Aridax International Journal of Research in Science & Technology
HJPARAM Publication | Open Access | Peer Reviewed

ISSN: XXXX-XXXX (Online)
Vol. 1, Issue 1 (Sep–Oct 2026), pp. 05–11
DOI: 10.XXXXX/hjparam.XXXX

RESEARCH ARTICLE Received: 03 October 2026 | Revised: 17 October 2026 | Accepted: 26 October 2026 | Published: 12 November 2026

A Multi-Modal Deep Learning Framework for Financial Fraud Detection Integrating Transactional, Behavioural and Network Data

Anurag Shrivastava[1,*] and Gourav Shrivastava[2]

[1] *Research Scholar, Department of Computer Science, SAGE University, Bhopal, Madhya Pradesh, India, ORCID: 0000-0002-6252-6724*
[2] *Associate Professor, Department of Computer Science, SAGE University, Bhopal, Madhya Pradesh, India*
*** Corresponding author:** *anurag.shri08@gmail.com*

Abstract

Background: Digital financial services face fraud that is adaptive, coordinated and increasingly hard to detect from transaction records alone. Research on sequence models, behavioural biometrics and graph neural networks has advanced separately, and systems that reason over all three kinds of evidence are scarce. This study specifies a unified framework that integrates transactional, behavioural and network data. **Methods:** Requirements were derived from the fraud detection literature and mapped to architectural components. Transaction histories are encoded by a recurrent or Transformer network, interaction signals such as keystroke and pointer dynamics by a temporal convolutional-recurrent network, and account–device–merchant relations by an inductive graph neural network. An attention layer assigns per-event weights to the three embeddings. Training uses a focal, cost-weighted loss for class imbalance and sliding-window retraining with threshold recalibration for concept drift. Shapley attributions and modality weights provide explanations. **Results:** The study delivers a modular architecture, an analytical comparison showing that none of eleven representative studies combines all three modalities, a requirement-to-component coverage map, and four falsifiable hypotheses with a time-aware evaluation protocol, baselines, ablations and statistical tests. No empirical performance is reported. **Conclusion:** The framework turns the case for multi-modal fraud detection into a testable design that addresses accuracy, adaptability and transparency together. Validation on linked multi-modal data is the required next step.

Keywords: *fraud detection; multi-modal deep learning; behavioural biometrics; graph neural networks; concept drift; explainable artificial intelligence*

1. Introduction

Digital payments, online banking and fintech platforms have moved most retail financial activity onto channels where the customer is never physically present. The same shift has widened the attack surface for fraud. Account takeover, synthetic identities and organised mule networks now sit alongside the stolen-card schemes that earlier detection systems were built for, and fraudsters change tactics as soon as a control starts to bite [1,2]. Detection is therefore a moving-target problem rather than a one-off classification exercise.

Most deployed systems still reason about one data source at a time. Rule engines and supervised classifiers score each transaction from its own attributes and a set of hand-built aggregates [3–5]. This design has three well-documented weaknesses. First, fraud is rare, so classifiers trained on raw data are biased toward the legitimate class [6]. Second, the statistical properties of both genuine and fraudulent activity drift over time, and labels arrive late because they depend on investigator feedback or customer disputes [7,8]. Third, a transaction that looks ordinary in isolation can be anomalous once the way the session was operated, or the accounts and devices it is connected to, are taken into account.

Three research streams address the third weakness, but largely in isolation from one another. Sequence models treat a customer's history as an ordered series and learn temporal spending patterns directly; Jurgovsky et al. showed that a long short-term memory (LSTM) network detects frauds that a random forest misses, particularly for card-present transactions [9,10]. Behavioural biometrics profile how a user types and moves a pointer, and have been shown to discriminate between individuals well enough to support continuous authentication [11,12]. Graph-based methods model relations among accounts, merchants and devices, first through network-derived features [13,14] and more recently through graph neural networks (GNNs) deployed at industrial scale [15–19].

Each stream captures evidence that the others cannot see. A sequence model has no view of collusion between accounts; a GNN has no view of whether the person at the keyboard is the account holder; a behavioural profile has no view of where the money goes. Multimodal machine learning offers a principled vocabulary for combining such heterogeneous signals, distinguishing the problems of representation, alignment and fusion [20], yet recent

© 2026 The Author(s). Published by HJPARAM Publication under CC BY 4.0

systematic reviews of fraud detection report that studies overwhelmingly rely on a single data type [2,21]. The practical obstacles are real: the three modalities differ in structure, sampling rate and availability, and any model that financial institutions can deploy must also be explainable to analysts and regulators [22].

This article addresses that gap. Its aim is to specify a unified multi-modal deep learning framework that jointly models transactional, behavioural and network data for fraud detection, and to set out an evaluation protocol rigorous enough to establish whether the integration actually helps. The contributions are fourfold: (i) a modular architecture with one encoder per modality and an attention-based fusion layer that weights modalities per decision; (ii) a training strategy that handles extreme class imbalance and concept drift together; (iii) an explanation layer that reports feature-level and modality-level attributions; and (iv) a time-aware evaluation design with explicit baselines, ablations and testable hypotheses. The article is a design and protocol study. It reports the framework specification and its analytical positioning against prior work; empirical performance figures are outside its scope and are not claimed.

2. Materials and Methods

2.1. Study Design

The framework was developed through design-science reasoning in three steps. Requirements were first derived from the fraud detection literature, covering foundational reviews [1–3,23], realistic operating conditions [7,24] and the three modality-specific research streams. Architectural components were then selected so that each requirement is met by a method with published evidence of effectiveness. Finally, an evaluation protocol was defined in advance so that later empirical work can confirm or refute the design claims. Peer-reviewed journal articles and proceedings of major data mining and machine learning venues were prioritised as sources.

2.2. Data Modalities and Preprocessing

Fraud detection is posed as cost-sensitive binary classification under distribution shift. Given the history available at time t, the model outputs a risk score $\hat{y} \in [0, 1]$ for the pending transaction, and an alert is raised when $\hat{y}$ exceeds a threshold τ chosen to respect a daily alert budget. The label $y \in \{0, 1\}$ becomes known only after a verification delay, and the joint distribution of inputs and labels is allowed to change between training and scoring [7,8].

The framework assumes that every scoring event is a transaction x initiated by an account at time t, for which three views are available. The transactional view contains amount, timestamp, merchant category, channel, geolocation and device identifier, together with the account's preceding T transactions. The behavioural view contains interaction signals recorded during the session that produced the transaction: key hold and flight times, pointer trajectories, typing speed, session duration and navigation order. The network view is a heterogeneous graph whose nodes are accounts, devices, IP addresses and merchants and whose edges record transfers, logins and shared attributes.

Preprocessing follows established practice. Missing numerical values are imputed with training-window medians and flagged with indicator variables. High-cardinality categorical fields such as merchant and device identifiers are mapped to learned embeddings rather than one-hot vectors. Amounts are log-transformed and standardised, and time-of-day is encoded cyclically, since periodic treatment of transaction time has been shown to add predictive value [5]. Behavioural streams are resampled to fixed-length windows and aligned to the transaction they precede. The graph is built incrementally: at scoring time it contains only edges with timestamps earlier than t, which prevents information from the future leaking into training.

No public dataset currently offers all three views for the same events. Empirical work will therefore require either data obtained under agreement with a financial institution or a carefully documented combination of sources, with public benchmarks such as the Elliptic transaction graph [25] used to validate individual branches. This constraint is treated as a limitation in Section 4.

2.3. Modality-Specific Encoders

Each modality is encoded by a network matched to its structure, producing embeddings of common dimension d. The transaction history is processed by a recurrent encoder,

$$h_t = LSTM(x_t, h_{t-1}), \; z_{tx} = h_T \tag{1}$$

where x_t is the embedded transaction at step t and the final hidden state summarises the sequence [10]. A Transformer encoder with self-attention [26] is specified as an interchangeable alternative for long histories, and the choice between the two is an experimental factor. The behavioural stream is encoded by one-dimensional convolutions that extract local timing motifs, followed by a gated recurrent layer whose final state gives z_{bh}; this branch learns an interaction signature for the account and measures deviation from it.

The network view is encoded by L layers of neighbourhood aggregation,

$$h_v^{(l+1)} = \sigma(W^{(l)} \cdot AGG\{h_u^{(l)} : u \in N(v) \cup \{v\}\}) \tag{2}$$

where $N(v)$ is the set of sampled neighbours of node v and AGG is a mean or attention-weighted aggregator. Inductive sampling in the manner of GraphSAGE [27] is preferred over transductive graph convolution [28] because new accounts and devices appear continuously, and attention over neighbours [29] allows the encoder to discount camouflage links that fraudsters create with legitimate entities [17]. The embedding of the transacting account after L layers is taken as z_{net}.

2.4. Adaptive Fusion

The three embeddings are combined by an attention mechanism that assigns each modality m a context-dependent weight:

$$e_m = w^T \tanh(W_m z_m), \; \alpha_m = \exp(e_m) / \Sigma_k \exp(e_k) \tag{3}$$

© 2026 The Author(s). Published by HJPARAM Publication under CC BY 4.0

$$z = \Sigma_m \alpha_m z_m, \quad \hat{y} = \sigma(w_o^\top \varphi(z) + b) \tag{4}$$

where φ is a two-layer feed-forward network and ŷ is the fraud risk score. Intermediate fusion of this kind is chosen over early concatenation of raw features, which ignores structural differences between modalities, and over late averaging of separate scores, which cannot model cross-modal interactions [20]. Because the weights α_m are computed per event, the model can lean on behavioural evidence for a suspected account takeover and on network evidence for a suspected mule transfer. A learned mask sets α_m to zero when a modality is absent, for example when a transaction arrives through a channel that records no interaction data, so the framework degrades gracefully instead of failing.

2.5. Imbalance-Aware and Drift-Aware Training

Fraud typically accounts for well under one per cent of transactions, so the network is trained with the focal loss [30],

$$FL(p_t) = -\alpha_t (1 - p_t)^\gamma \log(p_t) \tag{5}$$

where p_t is the predicted probability of the true class, α_t is a class weight and γ down-weights easy legitimate examples. Class weights are set from misclassification costs so that a missed fraud is penalised in proportion to its expected loss [5]. Synthetic over-sampling [31] is retained only as a baseline treatment for tabular models, because interpolating between minority examples is not well defined for sequences and graphs; for the graph branch, label-balanced neighbourhood sampling is used instead [18].

Training proceeds in two phases. Each encoder is first trained alone with its own classification head, which gives the unimodal baselines required by the evaluation and a stable initialisation. The encoders and the fusion layer are then fine-tuned jointly at a reduced learning rate. During joint training, whole modalities are dropped at random for a fraction of examples. This modality dropout serves two purposes: it prevents the fusion layer from relying on a single dominant branch, and it trains the mask that handles genuinely missing inputs. Early stopping on validation AUC-PR, weight decay and dropout within encoders guard against over-fitting to the small number of fraud cases.

Concept drift is handled at three levels, following the taxonomy of Gama et al. [8]. The model is retrained on a sliding window at a fixed cadence; the decision threshold is recalibrated on the most recent labelled window to hold the alert rate within investigator capacity; and a monitor compares the distribution of risk scores and modality weights between consecutive windows, triggering early retraining when divergence exceeds a preset bound. Delayed labels are respected by training only on transactions whose outcome would have been known at the retraining date [7].

2.6. Explainability

Two complementary explanations are produced for every alert. Shapley-value attributions [32] quantify the contribution of individual input features within each branch, and local surrogate explanations [33] are available as a cross-check. The fusion weights α_m provide a direct, model-intrinsic statement of which modality drove the decision. Attention weights are treated as supporting rather than conclusive evidence, and all explanations are evaluated for stability under small input perturbations before being shown to analysts.

2.7. Evaluation Protocol and Statistical Analysis

Data are split chronologically into training, validation and test periods separated by a gap equal to the label delay; random splitting is excluded because it lets a model learn from the future. The primary metric is the area under the precision–recall curve (AUC-PR), which is more informative than the ROC curve when classes are highly imbalanced [34,35]. Secondary metrics are AUC-ROC, precision, recall and F1 at the operating threshold, false-positive rate, and recall within the top k alerts per day, which reflects investigator workload [7].

The framework is compared with logistic regression, random forest and gradient-boosted trees [36] on engineered features [4], and with each unimodal branch alone. Ablations remove one modality at a time and replace attention fusion with concatenation and with score averaging. Uncertainty is estimated with 1000 stratified bootstrap resamples of the test period to give 95% confidence intervals. Differences in AUC-ROC between paired models are tested with the DeLong procedure [37], and comparisons across multiple test windows use the Wilcoxon signed-rank test with Holm correction [38], at a significance level of $p < 0.05$. Models are to be implemented in Python with PyTorch and PyTorch Geometric, with random seeds fixed and reported.

3. Results

The outputs of this study are a framework specification, an analytical comparison with existing approaches and a set of pre-specified hypotheses. They are presented in that order.

3.1. Framework Architecture

Figure 1 shows the resulting architecture. Data flow from left to right through four stages: three modality inputs, three structure-matched encoders, a single fusion layer and a decision stage that couples the classifier with explanation and drift monitoring. The stages are loosely coupled. Any encoder can be retrained or replaced without altering the others, provided it emits a d-dimensional embedding, and a new modality can be added by attaching another encoder to the fusion layer. The dashed path indicates that drift monitoring feeds back into training rather than into individual decisions.

The specification also fixes the scoring path. Sequence and behavioural encoding cost grows linearly with history length T and window length, and both run on data local to the account. Graph encoding is the expensive step, since sampling s neighbours over L layers touches on the order of s^L nodes. The design therefore bounds L at two or three and allows account embeddings to be refreshed asynchronously and cached, so that the synchronous path

© 2026 The Author(s). Published by HJPARAM Publication under CC BY 4.0

at authorisation time consists of two lightweight encoders, one cache lookup, the fusion layer and the classifier head.

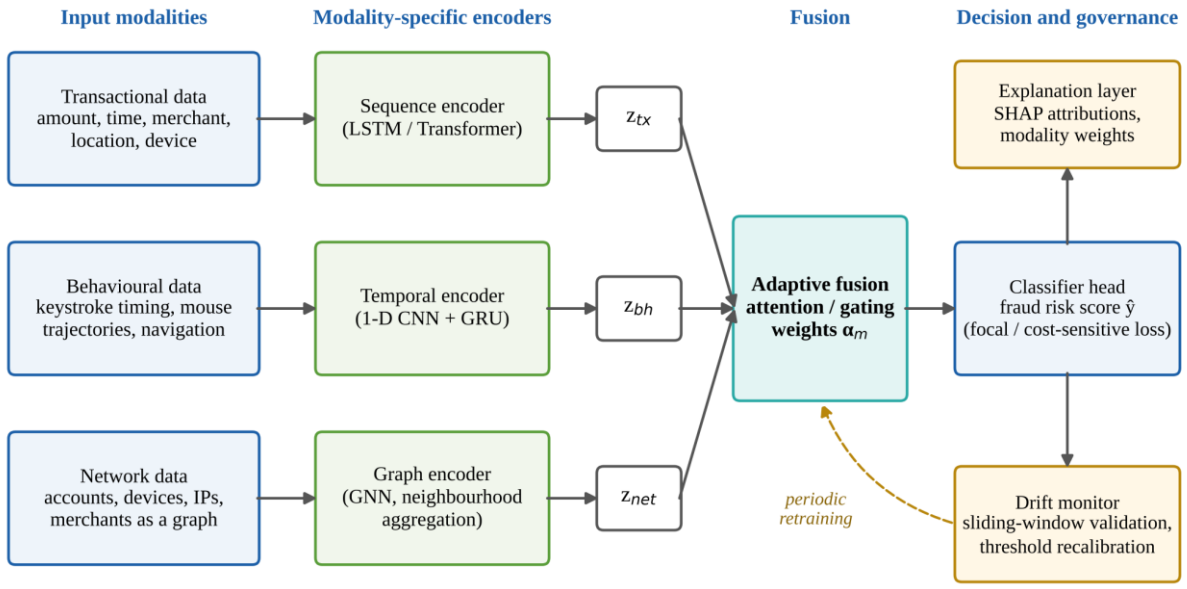

Figure 1. Architecture of the proposed multi-modal fraud detection framework. z_{tx}, z_{bh} and z_{net} are the transactional, behavioural and network embeddings; α_m are per-event modality weights; ŷ is the fraud risk score. Solid arrows show inference; the dashed arrow shows the retraining loop triggered by the drift monitor.

3.2. Positioning against Existing Approaches

Table 1 summarises representative studies from each research stream against the requirements derived in Section 2.1. The pattern is consistent. Transaction-only methods address imbalance and, in the best cases, drift, but have no relational or behavioural evidence. Graph-based detectors capture collusion and have begun to address camouflage and imbalance, yet they do not model the interaction behaviour of the session. Behavioural biometric studies establish that interaction signals identify users, but they were designed for authentication and have not been fused with payment data. None of the reviewed studies combines all three views, and only one explicitly offers interpretable output.

Table 1. Representative fraud detection studies by data modality, and the requirement each leaves open.

Study	Modality	Method and main finding	Requirement left open
Bhattacharyya et al. [4]	Transactional	Compared logistic regression, support vector machines and random forests on real card data; random forests performed best overall.	Static features only; no sequence, behavioural or relational evidence.
Correa Bahnsen et al. [5]	Transactional	Aggregated and periodic time features with cost-based evaluation improved savings over raw attributes.	Manual feature engineering; single modality.
Dal Pozzolo et al. [7]	Transactional	Formalised realistic operation with imbalance, drift and delayed labels; separate learners for feedback and delayed samples.	Addresses drift, but not relational or behavioural signals.
Jurgovsky et al. [9]	Transactional (sequence)	LSTM sequence classifier improved detection for card-present transactions and found different frauds from a random forest.	No graph or behavioural input; gains uneven across channels.
Ahmed and Traore [12]	Behavioural	Neural network on mouse dynamics distinguished users with low false acceptance and rejection rates.	Authentication setting; not linked to transactions.
Van Vlasselaer et al. [13]	Transactional + network features	Network-derived scores from the cardholder–merchant graph, added to intrinsic features, raised detection performance.	Hand-crafted network features; no behavioural view.
Liu et al. [15]	Network	Heterogeneous account–device GNN for malicious account detection in a large payment platform.	Topology and activity only; limited explanation.

© 2026 The Author(s). Published by HJPARAM Publication under CC BY 4.0

Study	Modality	Method and main finding	Requirement left open
Wang et al. [16]	Network (multi-view)	Semi-supervised GNN with hierarchical attention used labelled and unlabelled users and gave interpretable weights.	User-level prediction; no session behaviour or drift handling.
Dou et al. [17]	Network	Similarity-aware neighbour selection made GNN detectors robust to camouflaged fraudsters.	Graph only; static evaluation.
Liu et al. [18]	Network	Label-balanced sampling and neighbour selection addressed class imbalance in graph learning.	Graph only; no temporal drift treatment.
Cheng et al. [19]	Transactional + spatial-temporal graph	Spatial-temporal attention over transaction graph features outperformed baselines on card data.	No behavioural biometrics; one institution's data.

GNN, graph neural network; LSTM, long short-term memory. "Requirement left open" reflects the authors' assessment against Section 2.1.

3.3. Requirement Coverage

Table 2 maps each design requirement to the component that satisfies it and to the measurement that will verify it. Every requirement has both a mechanism and a metric, and no component is included without a requirement that motivates it.

Table 2. Design requirements, the framework component that addresses each, and the planned verification.

Requirement	Component	Verification
Temporal spending patterns	Sequence encoder (Eq. 1)	Ablation without z_{tx}
User interaction signature	Behavioural encoder	Ablation without z_{bh}
Collusion and shared entities	Graph encoder (Eq. 2)	Ablation without z_{net}
Context-dependent evidence	Attention fusion (Eqs. 3–4)	Versus concatenation and averaging
Class imbalance	Focal, cost-weighted loss (Eq. 5)	AUC-PR; recall at fixed alert rate
Concept drift	Sliding retraining; drift monitor	Metric decay across test windows
Missing modality	Learned fusion mask	Performance with masked inputs
Transparency	SHAP and modality weights	Attribution stability; analyst review

AUC-PR, area under the precision–recall curve; SHAP, Shapley additive explanations.

3.4. Pre-Specified Hypotheses

The evaluation protocol is organised around four hypotheses, stated here so that later results can be judged against them. H1: the full framework achieves higher AUC-PR than the strongest unimodal branch and than gradient-boosted trees on engineered features. H2: attention fusion outperforms both early concatenation and late score averaging at an equal parameter budget. H3: with sliding-window retraining and threshold recalibration, AUC-PR declines more slowly across successive test windows than for a model trained once. H4: removing any single modality lowers recall at a fixed false-positive rate, with the largest loss for the network view on collusive fraud and for the behavioural view on account takeover. Each hypothesis is falsifiable with the tests in Section 2.7, and a null result on any of them would be informative about where multi-modal integration does and does not pay.

4. Discussion

The central argument of this work is that the three dominant lines of fraud detection research are complementary rather than competing. The evidence for each is individually strong. Sequence models find frauds that static classifiers miss [9]; graph neural networks expose coordinated activity that no per-transaction model can represent [15,16,19]; and interaction dynamics identify the person operating an account independently of credentials [11,12]. What the literature does not yet show is how much is gained by combining them, because studies that hold all three views for the same events are absent from published reviews [21]. The framework specified here makes that question answerable.

Several design choices depart from common practice and deserve justification. Fusing learned embeddings, and not raw features or final scores, follows the multimodal learning literature, which finds that intermediate fusion best preserves modality-specific structure while still allowing interaction [20]. Computing fusion weights per event, in place of fixing them globally, reflects how fraud presents in operations: the informative signal differs by fraud type. Inductive graph encoding was preferred because production graphs change continuously [27], and temporal graph networks are a natural next step once event-level timestamps on edges are exploited more fully [39]. Building the graph strictly from past edges and splitting data chronologically are less a design preference than a requirement; without them, reported performance overstates what a deployed system would achieve [7].

The treatment of explainability is deliberately cautious. Post hoc attribution methods are widely used, but Rudin has argued that explanations of black-box models can be unfaithful to what the model computes and that inherently interpretable models should be preferred for high-stakes decisions [22]. The framework responds in two ways. The modality weights are part of the forward computation, so they describe the model's actual behaviour at a coarse level, and the evaluation includes an interpretable baseline so that the accuracy cost of transparency, if any, is measured and not assumed. Whether analysts find the

© 2026 The Author(s). Published by HJPARAM Publication under CC BY 4.0

explanations useful is an empirical question that the protocol addresses through structured review.

The practical implications concern cost and deployment as much as accuracy. False positives carry a direct cost in investigator time and an indirect cost in customer friction, which is why the protocol reports recall at a fixed alert budget alongside threshold-free metrics. Running three encoders adds inference latency; streaming architectures for fraud scoring show that such pipelines can operate at scale [24], but the latency of on-line neighbourhood sampling remains to be measured, and pre-computing account embeddings is the planned mitigation.

The study has clear limitations. It is a framework and protocol paper and contains no experimental results; the hypotheses in Section 3.4 may not hold, and the comparison in Table 1 is analytical. The framework assumes that behavioural and network data can be linked to transactions, which is true for app and web channels but not for all card-present payments. Behavioural signals are personal data, so collection requires a lawful basis, minimisation and safeguards against discriminatory proxies; fairness auditing across customer groups must accompany any deployment. Access to suitably linked data is the main practical obstacle, and results obtained at one institution may not transfer to another.

Three threats to the validity of the planned evaluation should also be acknowledged. Fraud labels are themselves the product of an existing detection process, so cases that current controls never flag are missing from the training data, and any model inherits that selection bias [7]. Fraudsters respond to deployed models; evidence that graph detectors can be evaded through camouflage [17] suggests behavioural signals may in time be imitated as well, so robustness under adaptive attack needs separate study. Finally, a framework with several interchangeable components invites over-tuning on the validation period, which is why the test period, metrics and hypotheses are fixed in advance.

These limitations indicate the direction of future work. Federated learning would allow institutions to train a shared model without exchanging raw records [40], which addresses both data access and privacy. Self-supervised pre-training on unlabelled transactions could reduce dependence on scarce fraud labels, and automated drift detection could replace the fixed retraining cadence. Extending the evaluation to adjacent domains such as insurance claims, digital lending and cryptocurrency transfers [25] would test how far the architecture generalises.

5. Conclusions

Fraud in digital finance is expressed simultaneously in what is transacted, how the session is operated and who is connected to whom, yet detection research has mostly examined one of these at a time. This article has specified a multi-modal deep learning framework that encodes transaction sequences, behavioural interaction signals and relational graphs with structure-matched networks and combines them through per-event attention fusion. The design pairs that architecture with focal, cost-weighted training for class imbalance, sliding-window retraining and monitoring for concept drift, and feature-level and modality-level explanations for regulatory transparency.

The contribution is a concrete, testable design, not a performance claim. Every requirement drawn from the literature is mapped to a component and to a verification measure, and four hypotheses with pre-specified statistical tests define what would count as evidence that integration helps. Empirical validation on linked multi-modal data under the time-aware protocol described here is the necessary next step, and its results, whether confirmatory or not, will clarify where multi-modal evidence adds most to financial fraud detection.

Author Contributions: Conceptualization, A.S. and G.S.; methodology, A.S.; writing—original draft, A.S.; writing—review and editing, G.S.; supervision, G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analysed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors thank the Department of Computer Science, SAGE University, Bhopal, for academic support.

Conflicts of Interest: The authors declare no conflicts of interest.

References

[1] Bolton, R.J.; Hand, D.J. Statistical fraud detection: A review. Stat. Sci. 2002, 17, 235–255. https://doi.org/10.1214/ss/1042727940

[2] West, J.; Bhattacharya, M. Intelligent financial fraud detection: A comprehensive review. Comput. Secur. 2016, 57, 47–66. https://doi.org/10.1016/j.cose.2015.09.005

[3] Ngai, E.W.T.; Hu, Y.; Wong, Y.H.; Chen, Y.; Sun, X. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decis. Support Syst. 2011, 50, 559–569. https://doi.org/10.1016/j.dss.2010.08.006

[4] Bhattacharyya, S.; Jha, S.; Tharakunnel, K.; Westland, J.C. Data mining for credit card fraud: A comparative study. Decis. Support Syst. 2011, 50, 602–613. https://doi.org/10.1016/j.dss.2010.08.008

[5] Correa Bahnsen, A.; Aouada, D.; Stojanovic, A.; Ottersten, B. Feature engineering strategies for credit card fraud detection. Expert Syst. Appl. 2016, 51, 134–142. https://doi.org/10.1016/j.eswa.2015.12.030

[6] He, H.; Garcia, E.A. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 2009, 21, 1263–1284. https://doi.org/10.1109/TKDE.2008.239

[7] Dal Pozzolo, A.; Boracchi, G.; Caelen, O.; Alippi, C.; Bontempi, G. Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Trans. Neural Netw. Learn. Syst. 2018, 29, 3784–3797. https://doi.org/10.1109/TNNLS.2017.2736643

[8] Gama, J.; Žliobaitė, I.; Bifet, A.; Pechenizkiy, M.; Bouchachia, A. A survey on concept drift adaptation. ACM Comput. Surv. 2014, 46, 44. https://doi.org/10.1145/2523813

[9] Jurgovsky, J.; Granitzer, M.; Ziegler, K.; Calabretto, S.; Portier, P.-E.; He-Guelton, L.; Caelen, O. Sequence classification for credit-card fraud detection. Expert Syst. Appl. 2018, 100, 234–245. https://doi.org/10.1016/j.eswa.2018.01.037

[10] Hochreiter, S.; Schmidhuber, J. Long short-term memory. Neural Comput. 1997, 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

 © 2026 The Author(s). Published by HJPARAM Publication under CC BY 4.0

[11] Monrose, F.; Rubin, A.D. Keystroke dynamics as a biometric for authentication. Future Gener. Comput. Syst. 2000, 16, 351–359. https://doi.org/10.1016/S0167-739X(99)00059-X

[12] Ahmed, A.A.E.; Traore, I. A new biometric technology based on mouse dynamics. IEEE Trans. Dependable Secur. Comput. 2007, 4, 165–179. https://doi.org/10.1109/TDSC.2007.70207

[13] Van Vlasselaer, V.; Bravo, C.; Caelen, O.; Eliassi-Rad, T.; Akoglu, L.; Snoeck, M.; Baesens, B. APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. Decis. Support Syst. 2015, 75, 38–48. https://doi.org/10.1016/j.dss.2015.04.013

[14] Akoglu, L.; Tong, H.; Koutra, D. Graph based anomaly detection and description: A survey. Data Min. Knowl. Discov. 2015, 29, 626–688. https://doi.org/10.1007/s10618-014-0365-y

[15] Liu, Z.; Chen, C.; Yang, X.; Zhou, J.; Li, X.; Song, L. Heterogeneous graph neural networks for malicious account detection. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM), Torino, Italy, 22–26 October 2018. https://doi.org/10.1145/3269206.3272010

[16] Wang, D.; Lin, J.; Cui, P.; Jia, Q.; Wang, Z.; Fang, Y.; Yu, Q.; Zhou, J.; Yang, S.; Qi, Y. A semi-supervised graph attentive network for financial fraud detection. In Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM), Beijing, China, 8–11 November 2019. https://doi.org/10.1109/ICDM.2019.00070

[17] Dou, Y.; Liu, Z.; Sun, L.; Deng, Y.; Peng, H.; Yu, P.S. Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. In Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM), Virtual Event, 19–23 October 2020. https://doi.org/10.1145/3340531.3411903

[18] Liu, Y.; Ao, X.; Qin, Z.; Chi, J.; Feng, J.; Yang, H.; He, Q. Pick and choose: A GNN-based imbalanced learning approach for fraud detection. In Proceedings of the Web Conference 2021 (WWW), Ljubljana, Slovenia, 19–23 April 2021. https://doi.org/10.1145/3442381.3449989

[19] Cheng, D.; Wang, X.; Zhang, Y.; Zhang, L. Graph neural network for fraud detection via spatial-temporal attention. IEEE Trans. Knowl. Data Eng. 2022, 34, 3800–3813. https://doi.org/10.1109/TKDE.2020.3025588

[20] Baltrušaitis, T.; Ahuja, C.; Morency, L.-P. Multimodal machine learning: A survey and taxonomy. IEEE Trans. Pattern Anal. Mach. Intell. 2019, 41, 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

[21] 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. https://doi.org/10.1016/j.dss.2020.113303

[22] Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 2019, 1, 206–215. https://doi.org/10.1038/s42256-019-0048-x

[23] Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. ACM Comput. Surv. 2009, 41, 15. https://doi.org/10.1145/1541880.1541882

[24] Carcillo, F.; Dal Pozzolo, A.; Le Borgne, Y.-A.; Caelen, O.; Mazzer, Y.; Bontempi, G. SCARFF: A scalable framework for streaming credit card fraud detection with Spark. Inf. Fusion 2018, 41, 182–194. https://doi.org/10.1016/j.inffus.2017.09.005

[25] Weber, M.; Domeniconi, G.; Chen, J.; Weidele, D.K.I.; Bellei, C.; Robinson, T.; Leiserson, C.E. Anti-money laundering in Bitcoin: Experimenting with graph convolutional networks for financial forensics. arXiv 2019, arXiv:1908.02591. https://doi.org/10.48550/arXiv.1908.02591

[26] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Advances in Neural Information Processing Systems 30 (NeurIPS 2017); Curran Associates: Red Hook, NY, USA, 2017. https://doi.org/10.48550/arXiv.1706.03762

[27] Hamilton, W.L.; Ying, R.; Leskovec, J. Inductive representation learning on large graphs. In Advances in Neural Information Processing Systems 30 (NeurIPS 2017); Curran Associates: Red Hook, NY, USA, 2017. https://doi.org/10.48550/arXiv.1706.02216

[28] Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. In Proceedings of the 5th International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017. https://doi.org/10.48550/arXiv.1609.02907

[29] Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph attention networks. In Proceedings of the 6th International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 30 April–3 May 2018. https://doi.org/10.48550/arXiv.1710.10903

[30] Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988. https://doi.org/10.1109/ICCV.2017.324

[31] 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. https://doi.org/10.1613/jair.953

[32] Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30 (NeurIPS 2017); Curran Associates: Red Hook, NY, USA, 2017. https://doi.org/10.48550/arXiv.1705.07874

[33] Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144. https://doi.org/10.1145/2939672.2939778

[34] Davis, J.; Goadrich, M. The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd International Conference on Machine Learning (ICML), Pittsburgh, PA, USA, 25–29 June 2006; pp. 233–240. https://doi.org/10.1145/1143844.1143874

[35] Saito, T.; Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE 2015, 10, e0118432. https://doi.org/10.1371/journal.pone.0118432

[36] 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. https://doi.org/10.1145/2939672.2939785

[37] DeLong, E.R.; DeLong, D.M.; Clarke-Pearson, D.L. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 1988, 44, 837–845. https://doi.org/10.2307/2531595

[38] Demšar, J. Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. 2006, 7, 1–30.

[39] Rossi, E.; Chamberlain, B.; Frasca, F.; Eynard, D.; Monti, F.; Bronstein, M. Temporal graph networks for deep learning on dynamic graphs. arXiv 2020, arXiv:2006.10637. https://doi.org/10.48550/arXiv.2006.10637

[40] McMahan, H.B.; Moore, E.; Ramage, D.; Hampson, S.; Agüera y Arcas, B. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), Fort Lauderdale, FL, USA, 20–22 April 2017. https://doi.org/10.48550/arXiv.1602.05629

 © 2026 The Author(s). Published by HJPARAM Publication under CC BY 4.0