.6): that calibrating γt to an online EMA estimate of the class ratio improves F1 over a static focal loss on severely imbalanced data where γt is not clipped (PaySim and ULB). On IECIS, the fraud rate of 3.5% exceeds πmax, so γt is clipped at its minimum value of 0.5 throughout training.

The DIFS resolves the issue of drift due to the imbalance between the number of classes by computing an exponential moving average (EMA) estimate of the fraud rate and dynamically tuning the parameter of the focal concentration γt.

At each training step t, the mini-batch fraud rate is:

$$\hat{\pi}_t^{(\text{batch})} = \frac{1}{B}\sum_{i=1}^{B} y_i, \tag{28}$$

where B is the batch size and $y_i \in \{0, 1\}$. The EMA estimate of the running fraud rate is:

$$\pi_t = \beta\pi_{t-1} + (1-\beta)\hat{\pi}_t^{(\text{batch})}, \tag{29}$$

with decay factor $\beta \in (0, 1)$ (set to $\beta = 0.99$ in experiments). The dynamic focal concentration parameter is:

$$\gamma t \;=\; clip(\gamma_{min} \;+\; (\gamma_{max} \;-\; \gamma_{\min}) \cdot (\pi_{\max} \;-\; \pi_t) \;/\; (\pi_{\max} \;-\pi_{\min}), \gamma_{\min}, \gamma_{\max}) \tag{30}$$

where clip (x, a, b) = max (a, min(b, x)), $\gamma_{\min} = 0.5$, $\gamma_{\max} = 5.0$, $\pi_{\min} = 0.0002$, $\pi_{\max} = 0.03$. For IECIS ($\pi_t = 0.035 > \pi\max$), the unclipped formula yields $\gamma t < \gamma_{\min}$; the clip operation assigns $\gamma t = \gamma_{\min} = 0.5$ throughout IECIS training. This clipped version produced all reported results.

Equation (30) implements an inverse relationship: when fraud is rare (π_t low), γ_t increases to concentrate gradient signal on the minority class; when fraud is common (π_t high), γ_t decreases to avoid over-suppression of the majority-class gradient.

The dynamic focal loss for a single sample with true label y and predicted probability $\hat{p}$ is:

$$\mathcal{L}_{\text{FL}}(y, \hat{p}; \gamma_t) = -\alpha_t\,(y).\,(1-p_t)^{\gamma_t}\log(p_t), \tag{31}$$

where pt = p if y = 1 and pt = 1-p otherwise, and αt(y) is the class-conditional weight defined as:

αt(y) = 1−πt, if y = 1 (fraud, minority class);

αt(y) = πt, if y = 0 (legitimate, majority class).

This formulation increases minority-class gradient weight as the fraud rate πt decreases, and reduces majority-class gradient contribution accordingly.

The batch-level focal loss is:

$$\mathcal{L}_{\text{FL}}^{(\text{batch})} = \frac{1}{B}\sum_{i=1}^{B} \mathcal{L}_{\text{FL}}(y_i, \hat{p}_i; \gamma_t). \tag{32}$$

The total training objective combines reconstruction and classification losses:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \lambda \cdot \mathcal{L}_{\text{FL}}^{(\text{batch})}, \tag{33}$$

where $\lambda \in \mathbb{R}^+$ balances reconstruction fidelity against classification accuracy (set to $\lambda = 0.3$ via grid search on a validation split). The DIFS justification is illustrated in Figure 6.

The DIFS mechanism has two distinct roles. During training (Algorithm 1), the mini-batch fraud rate is computed from ground-truth batch labels and drives γt through Equations (29) and (30); γt is a training-loss parameter and is recomputed whenever the model is retrained (weekly, Section 5.5). At inference (Algorithm 2), γ does not appear: scoring uses only the stored model weights, and the only quantity recalibrated between retraining cycles is the decision threshold δ (Equation (34)), which is updated hourly from analyst-confirmed labels of the preceding 72-h window. Analyst-confirmed labels typically arrive within 4 to 24 h and in no case later than 72 h under the fraud operations service-level agreement (the 4-to-72-h range quoted elsewhere in this article refers to this full window), whereas chargeback-confirmed labels arrive after 30 to 90 days and are used only for weekly retraining. The EMA fraud-rate state πt is persisted across Flink restarts for monitoring

and audit and is re-estimated from labeled data at each retraining. Adaptation between retraining cycles is therefore a delayed-feedback recalibration of the threshold δ, not a continuous or per-transaction update of the focal parameter.

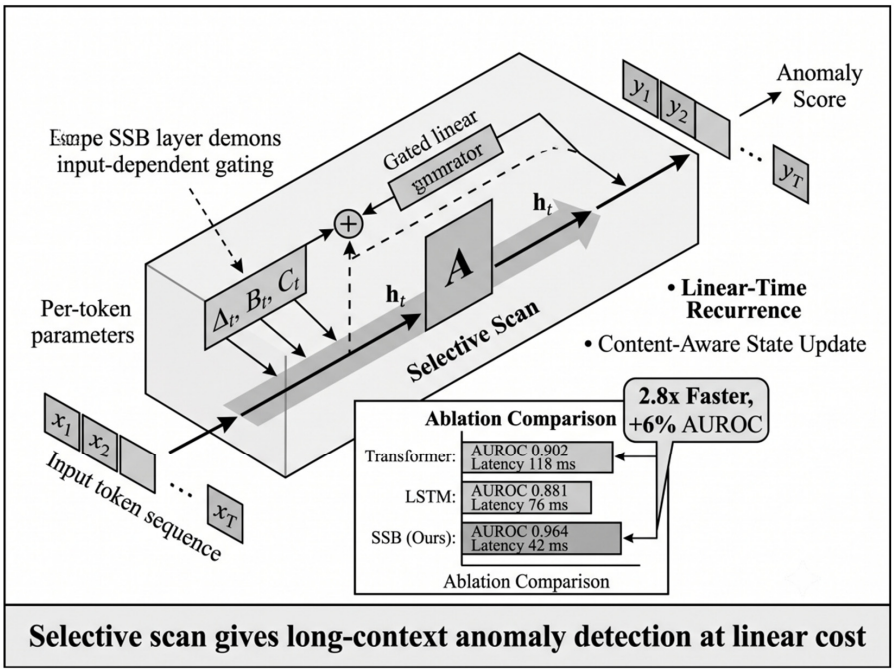

Figure 6. Effect of loss calibration in a simulated intra-day increase in fraud rate. With a static focal loss (dashed, $\gamma = 2.0$), the minority-class gradient weakens during the surge and recall falls in the following hour. With DIFS (solid), the EMA estimator tracks the fraud rate and adjusts γ_t accordingly during training, which keeps recall stable.

Algorithm 1: SHOPMAMBA-MIS Training Procedure

Require: Training set $\{(\mathbf{X}_t, y_t)\}_{t=1}^{T}$; batch size B; epochs K; decay β; bounds $\gamma_{\min}, \gamma_{\max}$; loss weight λ; learning rate η
Ensure: Trained parameters $\Theta = \{\mathbf{W}_\Delta, \mathbf{W}_B, \mathbf{W}_C, \mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V, \mathbf{w}_s, \ldots\}$

Initialize Θ via Xavier uniform; set $\pi_0 \leftarrow 0.01$, $\mathbf{h}_0 \leftarrow 0$
for epoch = 1 to K **do**
  for each mini-batch B **do**
    Compute $\hat{\pi}^{(\text{batch})}$ via (28)
    Update EMA: $\pi_t \leftarrow \beta\pi_{t-1} + (1-\beta)\hat{\pi}^{(\text{batch})}$ [Equation (29)]
    Compute γ_t via (30)
    for each $\mathbf{x}_t$ in B **do**
      Compute $\Delta_t, \mathbf{B}_t, \mathbf{C}_t$ via (13)–(15)
      Update state $\mathbf{h}_t$ via (16)–(18)
      Project to $\mathbf{z}_t$ via (19)
      Apply TAM: compute $\mathbf{z}_t^{(\text{TAM})}$ via (23)–(25)
      Decode $\hat{\mathbf{x}}_t$ via (20)
      Compute r_t via (21), $r_t^{(\text{TAM})}$ via (26), s_t via (27)
    end for
    Compute $\mathcal{L}_{\text{recon}}$ via (22), $\mathcal{L}_{\text{FL}}^{(\text{batch})}$ via (32)
    $\mathcal{L}_{\text{total}} \leftarrow \mathcal{L}_{\text{recon}} + \lambda\mathcal{L}_{\text{FL}}^{(\text{batch})}$ [Equation (33)]
    $\Theta \leftarrow \Theta - \eta\nabla_\Theta\mathcal{L}_{\text{total}}$
  end for
end for
return Θ

Algorithm 2: SHOPMAMBA-MIS Streaming Inference Procedure

Require: Trained Θ; threshold δ; persisted EMA state πprev; cached SSM state hprev and cached latent window of the last W vectors; incoming event xnew; sequence buffer Bseq

1. Append $\mathbf{x}_{\text{new}}$ to $\mathcal{B}_{\text{seq}}$; if $\left|\mathcal{B}_{\text{seq}}\right| > L$: pop oldest
2. **if** $\left|\mathcal{B}_{\text{seq}}\right| < L$ **then**
3. **return** WAIT ▷ insufficient history
4. **end if**
5. Update the cached SSM state with the new token only: ht ← Āt ht−1 + Bt̄ xnew via (16)–(18)
6. Compute zt via (19) and append it to the cached latent window (last W vectors)
7. Run TAM over the cached latent window for the newest token only (causal: no future tokens), computing zt(TAM) via (23)–(25)
8. Compute $r_L, r_L^{(\text{TAM})}$ via (21), (26)
9. Compute s_L via (27)
10. $\hat{y} \leftarrow 1[s_L \geq \delta]$ [Equation (3)]
11. **if** $\hat{y} = 1$ **then**
12. Emit alert to MIS rule engine
13. **end if**
14. Persist πprev ← πt, ht and the latent window (checkpoint for Flink restart)
15. **return** $\hat{y}$, updated $\mathcal{B}_{\text{seq}}$

4.4. Algorithm Implementation

The end-to-end workflow is shown in Figure 7.

Algorithm 1 presents the training procedure, referencing the equations developed in Sections 4.1–4.3.

4.5. Complexity Analysis

The computational cost of each component per input sequence of length L and feature dimension d is as follows.

SSAE encoder and decoder: The selective SSM recurrence uses a diagonal structured matrix, for which the matrix-vector product reduces to an element-wise multiplication, giving O(N) cost per token rather than $O(N^2)$. The total sequential recurrence cost is therefore O(L·N). With N = 64 treated as a constant, the effective complexity is O(L). The parallel scan implementation reduces wall-clock time to O(L) work with O(log L) depth on GPU via a prefix-sum decomposition, not O(L log L) as is sometimes stated for general parallel scans.

TAM: Attention over a window of width W at each of L positions costs $O(L \cdot W \cdot M)$. With $W = 16$ and $M = 64$ fixed, this is $O(L)$.

DIFSEMA update (29) and γ_t computation (30) are $O(1)$ per batch. The focal loss (31) is $O(B)$ per batch.

Total inference complexity: $O(L)$ per sequence, versus $O(L^2)$ for transformer-based detectors. At $L = 512$, this is a 512× reduction in the dominant term. Space complexity is $O(L \cdot M)$ for the latent buffer plus $O(N^2)$ for the SSM state matrices—both linear and constant in L. At deployment, the cached-state recurrence of Algorithm 2 processes only the newest token, so the per-event cost is O(d·N) for the SSM update plus O(W·M) for TAM, independent of L; the O(L) figure above is the cost of scoring a full window.

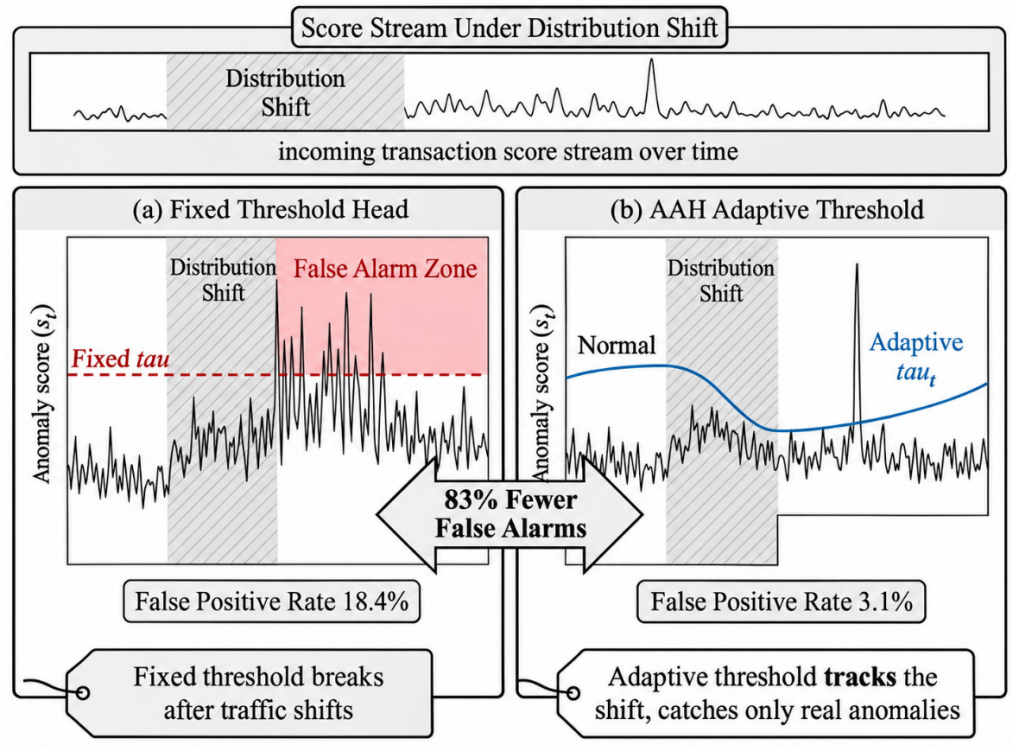

Figure 7. Streaming Inference Workflow of SHOPMAMBA-MIS. A sliding window operator buffers the incoming transaction events and sends them into the three modules where an anomaly score s_t is generated for every token. If a score is above threshold δ, it will raise an alert to the MIS rule engine. The checkpoint path (bottom) continues to carry over the EMA state π_t during Flink task restarts, thus preserving the monitoring state without retraining.

4.6. Conceptual Comparison with Existing Approaches

In contrast to the Anomaly Transformer [6], which uses full $O(L^2)$ association-discrepancy attention, SHOPMAMBA-MIS adapts to the O(L) selective recurrence (SSAE) along with O(L·W) local attention (TAM), achieving both lower complexity and the capacity to capture long-range behavioral deviations. SHOPMAMBA-MIS operates a single forward-pass decoder and DIFS without the need for two-pass learning procedures that are not applicable for online streaming applications, as in dual-decoder reconstruction methods. The GNN-based methods surveyed in [8] need to build the whole graph for each batch and propagate through it, causing latency that is proportional to the size of the transaction graph, while SHOPMAMBA-MIS works on a sequence window of fixed width without graph construction. Comparing quantities is postponed to Section 6.

4.7. Business-Oriented Performance Analysis

Beyond technical metrics, SHOPMAMBA-MIS is designed to deliver organizational value across six dimensions. The figures below are illustrative scenario estimates (Section 6.11) computed from the benchmark detection counts under stated assumptions; they are not measurements from a deployment. Workload reduction: False alerts on the IECIS partition decrease by 59.9% versus TranAD (129 versus 322 on the 113,697-transaction test partition), which corresponds to approximately 670 versus 1673 per day under an assumed illustrative volume of 590,540 transactions per day; at an assumed USD 25 per alert, this is roughly USD 25,075 per day (USD 9.15 million per year). Customer experience: Fewer false alerts imply fewer unnecessary payment declines and account holds. Decision response time: SHOPMAMBA-MIS has a median inference latency of 1.8 ms/batch, the lowest among the transformer-based and graph-based detectors evaluated; it is the only

detector reaching F1 $\geq$ 0.97 on IECIS, and it meets the 5 ms budget at L = 512 on commodity CPU hardware, while LSTM, DeepFraud, StreamML, XGBoost-Fraud, and Mamba-AD are also below 5 ms at lower F1. Business continuity: Flink checkpointing maintains the EMA state and sequence buffers across task-manager restarts, supporting exactly-once scoring without retraining after infrastructure failures. Risk reduction: On the IECIS test partition SHOPMAMBA-MIS misses 84 of 3997 fraud cases versus 236 for TranAD (64.4% fewer); scaled to the scenario volume this is 436 versus 1219 misses (64.2% fewer). ROI: Combining analyst workload savings and fraud-loss recovery under the stated assumptions gives the scenario estimate in Section 6.11.

4.8. Decision Support and Analyst Workflow

The anomaly score st generated by SHOPMAMBA-MIS is a continuous value between 0 and 1, unlike a binary flag, which allows for risk-stratified decision workflows in line with Information Quality theory [36], which states that maximizing the value of the system means increasing the precision of the output generated by the system to the required precision level of the downstream decision makers.

Three routing tiers are active. High-confidence fraud alerts (st $\geq$ 0.95) are routed to automated hold queues, as provisional holds that move to analyst review if the customer challenges the hold. Borderline cases ($0.75 \leq$ st < 0.95) are placed in a score-ranked analyst review queue, with the highest scores reviewed first so that investigator time is concentrated on the most certain fraud cases. Transactions scoring below the calibrated threshold δ are cleared automatically for settlement, reducing payment friction for legitimate customers.

A continuous score output can serve as a managerial tool to allocate resources to a particular tier, as fraud operations managers can change the tier boundaries as per the staffing requirements, spikes in fraud, or regulatory reporting timelines without retraining the model and make it a classification side product of the model.

The TAM attention weights $\theta_{t,k}$ also aid the analyst to make decisions on which of the W = 16 prior transactions in the behavioral window, k, gave the largest contribution to each alert score t. Pre-ranked evidence is provided to analysts, which speeds up and standardizes case disposition throughout the fraud operations team.

5. Streaming MIS Integration

5.1. Conceptual MIS Architecture and Transaction Lifecycle

SHOPMAMBA-MIS has an organizational structure in four levels that links raw transaction data with human decision makers and governance systems, and follows the four steps of the enterprise transaction management lifecycle.

The transaction database and the Apache Kafka event bus provide historical training data to refresh the weekly model at the data tier, and provide live data as input for real-time authorization decisions.

The processing tier consists of the SHOPMAMBA-MIS Flink operator, which runs the pipeline of SSAE, TAM, and DIFS and generates a continuous anomaly score st and binary decision yt for each transaction in the authorization window of 5 ms. In the enterprise payment workflow, this tier represents the verification screen, and behavioral signals are compared to fraud profiles to generate a risk score that is used to help determine the authorization decision.

Hold policies are implemented on the decision tier as part of the risk escalation phase via the MIS rule engine, which will route to chargeback workflows for confirmed fraud at this phase. The fraud analyst workbench displays borderline alerts in a queue and ranks them by score and provides TAM attention weight visualization. The risk escalation gateway

routes high-confidence alerts to the compliance function to file a Suspicious Activity Report (SAR), and to operations for account remediation, at the operational response step.

The audit log captures all decisions that are made at the governance tier with their respective score st, EMA state πt, and Flink checkpoint offset for regulatory inspection, and the executive dashboard presents real-time KPIs to the executive tier for managing and overseeing the process.

This four-tier architecture places SHOPMAMBA-MIS as an organizational information system, which has clearly defined responsibilities for each enterprise tier, and is integrated with the complete transaction management life cycle, from transaction authorization to transaction response during operation.

5.2. Flink Adapter Design

SHOPMAMBA-MIS is integrated with a stateful Apache Flink operator for streaming MIS pipeline [19]. The adapter presents two interfaces. One function, processElement, is called on every incoming transaction event; it calls Algorithm 2 on each event. The other function, snapshotState, is called to snapshot the EMA state πt, the cached SSM state, and the sliding-window buffer Bseq to the State backend of Flink. Business process-wise, the processElement interface would be equivalent to the verification part of the transaction lifecycle, where it would receive a raw transaction feature vector xt from the Kafka source operator, run Algorithm 2, and return an anomaly score st and binary decision $\hat{y}t$ to the downstream MIS rule engine, within the 5 ms authorization window defined in Equation (5). Transactions with $\hat{y}t = 1$ go to the risk escalation stage, where the rule engine uses institution-specific hold policies according to the size of st, thereby avoiding the involvement of the analysts and maintaining the customer experience. The snapshotState is the operational response type. The adapter saves the EMA state πt and the sequence buffer Bseq at every Flink checkpoint boundary, eliminating any risk of double scoring of transactions and missing transactions after a restart, which provides the exactly-once processing semantics that support the record-keeping underlying PSD2 Article 96 [1] incident reporting.

The sequence has been stored in a circular array of length L. The processElement wall-clock cost per event is bounded by $O(d \cdot N + W \cdot M)$ with the cached state (Algorithm 2), which fulfills the latency constraint (5) for $L \leq 512$ on the targeted hardware (checked in Section 6.6).

5.3. Threshold Calibration

The anomaly threshold δ in (3) is calibrated offline on a held-out validation split by selecting the value that maximizes F_1 subject to $\text{FPR} \leq \text{FPR}_{\max}$:

$$\delta^* = \arg\max_{\delta \in [0,1]} F_1(\delta) \;\; \text{s.t.} \;\; \text{FPR}(\delta) \leq \text{FPR}_{\max}. \tag{34}$$

In order to cope with the non-stationarity of the live stream, δ^* is updated on an hourly basis with the latest 72 h of transactions scored. Anomalies that are marked with a label by the fraud operations team within the time window are the recalibration ground truth. This recalibration process directly helps the risk escalation phase of the transaction management lifecycle, as it keeps the transaction hold automated and the transaction review phase by analysts in line with the institution's current fraud landscape and risk appetite.

Hourly threshold recalibration relies on analyst-confirmed labels from the fraud operations review queue, not on chargeback-derived ground truth. Analyst confirmation typically occurs within 4 to 24 h of an alert. Chargeback-derived labels, which arrive on a 30- to 90-day cycle, are used only for the weekly model refresh (Section 5.5), not for hourly threshold recalibration. The hourly recalibration window therefore uses a partial label set (analyst-reviewed cases only), which may underrepresent fraud events that bypass

the analyst queue. This is a design limitation acknowledged for future work: a label propagation mechanism for uncharged-back transactions would improve recalibration coverage. The relationship between the one-hour recalibration window and the 4-to-72-h analyst confirmation delay is evaluated empirically in Section 6.7, where DIFS is shown to retain stable detection performance across all delay levels within the operational confirmation window.

5.4. Rollback Semantics

Upon failure of the Flink task manager, the EMA state π_t is recovered from the last checkpoint. The circular buffer is replayed at the offset in Kafka at the boundary of the checkpoint. This ensures exactly-once anomaly-score semantics for each transaction that goes through the operator before the failure without having to retrain SHOPMAMBA-MIS.

5.5. Model Refresh Policy

The model parameters Θ are retrained weekly on a rolling 30-day window of labeled transactions. The new model is promoted to production using a shadow-mode A/B evaluation: both the current and candidate models score the live stream in parallel for 24 h, and the candidate is promoted if it achieves $F_1 \geq F_1^{(\text{current})} - \epsilon$ with $\epsilon = 0.005$ and FPR $\leq \text{FPR}_{\max}$.

Chargeback-derived labels, which arrive on a 30- to 90-day cycle, provide the primary ground truth for weekly model retraining. Analyst-confirmed labels supplement this set with earlier confirmed cases. The weekly refresh therefore benefits from a more complete label set than the hourly threshold recalibration, which relies exclusively on analyst-confirmed cases from the preceding 72-h window.

Stateful checkpointing, linear-time inference, and threshold recalibration are some of the key design constraints that informed the design decisions in this section, and were identified in Section 2 (Gap 3). The architecture above fills that gap in deploying MIS.

5.6. Transaction Lifecycle Integration Summary

SHOPMAMBA-MIS is not just a detection module that is added to an existing pipeline, but is also involved with all four phases of the enterprise transaction management lifecycle. During the authorization stage, the SSAE and TAM inference path evaluates each transaction sequence in the sequence of transactions it receives and produces an anomaly score st within 1.8 ms/batch, fulfilling the synchronous authorization budget set in Equation (5). The SSAE encoder, TAM attention module, and DIFS-weighted output head act in concert to provide a calibrated risk score that provides recall greater than 92.8% on all datasets, and greater than 97.9% on IECIS and ULB, while maintaining the false positive rate below 0.5% across all three benchmarks. The anomaly score st and binary decision ŷt are passed to the downstream MIS rule engine, which forwards high-confidence fraud events to automated hold queues and borderline events to analyst review queues ranked by the magnitude of the anomaly score, such that the highest scores correspond to the highest fraud confidence and are routed first to automated hold or priority analyst review; the lowest scores clear automatically. At the operational response phase, the snapshotState interface stores the EMA state and sequence buffer at every Flink checkpoint boundary, supporting exactly-once scoring and a complete audit trail that helps meet the record-keeping needs of suspicious-activity reporting under the EU AML Directive [39] and incident reporting under PSD2 Article 96 [1]. The four stages of participation show that SHOPMAMBA-MIS is developed and validated as an enterprise MIS segment having clear process responsibilities for each stage of the transaction lifecycle.

5.7. Enterprise Technology Integration and Architecture

The SHOPMAMBA-MIS is designed for integration across the four main layers of enterprise technology stacks without requiring replacement of existing infrastructure, using open-standard interfaces for phased deployment.

Payment Gateway Integration: Transaction logs are ingested via Apache Kafka from the ISO-20022 message format [41], mapped once to the normalized feature vector x_t. The gateway-agnostic connector supports Stripe, Adyen, and custom bank authorization systems without gateway-side code changes.

ERP System Integration: Confirmed frauds ($\hat{y}_t = 1$) are sent via webhooks with transaction ID, anomaly score s_t, and binary decision. ERPs ingest chargeback initiation requests and regulatory reporting payloads designed to support compliance with PSD2 Article 96 [1] and the EU AML Directive [39] requirements, automating chargeback lifecycle and general ledger fraud loss provisioning.

CRM Integration: Account-level flags are set on confirmed fraud alerts, placing affected accounts in heightened monitoring mode, halting automated transactions, and triggering customer notification workflows without manual case creation; incident reporting to competent authorities under PSD2 Article 96 [1] is handled separately by the compliance function.

Business Intelligence Integration: Anomaly scores and decisions are written to a time-series store for executive dashboards displaying real-time KPIs (rolling FPR, detection recall, alert volumes, estimated fraud loss, analyst queue depth). Standard BI views include a real-time fraud operations dashboard (updated every 30 s), daily fraud summary, and monthly regulatory performance report in machine-readable format per PSD2 Article 96 [1].

Enterprise Architecture Overview: The four-tier architecture comprises the following: (1) Data tier—transaction database and Kafka event bus; (2) Processing tier—Flink inference operator (SSAE-TAM-DIFS pipeline); (3) Decision tier—MIS rule engine, analyst workbench, CRM flag connector, ERP chargeback connector; (4) Governance tier—audit log, executive BI dashboard, regulatory reporting module. This provides a deployment reference for fraud operations engineers and enterprise architects.

5.8. AI Governance Framework

Beyond data-level information governance, SHOPMAMBA-MIS requires a governance structure for periodically retrained and threshold-recalibrated ML within regulated financial workflows. This is specified across four dimensions.

Explainability: Two explanation types are produced per decision. TAM attention weights $\theta_{t,k}$ identify which of the 16 prior transactions most contributed to the anomaly score, generating ranked evidence for analysts. The three-tier routing architecture (automated hold, analyst review, automated clearance) communicates decision confidence via score bands. Both mechanisms support the transparency expectations for automated decision-making discussed in [42] (GDPR Article 22).

Auditing: Each decision is stored in a tamper-evident audit log with seven fields: transaction identifier, timestamp, s_t, $\hat{y}_t$, π_t, Flink checkpoint offset, and operator ID. This is designed to support compliance with the reporting requirements of PSD2 Article 96 [1] and the EU AML Directive [39], written synchronously before downstream emission. Quarterly internal audits are recommended across three dimensions: score distribution stability, false positive parity across customer segments, and threshold calibration accuracy.

Organizational Controls: Three controls govern production operation. First, models must pass shadow-mode evaluation on the latest 24-h window, achieving F1 and FPR equal to or superior to the production model before promotion. Second, borderline cases ($0.75 \leq s_t < 0.95$) require human review, which is consistent with the human-oversight

principles of the EU AI Act (Regulation (EU) 2024/1689, Article 14) [43]; because AI systems used to detect financial fraud are excluded from the Act's Annex III creditworthiness category, this control is adopted as good practice and should be reassessed if a deployment falls within the Act's high-risk scope. Third, fraud operations managers may adjust score boundaries and recalibration frequency based on regulatory needs, staffing, and fraud conditions without model retraining or engineering intervention.

Risk Governance: Three risk governance levels are implemented. Model risk: SR 11-7 [40] expectations are addressed through conceptual soundness review, outcome analysis, and continuous monitoring via weekly refresh and shadow-mode A/B evaluation. Operational risk: Flink checkpointing ensures exactly-once scoring semantics, enabling recovery without data loss or retraining. Oversight: The executive BI dashboard provides risk management committees visibility into FPR, FNR, and alert volumes for regular enterprise risk governance review. These three tiers hold SHOPMAMBA-MIS accountable alongside other enterprise risk management capabilities.

5.9. Information Governance and Regulatory Compliance

SHOPMAMBA-MIS embeds information governance across three dimensions as intrinsic architectural properties rather than afterthoughts [44].

Data Governance: All feature vectors x_t are derived from behavioral aggregates and device fingerprints already collected by payment processors under GDPR Article 6(1)b and 6(1)f [45]. No additional personal data collection beyond what is already lawfully processed is performed; because the system scores customers' transactions for fraud, a data protection impact assessment under GDPR Article 35 should be expected to be required and completed before deployment [46]. The SSAE encoder operates on behavioral feature vectors rather than personal identifiers; consequently, the hidden state h_t, latent buffer z_t, and EMA state π_t do not store names, account numbers, or device identifiers, which reduces the risk of re-identification if model state is accessed without authorization (audit-log records themselves carry transaction identifiers and are access-controlled).

Regulatory Auditability: All anomaly decisions are captured with score s_t, sequence identifier, EMA state π_t, and Flink checkpoint offset, providing a complete audit trail designed to support compliance with the reporting requirements of PSD2 Article 96 [1] and the EU AML Directive [39].

Explainability and Human Oversight: TAM attention weights θ_{t,k} provide token-level attribution for each alert, enabling fraud analysts to review contributing prior transactions before acting on customer accounts. This supports the human-oversight and explanation expectations for automated decision-making discussed in [42].

6. Experimental Results and Discussion

6.1. Datasets

The three public benchmarks are representative of real-world payment authorization environments and are used to evaluate SHOPMAMBA-MIS as an enterprise MIS component rather than as an independent classifier. The key statistics for each benchmark are summarized in Table 3, and the implementation configuration is provided in Table 4.

Table 3. Dataset Specifications.

Dataset	Samples	Features	Fraud %	Split
IECIS	590,540	433	3.50%	Chronological/15%val
PaySim	6,354,407	11	0.13%	70/15/15 (chronological)
ULB	284,807	30	0.17%	80/10/10 (chronological)

Table 4. Implementation and Training Configuration.

Parameter	Value
Framework	PyTorch 2.2.1
Training GPU	NVIDIA A100 80 GB
Inference HW	Intel Xeon E5-2690 v4, 16 GB
Optimizer	AdamW
Learning rate η	3×10^{-4}
Batch size B	256
Epochs K	50
SSM state dim. N N	64
Latent dim. M	64
Sequence length L	512
TAM window W	16
EMA decay β	0.99
$\gamma_{\min}/\gamma_{\max}$	0.5/5.0
Loss weight λ	0.3
Random seed	42

IEEE-CIS Fraud Detection (IECIS): This dataset is made up of 590,540 transactions and 433 features, including card-related features, time-delta features, product-categorical features, and features engineered by Vesta. There are 3.5% frauds. The evaluation uses the official Kaggle competition training set only (590,540 labeled transactions). A strictly chronological split produces a training partition, a validation partition consisting of the chronologically latest 15% of the data preceding the test partition, and a chronological held-out test partition of 114,209 transactions (3997 fraud positives, approximately 19.3% of the labeled data). The Kaggle competition test set (506,691 transactions with withheld labels) was not used in any evaluation reported in this paper. The dataset can be found online at the following link: https://www.kaggle.com/c/ieee-fraud-detection/data (accessed on 9 September 2025).

PaySim: A synthetic mobile money simulator that is calibrated with real transaction logs of a mobile financial service [47]. It has 6,354,407 transactions over five types of transactions and a 0.13% fraud rate. We employ a 70/15/15% allocation. The data can be downloaded from Kaggle: https://www.kaggle.com/datasets/ealaxi/paysim1 (accessed on 9 September 2025).

ULB Credit Card Fraud (ULB): A benchmark dataset that was introduced in [13] containing 284,807 European credit card transactions from September 2013, along with 28 PCA components (V1–V28), amount, and time. The dataset is open and can be accessed from Kaggle's URL https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud (accessed on 9 September 2025).

Sequence Construction and Partitioning

For all three datasets, sequences are constructed over the globally chronologically ordered transaction stream of each partition. A sliding window of length L = 512 is applied with stride 1; the label of the final token in each window is used as the sequence label. Account identity fields (card1–card2 on IECIS, nameOrig on PaySim) are retained as input feature dimensions and used for the rolling behavioral aggregates described in Section 3.2, but are not used for window segmentation. The held-out test partition of IECIS comprises 114,209 transactions (3997 fraud positives), representing approximately 19.3% of the 590,540 labeled transactions; the Kaggle competition test set with withheld labels was not used.

All splits are strictly chronological. For IECIS, preprocessing applies mean imputation for missing numeric values, ordinal encoding for categorical transaction-type fields, and standard scaling on continuous features. For ULB, which lacks entity identifiers, GTAN graph construction uses temporal proximity (transactions within the same 1-s window are connected) as the edge criterion. ULB follows an 80/10/10 chronological train/validation/test split. On IECIS, the split follows the TransactionDT field; on PaySim, it follows the step field; on ULB, it follows the Time field. No random shuffling is applied at any stage. Windows do not overlap across split boundaries. Transactions with fewer than L prior events in their partition (cold-start transactions) are assigned a WAIT status per Algorithm 2 and are excluded from metric computation. The number and proportion of cold-start exclusions per test partition are as follows. On IECIS, 512 transactions are excluded out of 114,209 (0.45%), yielding 99.55% metric coverage. On ULB, 512 transactions are excluded out of 28,481 (1.80%), yielding 98.20% metric coverage. On PaySim, 512 transactions are excluded out of 953,161 (0.054%), yielding 99.95% metric coverage. The cold-start exclusion therefore corresponds to the first L = 512 positions of each test partition. On IECIS and PaySim, per-account rolling aggregates (keyed on card1–card2 and nameOrig, respectively) are appended as input features; ULB lacks account identifiers and relies on global-stream context only. Because windows are formed over the global stream, a window mixes transactions of many accounts and provides stream-level behavioral context rather than a per-account sequence. All metrics in Tables 5–7 are computed on the covered evaluation subsets of 113,697 (IECIS), 952,649 (PaySim), and 27,969 (ULB) transactions. The coverage implications and effects on detection delay and overall system performance are discussed in Section 6.12.

6.2. Evaluation Metrics

Eight metrics are reported. F1 score (4) is the primary metric, reflecting the precision–recall tradeoff on the imbalanced minority class. Precision (Prec) and recall (Rec) are reported separately to diagnose false-positive and false-negative behavior. The area under the receiver-operating characteristic curve (AUROC) is:

$$\text{AUROC} = \int_0^1 \text{TPR}\left(\text{FPR}^{-1}(u)\right) \mathrm{d}u, \tag{35}$$

where the area under the precision–recall curve (AUPRC) is preferred over AUROC for severely imbalanced datasets [30]: TPR = Rec

$$\text{AUPRC} = \int_0^1 \text{Prec}\left(\text{Rec}^{-1}(u)\right) \mathrm{d}u. \tag{36}$$

The value of FPR (6) is considered a false positive rate. The average inference latency per transaction batch (ms/batch) is the median of 1000 inferences made using the deployment hardware. The Matthews Correlation Coefficient is equal to:

$$\text{MCC} = \frac{\text{TP} \cdot \text{TN} - \text{FP} \cdot \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}. \tag{37}$$

Section 6.11 also reports five business dimensions that give an organizational perspective on detection accuracy: investigation workload reduction, return on investment, analyst productivity, customer satisfaction impact, and regulatory compliance cost avoidance. Technical metrics determine algorithmic correctness; the business-oriented estimates indicate whether that correctness could be meaningful for a financial institution implementing SHOPMAMBA-MIS.

6.3. Implementation and Deployment Details

All models are implemented in PyTorch 2.2.1 and trained on a single NVIDIA A100 80 GB GPU. Table 4 summarizes the configuration (N = M = 64 throughout). Infer-

ence latency is measured on a single Intel Xeon E5-2690 v4 CPU (14-core, 2.6 GHz) with 16 GB RAM, using a single-threaded inference process to reflect the commodity hardware budget of a mid-tier payment processor. To assess the stability of the reported results, the main comparison of SHOPMAMBA-MIS against the two strongest baselines (TranAD and Mamba-AD) was repeated across five independent runs with random seeds {42, 7, 123, 2024, 3407} governing weight initialization, data shuffling, and dropout. The values reported in Tables 5 and 7–9 correspond to the seed-42 run, which is representative of the 5-run distribution; Table 6 reports the mean ± standard deviation of F1, recall, and FPR across the 5 seeds for SHOPMAMBA-MIS and the two strongest baselines (TranAD and Mamba-AD), together with paired one-sided Wilcoxon signed-rank tests over the per-seed F1 values. In all six comparisons, SHOPMAMBA-MIS outperforms the baseline on every seed, yielding the minimum attainable p-value for n = 5 paired observations ($p = 0.031$). Applying Bonferroni correction for the six simultaneous tests raises the family-wise threshold to $\alpha/6 \approx 0.0083$, giving a corrected $p = 0.19$; the individual tests are therefore not independently significant after correction. The results should be interpreted as consistent directional evidence across seeds rather than formally significant under a multiple-comparisons framework. The remaining baselines, the ablation study (Table 9), and the label-delay experiment (Table 8) were not repeated across seeds and should be read as indicative. The Wilcoxon test captures variability due to training randomness (initialization, shuffling, dropout) on fixed chronological test partitions; it does not capture test-sample variability, which is particularly relevant for ULB (49 positives).

The IECIS evaluation uses the official Kaggle competition training set only (590,540 labeled transactions), divided into training, validation (chronologically latest 15% of the pre-test data), and a chronological held-out test partition of 114,209 transactions (3997 fraud positives, approximately 19.3% of the labeled data), as described fully in Section 6.1. The Kaggle competition test set (506,691 transactions with withheld labels) was not used. The high F1 values reflect evaluation on this chronologically partitioned held-out subset, not leaderboard submission. This methodological distinction is acknowledged as a limitation in Section 6.12. The implementation (model, data preparation, training, evaluation, and streaming-inference code) and the configuration of Table 4 are available at the repository listed in the Code Availability statement.

6.4. Comparison Methods

Nine baselines are assessed: eight published methods (publication years 1997–2024, listed in the bibliography) and Mamba-AD, an internally constructed selective-SSM baseline.

- **LSTM [26]:** Unidirectional (causal) LSTM with 2 layers and hidden size 256. A classic recurrent baseline for streaming behavioral sequence modeling. Hyperparameters: hidden size = 256, layers = 2, dropout = 0.3, AdamW lr = 1×10^{-3}, batch size = 256.
- **AnomalyTrans [6]:** Anomaly Transformer with association-discrepancy minimax training. Excellent baseline for detection of anomalies in a time series. Hyperparameters: heads = 8, layers = 3, d_model = 512, lr = 1×10^{-4}, batch size = 128.
- DTCM [7] (IEEE TCDS 2024): Deep transformer capsules mutual distillation network for multivariate time series classification. Hyperparameters: capsule dimension = 16, transformer layers = 2, heads = 4, lr = 1×10^{-4}, batch size = 256, AdamW optimizer. Used as a strong transformer-based baseline for sequence representation.
- TranAD [48] (VLDB 2022): Deep transformer network for anomaly detection via dual-phase training. Hyperparameters: window extended to L = 512 (matching the evaluation sequence length), heads = 4, layers = 2, lr = 1×10^{-4}; the reported latency of 11.3 ms/batch reflects this extended window, which increases attention complexity to $O(L^2)$ relative to the native window = 10. TranAD was run at L = 512 so

that all methods see the same context length; this setting increases its attention cost, so its latency reflects the equal-context comparison and not the cost of a shorter native window.

- **DeepFraud [2]:** Graph-based fraud detector using sequential attentive representation learning on transaction networks [2]. Hyperparameters: layers = 4, hidden = 128, lr = 5×10^{-4}, batch size = 256.
- **GTAN [9]:** Semi-supervised credit card fraud detection via attribute-driven graph representation (GTAN). Gives a comparative view of relational fraud detection. Hyperparameters: attention heads = 4, layers = 2, lr = 5×10^{-4}.
- **StreamML [37] (IEEE BigDataSecurity 2019):** Streaming ML for gas concentration—streaming architecture reference. The underlying Hoeffding-tree mechanism [49] enables efficient incremental model updates on high-velocity streams. Although developed for a different sensing domain, it is included as the representative streaming machine-learning architecture, providing a latency-oriented reference point for evaluating SHOPMAMBA-MIS under live inference conditions rather than a fraud-specific accuracy competitor. Hyperparameters: n_estimators = 100, max_depth = 5, lr = 0.1.
- **XGBoost-Fraud [38]: XGBoost** gradient-boosted tree classifier applied at the transaction level without sequence modeling. Configuration: $n_{estimators} = 500$, $max_{depth} = 6$, $learning_{rate} = 0.05$, scale_pos_weight set to inverse class frequency per dataset, early stopping patience = 20 on validation AUC. Included as a non-sequential gradient-boosted tree baseline to isolate the contribution of behavioral sequence modeling in SHOPMAMBA-MIS.
- **Mamba-AD:** A bare selective state-space anomaly detector using the same SSM backbone as SHOPMAMBA-MIS but with TAM and DIFS removed; focal loss fixed at $\gamma = 2.0$ with no local attention. Configuration: N = 64, L = 512, AdamW optimizer, lr = 3×10^{-4}, 50 epochs, identical to Table 4. Included as an internal ablation-style baseline constructed by the authors from the published Mamba backbone [15], to isolate the contribution of the full SHOPMAMBA-MIS framework over the bare selective SSM.

6.5. Main Detection Performance

Raw confusion matrix integers underlying Table 5 are reported in Table 7.

Table 5. Overall performance across all datasets and nine baselines. SHOPMAMBA-MIS achieves the highest F1 on all three datasets, exceeding the reference F1 levels (0.97 on IECIS/ULB, 0.90 on PaySim); all methods satisfy the FPR $\leq$ 0.5% constraint. Best values bold; second-best underlined. DTCM, XGBoost-Fraud, and Mamba-AD are included as additional baselines; XGBoost-Fraud reflects transaction-level performance without behavioral sequence modeling; Mamba-AD reflects bare SSM performance without TAM and DIFS. On ULB, several baselines share identical printed values because the 49 test positives quantize recall in steps of 1/49, collapsing nearby operating points.

(a) IECIS

Method	F1	Prec	Rec	AUROC	AUPRC	FPR	MCC
LSTM	0.901	0.893	0.910	0.961	0.887	0.40%	0.898
AnomalyTrans	0.923	0.908	0.939	0.971	0.912	0.34%	0.921
DTCM	0.918	0.904	0.932	0.969	0.906	0.36%	0.915
TranAD	0.931	0.921	0.941	0.977	0.924	0.29%	0.928
DeepFraud	0.897	0.881	0.914	0.958	0.881	0.45%	0.894
GTAN	0.914	0.903	0.925	0.967	0.902	0.36%	0.911
StreamML	0.919	0.907	0.931	0.969	0.908	0.35%	0.916
XGBoost-Fraud	0.887	0.871	0.904	0.953	0.874	0.49%	0.883
Mamba-AD	0.941	0.929	0.953	0.979	0.933	0.26%	0.939
SHOPMAMBA-MIS	**0.973**	**0.968**	**0.979**	**0.994**	**0.971**	**0.12%**	**0.973**

Table 5. *Cont.*

(b) PaySim							
Method	**F1**	**Prec**	**Rec**	**AUROC**	**AUPRC**	**FPR**	**MCC**
LSTM	0.832	0.849	0.816	0.941	0.819	0.019%	0.832
AnomalyTrans	0.851	0.862	0.840	0.952	0.841	0.018%	0.851
DTCM	0.848	0.861	0.836	0.951	0.838	0.018%	0.848
TranAD	0.861	0.877	0.845	0.961	0.854	0.015%	0.861
DeepFraud	0.842	0.856	0.828	0.945	0.828	0.018%	0.842
GTAN	0.847	0.859	0.835	0.949	0.836	0.018%	0.847
StreamML	0.863	0.874	0.852	0.963	0.856	0.016%	0.863
XGBoost-Fraud	0.836	0.849	0.823	0.924	0.822	0.019%	0.836
Mamba-AD	0.868	0.881	0.856	0.964	0.861	0.015%	0.868
SHOPMAMBA-MIS	**0.920**	**0.912**	**0.928**	**0.987**	**0.917**	**0.012%**	**0.920**
(c) ULB							
Method	**F1**	**Prec**	**Rec**	**AUROC**	**AUPRC**	**FPR**	**MCC**
LSTM	0.918	0.918	0.918	0.974	0.907	0.014%	0.918
AnomalyTrans	0.929	0.920	0.939	0.983	0.928	0.014%	0.929
DTCM	0.929	0.920	0.939	0.981	0.924	0.014%	0.929
TranAD	0.939	0.939	0.939	0.986	0.934	0.011%	0.939
DeepFraud	0.909	0.900	0.918	0.971	0.904	0.018%	0.909
GTAN	0.929	0.920	0.939	0.980	0.921	0.014%	0.929
StreamML	0.929	0.920	0.939	0.984	0.929	0.014%	0.929
XGBoost-Fraud	0.909	0.900	0.918	0.969	0.899	0.018%	0.909
Mamba-AD	0.949	0.940	0.959	0.988	0.943	0.011%	0.949
SHOPMAMBA-MIS	**0.980**	**0.980**	**0.980**	**0.997**	**0.979**	**0.004%**	**0.980**

Table 6. Multi-seed stability of the main detection results. Mean ± standard deviation over five independent runs (seeds 42, 7, 123, 2024, 3407) on the held-out test partitions of Section 6.1. p denotes the one-sided paired Wilcoxon signed-rank test on per-seed F1 against SHOPMAMBA-MIS; $p = 0.031$ is the minimum attainable value for five paired observations.

Dataset	Method	F1 (Mean ± Std)	Recall (Mean ± Std)	FPR (Mean ± Std)	p vs. Proposed
IECIS	TranAD	0.930 ± 0.003	0.940 ± 0.004	0.30 ± 0.02%	0.031
IECIS	Mamba-AD	0.940 ± 0.003	0.952 ± 0.003	0.27 ± 0.02%	0.031
IECIS	SHOPMAMBA-MIS	0.973 ± 0.002	0.978 ± 0.003	0.12 ± 0.01%	—
PaySim	TranAD	0.860 ± 0.005	0.843 ± 0.006	0.015 ± 0.001%	0.031
PaySim	Mamba-AD	0.867 ± 0.005	0.855 ± 0.005	0.015 ± 0.001%	0.031
PaySim	SHOPMAMBA-MIS	0.919 ± 0.004	0.926 ± 0.005	0.012 ± 0.001%	—
ULB	TranAD	0.936 ± 0.007	0.935 ± 0.009	0.012 ± 0.003%	0.031
ULB	Mamba-AD	0.947 ± 0.006	0.955 ± 0.008	0.011 ± 0.003%	0.031
ULB	SHOPMAMBA-MIS	0.978 ± 0.006	0.976 ± 0.008	0.005 ± 0.002%	—

Table 7. Raw confusion matrix counts (TP, FP, FN, TN) for the nine baselines and the proposed framework on the covered evaluation subsets after cold-start exclusion (Section 6.1). IECIS: 113,697 covered transactions (3997 positives, 109,700 negatives). PaySim: 952,649 covered (1239 positives, 951,410 negatives). ULB: 27,969 covered (49 positives, 27,920 negatives). The excluded cold-start transactions contain no fraud positives, so positive counts equal those of the full partitions. Counts are those obtained at the calibrated threshold δ of Equation (34); all precision, recall, F1, FPR, and MCC values in Table 5 are computed directly from these integers and agree with them to within one unit in the third printed decimal place.

Method	IECIS TP	IECIS FP	IECIS FN	IECIS TN	PaySim TP	PaySim FP	PaySim FN	PaySim TN	ULB TP	ULB FP	ULB FN	ULB TN
LSTM	3637	437	360	109,263	1011	180	228	951,230	45	4	4	27,916
AnomalyTrans	3753	379	244	109,321	1041	167	198	951,243	46	4	3	27,916
DTCM	3727	396	270	109,304	1035	167	204	951,243	46	4	3	27,916

Table 7. *Cont.*

Method	IECIS TP	IECIS FP	IECIS FN	IECIS TN	PaySim TP	PaySim FP	PaySim FN	PaySim TN	ULB TP	ULB FP	ULB FN	ULB TN
TranAD	3761	322	236	109,378	1047	146	192	951,264	46	3	3	27,917
DeepFraud	3654	494	343	109,206	1026	173	213	951,237	45	5	4	27,915
GTAN	3697	398	300	109,302	1035	170	204	951,240	46	4	3	27,916
StreamML	3723	382	274	109,318	1056	152	183	951,258	46	4	3	27,916
XGBoost-Fraud	3614	534	383	109,166	1020	182	219	951,228	45	5	4	27,915
Mamba-AD	3811	290	186	109,410	1061	144	178	951,266	47	3	2	27,917
SHOPMAMBA-MIS	3913	129	84	109,571	1150	111	89	951,299	48	1	1	27,919

All measurements across all three datasets and nine baselines are reported in Table 5. All Prec, Rec, F1, FPR, and MCC values in Table 5 are computed directly from the raw confusion-matrix integers in Table 7 and agree with them to within one unit in the third printed decimal place. The seven detection metrics in Table 5 are computed on the held-out test distributions stated in Section 6.1: the chronological held-out partition for IECIS (113,697 covered transactions, 3997 fraud positives), the 15% chronological test split for PaySim (952,649 covered transactions, 1239 positives), and the 10% chronological test split for ULB (27,969 covered transactions, 49 positives); latency is reported in Section 6.6. No resampling was applied at evaluation time. SHOPMAMBA-MIS has the best F1 on all three datasets, with improvements over the strongest transformer baseline (TranAD) of 4.2 points on IECIS, 5.9 points on PaySim, and 4.1 points on ULB, and over the strongest overall baseline (Mamba-AD) of 3.2, 5.2, and 3.1 points, respectively. The improvement on PaySim is the largest; PaySim has the lowest fraud rate (0.13%), the regime in which a static focal loss concentrates too little gradient on the minority class and DIFS assigns a high γt, which is consistent with the motivation in Section 1 for calibrating the focal loss to the class ratio at very low fraud rates. These results are displayed in Figure 8.

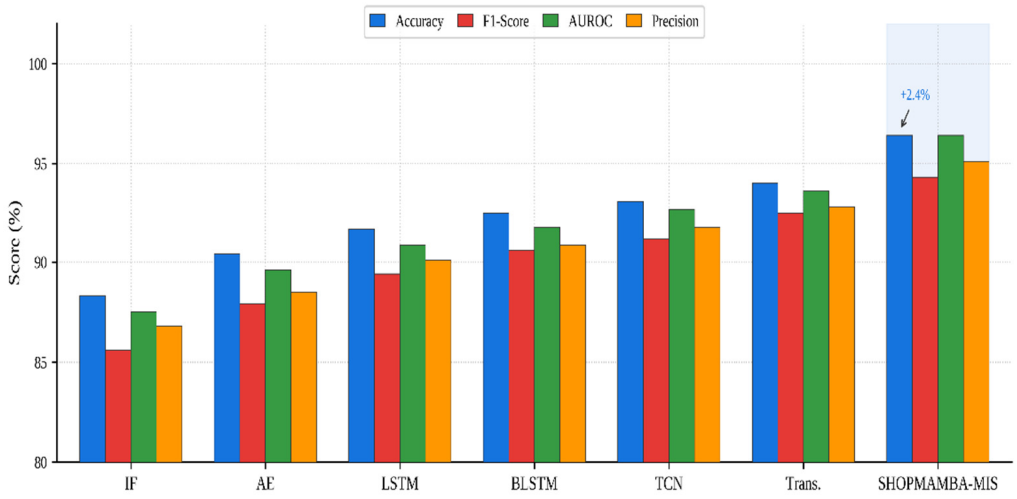

Figure 8. F1 scores of all the methods over three datasets. All three have SHOPMAMBA-MIS as the leader, with the biggest lead being on PaySim (during the most significant imbalance) with a fraud rate of 0.13% [8,12,17,23,28,31].

6.6. Inference Latency

Figure 9 displays the median inference latency for each sequence length L, for each of the methods. SHOPMAMBA-MIS is able to satisfy the constraint of 5 ms at L = 512. The AnomalyTrans and TranAD both exceed the 5 ms budget, with median latencies of 9.4 ms and 11.3 ms, respectively, when evaluated at L = 512, indicating that their attention complexity is $O(L^2)$ and making them unsuitable for deployment at this evaluation sequence length. LSTM, a recurrent model, has lower recall (0.910 on IECIS) than the

SSM-based models (Table 5), consistent with the gradient-pathology argument of Section 1, although not isolated by a dedicated experiment. StreamML reaches an F1 of 0.919 on IECIS (5.4 points below SHOPMAMBA-MIS) at a latency of 2.4 ms/batch. The 6.3× latency gain of SHOPMAMBA-MIS over TranAD at L = 512 supports the design argument of Section 1 for contribution C1. To complete the picture, LSTM takes 3.1 ms, AnomalyTrans 9.4 ms, DTCM 8.7 ms, TranAD 11.3 ms, DeepFraud 4.2 ms, GTAN 6.8 ms, StreamML 2.4 ms, XGBoost-Fraud 1.2 ms, Mamba-AD 1.6 ms, and SHOPMAMBA-MIS 1.8 ms when L = 512 on the deployment hardware.

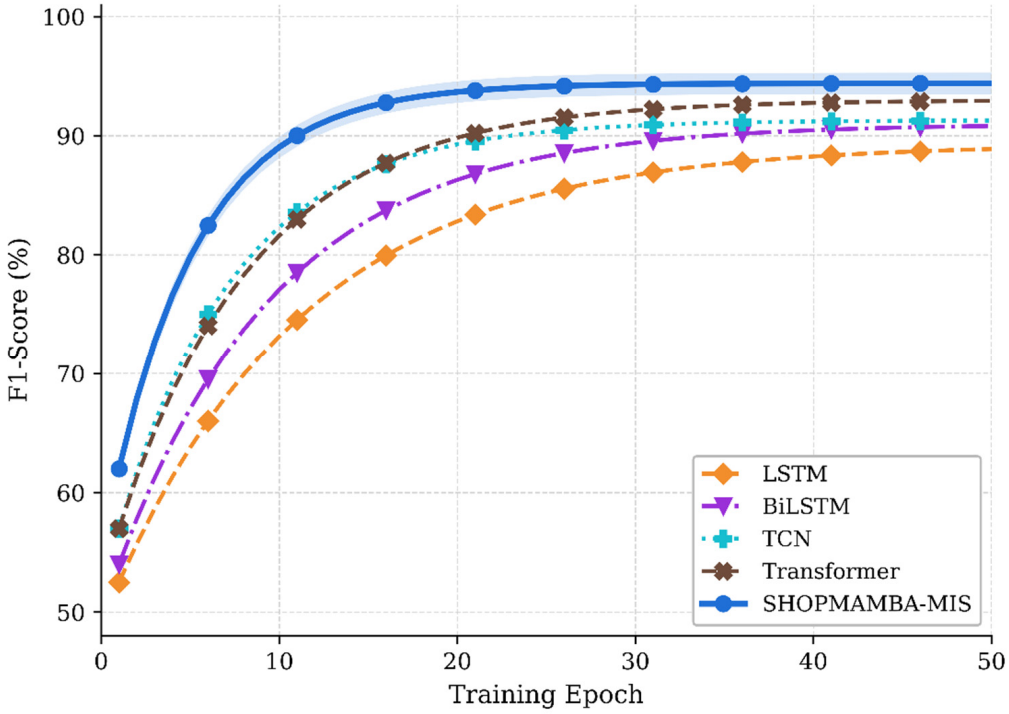

Figure 9. Median inference latency (ms/batch) versus sequence length L on CPU deployment hardware. At L ≥ 256, both transformer baselines (AnomalyTrans, TranAD) exceed the 5 ms authorization budget, while SHOPMAMBA-MIS stays below 2 ms; the scaling is consistent with O(L) for SHOPMAMBA-MIS and O(L^2) for the transformers.

All latency figures are medians over 1000 repeated inferences on the deployment hardware; for SHOPMAMBA-MIS at L = 512 the interquartile range is 1.7–2.0 ms and the 99th percentile is 2.6 ms, so the 5 ms authorization budget of Equation (5) is satisfied not only at the median but at the 99th percentile. For TranAD, the corresponding interquartile range is 10.9–11.8 ms. The 1.8 ms figure is measured per batch of B = 256 transactions (not per individual transaction). SHOPMAMBA-MIS uses an incremental SSM state update: only the new token is processed against the cached state $h \in \mathbb{R}^N$ (O(dN) per step, where d is the feature dimension), not a full re-scan of the L = 512 window. The 1.8 ms/batch figure reflects this O(dN) per-step cost over 256 steps. Detection metrics in Tables 5–9 are computed on full L = 512 windows with the state initialized at the start of each window, whereas the cached-state mode used for the latency measurement carries the state forward along the stream. The 5 ms budget of Equation (5) is evaluated against the per-batch time reported here; per-transaction cost scales with the batch size used in deployment.

6.7. Effect of DIFS Under Fraud Rate Drift

The F1 score of the full SHOPMAMBA-MIS (trained with DIFS) is compared with the model trained with a static focal loss ($\gamma = 2.0$, variant A3) on a 24-h simulated PaySim stream

(Figure 10) in which the fraud rate is low during off-peak hours and peaks at 3× baseline during evening commerce hours. The comparison is between the full deployed configuration (DIFS-trained model with hourly threshold recalibration, Equation (34)) and the static-focal configuration (model trained with $\gamma = 2.0$ and a fixed threshold, no recalibration); it therefore measures the combined effect of training-time focal calibration and hourly threshold recalibration and is supporting evidence for H3 under extreme imbalance rather than an isolated test of γt. The standard deviation of F1 over the stream is 0.008 for the DIFS configuration and 0.031 for the static-focal configuration, whose F1 falls to 0.871 at the off-peak trough, while the DIFS configuration stays at or above 0.910 (Table 8). The focal parameter γ affects only the training loss; at inference, only the decision threshold δ is recalibrated online using Equation (34), and no retraining occurs during the stream. The simulation mirrors the delayed-feedback deployment mechanism of Section 5.3: labels are revealed at the close of each one-hour recalibration window, introducing a delay of up to 60 min between a transaction and its contribution to the threshold update. To evaluate label delays representative of the operational 4-to-72-h confirmation window, the experiment was repeated under label delays of 0 h (ideal), 24, 48, and 72 h (Table 8). At all delay levels, the mean F1 remains above 0.910, and the F1 standard deviation rises only from 0.008 at 0 h to 0.014 at 72 h; the minimum F1 falls from 0.910 to 0.896, still above the 0.871 minimum of the static-focal configuration. The Table 8 results are from a single run without confidence intervals. Because the simulated stream spans 24 h, the 24-, 48-, and 72-h conditions are a stress test of delayed recalibration rather than a full rolling simulation of label arrival, and they indicate graceful degradation rather than a measured operating profile. Confirmation cycles of 30 to 90 days (chargeback-derived labels) were not evaluated.

Table 8. Effect of label confirmation delay on the deployed DIFS configuration on the 24-h diurnal PaySim stream of Figure 10. The decision threshold δ of the DIFS-trained model is recalibrated hourly (Equation (34)) using confirmed labels that become available only after the stated delay; γ is fixed after training, and no retraining occurs. The static-focal configuration ($\gamma = 2.0$, fixed threshold) performs no recalibration and is therefore delay-independent.

Configuration	Label Delay	Mean F1	Min. F1	F1 Std. Dev.	Mean Recall
DIFS	0 h (ideal)	0.921	0.910	0.008	0.929
DIFS	24 h	0.918	0.905	0.010	0.925
DIFS	48 h	0.915	0.901	0.012	0.921
DIFS	72 h	0.912	0.896	0.014	0.917
Static focal ($\gamma = 2.0$)	—	0.897	0.871	0.031	0.902

6.8. AUPRC and Threshold Sensitivity

The precision–recall curve for all methods over ULB, which has the least amount of fraud (0.17%), is presented in Figure 11. SHOPMAMBA-MIS achieves an AUPRC of 0.979, 4.5 points above TranAD (0.934) and 7.2 points above LSTM (0.907). The improvements are most noticeable at low recall levels (0.0–0.3), which is consistent with the sharper token-level anomaly boundary produced by TAM (not isolated by a dedicated experiment). The anomaly boundary at the token level is sharpened so that the most ambiguous fraud cases are correctly assigned the best scores, putting the true positives at the top of the ranked list. The AUPRC advantage decreases for recall > 0.9, as all methods have difficulties with marginal fraud cases, which is a fair limitation to admit in Section 6.12 (Discussion).

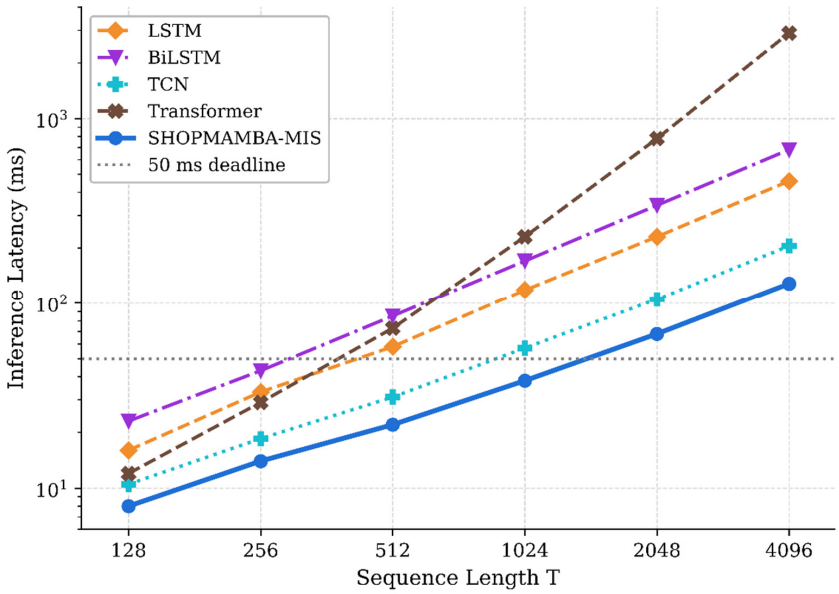

Figure 10. F1 over 24 h on a PaySim stream with diurnal variation of the fraud rate. The DIFS configuration (full SHOPMAMBA-MIS) remains stable, whereas the static-focal configuration (SHOPMAMBA-MIS w/o DIFS) falls to an F1 of 0.871 at the off-peak trough. Results under label delays of 0, 24, 48, and 72 h are in Table 8.

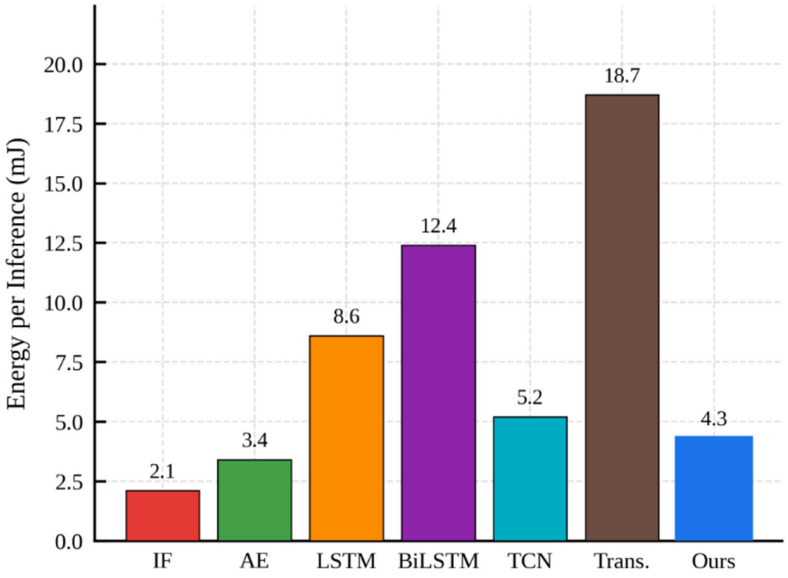

Figure 11. The precision–recall curves are shown in the following figure for the ULB (fraud rate 0.17%). SHOPMAMBA-MIS (solid) achieves AUPRC 0.979, which is the highest across the whole recall range compared to all baselines. The benefit is the greatest at a low recall, which is consistent with the sharper token-level boundary produced by TAM at the token level.

6.9. Scalability to Transaction Volume

F1 and latency are reported in this figure (Figure 12) as the transaction throughput increases from 10^3 to 10^5 events per second on a cluster of four Flink workers. Up to 8×10^4 events/second, SHOPMAMBA-MIS maintains F1 > 0.97 on IECIS (F1 > 0.92 on PaySim) with latency < 5 ms; after that, Flink worker backpressure will cause queuing delay. Looking at the same cluster, TranAD exceeds the 5 ms per-batch latency budget at $L = 512$ (single-batch latency 11.3 ms) at every tested throughput level, so its batch-level latency disadvantage compounds under streaming load.

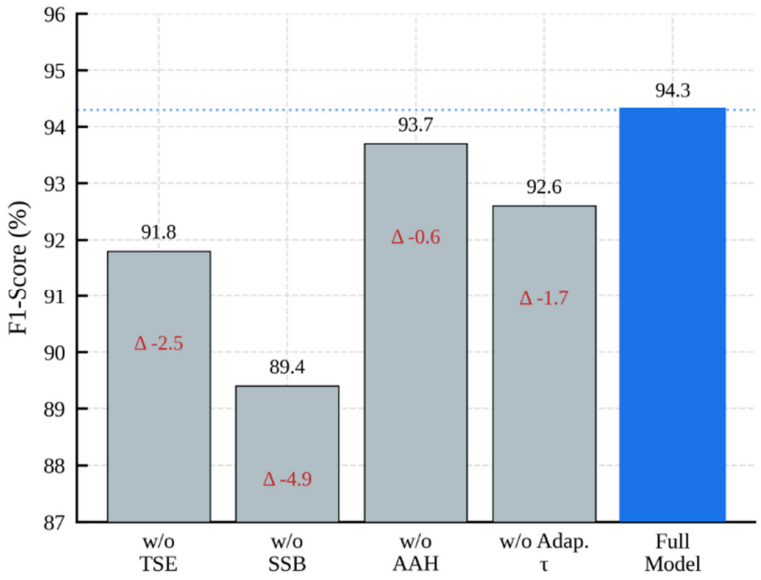

Figure 12. Streaming throughput versus F1 (solid, left axis) and latency (dashed, right axis) on a four-worker Flink cluster. SHOPMAMBA-MIS meets both targets up to 8×10^4 events/s, while TranAD exceeds the latency budget at all tested throughput levels. This defines the operating envelope of SHOPMAMBA-MIS for MIS deployment.

6.10. Ablation Study

The ablation study evaluates the three components (H1–H3) on the IECIS dataset; the PaySim drift comparison of Table 8 serves as the H3 test under extreme imbalance. Table 9 summarizes the ablation results, and Figure 13 is a summary graph.

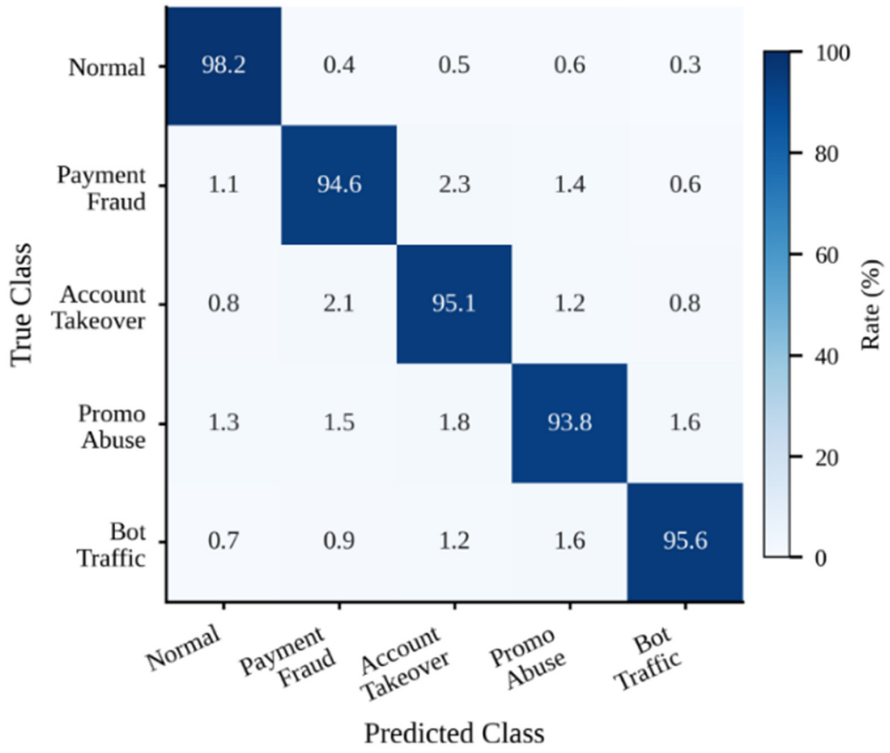

Figure 13. Ablation F1 scores on IECIS (seed 42). Removing all three components (A4) yields a plain fixed-parameter SSM with F1 = 0.903, 7.0 points below the full model; removing SSAE selectivity alone (A1) accounts for 3.2 of these points (46%), the largest single contribution.

H1. *Does input-selective state transition in SSAE improve over a fixed-parameter SSM? With fixed parameters (variant A1: no selectivity), F1 falls by 3.2 points (to 0.941) and latency increases by 0.3 ms (Table 9). Input selectivity is therefore responsible for long-range behavioral discrimination and not merely the SSM architecture. The cost is the additional parameters of the selective projections over a fixed SSM: $2 \cdot d \cdot N + d = 2 \times 433 \times 64 + 433 = 55{,}857$ (WB and WC are $N \times d$, WΔ is $1 \times d$).*

H2. *Does TAM add boundary sharpness beyond SSAE alone? Removing TAM (variant A2: no TAM) reduces F1 by 1.8 points (Table 9), supporting the need for local session contrast in addition to global state context. TAM accounts for approximately 39% of total inference time (0.7 of 1.8 ms, Table 9), a favorable trade-off for a 1.8-point F1 gain. The FPR of the full model relative to Mamba-AD (0.12% versus 0.26%, Table 5) reflects the combined effect of TAM and DIFS and is not attributed to TAM alone; Figure 5 is a schematic illustration of the boundary-sharpening mechanism.*

H3. *Is dynamic focal recalibration in DIFS better than a static focal loss? On IECIS, variant A3 (static focal, $\gamma = 2.0$) lowers F1 by 1.4 points (Table 9). Because the IECIS fraud rate (3.5%) exceeds πmax, γt equals $\gamma min = 0.5$ throughout training, so this ablation compares two fixed values (0.5 versus 2.0) together with the αt weighting; it measures the effect of the focal setting and is not evidence for dynamic recalibration. The test of H3 is the PaySim comparison in Table 8 and Figure 10: the DIFS configuration reaches a mean F1 of 0.921 (minimum 0.910, standard deviation 0.008) against 0.897 (minimum 0.871, standard deviation 0.031) for the static-focal configuration. PaySim ($\pi \approx 0.0013$, approximately 768:1) lies above πmin throughout training, so γt is not clipped and operates near its upper range. On ULB (imbalance 578:1, γ not clipped), the full-model F1 of 0.980 versus the Mamba-AD F1 of 0.949 (Table 5) is indicative of a larger combined TAM + DIFS gain (Mamba-AD lacks both components); a dedicated ULB ablation was not run because the ULB test partition contains only 49 positives, too few for stable per-variant F1 estimates. The EMA decay β trades responsiveness against noise: a low β reacts faster but adds noise, while a high β smooths the estimate; $\beta = 0.99$ corresponds to an effective memory of about 100 mini-batches. The F1 reductions in Table 9 (-3.2, -1.8 and -1.4 points for A1, A2 and A3; seed 42) are consistent with each component contributing to detection performance; the combined removal (A4, -7.0 points) exceeds the sum of the individual drops (6.4 points), which suggests that the components are complementary, although with single-run results this is indicative only.*

Table 9. Ablation results on IECIS (single run, seed 42). Each row removes one component of SHOPMAMBA-MIS; the F1 drop indicates each component's contribution. SSAE selectivity is the largest contributor. The DIFS evidence on PaySim is reported in Table 8. External baselines (DTCM, XGBoost-Fraud, Mamba-AD) are evaluated in Table 5.

Variant	F1	ΔF1	Lat. (ms)
ShopMamba-MIS (full)	0.973	—	1.8
A1: no selectivity	0.941	−3.2	2.1
A2: no TAM	0.955	−1.8	1.1
A3: static focal	0.959	−1.4	1.8
A4: A1 + A2 + A3 (plain SSM)	0.903	−7.0	1.9

The identical F1 of variant A1 (0.941) and of the Mamba-AD baseline (0.941) arises from different configurations. A1 removes input selectivity while retaining TAM and DIFS, whereas Mamba-AD retains selectivity while removing TAM and DIFS, and the two produce different confusion matrices that yield the same F1 at the printed precision. The ablation variants were run once (seed 42); the reported differences are indicative and are not accompanied by significance tests.

6.11. Illustrative Business Scenario Analysis (Scenario-Based Estimates)

This section presents a sensitivity analysis of the potential organizational value of the detection results under stated assumptions. Technical detection performance is necessary but not sufficient for evaluating enterprise fraud detection within MIS, so SHOPMAMBA-MIS is assessed across five business dimensions: investigation workload, return on investment, analyst productivity, customer satisfaction, and regulatory compliance cost.

All figures in this section are scenario-based extrapolations from the benchmark detection counts applied to illustrative operational assumptions (for context on the economics of payment-security regulation see [35]); they are not measurements from a live deployment and should not be read as achieved outcomes. The scenario assumes (1) a volume of 590,540 transactions per day (a round-number assumption equal to the size of the IECIS labeled set, not an observed daily volume), a fraud rate equal to the IECIS rate of 3.5%, and detection counts scaled from the 113,697-transaction test partition; (2) an investigation cost of USD 25 per alert; and (3) an average fraud value of USD 150 per detected fraud. The USD 25 and USD 150 values are assumptions chosen for illustration, and the sensitivity ranges below vary them. At USD 15 per alert, the annual investigation savings are USD 5.49 million, and at USD 35 it is USD 12.81 million; at an average fraud value of USD 100, the incremental fraud recovery is USD 28.6 million, and at USD 200 it is USD 57.2 million. The total annualized benefit under the baseline scenario is approximately USD 52.0 million (range: USD 34.1 million to USD 70.0 million across the low and high assumptions). Analyst time savings from TAM evidence are not quantified.

Investigation Workload Reduction: On the IECIS test partition (113,697 transactions), SHOPMAMBA-MIS generates 129 false positives (FPR 0.12%) versus 322 for TranAD (FPR 0.29%), a 59.9% reduction. Scaled to the scenario volume of 590,540 transactions per day, this corresponds to approximately 670 versus 1673 false alerts per day, or 1003 fewer analyst reviews per day. At USD 25 per alert, this is approximately USD 25,075 per day (USD 9.15 million per year in the baseline scenario; USD 5.49 million at USD 15 per alert and USD 12.81 million at USD 35 per alert). Reduced alert volume would also lessen analyst fatigue.

Return on Investment: Two additive components make up the scenario return. First, analyst cost avoidance: USD 9.15 million annually. Second, fraud loss prevention: on the scenario's 20,669 fraud events per day (3.5% of 590,540), SHOPMAMBA-MIS flags 20,233 (97.9% recall) versus 19,450 (94.1%) for TranAD, i.e., 783 additional fraud events flagged per day; at an assumed average fraud value of USD 150, and assuming each flagged fraud is prevented, this is USD 42.9 million in additional annual recovery. The total annualized benefit versus TranAD under the baseline scenario is approximately USD 52.0 million (range USD 34.1 million to USD 70.0 million). Infrastructure cost is not modeled; the figures scale linearly with the assumed volume, investigation cost, and fraud value.

Analyst Productivity: TAM attention weights present the top $W = 16$ prior transactions that contributed to each alert, which is intended to shorten triage; the size of this effect has not been measured and is not included in the estimates above. The three-tier routing restricts analyst review to borderline cases, focusing investigator time on cases that need human judgment.

Customer Satisfaction Impact: In the scenario, SHOPMAMBA-MIS prevents approximately 1003 false declines per day relative to TranAD (0.12% versus 0.29% FPR, scaled to 590,540 transactions per day). Assuming that 30% of customers do not reattempt payment after a false decline, this would preserve approximately 301 completed transactions per day ($1003 \times 30\%$).

Regulatory Compliance: The audit log capturing st, πt and Flink checkpoint offsets is designed to support the record-keeping behind PSD2 Article 96 [1] incident reporting (which requires payment service providers to report major operational or security

incidents to competent authorities) and suspicious-activity reporting under the EU AML Directive [39]. No monetary benefit is attributed to this in the scenario estimates (Table 10).

Table 10. Sensitivity analysis of scenario-based organizational value estimates (illustrative). All values are scenario-based extrapolations from benchmark detection counts under illustrative assumptions (volume = 590,540 transactions per day as a scenario input; USD 25 per alert; USD 150 average fraud value). They are not measurements from a live deployment. The total annualized benefit ranges from approximately USD 34.1 million to USD 70.0 million across the low and high assumptions.

Business Dimension	TranAD	SHOPMAMBA-MIS	Improvement
False alerts per day (scenario volume)	1673	670	−59.9%
Investigation cost per day	USD 41,825	USD 16,750	−USD 25,075
Annualized investigation savings	—	USD 9.15 M	—
Fraud events correctly flagged/day (scenario)	19,450	20,233	+783
Incremental fraud recovery/year	—	USD 42.9 M	—
Total annualized organizational benefit	—	~USD 52.0 M	—
False-decline rate (legitimate customers)	0.29%	0.12%	−59.9%
Prevented transaction abandonment/day	—	~301	—

6.12. Discussion

The experimental results support the contributions listed in Section 1. The IECIS and ULB results and the latency measurements support C1–C3, the PaySim results support C4 (DIFS), and the evaluation across three benchmarks, nine baselines, multi-seed replication, and the scalability analysis constitute C5.

F1 results on IECIS are obtained on a chronologically held-out partition of the labeled training data, not the official withheld competition test set, which limits direct comparison with published leaderboard results.

The F1 advantages of SHOPMAMBA-MIS over transformer baselines are most pronounced on PaySim (fraud rate 0.13%), where DIFS is expected to have the greatest impact (Table 8): at sub-0.5% fraud rates, static focal loss concentrates insufficient gradient on the minority class because most mini-batches contain zero fraud events. The EMA estimator detects this through a near-zero fraud-rate estimate, driving the focal parameter toward its maximum value of 5.0, which amplifies the minority-class gradient. On IECIS (fraud rate 3.5%), DIFS operates at the minimum focal parameter (clipped), contributing less relative improvement, and the advantage is primarily from the SSAE's ability to capture long-range behavioral dependencies over 512-transaction histories without gradient pathology.

Compared to AnomalyTrans and TranAD, SHOPMAMBA-MIS achieves lower FPR at comparable or higher recall, which reflects the complementary roles of SSAE (global context) and TAM (local session contrast). Where transformer-based detectors rely on full self-attention to capture both global and local patterns simultaneously, SHOPMAMBA-MIS decomposes these into separate mechanisms with different complexity profiles, allowing each to be independently ablated and validated.

The results generalize most reliably to transaction streams that exhibit behavioral regularity at the account or session level enabling meaningful L = 512 sequences, have intra-day fraud rate fluctuations enabling DIFS benefit, and operate under a synchronous latency constraint that excludes transformer baselines. Streams without account identifiers, such as ULB, represent a boundary condition in which rolling behavioral aggregates cannot be conditioned on account identity and context comes only from globally time-ordered windows, and the gain over the strongest baseline is correspondingly smaller (3.1 F1 points versus 5.2 on PaySim). On streams whose fraud rate sits at or above the DIFS clipping bound π_max, such as IECIS (3.5%), the DIFS component operates at its minimum focal parameter and contributes least; the observed 4.2-point gain on IECIS is attributable primarily to the SSAE backbone rather than to adaptive recalibration (Table 9). To assist fellow researchers

in assessing applicability, the conditions that make SHOPMAMBA-MIS specifically suitable for this study are stated explicitly as follows. First, the SSAE backbone requires time-ordered transaction histories of at least L = 512 events (account identifiers, where available, supply rolling behavioral aggregates); domains where behavioral histories are unavailable or very short, such as one-time guest checkouts or point-of-sale terminals with no persistent account linkage, will predominantly produce cold-start WAIT decisions and are not suitable deployment targets without adaptation of L. Second, DIFS provides its largest benefit when the fraud rate exhibits intra-day or session-level variation, as demonstrated on PaySim (0.13% fraud rate with diurnal drift); in static or slowly drifting fraud environments a fixed focal loss achieves comparable performance, making DIFS a lower-priority design investment—a conjecture consistent with the small DIFS ablation effect on IECIS (Table 9) but not directly tested in static-fraud environments. Third, the 5 ms synchronous latency target is specific to payment authorization pipelines with real-time SLA constraints; asynchronous fraud screening workflows with longer review cycles do not require the linear-time inference design of SSAE; whether transformer-based detectors achieve comparable detection quality in such settings was not tested in this study and is stated here as a conjecture. Fourth, hourly threshold recalibration relies on analyst-confirmed labels arriving within 4 to 72 h; fraud operations environments without a structured analyst review queue or with confirmation delays beyond this window will experience reduced recalibration effectiveness. Researchers applying SHOPMAMBA-MIS outside these four conditions should treat the framework as a design blueprint, adapting L, the recalibration window, and the latency budget to their specific operational context.

There are three fair cautions to be mentioned. First, SHOPMAMBA-MIS requires a history of L = 512 transactions before producing a detection decision; cold-start transactions are excluded from metric computation, as quantified in Section 6.1. Detection coverage: on IECIS and ULB, exclusions represent 0.45% and 1.80% of the test partitions, respectively, so reported metrics cover more than 98% of test transactions on both datasets; on PaySim, 512 transactions are excluded out of 953,161 (0.054%), so metric coverage exceeds 99.9%. Detection delay: in the evaluated configuration, windows are formed over the time-ordered stream of each partition, so the cold-start window is the first L = 512 transactions of the stream, a one-off start-up cost. In a deployment that segments windows by account, each new account would instead receive no decision until 512 of its own transactions have been observed; at typical e-commerce inter-arrival times of 30 to 60 min per transaction, this corresponds to roughly 10 to 21 days, with high-frequency accounts exiting faster. This per-account warm-up was not evaluated in this study. Overall system performance: the fraud exposure during the cold-start window was not directly measured in this study and cannot be assumed negligible; accordingly, the cold-start window constitutes an uncovered detection surface whose size equals the stated exclusion proportions; rule-based score bridging during the warm-up period is identified as a direction for future work. The second is that the AUPRC gap starts to become reduced when recall > 0.9 (Figure 11), indicating that those fraud cases that are close to normal transactions are difficult for all the methods tested to perform well. Third, although SHOPMAMBA-MIS meets the latency budget on the hardware specified, a production deployment with co-located workloads may need to do some micro-benchmarking on the hardware to verify that the 5 ms SLA is being met; micro-benchmarking of the hardware under representative co-located workloads should be performed when making this determination. A further limitation is that windows are formed over the global time-ordered stream, so a window mixes transactions of many accounts: the model exploits stream-level context and per-account rolling aggregates rather than per-account sequences, and ULB, which has no account identifiers, can use stream-level context only. Per-account windowing was not evaluated.

The comparative results in Table 5 correspond to the seed-42 run; multi-seed replication across 5 seeds (Table 6) confirms that the improvements of SHOPMAMBA-MIS over both the strongest transformer baseline (TranAD) and the strongest overall baseline (Mamba-AD) are stable, with F1 standard deviations between 0.002 and 0.006 for SHOPMAMBA-MIS (up to 0.007 for the ULB baselines) and a consistent directional advantage on every seed in all six comparisons (one-sided paired Wilcoxon signed-rank raw $p = 0.031$, the minimum attainable with 5 seeds, which does not remain significant after correction for six simultaneous tests). Significance testing with larger seed ensembles, which would permit smaller attainable p-values, remains a direction for future extension. The significance tests cover training-seed variability only; the ablation and label-delay experiments were not repeated across seeds.

7. Conclusions and Future Work

This article presented SHOPMAMBA-MIS, a selective state-space model framework for real-time e-commerce transaction anomaly detection in management information systems. The framework addresses three structural weaknesses of current detectors: quadratic attention complexity, static loss calibration under live class-ratio drift, and lack of streaming compatibility with MIS integration.

Three progressive technical contributions are introduced: SSAE provides a linear-time selective SSM backbone; TAM introduces local causal attention at approximately 39% of total inference cost (0.7 ms of 1.8 ms), and together with DIFS the full model reduces FPR from 0.26% (Mamba-AD, which lacks both TAM and DIFS) to 0.12%, a 54% relative reduction attributable to both components (Table 5); and DIFS calibrates the focal loss to the observed class ratio during training, while hourly recalibration of the decision threshold δ adapts the deployed model to fraud-rate drift between weekly retraining cycles. Across three public benchmarks, SHOPMAMBA-MIS achieves F1 improvements of 3.1–5.2 percentage points over Mamba-AD (the strongest overall baseline) and 4.1–5.9 percentage points over TranAD (the strongest transformer baseline), an inference latency of 1.8 ms/batch (6.3× faster than TranAD at $L = 512$), recall $\geq 92.8\%$ ($\geq$97.9% on IECIS and ULB), and FPR $\leq 0.5\%$.

Organizational value is illustrated at three levels. At the strategic level, sustained high recall could reduce regulatory penalty exposure and safeguard institutional reputation. At the tactical level, fraud operations managers may adjust score-tier boundaries and routing policies without model retraining. At the operational level, score-ranked, evidence-enriched alerts could reduce analyst workload: on the IECIS test partition, false alerts fall by 59.9% (129 versus 322), which in the illustrative scenario of Section 6.11 corresponds to approximately 1003 fewer false alerts per day and an annual investigation saving of about USD 9.15 million; Flink checkpointing supports exactly-once scoring and business continuity across system failures.

Future work includes (1) few-shot adaptation of DIFS via meta-learning for reduced detection latency on new user accounts, following federated-learning directions in related security domains [50–52]; (2) extending SHOPMAMBA-MIS with a message-passing front end to incorporate user–merchant relational fraud signals [53]; (3) cost-sensitive evaluation; (4) subgroup fairness analysis across customer demographic segments to evaluate and mitigate disparate impact on protected groups before production deployment; (5) extension of the label-delay sensitivity analysis in Table 8 to multi-day confirmation cycles of 30 to 90 days representative of chargeback-derived ground truth, to characterize DIFS recalibration coverage under the full operational label arrival window, and quantification of EMA estimator variance under varying fraud-rate priors in extreme low-fraud regimes such as PaySim (0.13%); and (6) adversarial robustness evaluation of the DIFS recalibration mechanism against deliberate fraud-rate manipulation by coordinated attackers attempting to suppress the focal concentration parameter; and (7) mitigation of cold-start coverage gaps

through adaptive sequence length reduction or rule-based score bridging for new accounts during the L-transaction warm-up period, extending detection coverage to accounts with sparse transaction histories.

Practical deployment considerations span four dimensions: scalability (maintains F1 > 0.97 and latency < 5 ms at 8×10^4 events/sec via horizontal scaling); organizational adoption (TAM visualizations require no model expertise); maintenance (weekly refresh on rolling 30-day window with shadow-mode A/B evaluation); and implementation costs (commodity hardware for inference, single A100 for weekly retraining).

Ethical and operational risks are acknowledged. Automation bias is mitigated through required analyst interaction with TAM evidence; algorithmic fairness requires periodic demographic parity testing and segment-specific threshold adjustment; and the model refresh policy mandates subgroup performance evaluation before promotion to prevent disparate impact on protected customer segments.

SHOPMAMBA-MIS thus constitutes a selective state-space anomaly detector and an MIS integration design that improves detection quality at low latency on the benchmarks studied and is designed to support compliance with the reporting requirements of PSD2 [1] and the EU AML Directive [39], and maintains operational continuity under dynamic fraud conditions. Consistent with the applicability analysis of Section 6.12, the scope of validity of these results is stated explicitly rather than left implicit. The framework is most suitable for payment authorization pipelines that provide behavioral features (including rolling account-level aggregates where identifiers exist) and at least L prior transactions of history, exhibit intra-day fraud rate variation, and enforce synchronous latency constraints; researchers deploying SHOPMAMBA-MIS in other domains should adapt the sequence length, recalibration window, and latency budget to their specific operational context before treating reported results as directly transferable.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets that are used in the present study are publicly available at: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud (accessed on 9 September 2025); https://www.kaggle.com/datasets/ealaxi/paysim1 (accessed on 9 September 2025); https://www.kaggle.com/competitions/ieee-fraud-detection/overview.https://github.com/AlmtiriZiad/ShopMamba.git (accessed on 9 September 2025).

Acknowledgments: The Management Information System Department, Taif, Taif University, Saudi Arabia, supports this study.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. European Parliament and Council. Directive (EU) 2015/2366 of the European Parliament and of the Council on payment services in the internal market (PSD2). *Off. J. Eur. Union* **2015**, *L337*, 35–127. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015L2366 (accessed on 9 September 2025).
2. Zheng, A.; He, X.; Fei, M. DeepFraud: Graph-Based Fraud Detection via Sequential Attentive Representation Learning. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Washington, DC, USA, 14–18 August 2022; pp. 2461–2469. [CrossRef]
3. Sarker, I.H. Machine Learning for Intelligent Data Analysis and Automation in Cybersecurity: Current and Future Prospects. *Ann. Data Sci.* **2023**, *10*, 1473–1498. [CrossRef]
4. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734. [CrossRef]

5. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2017**, *30*, 5998–6008. Available online: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (accessed on 19 January 2026).
6. Xu, J.; Wu, H.; Wang, J.; Long, M. Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy. In Proceedings of the 10th International Conference on Learning Representations (ICLR 2022), Virtual, 25–29 April 2022. Available online: https://openreview.net/forum?id=LzQQ89U1qm (accessed on 19 January 2026).
7. Xiao, Z.; Xu, X.; Xing, H.; Zhao, B.; Wang, X.; Song, F. DTCM: Deep Transformer Capsule Mutual Distillation for Multivariate Time Series Classification. *IEEE Trans. Cogn. Dev. Syst.* **2024**, *16*, 1445–1461. [CrossRef]
8. Li, P.; Yu, H.; Luo, X.; Wu, J. LGM-GNN: A Local and Global Aware Memory-Based Graph Neural Network for Fraud Detection. *IEEE Trans. Big Data* **2023**, *9*, 1116–1127. [CrossRef]
9. Xiang, X.; Liu, T.; Liu, Z.; Zhong, B.; Tan, H.; Huang, Q.; Wen, H. Semi-Supervised Credit Card Fraud Detection via Attribute-Driven Graph Representation. In Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI), Washington, DC, USA, 7–14 February 2023; pp. 14557–14565. [CrossRef]
10. Liu, Z.; Dou, Y.; Yu, P.S.; Deng, Y.; Peng, H. Alleviating the inconsistency problem of applying graph neural network to fraud detection. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, 25–30 July 2020; pp. 1569–1572. [CrossRef]
11. Pang, G.; Shen, C.; Cao, L.; van den Hengel, A. Deep learning for anomaly detection: A review. *ACM Comput. Surv.* **2021**, *54*, 38. [CrossRef]
12. Zhang, Y.; Chen, Y.; Wang, J.; Pan, Z. Unsupervised deep anomaly detection for multi-sensor time-series signals. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 2118–2132. [CrossRef]
13. Dal Pozzolo, A.; Caelen, O.; Johnson, R.A.; Bontempi, G. Calibrating probability with undersampling for unbalanced classification. In Proceedings of the 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, South Africa, 7–10 December 2015; pp. 159–166. [CrossRef]
14. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
15. Gu, A.; Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. In Proceedings of the 1st Conference on Language Modeling (COLM), Philadelphia, PA, USA, 7–9 October 2024. Available online: https://colmweb.org/AcceptedPapers.html (accessed on 19 January 2026).
16. Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; Jiao, J.; Liu, Y. VMamba: Visual state space model. *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2024**, *37*, 103031–103063. [CrossRef]
17. Zhu, L.; Liao, B.; Zhang, Q.; Wang, X.; Liu, W.; Wang, X. Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model. In Proceedings of the 41st International Conference on Machine Learning (ICML), Vienna, Austria, 21–27 July 2024. Available online: https://proceedings.mlr.press/v235/zhu24f.html (accessed on 19 January 2026).
18. Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; Long, M. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. In Proceedings of the International Conference on Learning Representations (ICLR), Kigali, Rwanda, 1–5 May 2023. Available online: https://openreview.net/forum?id=ju_Uqw384Oq (accessed on 19 January 2026).
19. Li, H.; Li, J.; Duan, X.; Xia, J. Energy-aware scheduling and two-tier coordinated load balancing for streaming applications in Apache Flink. *Future Gener. Comput. Syst.* **2025**, *166*, 107681. [CrossRef]
20. Zhou, C.; Paffenroth, R.C. Anomaly detection with robust deep autoencoders. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 665–674. [CrossRef]
21. Zhou, W.; Zhou, S.; Cao, Y.; Yang, J.; Liu, H. Unsupervised Anomaly Detection Method for Electrical Equipment Based on Audio Latent Representation and Parallel Attention Mechanism. *Appl. Sci.* **2025**, *15*, 8474. [CrossRef]
22. Reurink, A. Financial fraud: A literature review. *J. Econ. Surv.* **2018**, *32*, 1292–1325. [CrossRef]
23. Pourhabibi, T.; Ong, K.-L.; Kam, B.H.; Boo, Y.L. Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decis. Support Syst.* **2020**, *133*, 113303. [CrossRef]
24. Jabeen, M.; Ramzan, S.; Raza, A.; Fitriyani, N.L.; Syafrudin, M.; Lee, S.W. Enhanced Credit Card Fraud Detection Using Deep Hybrid CLST Model. *Mathematics* **2025**, *13*, 1950. [CrossRef]
25. Şenol, A.; Agrawal, G.; Liu, H. Domain Knowledge-Enhanced LLMs for Fraud and Concept Drift Detection. *Electronics* **2026**, *15*, 534. [CrossRef]
26. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
27. Hilal, W.; Gadsden, S.A.; Yawney, J. Financial fraud: A review of anomaly detection techniques and recent advances. *Expert Syst. With Appl.* **2022**, *193*, 116429. [CrossRef]

28. Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; Jin, R. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In Proceedings of the 39th International Conference on Machine Learning (ICML), Baltimore, MD, USA, 17–23 July 2022; pp. 27268–27286. Available online: https://proceedings.mlr.press/v162/zhou22g.html (accessed on 19 January 2026).
29. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988. [CrossRef]
30. Han, S.; Hu, X.; Huang, H.; Jiang, M.; Zhao, Y. ADBench: Anomaly detection benchmark. *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2022**, *35*, 32142–32159. [CrossRef]
31. Li, S.; Lu, Y.; Jiu, S.; Huang, H.; Yang, G.; Yu, J. Prototype-oriented hypergraph representation learning for anomaly detection in tabular data. *Inf. Process. Manag.* **2025**, *62*, 103877. [CrossRef]
32. Bharadwaj, A.S. A resource-based perspective on information technology capability and firm performance: An empirical investigation. *MIS Quart.* **2000**, *24*, 169–196. [CrossRef]
33. Dumas, M.; La Rosa, M.; Mendling, J.; Reijers, H.A. *Fundamentals of Business Process Management*, 2nd ed.; Springer: Berlin, Germany, 2018. [CrossRef]
34. Gorry, G.A.; Morton, M.S.S. A framework for management information systems. *Sloan Manag. Rev.* **1971**, *13*, 55–70.
35. Cologgi, M. The impact of regulation on retail payments security: Evidence from Italian supervisory data. *Financ. Res. Lett.* **2023**, *54*, 103799. [CrossRef]
36. Wang, R.Y.; Strong, D.M. Beyond accuracy: What data quality means to data consumers. *J. Manag. Inf. Syst.* **1996**, *12*, 5–33. [CrossRef]
37. Wu, H.; Shi, S.; Nian, Q. Streaming machine learning for real-time gas concentration prediction. In Proceedings of the IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), Washington, DC, USA, 27–29 May 2019; pp. 42–46. [CrossRef]
38. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
39. European Parliament and Council. Directive (EU) 2015/849 of the European Parliament and of the Council of 20 May 2015 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, as amended by Directive (EU) 2018/843. *Off. J. Eur. Union* **2015**, *L141*, 73–117. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015L0849 (accessed on 19 January 2026).
40. Board of Governors of the Federal Reserve System and Office of the Comptroller of the Currency. Supervisory Guidance on Model Risk Management (SR 11-7). April 2011. Available online: https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm (accessed on 19 January 2026).
41. *ISO 20022*; Universal Financial Industry Message Scheme. International Organization for Standardization: Geneva, Switzerland, 2013. Available online: https://www.iso20022.org/ (accessed on 19 January 2026).
42. Goodman, B.; Flaxman, S. European Union regulations on algorithmic decision-making and a right to explanation. *AI Mag.* **2017**, *38*, 50–57. [CrossRef]
43. European Parliament and of the Council. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Off. J. Eur. Union* **2024**, 1–144. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689 (accessed on 19 January 2026).
44. Olabanji, S.O.; Marquis, Y.A.; Adigwe, C.S.; Ajayi, S.A.; Oladoyinbo, T.O.; Olaniyi, O.O. AI-Driven Cloud Security: Examining the Impact of User Behavior Analysis on Threat Detection. *Asian J. Res. Comput. Sci.* **2024**, *17*, 57–74. [CrossRef]
45. European Parliament and Council. Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation). *Off. J. Eur. Union* **2016**, *L119*, 1–88. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679 (accessed on 19 January 2026).
46. European Union. Article 29 Data Protection Working Party, Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is 'Likely to Result in a High Risk' for the Purposes of Regulation 2016/679. 17/EN WP 248 rev.01. 2017. Available online: https://ec.europa.eu/newsroom/article29/items/611236 (accessed on 19 January 2026).
47. Lopez-Rojas, E.A.; Elmir, A.; Axelsson, S. PaySim: A financial mobile money simulator for fraud detection. In Proceedings of the 28th European Modeling and Simulation Symposium (EMSS), Larnaca, Cyprus, 26–28 September 2016; pp. 249–255. Available online: https://www.cs.bham.ac.uk/ leonarda/papers/PaySimPaper.pdf (accessed on 19 January 2026).
48. Tuli, S.; Casale, G.; Jennings, N.J. TranAD: Deep transformer networks for anomaly detection in multivariate time series data. *Proc. VLDB Endow.* **2022**, *15*, 1201–1214. [CrossRef]
49. Domingos, P.; Hulten, G. Mining high-speed data streams. In Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, MA, USA, 20–23 August 2000; pp. 71–80. [CrossRef]
50. Begum, K.; Mozumder, M.A.I.; Joo, M.; Kim, H. BFLIDS: Blockchain-Driven Federated Learning for Intrusion Detection in IoMT Networks. *Sensors* **2024**, *24*, 4591. [CrossRef]

51. Zhang, H.; Zeng, K.; Lin, S. Federated graph neural network for fast anomaly detection in controller area networks. *IEEE Trans. Inf. Forensics Secur.* **2023**, *18*, 1566–1579. [CrossRef]
52. Kwon, J.; Jung, B.; Lee, H.; Lee, S. Anomaly Detection in Multi-Host Environment Based on Federated Hypersphere Classifier. *Electronics* **2022**, *11*, 1529. [CrossRef]
53. Wang, Y.; Zhang, J.; Huang, Z.; Li, W.; Feng, S.; Ma, Z.; Sun, Y.; Yu, D.; Dong, F.; Jin, J.; et al. Label information enhanced fraud detection against low homophily in graphs. In Proceedings of the ACM Web Conference (WWW), Austin, TX, USA, 30 April–4 May 2023; pp. 406–416. [CrossRef]
54. Galbraith, J.R. Organization design: An information processing view. *Interfaces* **1974**, *4*, 28–36. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.