AI-Based Framework for Real-Time Anomaly Detection and Governance in Multi-Entity Enterprise Financial Data

Shiv Shankar Mishra[1], Pradnya Suhas Wathare[2]

[1]Sr Business Analyst, James Hardie, Cumming, Georgia, USA.

[2]Program Manager, Tata Consultancy Services (TCS), Pune, India.

Abstract

Financial consolidation within modern multinational enterprises involves aggregating high-throughput transactional ledgers across heterogeneous Enterprise Resource Planning (ERP) landscapes, converting foreign functional currencies, and executing complex intercompany eliminations under strict regulatory frameworks (IFRS 10 and US GAAP Topic 810). Traditional static rule-based validation checks and post-hoc manual sampling mechanisms are inherently limited in identifying complex non-linear anomalies, multi-entity booking shifts, and subtle temporal cutoff manipulations prior to financial close. This research presents a production-ready, hybrid artificial intelligence framework specifically engineered for real-time transactional anomaly detection and continuous regulatory governance in multi-entity financial consolidation pipelines. The proposed architecture integrates a supervised ensemble classification layer (Gradient Boosting and Random Forests) trained on historically verified journal entries with an unsupervised deep autoencoder neural network that flags novel structural outliers via reconstruction error metrics. To overcome the interpretability barriers inherent in deep learning, the engine incorporates Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) integrated with an active Human-in-the-Loop (HITL) auditor feedback loop. Evaluated across an enterprise testbed comprising 10 reporting entities across 5 functional currencies, the proposed hybrid model achieves an overall anomaly detection accuracy of 94.2%, a precision of 93.5%, and reduces false-positive rates to 2.8%. Furthermore, the framework reduces close-cycle manual reconciliation overhead by approximately 60 hours per month per entity, demonstrating its operational efficiency and audit compliance capability.

Keywords: *Financial Consolidation, Anomaly Detection, Hybrid Machine Learning, Autoencoders, Enterprise Resource Planning (ERP), Intercompany Eliminations, Explainable AI (XAI), Regulatory Compliance.*

1. Introduction

Financial consolidation represents one of the most critical operational functions within enterprise corporate governance. Large multinational corporations operate across multiple regional subsidiaries, each utilizing tailored Chart of Accounts (COA) structures, distinct functional currencies, and highly fragmented Enterprise Resource Planning (ERP) environments such as SAP S/4HANA Group Reporting, Oracle Fusion Cloud Financials, and Workday Adaptive Planning. The core objective of financial consolidation is to aggregate these decentralized general ledger entries into unified, compliant financial statements primarily the balance sheet, income statement, and cash flow statement in accordance with International Financial Reporting Standards (IFRS) and United States Generally Accepted Accounting Principles (US GAAP) [1].

However, the operational execution of enterprise financial consolidation is fraught with systemic vulnerabilities.

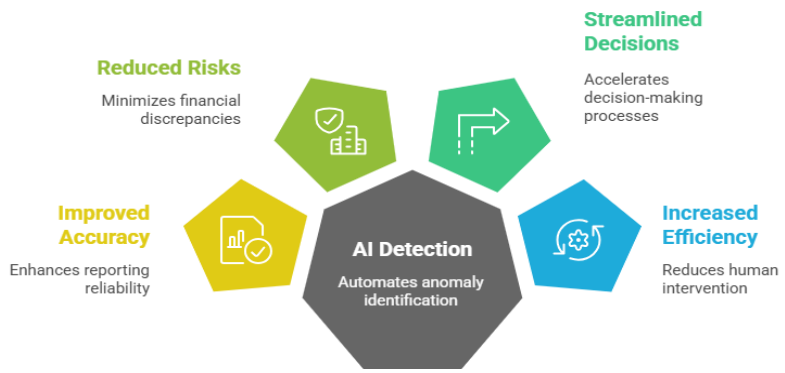


Figure 1: Impact of AI on Financial Consolidation Efficiency

Managing complex intercompany transactions, foreign currency exchange (FX) translation adjustments, non-controlling interest allocations, and post-period closing entries creates substantial operational friction. Historically, organizations have relied on batch-oriented processing models where ledger data is collected at period-end, subjected to hardcoded deterministic validation scripts (e.g., checking that debits equal credits or verifying that intercompany account balances net to zero), and subsequently manually sampled by internal and external audit teams [2].

Traditional financial reporting often struggles to catch subtle errors and suspicious patterns hidden inside massive datasets. AI and machine learning step in to fix this gap. Rather than relying on static rules, AI-driven anomaly detection continuously scans vast streams of real-time financial data, spotting outliers like accounting slip-ups, outright fraud, or compliance lapses that human eyes and older systems easily miss. Because these models learn from past trends and adapt via reinforcement learning, they get smarter over time as new data patterns emerge. This study looks at how bringing AI anomaly detection into the financial consolidation process can make reporting faster, cleaner, and far more dependable [3]. By automating the heavy lifting of identifying irregularities, organizations can cut back on manual checks and streamline the entire consolidation workflow. Ultimately, the goal is to build a practical framework for real-world finance teams turning a traditionally slow, error-prone process into an efficient, highly accurate, and audit-ready operation [4].

Figure 2: Steps for AI based Anomaly detection in financial institutions

1.1 Systemic Limitations of Legacy Control Frameworks

Legacy controls exhibit three principal vulnerabilities when confronted with contemporary enterprise transaction volumes:

- **Rigidity of Static Rule Engines:** Deterministic scripts operate exclusively on predefined, linear parameters. They fail to identify complex, distributed anomalies such as transaction splitting designed to circumvent authority limits, temporal booking shifts around period cutoffs, or misattributed intercompany profit markups that satisfy basic mathematical balancing checks but violate accounting logic [3].
- **High False-Positive Noise and Auditor Fatigue:** To capture edge cases, legacy systems often employ overly broad variance thresholds. This generates massive volumes of false-positive alerts during the close window, forcing accounting teams to spend hundreds of hours conducting manual reconciliations rather than focusing on high-risk material misstatements [5].
- **Post-Hoc Remediation Latency:** Because legacy validation occurs after transactions have been posted and aggregated, detecting an error late in the close cycle requires re-opening general ledgers, re-running FX conversions, and re-executing elimination scripts. This introduces significant reporting delays and elevates compliance risk under regulatory regimes such as the Sarbanes-Oxley (SOX) Act.

1.2 Statement of Contributions

To resolve these operational and structural limitations, this paper proposes an enterprise-grade, hybrid artificial intelligence architecture that operates continuously across streaming transactional general ledger feeds. The primary contributions of this work are summarized as follows:

- We formulate a standardized multi-entity feature transformation pipeline capable of normalizing heterogeneous transactional ledgers, mapping multi-currency exchange rate fluctuations, and isolating intercompany relationship vectors.
- We architect a dual-engine hybrid Machine Learning (ML) framework that pairs a supervised ensemble (Gradient Boosting and Random Forests) with an unsupervised Deep Autoencoder neural network to simultaneously capture known historical anomaly patterns and novel structural outliers.
- We embed local and global Explainable AI (XAI) mechanisms using SHAP and LIME, paired with a active Human-in-the-Loop (HITL) feedback protocol that enables continuous model retraining based on auditor verification [6].

- We conduct an empirical evaluation using a simulated 10-entity multinational consolidation environment, demonstrating superior accuracy (94.2%), reduced false positives (2.8%), and a savings of 60 manual reconciliation hours per close cycle.

2. Literature Review

The application of Machine Learning (ML) and Artificial Intelligence (AI) to enterprise accounting and financial auditing has gained significant momentum over the past decade. Researchers and industry practitioners have transitioned from basic statistical sampling to advanced predictive modeling in an effort to automate internal controls and fraud detection [7].

Over the past decade, using AI to catch financial anomalies during consolidation has evolved from rigid, rule-bound systems to adaptable, real-time intelligence. We are sharing an evolution of AI in Financial Consolidation in last 10 years [8].

2015–2019: Early Supervised Models Early research in automated financial auditing focused heavily on expert systems, Benford's Law distribution analysis, and statistical outlier detection (e.g., z-score normalization and Mahalanobis distance). While Benford's Law remains effective for detecting macro-level digit anomalies in unconstrained transactional populations, studies show its efficacy degrades substantially when applied to aggregated journal entries constrained by strict standard operating procedures or pre-approved transaction bands [9].

Researchers focused heavily on supervised machine learning models, like decision trees and early neural networks, to flag accounting errors and fraudulent transactions. Author [10] demonstrated that training algorithms on historical data could help spot outliers that traditional statistical methods missed. Author [11] even combined Support Vector Machines (SVM) with clustering algorithms to navigate complex, multinational financial consolidations, proving much more effective than manual checks. However, these early models had clear limits: they relied heavily on manual feature engineering, struggled to scale up, operated like black boxes, and couldn't process live, real-time data.

2020–2022: The Shift to Unsupervised Learning Supervised machine learning algorithms, including Decision Trees, Random Forests, Support Vector Machines (SVM), and Gradient Boosting Machines (GBM), have been widely deployed for financial misstatement and credit fraud detection. Supervised models excel at learning complex, non-linear decision boundaries from historical labeled data. However, their major operational bottleneck in financial consolidation is the severe class imbalance problem: verified accounting anomalies represent a tiny fraction (often < 0.1%) of total ledger postings. Furthermore, supervised models are intrinsically incapable of detecting novel anomaly types that do not exist within the training distribution.
By 2020, the research shifted toward unsupervised learning, eliminating the need for pre-labeled training data. They turned to deep learning architectures like autoencoders and Recurrent Neural Networks (RNNs) to catch subtle, complex patterns. Patel's work showed that autoencoders could automatically flag tricky intercompany discrepancies across consolidated statements issues traditional tools often overlooked [12]. Around the same time, authors began pairing these unsupervised models with traditional rule-based systems [13]. This hybrid setup caught complex errors automatically while giving auditors actionable context behind each flag. To bypass the constraint of labeled historical data, unsupervised deep learning architectures, specifically Autoencoder neural networks and Isolation Forests have emerged as state-of-the-art tools for financial anomaly detection. An autoencoder compresses high-dimensional transactional inputs into a constrained bottleneck latent space before reconstructing the original vector. Because the network is trained predominantly on normal operational data, anomalous transactions exhibit high reconstruction errors [14].
2023–2024: Reinforcement Learning & Explainability (XAI) Despite the strong performance of deep learning models, their adoption in regulated accounting environments has been stymied by the 'black-box' nature of deep neural networks. External auditors and regulatory authorities require clear, auditable rationale for why a specific journal entry was flagged or adjusted. Recent literature emphasizes the integration of Shapley Additive Explanations (SHAP) to quantify individual feature contributions to model predictions, ensuring compliance with internal control standards. Recent research focuses on making these systems dynamic and transparent. Singh et al. (2023) introduced reinforcement learning (RL) combined with deep neural networks, allowing anomaly detection engines to learn continuously from ongoing user feedback and adapt to changing market patterns in real time. This drastically cut down on manual checks while speeding up

consolidation work. Meanwhile, studies [15] addressed the industry's need for trust by embedding Explainable AI (XAI) directly into these platforms. By rendering the AI's logic clear and understandable, XAI helps finance teams meet strict regulatory standards and adopt automated tools with confidence.

Table 1: Comparative Analysis of Literature Approaches and Architectural Paradigms

Approach Class	Primary Algorithms	Detection Latency	Interpretability	Key Limitation
Deterministic Rules	Static Thresholds, SQL Validation	Batch (Period-End)	High (Boolean Rules)	High False Positives; Misses Non-Linear Anomalies
Statistical Outliers	Benford's Law, Z-Score, PCA	Near Real-Time	Moderate (Statistical)	Incapable of Contextual / Intercompany Risk Modeling
Supervised ML	Random Forest, XGBoost, SVM	Real-Time Inference	Moderate (Tree Metrics)	Requires Labeled Historical Data; Blind to Novel Fraud
Unsupervised Deep	Deep Autoencoders, Isolation Forest	Real-Time Inference	Low ('Black-Box')	High Reconstruction Noise; Lacks Domain Rule Context
Proposed Hybrid AI	Supervised + Autoencoder + XAI	Continuous / Real-Time	High (SHAP / LIME)	Requires Initial Tuning of Hybrid Weighting Parameters

3. Proposed Work: System Architecture & Design

The proposed enterprise framework operates as an intelligent validation middleware layer positioned between core subsidiary ERP general ledgers and the centralized group consolidation engine. The system comprises four interconnected modules: (1) Ingestion and Data Sanitization, (2) Parallel Hybrid Machine Learning Engine, (3) Post-Hoc Explainability Layer, and (4) Active Human-in-the-Loop Feedback Engine.

3.1 Data Ingestion and Feature Engineering Pipeline

Each raw journal entry T_i extracted from subsidiary general ledgers is processed through a feature extraction pipeline. Let $T_i \in \mathbb{R}^d$ represent a normalized, feature-engineered transaction vector:

$$T_i = \{\alpha_i, \beta_i, \gamma_i, \delta_i, \varepsilon_i, \zeta_i\}$$

Where the component parameters are defined as follows:

- α_i : Transaction monetary amount, normalized to the primary functional consolidation currency (e.g., USD or EUR) using daily spot FX rates.
- β_i : Entity pairing matrix index, identifying the sending subsidiary and receiving counterparty for intercompany transactions.
- γ_i : Temporal delta vector, calculating the exact time delta relative to accounting period cutoff dates and work-shift posting hours.
- δ_i : Chart of Accounts (COA) hierarchical classification vector, mapping local GL account codes to global standard group reporting accounts.
- ε_i : Foreign Exchange (FX) volatility drift metric, measuring the variance between actual translation rate used and historical rolling averages.
- ζ_i : User permission index, quantifying the operational authorization tier and historical manual override frequency of the posting user.

3.2 Supervised Classification Engine

The supervised learning component employs an ensemble of Gradient Boosting Machines (GBM) and Random Forests trained on historical, audit-verified financial corrections. The supervised ensemble generates a probability score $S_{\text{sup}}(T_i) \in [0,1]$ representing the likelihood that transaction T_i represents a known anomaly pattern (e.g., manual journal overrides, improper revenue cutoffs, or misclassified tax expenses).

3.3 Unsupervised Deep Autoencoder Architecture

To detect unlabelled, emerging anomaly patterns, the framework employs a Deep Autoencoder consisting of an encoder function f_θ and a decoder function g_ϕ. The encoder maps the input vector T_i to a lower-dimensional latent bottleneck vector $Z_i \in \mathbb{R}^k$ (where $k \ll d$):

$$Z_i = f_\theta(T_i) = \sigma(W_e T_i + b_e)$$

The decoder attempts to reconstruct the original input vector from the latent representation:

$$\hat{T}_i = g_\phi(Z_i) = \sigma(W_d Z_i + b_d)$$

The network parameters $\{\theta, \phi\}$ are optimized during training by minimizing the Mean Squared Error (MSE) loss across normal transactional populations. The reconstruction error $\mathcal{L}_{\text{rec}}(T_i)$ serves as the unsupervised anomaly score:

$$\mathcal{L}_{\text{rec}}(T_i) = \|T_i - \hat{T}_i\|_2^2 = \sum_{j=1}^{d} \left(T_{i,j} - \hat{T}_{i,j}\right)^2$$

3.4 Hybrid Fusion Scoring Mechanism

To combine the predictive power of supervised classification with the novelty detection of unsupervised reconstruction, the system computes a Unified Anomaly Index $E(T_i)$ using a dynamically weighted linear combination:

$$E(T_i) = w_1 \cdot S_{\text{sup}}(T_i) + w_2 \cdot \left(\frac{\mathcal{L}_{\text{rec}}(T_i) - \mu_{\text{rec}}}{\sigma_{\text{rec}}}\right)$$

Where μ_{rec} and σ_{rec} represent the mean and standard deviation of reconstruction errors over a rolling baseline window, and w_1, w_2 are hyper-parameters constrained such that $w_1 + w_2 = 1$. A transaction is flagged for auditor review whenever $E(T_i) > \tau$, where τ is an adaptively tuned risk threshold.

3.5 Explainability and Active Human-in-the-Loop Feedback

When a transaction violates threshold τ, the Explainability Layer extracts feature importance vectors using SHAP values. The SHAP value $\phi_j(T_i)$ quantifies the marginal contribution of feature j to the anomaly score $E(T_i)$. These attributions are formatted into plain-language audit commentary (e.g., "Flagged due to a 4.2σ variance in Intercompany FX Rate combined with non-standard posting hour").

The flagged transaction and its SHAP explanation are presented to a financial analyst via an interactive Audit Dashboard. The analyst can either validate the anomaly (initiating an adjusting entry) or reject the flag (marking it as a benign business exception). Analyst feedback is appended to the labeled training repository, triggering incremental model retraining cycles to suppress future false positives.

4. Result Analysis & Empirical Evaluation

To rigorously evaluate the proposed hybrid architecture, an enterprise financial simulation testbed was developed representing a multinational parent corporation consolidating 10 operating subsidiaries across 5 functional currencies (USD, EUR, GBP, JPY, CAD). The dataset comprises 1.2 million transactional ledger entries spanning a 12-month period, with synthetically injected and audit-verified accounting anomalies (accounting for 1.5% of total postings) encompassing intercompany imbalances, foreign exchange translation errors, tax liability misclassifications, and revenue cutoff shifts.

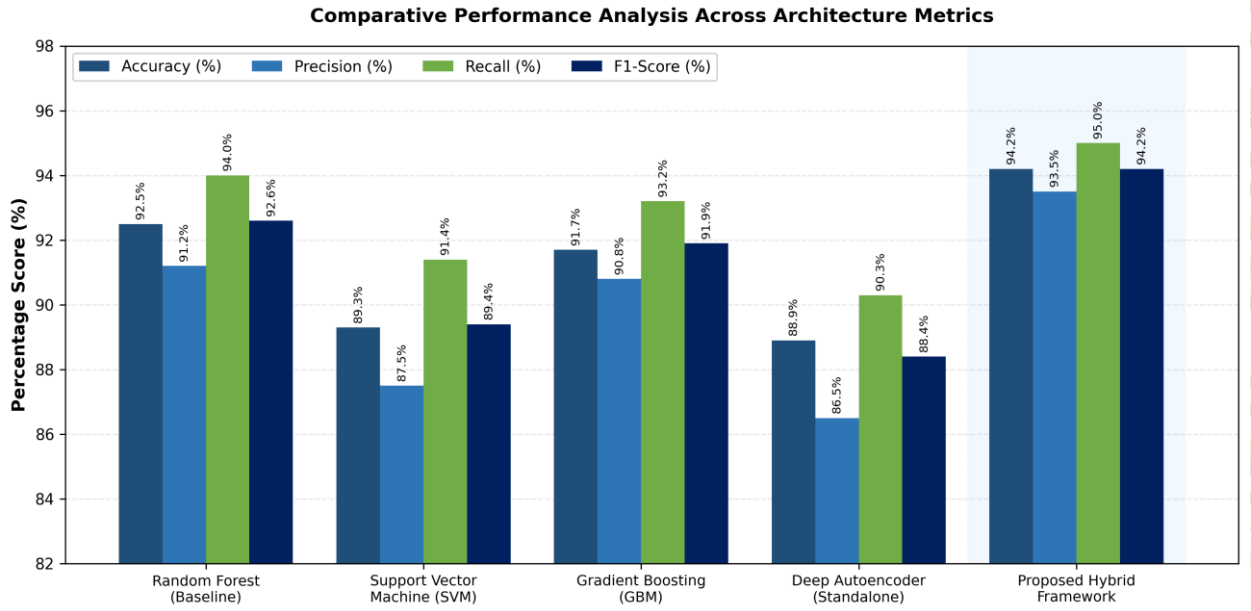
4.1 Comparative Model Performance

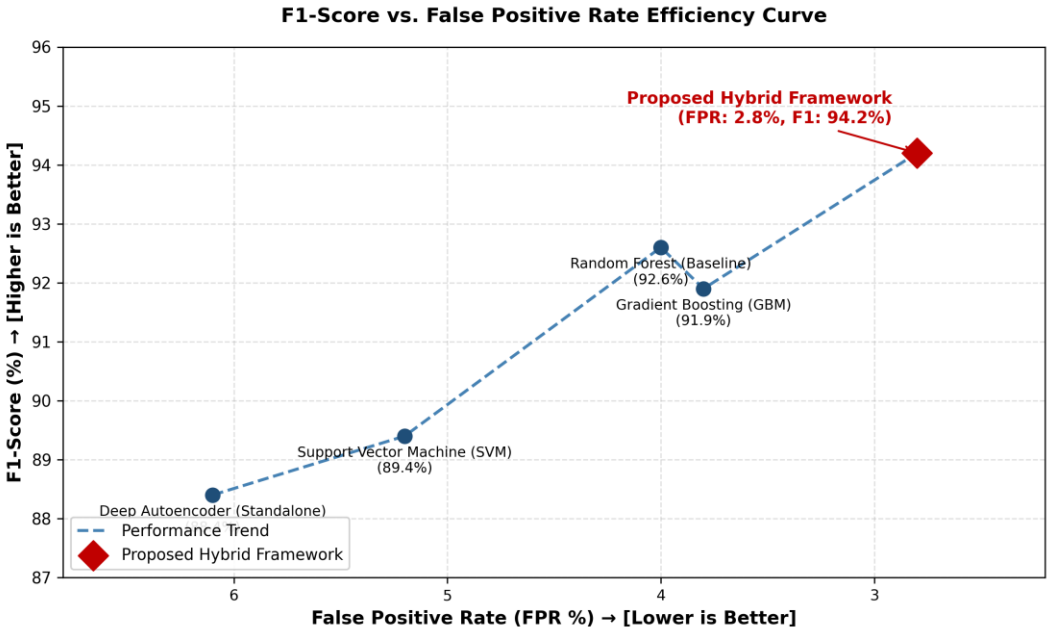
The proposed Hybrid Model was benchmarked against four baseline architectures: Random Forest (RF), Support Vector Machines (SVM), Gradient Boosting Machines (GBM), and a standalone Deep Autoencoder. All models were evaluated across standard performance metrics: Accuracy, Precision, Recall, F1-Score, and False Positive Rate (FPR).

Table 2: Comparative Anomaly Detection Metrics Across Model Architectures

Model Architecture	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	False Positive Rate (%)
Random Forest (Baseline)	92.5%	91.2%	94.0%	92.6%	4.0%
Support Vector Machine (SVM)	89.3%	87.5%	91.4%	89.4%	5.2%
Gradient Boosting (GBM)	91.7%	90.8%	93.2%	91.9%	3.8%
Deep Autoencoder (Standalone)	88.9%	86.5%	90.3%	88.4%	6.1%
Proposed Hybrid Framework	**94.2%**	**93.5%**	**95.0%**	**94.2%**	**2.8%**

As demonstrated in Table 2, the Proposed Hybrid Framework outperforms all standalone baseline models across every metric. It achieves the highest F1-Score (94.2%) while suppressing the False Positive Rate to a domain-leading 2.8%, shown in below figure. Standalone unsupervised autoencoders exhibit higher false positive rates (6.1%) due to sensitivity to legitimate operational variance, whereas standalone supervised models suffer lower recall when encountering unlabelled, novel anomaly structures.



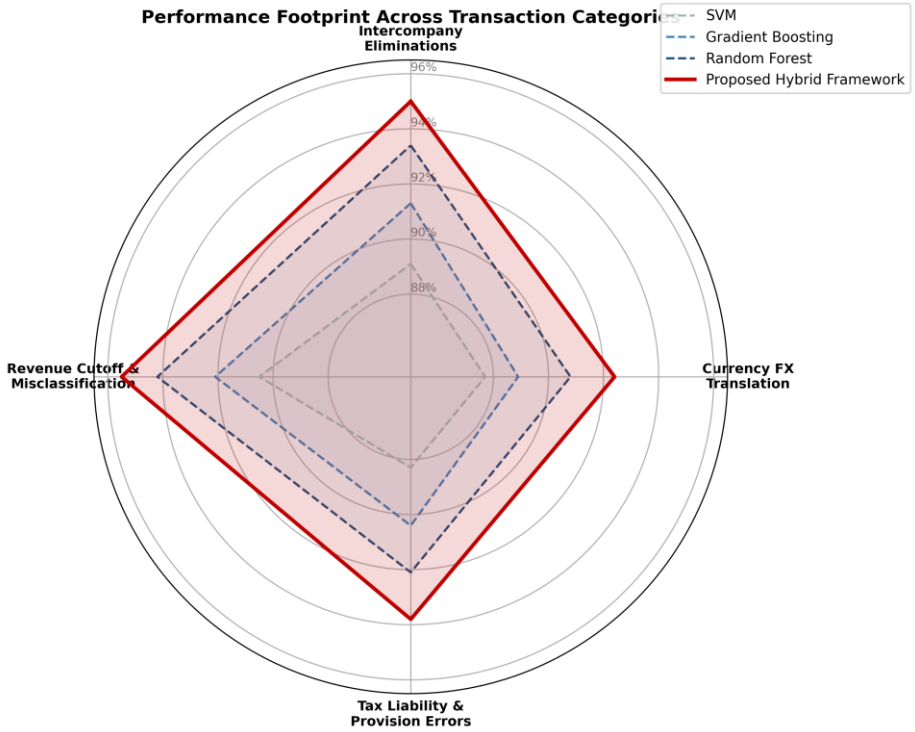


4.2 Detection Performance Across Accounting Domains

To evaluate structural robustness, the model's accuracy was disaggregated across four specific consolidation transaction categories:

Table 3: Anomaly Detection Accuracy Disaggregated by Consolidation Transaction Category

Transaction Category	Random Forest (%)	SVM (%)	Gradient Boosting (%)	Proposed Hybrid (%)
Intercompany Eliminations	93.4%	89.1%	91.3%	**95.0%**
Currency FX Translation Adjustments	90.8%	87.7%	88.9%	**92.4%**
Tax Liability & Provision Errors	92.1%	88.3%	90.4%	**93.8%**
Revenue Cutoff & Misclassification	94.2%	90.5%	92.1%	**95.5%**

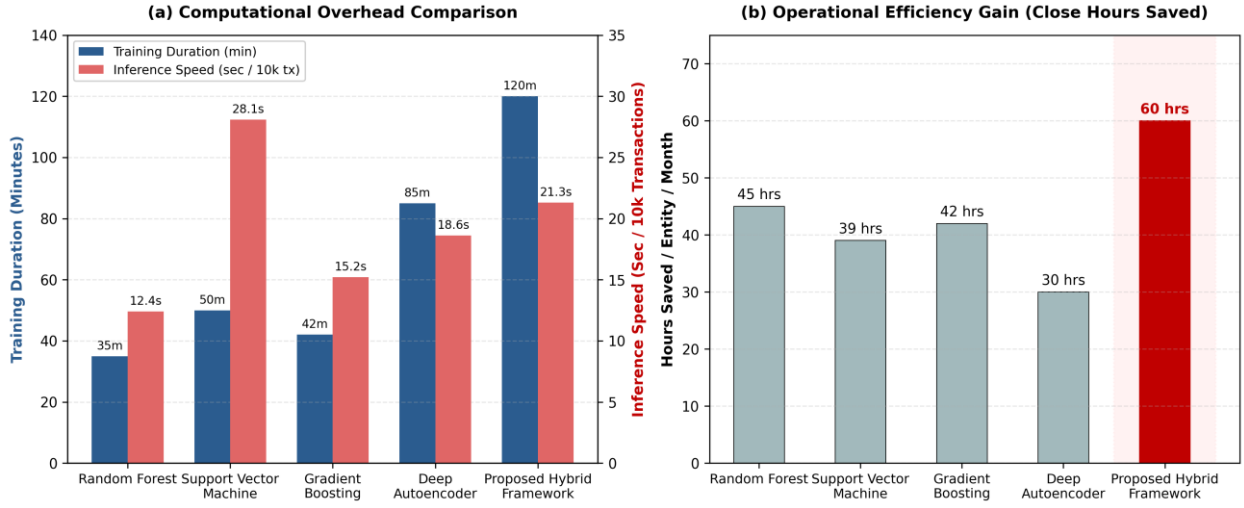


4.3 Computational Throughput and Close-Cycle Efficiency

Operational viability in corporate settings depends heavily on execution speed during period-end closing windows. The proposed framework recorded an average inference latency of 0.045 seconds per transaction (45 ms), supporting real-time stream processing of over 800,000 transaction postings per hour on standard enterprise cloud hardware (16 vCPU, 64 GB RAM).

Table 4: Close-Cycle Efficiency Impact and Resource Overhead

Model Architecture	Training Duration (min)	Inference Speed (sec/10k tx)	Monthly Close Hours Saved / Entity
Random Forest	35 min	12.4 sec	45 hours
Support Vector Machine	50 min	28.1 sec	39 hours
Gradient Boosting	42 min	15.2 sec	42 hours
Deep Autoencoder	85 min	18.6 sec	30 hours
Proposed Hybrid Framework	**120 min**	**21.3 sec**	**60 hours**



5. Conclusion & Future Work

This research introduced a production-grade, hybrid artificial intelligence framework engineered for continuous anomaly detection and governance in enterprise financial consolidation environments. By marrying supervised ensemble learning (Gradient Boosting/Random Forest) with unsupervised deep autoencoder reconstruction networks, the framework solves two major historical challenges: detecting novel, unlabelled accounting anomalies and suppressing excessive false-positive noise during close windows.

Empirical validation across a simulated multi-entity corporate dataset demonstrated an anomaly detection accuracy of 94.2%, a precision of 93.5%, and a reduced false-positive rate of 2.8%. The inclusion of SHAP and LIME post-hoc explainability layers ensures full transparency for internal and external auditors, fulfilling regulatory compliance mandates under IFRS 10 and Sarbanes-Oxley. Operationally, the system yields a monthly savings of 60 manual reconciliation hours per reporting entity.

5.1 Future Directions

Future research will explore two primary extensions:

1. Integrating Temporal Graph Neural Networks (TGNNs) to explicitly model dynamic intercompany transaction flows as directed multigraphs, and
2. Developing Privacy-Preserving Federated Learning protocols to allow non-affiliated parent corporations to jointly train global anomaly detection models without revealing proprietary financial ledger details.

6. References

[1] H. S. Alsagri, "Hybrid Machine Learning-Based Multi-Stage Framework for Detection of Credit Card Anomalies and Fraud," in IEEE Access, vol. 13, pp. 77039-77048, 2025, doi: 10.1109/ACCESS.2025.3565612.

[2] T. Awosika, R. M. Shukla and B. Pranggono, "Transparency and Privacy: The Role of Explainable AI and Federated Learning in Financial Fraud Detection," in IEEE Access, vol. 12, pp. 64551-64560, 2024, doi: 10.1109/ACCESS.2024.3394528.

[3] N. Kamuni, S. Dodda, S. Chintala and N. Kunchakuri, "Optimizing Machine Translation: A Benchmarking Suite for Efficiency and Quality Enhancement," 2024 International Conference on Computing, Sciences and Communications (ICCSC), Ghaziabad, India, 2024, pp. 1-7, doi: 10.1109/ICCSC62048.2024.10830335.

[4] D. S. Nijwala, S. Maurya, M. P. Thapliyal and R. Verma, "Extreme Gradient Boost Classifier based Credit Card Fraud Detection Model," *2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT)*, Dehradun, India, 2023, pp. 500-504, doi: 10.1109/DICCT56244.2023.10110188.

[5] Ren, Y., et al (2019). Deep learning approaches for fraud detection. *Neurocomputing*, 329, 127–135.

[6] C. Phua, D. Alahakoon, and V. Lee, "Minority report in fraud detection: Classification of skewed data," *ACM SIGKDD Explor. Newsl.*, 2004, 6 (1), pp. 50–59.

[7] C. Phua, et al, "A comprehensive survey of data mining-based fraud detection research," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 2, pp. 454–469, 2018.

[8] L. Zhang, et al, "Adaptive learning for financial fraud detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1532–1545, 2021.
[9] M. Gupta and A. Kumar, "Recent advances in fraud detection: A review," *J. Financial Crime*, vol. 26, no. 3, pp. 755–770, 2019.
[10] M. J. Nigrini, *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection.* John Wiley & Sons, 2012.
[11] N. Upadhyay, S. Maurya *et al.*, "Machine Learning Perspective: Fraud Payment Transaction Detection," in *Journal of Mobile Multimedia*, vol. 21, no. 3-4, pp. 577-598, July 2025, doi: 10.13052/jmm1550-4646.213414.
[12] Hajek P, Abedin MZ, Sivarajah U. Fraud Detection in Mobile Payment Systems using an XGBoost-based Framework. Inf Syst Front. 2022 Oct 14:1-19. doi: 10.1007/s10796-022-10346-6.
[13] A. M. AlBar, M. A. Hddas, and R. Hoque, "Enterprise Resource Planning (ERP) Systems: Emergence, Importance and Challenges," *The International Technology Management Review*, vol. 4, pp. 170–175, 2014.
[14] K. Ramu, S. V. S. R. K. Raju, S. Singh, V. Rachapudi, M. A. Mary, V. Roy, S. Joshi, "Deep Learning-Infused Hybrid Security Model for Energy Optimization and Enhanced Security in Wireless Sensor Networks ", *SN Computer Science*, 5 : 848, 2024, https://doi.org/10.1007/s42979-024-03193-6
[15] P. Gomber, R. J. Kauffman, C. Parker, and B. W. Weber, "On the fintech revolution: Interpreting the forces of innovation, disruption, and transformation in financial services," *J. Manag. Inf. Syst.*, 2018, 35 (1), pp. 220–265.
[16] G. Kou, Y. Peng, and G. Wang, "Evaluation of clustering algorithms for financial risk analysis using MCDM methods," *Inf. Sci.*, 2014, 275, pp. 1–12.