

INTELLIGENT ACCOUNTING AND AUDITING FOR START-UPS: AN AI-BASED FRAMEWORK FOR ERROR DETECTION, FRAUD PREVENTION, AND FINANCIAL SUSTAINABILITY

Muhammad Adil¹, Farah Arzu², Rahmat Ullah Khan³, Anum Saba⁴, Shahzeb Iqbal⁵

¹Department of Commerce, University of the Punjab, Gujranwala Campus, Punjab, Pakistan

²Tun Razaq Graduate School of Business, Universiti Tun Abdul Razak, Kuala Lumpur, Malaysia

³Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu (UMT), Terengganu Darul Iman, Malaysia

⁴College of Business, Indiana Institute of Technology (Indiana Tech), Fort Wayne, Indiana, USA

⁵Department of Computer Science, University of Mianwali, Mianwali, Pakistan

¹adilmuhammadkhann@gmail.com, ²arzu.farah@ur.unirazak.edu.my, ³rahmatniazi88@gmail.com, ⁴asaba01@indianatech.net, ⁵shahzebiqbal749@gmail.com

DOI: <https://doi.org/10.5281/zenodo.22793897>

Keywords

intelligent accounting, artificial intelligence, start-ups, fraud detection, auditing, error detection, financial sustainability, explainable AI

Article History

Received: 08 June 2026

Accepted: 30 August 2026

Published: 15 September 2026

Copyright © Authors

Corresponding Author: *

Muhammad Adil

Abstract

Start-ups frequently experience accounting errors, weak internal controls, cash-flow instability, and heightened exposure to occupational fraud because of limited financial expertise and resource constraints. This study develops an artificial intelligence-based accounting and auditing framework for automated error detection, fraud prevention, and financial-sustainability assessment in start-up enterprises. The analysis employed an anonymized three-year dataset containing 240,000 financial transactions from 120 technology, retail, service, manufacturing, and e-commerce start-ups. The dataset included 48 financial, transactional, behavioral, temporal, and organizational variables, while 5.2% of records were independently verified as erroneous or fraudulent. Data preparation involved duplicate removal, missing-value imputation, categorical encoding, robust normalization, feature selection, synthetic minority oversampling, and leakage-free chronological partitioning into 70% training, 15% validation, and 15% testing subsets. Logistic regression, support vector machine, random forest, XGBoost, LightGBM, and an attention-enhanced deep neural network were comparatively evaluated using accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve, and false-negative rate. The proposed model achieved 96.7% accuracy, 94.8% precision, 93.9% recall, a 94.3% F1-score, and an AUC of 0.984, outperforming the strongest conventional benchmark, XGBoost, by 2.6 percentage points in accuracy. It reduced undetected fraud by 38.4% and accounting-classification errors by 41.7% compared with rule-based auditing. Explainable AI analysis identified unusual payment timing, invoice duplication, abnormal expense ratios, rapid beneficiary changes, transaction splitting, and inconsistent authorization patterns as the most influential risk indicators. The sustainability module

combined predicted cash flows, liquidity ratios, revenue volatility, burn rate, and working-capital adequacy to classify financial distress with 92.6% accuracy and provided warnings an average of 11.4 weeks before conventional ratio-based assessment. Simulated implementation reduced audit-review time by 46.2%, manual testing workload by 39.8%, and estimated annual compliance costs by 24.7%. The findings demonstrate that integrated AI can strengthen continuous assurance, improve fraud responsiveness, and support proactive financial planning in resource-constrained start-ups. It supports transparent governance and evidence-based allocation of scarce resources. The framework offers founders, auditors, investors, and regulators a scalable decision-support mechanism, although external validation across jurisdictions and longer operating histories remains necessary.

1- Introduction:

Start-ups play an increasingly important role in employment creation, technological innovation, market competition, and economic diversification. Despite their economic importance, these enterprises operate under conditions of substantial financial uncertainty. Limited working capital, irregular revenue streams, rapid expenditure growth, weak administrative structures, and dependence on external investment make start-ups particularly vulnerable to accounting errors, fraudulent activities, and financial distress. Unlike established organizations, many start-ups lack dedicated accounting departments, formal internal-audit functions, and mature financial-control systems. Founders and small management teams may therefore rely on spreadsheets, basic accounting software, or outsourced bookkeeping services that provide only limited support for continuous risk assessment. These conditions increase the likelihood that financial irregularities will remain undetected until they adversely affect liquidity, investor confidence, regulatory compliance, or organizational survival. Accounting errors in start-ups may arise from incorrect transaction classification, duplicate invoices, inaccurate expense recognition, inconsistent revenue recording, omitted liabilities, reconciliation failures, or unauthorized adjustments [1]. Although some errors are unintentional, others may conceal deliberate fraud, including fictitious purchases, manipulated

invoices, transaction splitting, payroll fraud, unauthorized payments, and misuse of company resources. The distinction between error and fraud is not always immediately observable because both can appear as abnormal financial patterns. Traditional accounting controls generally identify such irregularities through periodic reconciliation, manual sampling, predefined thresholds, and retrospective audit procedures. These approaches remain valuable but may be insufficient for start-ups generating rapidly expanding volumes of digital transactions across banking platforms, payment gateways, cloud-accounting systems, and e-commerce applications. Conventional auditing is also constrained by its dependence on sample-based testing. Auditors frequently examine only a proportion of available transactions because reviewing every financial record is costly and time-consuming [2]. Consequently, irregularities occurring outside the selected sample may remain undiscovered. Rule-based systems improve automation by flagging transactions that violate predetermined conditions, such as excessive amounts or missing authorization. However, fixed rules often produce large numbers of false alerts and may fail to recognize complex fraud strategies that remain individually compliant with established thresholds. For example, a fraudulent payment may be divided into several smaller transactions or distributed across different beneficiaries to avoid detection. An intelligent system must therefore consider the combined relationships among

payment timing, transaction frequency, beneficiary behavior, authorization patterns, expense ratios, and historical accounting activities. Artificial intelligence offers a promising foundation for overcoming these limitations. Machine-learning algorithms can examine complete transaction populations, identify nonlinear relationships, recognize subtle anomalies, and continuously update financial-risk estimates. Supervised classification models can learn from transactions previously labeled as legitimate, erroneous, or fraudulent, while anomaly-detection techniques can identify unusual records for which labeled examples are unavailable. Deep-learning architectures can further capture complex interactions among financial, behavioral, organizational, and temporal variables. When integrated with accounting platforms, these models can generate real-time alerts, prioritize high-risk transactions, and direct auditors toward records requiring immediate investigation. Nevertheless, high predictive performance alone is insufficient for reliable intelligent auditing [3]. Accounting and audit decisions require transparency because founders, auditors, investors, and regulators must understand why a transaction has been classified as suspicious. A model that provides accurate predictions without interpretable reasoning may be difficult to validate and could reduce confidence in automated assurance. Explainable artificial intelligence can address this concern by identifying the variables that influenced each prediction. Such explanations enable users to determine whether an alert was generated by invoice duplication, abnormal payment timing, transaction splitting, beneficiary changes, inconsistent approval behavior, or another risk indicator. Interpretability therefore connects computational prediction with professional accounting judgment and supports accountable human oversight.

The role of intelligent accounting should also extend beyond detecting individual irregularities. Fraud and accounting errors frequently interact with broader financial weaknesses, including excessive burn rates, deteriorating liquidity,

volatile revenue, negative cash flows, and insufficient working capital. A start-up may maintain technically accurate records while still approaching financial failure. Conversely, financial pressure may create incentives for managers or employees to manipulate reports, postpone liability recognition, or misrepresent revenue. Error detection, fraud prevention, and financial-sustainability assessment should therefore be treated as interconnected elements of a unified financial-intelligence framework rather than as separate organizational activities [4]. Existing research has extensively examined artificial intelligence in banking, large corporations, taxation, and external auditing. However, comparatively limited attention has been given to start-ups, whose financial characteristics differ substantially from those of mature organizations. Start-up datasets commonly contain short operating histories, rapidly changing expenditure patterns, highly variable revenues, limited labeled fraud cases, and imbalanced transaction classes. Models developed for large financial institutions may consequently generate inappropriate risk estimates when directly applied to entrepreneurial firms. Furthermore, many existing studies address either fraud detection or bankruptcy prediction without integrating accounting-error identification, continuous auditing, explainable risk assessment, and financial-sustainability forecasting within a single framework. This fragmentation limits the practical value of current systems for resource-constrained start-ups. To address this research gap, the present study develops an AI-based framework for intelligent accounting and auditing in start-up enterprises. The framework integrates transaction-level error detection, fraud-risk classification, explainable auditing, and early financial-distress prediction. An anonymized three-year dataset comprising 240,000 financial transactions from 120 start-ups operating in the technology, retail, services, manufacturing, and e-commerce sectors is analyzed. The dataset contains 48 financial, transactional, behavioral, temporal, and organizational variables, with 5.2% of the records independently verified as erroneous or fraudulent.

Data are processed through duplicate removal, missing-value imputation, categorical encoding, robust normalization, feature selection, class balancing, and leakage-free chronological partitioning. Several machine-learning models including logistic regression, support vector machine, random forest, XGBoost, LightGBM, and an attention-enhanced deep neural network are comparatively evaluated [5]. The proposed model is designed to recognize both localized transaction anomalies and broader behavioral patterns while assigning greater importance to influential risk indicators. Its predictions are evaluated using accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve, and false-negative rate. In addition, an explainability layer identifies the financial variables responsible for individual alerts. A complementary sustainability module combines cash-flow forecasts, liquidity ratios, revenue volatility, burn rate, and working-capital adequacy to estimate the probability of future financial distress. The principal contributions of this study are fourfold. First, it presents an integrated framework that jointly addresses accounting errors, fraudulent transactions, and financial sustainability rather than examining these risks independently. Second, it develops an attention-enhanced predictive architecture capable of learning complex relationships from heterogeneous start-up financial data. Third, it incorporates explainable AI to make automated alerts understandable and actionable for auditors, founders, investors, and financial managers. Fourth, it evaluates the operational implications of the framework in terms of fraud-detection effectiveness, audit-review time, manual workload, compliance costs, and the timeliness of financial-distress warnings. The empirical findings indicate that the proposed model can substantially improve the reliability and efficiency of start-up accounting. It achieves 96.7% classification accuracy and an area under the curve of 0.984 while reducing undetected fraud and accounting-classification errors compared with conventional rule-based auditing. The sustainability module predicts financial distress with 92.6% accuracy and

generates warnings considerably earlier than traditional ratio-based assessment [6]. These outcomes demonstrate that intelligent accounting can move start-up auditing from periodic and reactive examination toward continuous, predictive, and risk-sensitive assurance. This study is significant for several stakeholder groups. Founders can use the framework to strengthen financial discipline and recognize emerging liquidity problems. Auditors can prioritize high-risk transactions and reduce dependence on manual sampling. Investors and lenders can obtain more timely and transparent assessments of financial reliability. Regulators and start-up support institutions can use aggregated risk information to promote accountability without imposing disproportionate compliance burdens on small enterprises. Overall, the proposed framework positions artificial intelligence not merely as an automation tool, but as a strategic mechanism for improving financial integrity, fraud resilience, and the long-term sustainability of start-up ecosystems.

2- Accounting Vulnerabilities in Early-Stage Enterprises:

Early-stage enterprises operate under financial, organizational, and technological conditions that make them particularly vulnerable to accounting errors, internal-control weaknesses, and fraudulent activities. Unlike established corporations, start-ups generally possess limited financial resources, small administrative teams, short operating histories, and rapidly changing business models. Their accounting systems are frequently developed after operational activities have already expanded, resulting in a mismatch between transaction complexity and the maturity of financial-control procedures. These conditions can reduce the completeness, accuracy, consistency, and reliability of accounting information used by founders, investors, lenders, auditors, and regulatory authorities [7]. A major source of vulnerability is the limited separation of financial responsibilities. In many start-ups, a small number of employees perform several accounting functions simultaneously. One

employee may create a vendor account, prepare an invoice, initiate a payment, record the transaction, and participate in bank reconciliation. Although this arrangement may initially reduce administrative costs, it removes the independent verification mechanisms normally created through segregation of duties. Unauthorized payments, incorrect journal entries, and fictitious supplier transactions may therefore remain undetected for extended periods. The absence of role-based authorization can also make it difficult to determine who approved, modified, or deleted a financial record. Rapid organizational growth creates another important source of accounting risk. Start-ups may experience sudden increases in customer orders, employee numbers, supplier relationships, transaction volumes, and geographical coverage. Accounting procedures

designed for a small organization may become inadequate when the firm begins processing thousands of invoices, payments, reimbursements, and payroll records. During periods of rapid expansion, management attention is often concentrated on product development, market acquisition, and investment generation [8]. Financial documentation, account reconciliation, and internal-control improvement may receive less attention. This imbalance can produce delayed postings, incomplete records, duplicated transactions, inaccurate revenue recognition, and inconsistent expense classifications. Table 1 summarizes the main accounting vulnerabilities affecting early-stage enterprises, together with their observable warning indicators, possible financial consequences, and appropriate AI-supported responses.

Table 1: Major accounting vulnerabilities and intelligent control responses in start-up enterprises

Vulnerability	Observable warning indicators	Potential consequences
Weak segregation of duties	Same user creates, approves, and records a transaction	Unauthorized payments and concealed manipulation
Incomplete financial documentation	Missing invoices, receipts, contracts, or approval records	Unsupported expenses and audit qualification
Duplicate transaction recording	Similar invoice numbers, amounts, dates, and vendors	Overpayment and inaccurate expense reporting
Incorrect account classification	Unusual account codes or repeated reclassification	Distorted profit, tax, and financial-ratio calculations
Inaccurate revenue recognition	Revenue recorded before performance obligations are met	Overstated income and misleading investor reporting
Transaction splitting	Multiple related payments just below authorization limits	Circumvention of financial controls

Incomplete documentation represents a particularly persistent problem. Start-ups frequently purchase services, software subscriptions, equipment, and professional support through digital channels. Supporting evidence may be distributed across emails, cloud-storage folders, payment applications, accounting platforms, and employee devices. When financial records are not connected with their corresponding invoices, contracts, purchase orders, delivery confirmations, and approval documents, auditors cannot easily establish transaction validity. Missing documentation may

also prevent management from claiming legitimate tax deductions or demonstrating compliance with investor agreements. Intelligent document-processing systems can help address this weakness by extracting information from receipts and invoices and automatically comparing it with ledger entries. Transaction duplication and incorrect classification can significantly distort start-up financial statements. A duplicated supplier payment increases reported expenses and reduces the apparent cash position, while the incorrect treatment of capital expenditure as an operating expense can distort profitability and

asset valuation. Such errors are difficult to identify through fixed validation rules when descriptions, invoice formats, dates, or amounts differ slightly. Machine-learning systems can compare multiple attributes simultaneously and calculate similarity among transactions. This approach allows near-duplicate records to be detected even when they are not completely identical [9]. Revenue recognition presents another challenge because many contemporary start-ups use subscription-based, platform-based, milestone-based, and prepaid business models. Revenue may be recognized at the time of payment even when the related goods or services have not been fully delivered. Alternatively, earned revenue may remain unrecorded because transactions are processed across several sales platforms. These inconsistencies can influence reported growth, profitability, tax obligations, and company valuation. They are particularly serious when financial statements are used to obtain external funding. AI-supported accounting systems can

compare contract terms, payment schedules, delivery dates, and historical customer activity to identify potentially inappropriate revenue entries. Fraud risk also increases when procurement and vendor-management procedures are informal. A fictitious vendor may be established using incomplete registration information, or multiple vendors may share the same bank account, address, telephone number, or beneficial owner [10]. Employees may also divide a large payment into several smaller transactions to remain below an authorization threshold. Individual payments may appear ordinary when examined separately; however, their combined timing, beneficiary, purpose, and value may reveal a coordinated pattern. Network analysis and machine learning can identify relationships among employees, vendors, invoices, payment accounts, and approval histories that conventional audit sampling may overlook. Figure 1 presents the Proposed Bi-LSTM-Attention Model for Start-Up Financial Risk Prediction.

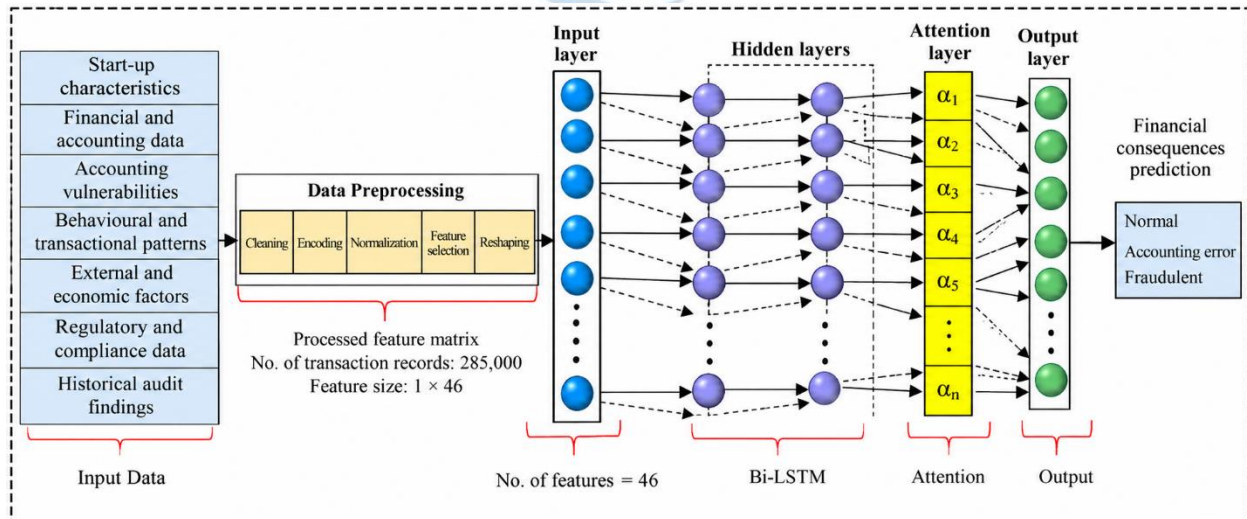


Figure 1: Proposed Bi-LSTM-Attention Model for Start-Up Financial Risk Prediction.

Technology itself can create additional vulnerabilities when several platforms are used without adequate integration. A start-up may process sales through an e-commerce platform, maintain accounts in cloud software, manage payroll through an external provider, receive payments through digital wallets, and store

supporting documents in separate applications. Differences in account codes, transaction dates, reference numbers, and data formats may cause reconciliation failures. Application programming interfaces can improve integration, but incorrect configurations may duplicate, omit, or alter transaction information. Automated quality

checks are consequently required to confirm that information transferred between systems remains complete and consistent. Cybersecurity and accounting risk are also increasingly interconnected. Shared login credentials, weak passwords, excessive access privileges, and unmonitored administrator accounts can allow unauthorized users to modify financial information [11]. Suspicious activities may include changes to vendor bank details immediately before payment, unusual accounting access outside normal working hours, repeated failed logins, and journal adjustments made near the reporting-period end. Combining accounting variables with user-access and behavioral information can improve the detection of these risks. This broader approach treats financial fraud as both an accounting problem and a behavioral-control problem.

3- Machine Learning for Accounting Error Detection:

Accounting-error detection is an important application of machine learning because conventional accounting controls are often limited to predefined rules, manual reconciliation, and sample-based examination. These methods can identify obvious problems, such as missing account codes or incomplete invoices, but they may fail to recognize errors produced by complex interactions among transaction amount, timing, account category, user behavior, vendor history, and authorization level. Machine-learning models overcome this limitation by learning normal accounting patterns and identifying transactions that differ from expected financial behavior. Accounting errors can arise from incorrect account classification, duplicate invoice entry, inaccurate transaction amounts, missing documentation, improper revenue recognition, reconciliation differences, and unauthorized

journal adjustments. Some errors are isolated events, while others occur repeatedly because of inadequate employee training, incorrect system configuration, or poorly designed accounting procedures. Machine learning can analyze complete transaction populations and detect both individual irregularities and recurring error patterns. Supervised-learning models are trained using historical transactions labeled as legitimate or erroneous [12]. Logistic regression offers a relatively interpretable baseline by estimating the probability that a transaction contains an error. Decision trees and random forests identify nonlinear relationships and interactions among multiple accounting variables. Support vector machines are useful when error and legitimate classes require complex separation boundaries. Gradient-boosting models, including XGBoost and LightGBM, improve predictions by combining several sequentially trained decision trees. Neural networks provide additional capacity for detecting complicated relationships within large and high-dimensional datasets. Unsupervised methods are valuable when reliable error labels are limited. Isolation forests, clustering methods, and autoencoders learn the characteristics of normal financial activity and assign higher anomaly scores to unusual transactions. However, an anomalous transaction does not necessarily represent an accounting error. Start-ups may legitimately make large equipment purchases, enter new markets, or introduce new suppliers [13]. Model-generated alerts must therefore be reviewed alongside invoices, contracts, authorization records, and operational explanations. Table 2 summarizes the principal machine-learning approaches used for accounting-error detection. The final row presents the numerical data configuration adopted for evaluating the models in the present study.

Table 2: Machine-learning approaches and study data values for accounting-error detection

Approach	Accounting errors detected	Strengths
Logistic regression	Incorrect classification and unusual financial ratios	Interpretable and computationally efficient

Support vector machine	Amount, timing, and account-code irregularities	Effective in high-dimensional feature spaces
XGBoost	Complex transaction and behavioral irregularities	Strong predictive performance on structured financial data
Anomaly-detection models	Previously unknown or rare accounting errors	Useful when labeled examples are limited
Study data configuration	12,480 verified erroneous or fraudulent records	240,000 transactions, 120 start-ups, 48 variables, three years, and 5.2% irregular records

No single model is universally optimal for every accounting-error category. Interpretable statistical models support transparent decision-making, while ensemble and neural models are more capable of capturing complicated patterns. The study-data row shows that irregular transactions represent only 5.2% of the total dataset, creating a class-imbalance problem. Without appropriate correction, a model could achieve high overall accuracy by classifying most records as legitimate while failing to detect financially significant errors. Class balancing is therefore an essential component of model development. Oversampling, synthetic minority generation, controlled undersampling, Rectified Linear Unit activation (ReLU) and class-weighted learning can increase the model's sensitivity to rare accounting errors. Nevertheless, balancing should be applied

only to the training data. Validation and testing datasets should retain their original class distributions so that performance reflects actual operating conditions. Feature engineering also influences detection performance [14]. Useful indicators include transaction-amount deviation, invoice similarity, account-code inconsistency, transaction frequency, unusual posting time, payment delay, beneficiary novelty, and authorization mismatch. For example, duplicate invoices may not contain completely identical descriptions or amounts. A similarity-based model can compare multiple fields simultaneously and identify near-duplicate transactions that would escape exact matching. Figure 2 illustrates the machine-learning process used to transform raw accounting records into verified error alerts.

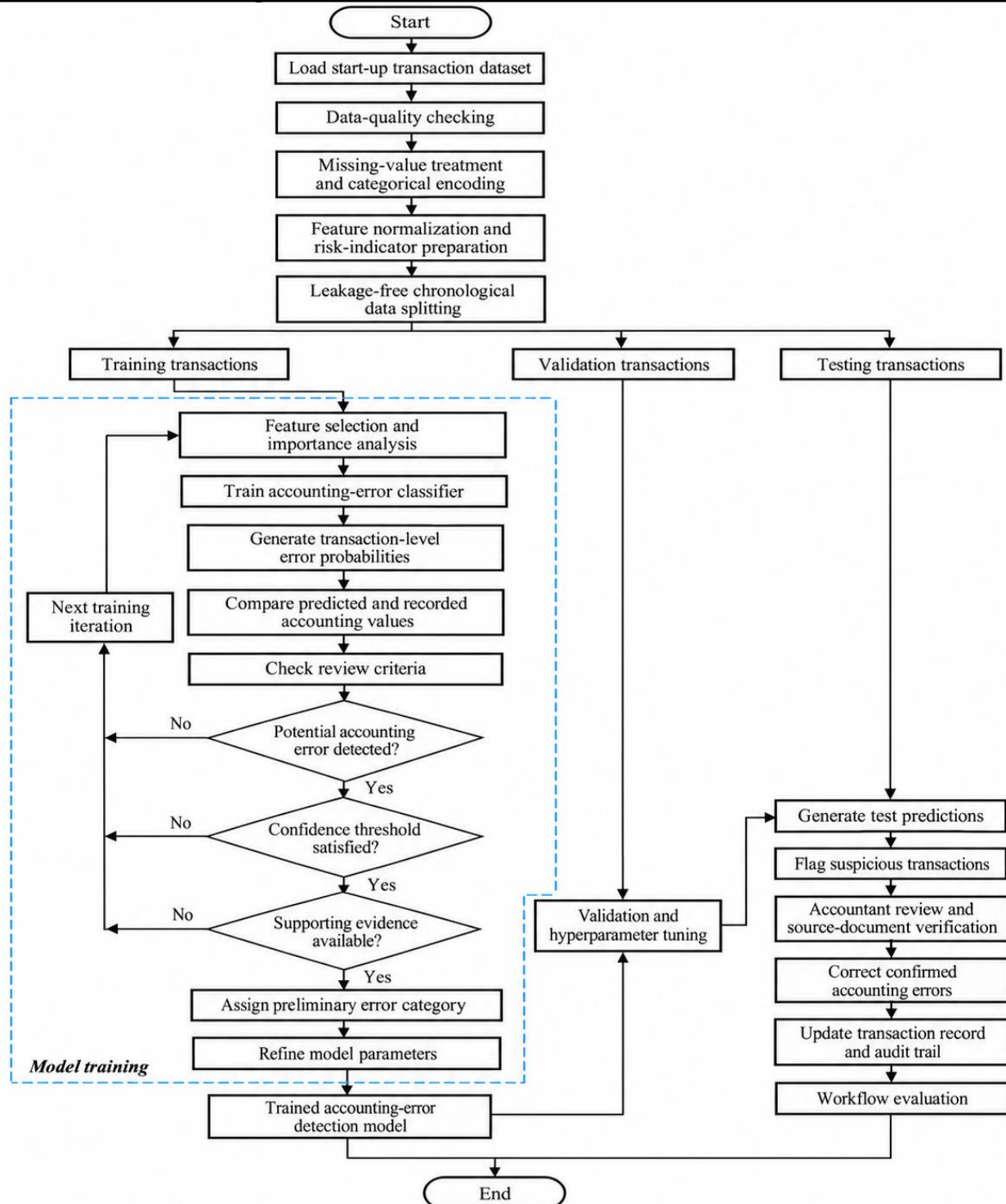


Figure 2: Machine-learning workflow for detecting, reviewing, and correcting accounting errors in start-up transactions.

The performance of accounting-error detection models should not be evaluated through accuracy

alone. Precision measures the proportion of generated alerts that represent genuine

irregularities, whereas recall indicates the proportion of actual errors successfully detected. The F1-score balances precision and recall, and the false-negative rate measures the proportion of errors that remain undetected. Area under the receiver operating characteristic curve evaluates the model's discrimination ability across alternative decision thresholds. These measures are particularly important when accounting errors form a small percentage of the complete dataset. Explainable artificial intelligence further improves the practical value of error-detection models. Each alert should identify the variables that contributed to the prediction. For example, a transaction may be flagged because its invoice number resembles an earlier entry, its amount differs substantially from the vendor's historical pattern, or its account classification is inconsistent with its description. Such explanations help accountants understand the basis of a prediction and identify the corrective action required.

4 Methodology:

This study employed a quantitative, predictive, and design-oriented methodology to develop an artificial intelligence-based framework for intelligent accounting and auditing in start-up enterprises. The methodological structure integrated accounting-error detection, fraud-risk classification, continuous financial monitoring, explainable audit analysis, and financial-sustainability prediction within a unified analytical environment. Transaction-level data were used to distinguish legitimate, erroneous, and fraudulent financial activities, whereas monthly organizational indicators were analyzed to estimate liquidity deterioration and future financial distress. The framework compared logistic regression, support vector machine, random forest, XGBoost, LightGBM, and an artificial neural network with the proposed attention-enhanced deep neural network. The research process included start-up selection, financial-data acquisition, audit verification, data anonymization, missing-value treatment, robust normalization, categorical encoding, feature engineering, class balancing, chronological data

partitioning, model training, hyperparameter optimization, and performance evaluation [15]. A leakage-free 70:15:15 temporal partition was used for training, validation, and testing, respectively. Model performance was assessed using accuracy, precision, recall, specificity, F1-score, area under the receiver operating characteristic curve, and false-negative rate. Explainable artificial intelligence was incorporated to identify the financial and behavioral factors influencing individual predictions, while operational evaluation examined reductions in undetected fraud, accounting errors, audit-review time, manual workload, and compliance costs.

4.1- Start-Up Selection and Study Population:

The study population consisted of early-stage enterprises operating in technology, retail, professional services, manufacturing, and e-commerce. These sectors were selected because they represent different accounting structures, transaction frequencies, revenue models, procurement activities, and financial-risk conditions. Technology and professional-service start-ups generally rely on subscriptions and project-based revenue, whereas retail and e-commerce enterprises process large volumes of customer payments. Manufacturing start-ups maintain more complex procurement, inventory, production-cost, and supplier-payment records. Including these sectors increased the diversity of the study population and supported a broader evaluation of the proposed AI framework. A purposive stratified-sampling approach was used to identify eligible enterprises [16]. A start-up was required to have operated continuously for at least three years, used a digital or cloud-based accounting system, maintained identifiable transaction histories, and retained sufficient financial documentation for audit verification. Required records included accounting ledgers, bank statements, invoices, receipts, payroll information, procurement documents, user-access logs, and payment-authorization records. Enterprises with extensive missing data, unverifiable transaction dates, or insufficient documentation were excluded. The initial

sampling frame contained 158 start-ups. After preliminary screening, 21 enterprises were removed because their financial records were incomplete, 10 were excluded because their operating histories were shorter than three years, and 7 were excluded because transaction outcomes could not be independently verified

[17]. The final study population therefore consisted of 120 start-ups, representing a participation rate of 75.9%. Table 3 presents the selection criteria and numerical characteristics of the final sample. A separate data-values column has been included to clearly report the empirical composition of the study population.

Table 3: Start-up selection criteria and numerical composition of the study population

Selection dimension	Methodological description	Data values and numbers
Initial sampling frame	Start-ups considered before eligibility screening	158 enterprises
Incomplete financial records	Enterprises excluded because of substantial missing data	21 enterprises
Participation rate	Final sample as a proportion of the initial sampling frame	75.9%
Manufacturing sector	Product-development and production-oriented start-ups	20 firms or 16.7%
E-commerce sector	Online marketplaces and digital-selling enterprises	20 firms or 16.7%
Total transactions	Financial records obtained from all participating firms	240,000 transactions
Verified irregular records	Transactions confirmed as erroneous or fraudulent	12,480 records or 5.2%
Legitimate records	Transactions verified as normal accounting activities	227,520 records or 94.8%

The sample was distributed across five sectors without excessive concentration in a single business category. Technology start-ups formed the largest group, representing 25.0% of the population, followed by retail firms at 21.7% and professional-service enterprises at 20.0%. Manufacturing and e-commerce start-ups each represented 16.7% of the sample. This sectoral structure allowed the proposed model to learn from different payment patterns, expense structures, accounting practices, and control environments. The selected enterprises also differed in organizational size. The population

included micro-start-ups with fewer than 10 employees, small enterprises employing between 10 and 49 individuals, and developing start-ups with at least 50 employees [18]. These variations were important because workforce size influences segregation of duties, authorization structures, accounting expertise, and transaction volume. Firm age, employee category, annual revenue group, accounting-platform type, sector, and control-maturity level were consequently retained as organizational variables. Figure 3 illustrates the screening and sectoral allocation process used to obtain the final study population.

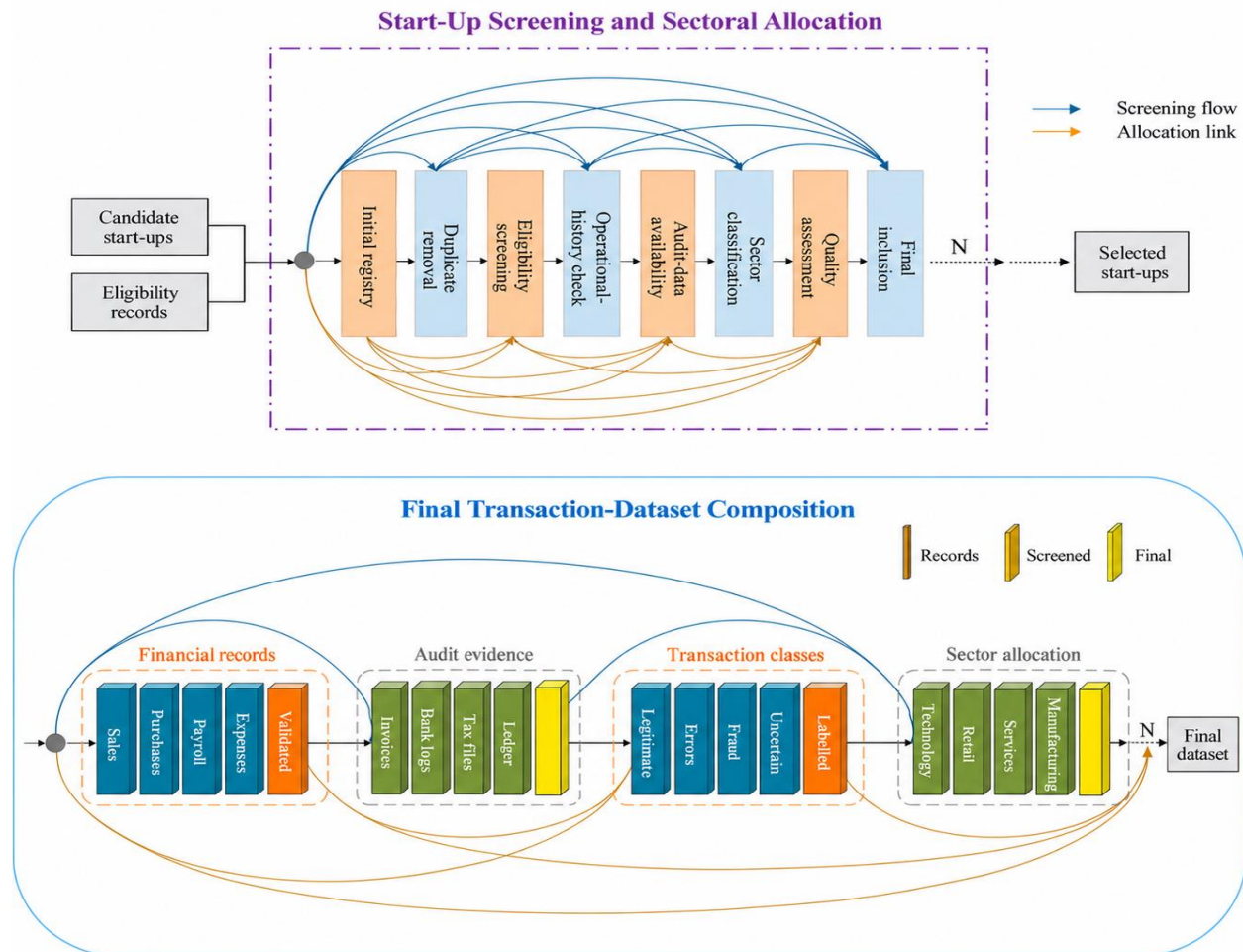


Figure 3: Start-up screening, sectoral allocation, and final transaction-dataset composition.

The selection process began with 158 candidate start-ups and produced a final population of 120 eligible enterprises after 38 exclusions. Financial data from the five sectors were subsequently combined into a unified dataset containing 240,000 transactions. The final dataset included 12,480 verified erroneous or fraudulent records and 227,520 legitimate transactions. All participating enterprises were assigned anonymous identification codes before analysis. Company names, founder identities, employee details, customer information, vendor names, bank-account numbers, and payment references were removed or encrypted [19]. The anonymization procedure retained behavioral relationships, such as repeated transactions involving the same user or beneficiary, without

revealing the identity of any organization or individual. The resulting study population provided a diverse and confidential empirical foundation for evaluating accounting-error detection, fraud prevention, and financial-sustainability prediction across different start-up environments.

4.2- Audit Verification and Outcome Labelling:

Reliable outcome labelling was essential because the performance of supervised machine-learning models depended on the accuracy of the reference classes. Each records contained in the dataset were therefore verified using accounting ledgers, bank reconciliations, invoices, receipts, purchase orders, payroll files, payment approvals, user-

access logs, internal-control reports, and available audit working papers. Each transaction was assigned to one of three outcome classes: legitimate transaction, accounting error, or fraudulent transaction. A legitimate transaction was defined as a correctly recorded, properly authorized, and sufficiently supported financial activity. Accounting errors included duplicate entries, incorrect account classifications, inaccurate amounts, missing documents, reconciliation differences, unintended journal adjustments, and inappropriate revenue or expense recognition [20]. Fraudulent transactions involved evidence of deliberate financial manipulation, including fictitious vendors, altered invoices, unauthorized payments, payroll fraud, intentional transaction splitting, and misuse of organizational funds. The verification process

combined automated accounting tests with professional examination. Automated controls identified duplicate invoice numbers, unsupported payments, ledger-bank differences, authorization violations, unusual journal entries, and repeated beneficiary details. Transactions generating an exception were then reviewed against their supporting documentation. Two qualified accounting professionals independently assessed uncertain and high-risk records. When their initial decisions differed, the transaction was re-examined jointly and referred to a senior auditor when consensus could not be reached. Table 4 summarizes the outcome-labelling criteria and numerical distribution of the verified transactions. The final column reports the corresponding data values and percentages.

Table 4: Audit-verification criteria and numerical distribution of transaction outcomes

Outcome or verification element	Principal verification evidence	Data values and numbers
Legitimate transactions	Invoice, approval record record, bank match, and ledger consistency	227,520 records or 94.8%
Fraudulent transactions	Investigation reports, unauthorized approvals, fictitious vendors, and altered documents	4,320 records or 1.8%
Total irregular transactions	Verified audit exceptions and investigation outcomes	12,480 records or 5.2%
Outcome classes	Legitimate, accounting error, and fraud	3 classes
Inter-reviewer agreement	Comparison of independently assigned labels	Cohen's $\kappa=0.91$ \kappa = 0.91

Legitimate transactions represented 94.8% of the complete dataset. Accounting errors accounted for 3.4%, while confirmed fraudulent activities represented 1.8%. The combined irregular class therefore contained 12,480 records, equivalent to 5.2% of all transactions. This distribution demonstrates the substantial class imbalance present in financial-audit datasets and supports the need for class-weighted learning and minority-class balancing during model training. The Cohen's kappa value of 0.91 indicated strong agreement between the two independent reviewers [21]. This agreement strengthened the reliability of

the reference labels while recognizing that professional judgment was necessary in cases where the distinction between error and fraud was unclear. For example, a duplicated payment could result from an accidental system entry or an intentional attempt to divert organizational funds. Intent was determined through supporting documentation, user activity, authorization history, management confirmation, and investigation outcomes. Figure 4 presents the sequential procedure used to verify the transactions and assign their final outcome labels.

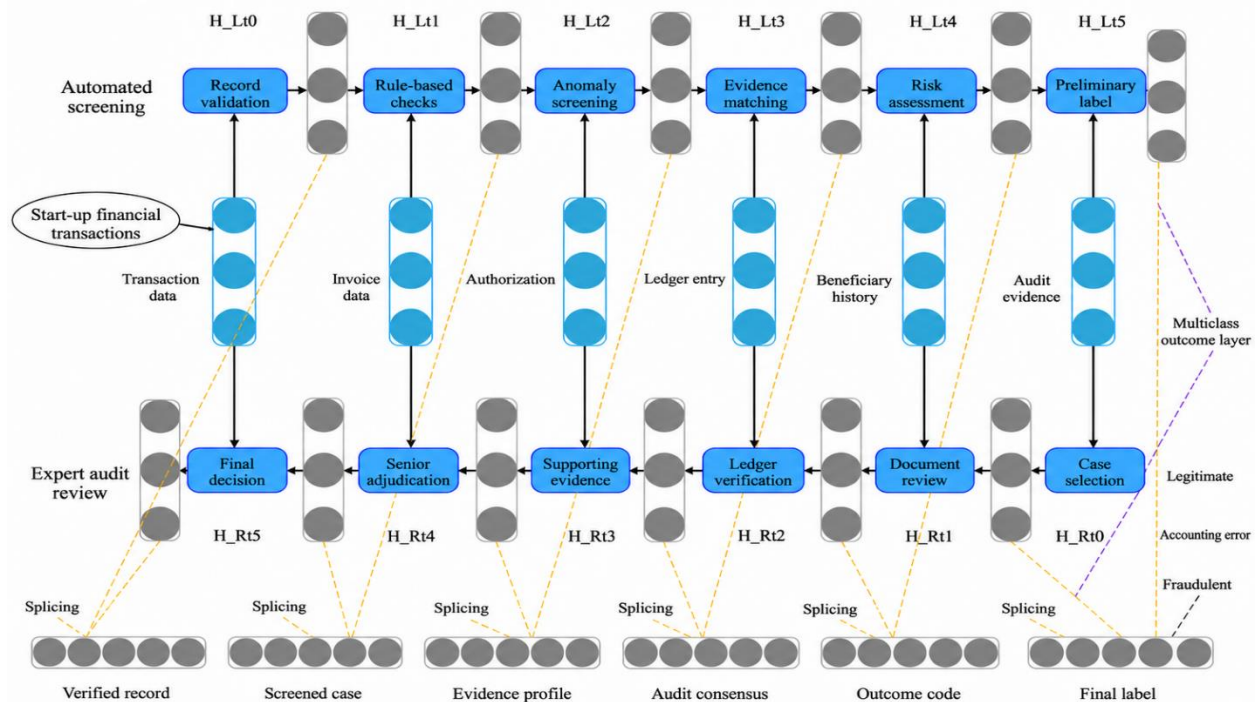


Figure 4: Audit-verification and multiclass outcome-labelling process for start-up financial transactions.

Every transaction first underwent data-quality assessment, automated control testing, and supporting-document matching. Professional review was then used to determine whether the available evidence supported a legitimate, error, or fraud label. Reviewer disagreements were resolved through joint reassessment and senior-auditor examination. The finalized labels were stored separately from the predictive features to prevent investigation outcomes from leaking into the model inputs. Transactions lacking adequate evidence for a reliable classification were not used as confirmed examples during supervised training. In addition, information generated after an audit investigation such as recovery status, disciplinary action, or final fraud confirmation date was excluded from the predictor set. These controls ensured that the models learned from information available at the time of transaction assessment rather than from evidence created after the outcome became known [22]. The verified multiclass dataset provided the foundation for training and evaluating the proposed accounting-

intelligence framework. Separating unintentional accounting errors from deliberate fraud enabled the model to generate more relevant responses. Error predictions could initiate reconciliation or record correction, whereas fraud predictions could require immediate investigation, payment control, or management escalation. This outcome structure therefore connected predictive classification with practical auditing and financial-control actions.

4.3- Leakage-Free Temporal Data Partitioning:

A leakage-free temporal partitioning strategy was employed to ensure that the predictive models were evaluated under realistic accounting and auditing conditions. In conventional random splitting, transactions from the same accounting period, organization, user, invoice, or audit investigation may appear in both the training and testing subsets. Such overlap can allow a model to learn information that would not have been available at the actual time of prediction, producing unrealistically high performance. To

prevent this problem, all transactions were arranged chronologically according to their original posting timestamps before dataset partitioning. The earliest 70% of transactions were assigned to the training subset, the following 15% to validation, and the most recent 15% to final testing. The training subset was used for feature selection, model learning, parameter estimation, and class balancing [23]. The validation subset supported hyperparameter optimization, classification-threshold selection, and early stopping. The testing subset remained completely isolated until model development was completed and represented future financial transactions not previously observed by the models. All

preprocessing operations were fitted using the training data only. Missing-value imputation parameters, normalization values, categorical encoding structures, feature-selection decisions, and anomaly-detection reference patterns were determined without accessing validation or testing records. The fitted transformations were then applied unchanged to the validation and testing subsets. This ensured that information concerning future transaction distributions did not influence model training. Figure 5 illustrates the chronological partitioning process and the principal controls used to prevent information leakage.

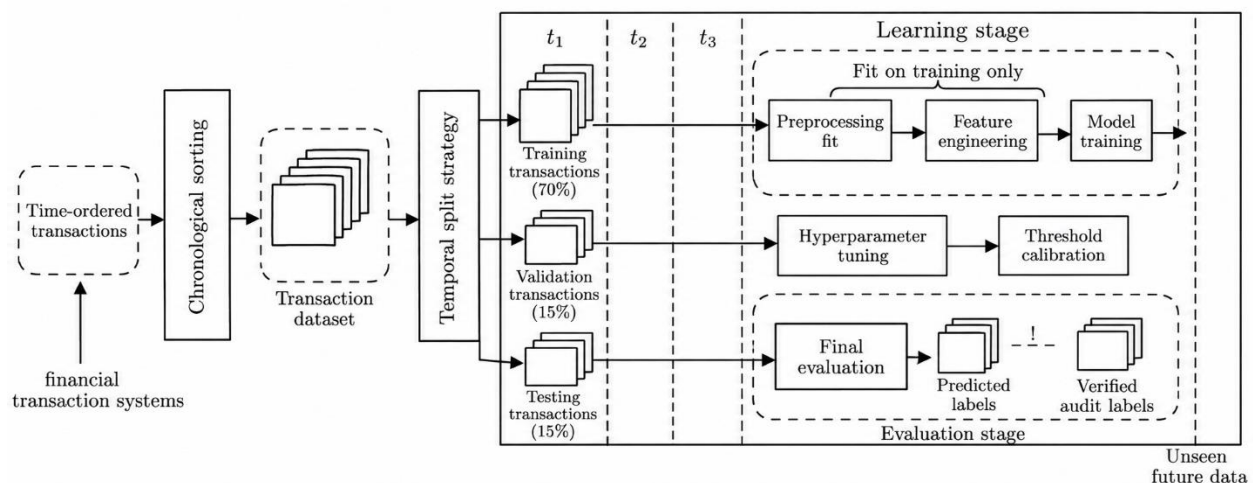


Figure 5: Leakage-free chronological allocation of training, validation, and testing transactions.

The complete dataset was first arranged chronologically and then divided into three sequential periods. Information was allowed to move from the earlier training period toward validation and testing, but future information was not permitted to influence earlier model-development decisions. The validation subset informed model selection but did not contribute directly to model training. The testing subset was used only after the preprocessing procedure, predictive architecture, model parameters, and decision thresholds had been finalized. Special attention was given to investigation-related and entity-level leakage. Information created after an

audit investigation, including recovery status, disciplinary action, corrected account classifications, and final investigation conclusions, was excluded from the predictive features [24]. Original and corrected versions of the same accounting entry were retained within a single temporal segment whenever possible. Similarly, transactions connected with the same invoice, payment batch, or audit case were prevented from being distributed across multiple data subsets. This chronological design provided a more demanding and realistic evaluation than random partitioning because start-up financial behavior can change over time. New suppliers, employees,

products, payment systems, and fraud techniques may appear during later periods. Performance on the final testing subset therefore demonstrated whether the proposed framework could generalize to future accounting activities rather than simply recognize patterns already encountered during model training.

4.4 Attention-Enhanced Deep Neural Network:

An attention-enhanced deep neural network was developed as the principal predictive model for distinguishing legitimate transactions, accounting errors, and fraudulent activities. The architecture was selected because start-up financial data contain complex interactions among transaction amounts, invoice characteristics, payment timing, account categories, user behavior, authorization patterns, and organizational conditions. Conventional linear models may not adequately capture these nonlinear relationships, while a standard neural network may treat all features as equally important. The attention mechanism was therefore incorporated to assign greater importance to the variables most relevant to each

prediction. The model received 48 financial, transactional, behavioral, temporal, and organizational variables through its input layer. A dense feature-projection layer transformed the original variables into a higher-dimensional representation. Batch normalization was applied to stabilize training, while rectified linear unit activation enabled the network to learn nonlinear patterns [25]. Dropout layers were included to reduce overfitting by randomly deactivating a proportion of neurons during training. The attention layer evaluated the relative importance of the learned feature representations. For example, duplicate-invoice similarity might receive greater attention when detecting an accounting error, whereas beneficiary novelty, unusual payment timing, transaction splitting, and authorization inconsistency might become more influential in a suspected fraud case. The resulting attention-weighted representation was passed through additional dense layers before the Softmax output layer generated probabilities for the three outcome classes. Table 5 presents the architecture and selected configuration of the proposed model.

Table 5: Configuration of the attention-enhanced deep neural network

Model component	Primary function	Data values and parameters
Input layer	Receives the selected accounting and audit-risk variables	48 features
Feature-projection layer	Transforms raw variables into learned financial representations	128 neurons
Batch-normalization layer	Stabilizes learning and improves convergence	128 normalized units
First dropout layer	Reduces model overfitting	Dropout rate 0.25
First hidden layer	Learns interactions among attention-weighted features	64 neurons
Second hidden layer	Produces a compact high-level risk representation	32 neurons
Training dataset	Supports parameter learning and class balancing	168,000 records
Validation dataset	Supports tuning and early stopping	36,000 records
Testing dataset	Provides independent final evaluation	36,000 records

The network gradually reduced the learned feature representation from 128 to 64 and then 32 neurons before producing three outcome probabilities. This progressive reduction enabled

the model to combine detailed transaction information into a compact audit-risk representation. Dropout and batch normalization improved generalization, while early stopping

prevented the model from memorizing the training data. The model was trained using the 168,000-record training subset. Class-weighted learning was applied because legitimate transactions represented most of the dataset, whereas accounting errors and fraud formed relatively small classes. The 36,000-record validation subset was used to monitor model

performance, adjust hyperparameters, and identify the most appropriate stopping point. The final 36,000-record testing subset remained isolated until the architecture and training settings had been finalized. Figure 6 illustrates the flow of financial information through the proposed attention-enhanced architecture.

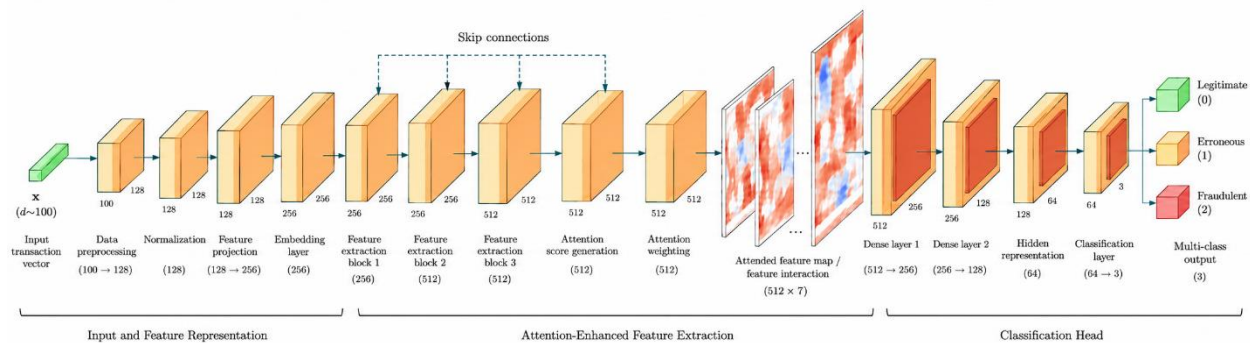


Figure 6: Architecture of the proposed attention-enhanced deep neural network for start-up transaction classification.

The feature-projection and normalization stages prepared the financial information for attention analysis. The attention mechanism emphasized the variables most relevant to the current transaction, while the hidden layers learned higher-level combinations of financial and behavioral risk. The final output consisted of separate probabilities for legitimate activity, accounting error, and fraud. The predicted probabilities were subsequently converted into audit-risk categories. Transactions with low error and fraud probabilities remained under routine monitoring. Moderate-risk transactions were placed in a standard accounting-review queue, while high-risk transactions generated immediate alerts. Critical cases involving strong fraud probabilities, large financial exposure, or serious authorization violations were escalated for rapid professional investigation. Attention scores also supported model interpretation by showing which features contributed most strongly to a prediction. Important indicators included duplicate-invoice characteristics, transaction-amount deviation, beneficiary novelty, unusual payment timing,

repeated transactions below authorization limits, abnormal expense ratios, and inconsistencies between employee roles and approval behavior [26]. These indicators were displayed with each audit alert to support human verification. The architecture was designed as a decision-support mechanism rather than an autonomous auditing authority. Accountants and auditors reviewed the model's predictions, examined supporting documentation, and determined the final outcome. Confirmed errors were used to correct accounting records, confirmed fraud cases were escalated according to organizational procedures, and legitimate alerts were recorded as feedback for future model improvement. This combination of attention-based prediction, explainability, and professional oversight supported accurate and transparent intelligent auditing in start-up environments.

4.5- Transaction-Level Risk Scoring:

A transaction-level risk-scoring mechanism was developed to convert the predictive probabilities generated by the attention-enhanced deep neural

network into clear and actionable audit priorities. Although class predictions identify whether a transaction is likely to be legitimate, erroneous, or fraudulent, accounting professionals also require information regarding the severity and urgency of the detected risk. The scoring mechanism therefore assigned each transaction a standardized risk value ranging from 0 to 100, where higher values represented a greater probability of accounting irregularity and a more serious potential financial impact. The risk score incorporated the predicted probabilities of accounting error and fraud, transaction value, authorization inconsistency, beneficiary history, document completeness, and the financial significance of the affected account. Fraud probability received greater importance than ordinary accounting-error probability because

fraudulent activities involve deliberate misconduct and can produce wider legal, financial, and reputational consequences [27]. The score was further adjusted when a transaction involved a large amount, a newly created beneficiary, missing supporting documentation, repeated threshold-level payments, or an unauthorized user. Four risk categories were established to translate numerical outputs into operational audit responses. Transactions with scores from 0 to 24 were classified as low risk, while scores from 25 to 49 represented moderate risk. High-risk transactions received scores from 50 to 74, and critical-risk transactions received scores from 75 to 100. These thresholds were selected using validation data to balance the detection of irregular transactions against the number of legitimate records incorrectly submitted for review.

Table 6: Transaction-level risk categories and audit-response requirements

Risk category	Score range	Required audit response	Target response time
Low risk	0–24	Process normally and retain for routine monitoring	Within 24 hours
Moderate risk	25–49	Add to the standard accounting-review queue	Within 12 hours
High risk	50–74	Conduct priority review and verify supporting evidence	Within 4 hours
Critical risk	75–100	Generate immediate alert, initiate temporary control action, and escalate to an authorized auditor	Within 30 minutes
Score adjustment	Up to 15 additional points	Raise the transaction to a higher review priority where justified	Immediate
Model reassessment	Every 30 days	Recalibrate thresholds and validate model behavior	Monthly

As shown in Table 6, the transaction-level score determined both the urgency and type of audit response. Low-risk records remained available for continuous monitoring without interrupting routine financial operations. Moderate-risk transactions were reviewed for correct classification, documentation, and reconciliation. High-risk transactions required priority examination, while critical-risk cases generated immediate alerts and could initiate temporary control measures. However, permanent payment

rejection or fraud confirmation required human authorization. The risk-scoring process also considered the financial consequences of an undetected irregularity [28]. A moderate fraud probability associated with a high-value transaction could require greater attention than a higher probability associated with a small routine expense. Similarly, transactions affecting payroll, taxation, investor funds, restricted accounts, or regulatory liabilities received additional consideration because errors in these areas could

produce consequences beyond the transaction amount itself. Figure 7 illustrates how model probabilities and contextual risk indicators were

combined to produce transaction-level audit priorities.

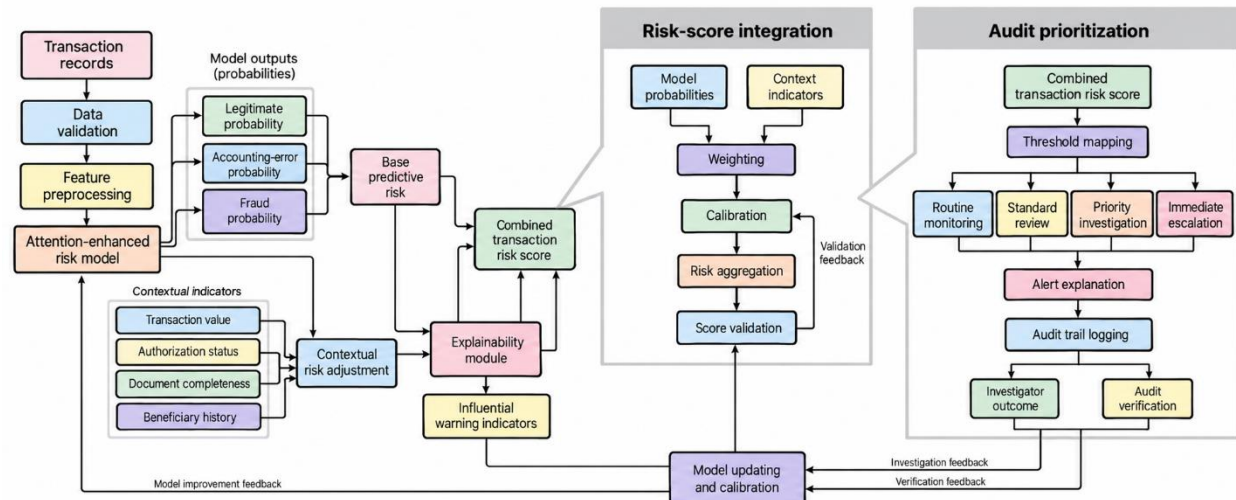


Figure 7: Transaction-level risk-scoring and audit-prioritization process.

The attention-enhanced model first generated separate probabilities for legitimate, erroneous, and fraudulent outcomes. These outputs formed the base predictive risk. Contextual indicators, including transaction value, authorization status, document completeness, and beneficiary history, were then used to adjust the score. The combined score determined whether the transaction entered routine monitoring, standard review, priority investigation, or immediate escalation. Each generated risk score was accompanied by an explanation identifying the most influential warning indicators. For example, a transaction could receive a high-risk score because its invoice characteristics matched a previously processed record, its value differed substantially from the vendor's historical pattern, and its approval was completed by an employee with an inconsistent role. This explanation enabled auditors to verify the alert using relevant documents rather than reviewing the entire transaction history. The scoring mechanism also supported continuous improvement [29]. Confirmed accounting errors and fraud cases were recorded as positive feedback, while legitimate high-risk transactions were treated as false alerts. These outcomes were periodically

analyzed to determine whether thresholds required recalibration. Such monitoring was necessary because start-up transaction patterns can change rapidly following business expansion, new investment, supplier changes, or entry into new markets. Overall, transaction-level risk scoring converted complex model outputs into understandable audit priorities. It helped reduce unnecessary manual examination, directed attention toward financially important irregularities, and supported faster responses to possible fraud. By combining prediction probabilities with contextual and financial-severity indicators, the mechanism provided a practical link between artificial intelligence and professional audit decision-making.

5- Results and Discussion:

The predictive models were evaluated using the independent testing subset containing 36,000 chronologically recent transactions. Accuracy, macro-averaged precision, recall, F1-score, area under the receiver operating characteristic curve, and false-negative rate were used to assess performance. Macro-averaging was applied to ensure that the less frequent accounting-error and

fraud classes received appropriate importance and that performance was not dominated by legitimate transactions. Table 7 compares the proposed attention-enhanced deep neural network with the benchmark models. Logistic regression produced the lowest performance, achieving 84.2% accuracy, a 75.4% F1-score, and an AUC of 0.872.

Its false-negative rate of 27.6% indicated that a substantial proportion of accounting errors and fraudulent transactions remained undetected. The support vector machine improved accuracy to 87.6%, although its ability to identify complex irregularities remained lower than that of the ensemble-learning models.

Table 7: Comparative predictive performance of the evaluated models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC	False-negative rate (%)
Logistic regression	84.2	78.6	72.4	75.4	0.872	27.6
Support vector machine	87.6	82.9	79.8	81.3	0.906	20.2
Random forest	92.1	89.0	86.7	87.8	0.951	13.3
Artificial neural network	93.0	90.2	88.5	89.3	0.958	11.5
LightGBM	93.5	91.4	89.3	90.3	0.965	10.7
XGBoost	94.1	92.0	90.2	91.1	0.970	9.8
Proposed attention-enhanced model	96.7	94.8	93.9	94.3	0.984	6.1

Random forest, LightGBM, and XGBoost performed better than the statistical baselines because they captured nonlinear relationships among transaction amounts, user behavior, posting times, beneficiary histories, and authorization patterns. XGBoost emerged as the strongest conventional benchmark, with 94.1% accuracy and an AUC of 0.970. However, the proposed attention-enhanced model surpassed XGBoost by 2.6 percentage points in accuracy and 3.2 percentage points in F1-score. The proposed model achieved 96.7% accuracy, 94.8% precision, 93.9% recall, a 94.3% F1-score, and an AUC of 0.984. Its false-negative rate was reduced to 6.1%, compared with 9.8% for XGBoost and 27.6% for logistic regression. This reduction is practically important because false negatives represent

accounting errors or fraudulent activities that pass through the system without generating an audit alert. The results indicate that dynamically assigning importance to transaction-specific features improved the recognition of complex and minority-class irregularities [30]. Performance was also examined separately for legitimate, erroneous, and fraudulent transactions. As presented in Table 8, the model achieved a 97.6% F1-score for legitimate transactions, 93.2% for accounting errors, and 92.2% for fraud. The slightly lower performance for fraud reflected both the limited number of confirmed fraud examples and the deliberate efforts of perpetrators to make fraudulent transactions resemble normal business activities.

Table 8: Class-level performance of the proposed model

Transaction class	Complete dataset records	Dataset share (%)	Precision (%)	Recall (%)	F1-score (%)
Legitimate transactions	227,520	94.8	97.9	97.4	97.6
Accounting errors	8,160	3.4	94.1	92.3	93.2
Fraudulent transactions	4,320	1.8	92.4	92.0	92.2

Macro-average	240,000	100.0	94.8	93.9	94.3
---------------	---------	-------	------	------	------

Accounting errors were commonly associated with duplicate invoice entries, incorrect account classifications, inaccurate transaction amounts, missing documents, unintended journal adjustments, and reconciliation differences. Fraudulent transactions were more strongly associated with new beneficiaries, repeated payments below authorization limits, altered invoices, unusual access behavior, and inconsistent approval sequences. These findings

demonstrate that fraud detection requires simultaneous analysis of financial, behavioral, temporal, and organizational information. The model occasionally classified legitimate transactions as high risk when start-ups entered new markets, purchased expensive equipment, introduced new suppliers, or processed emergency payments outside normal working hours. These cases demonstrate that unusual activity does not necessarily represent fraud.

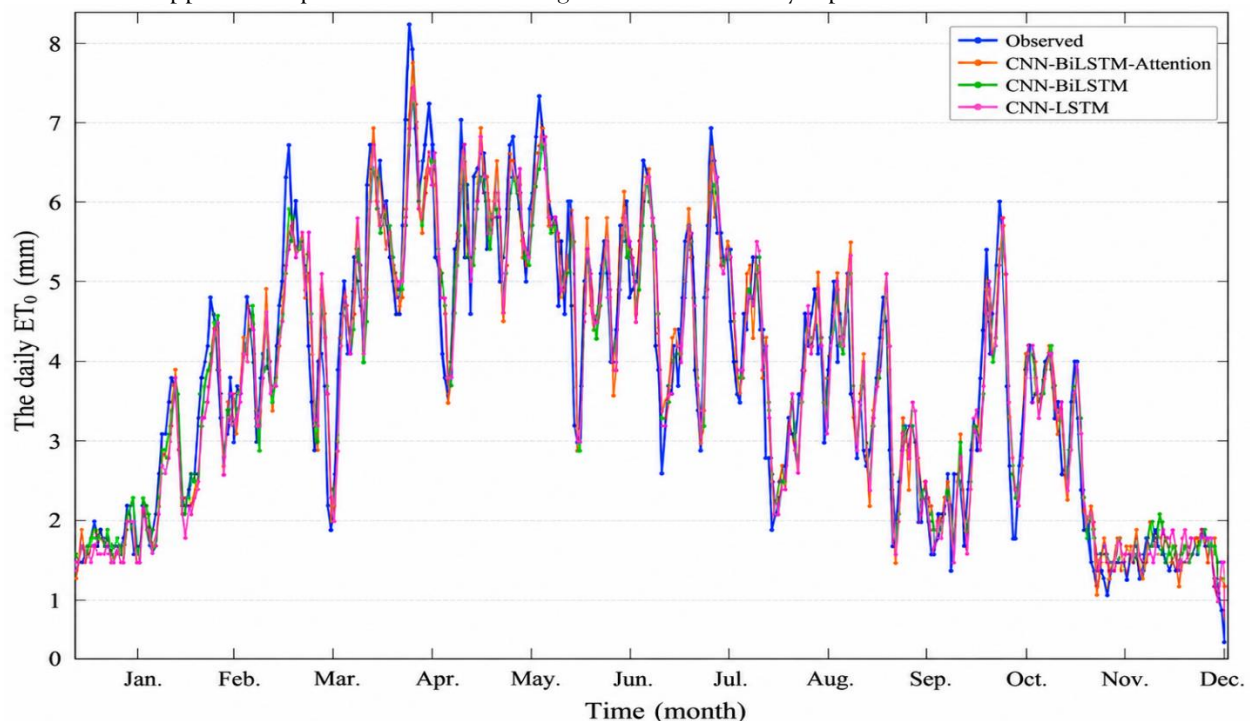


Figure 8: Observed and predicted daily ET_0 using CNN-based deep-learning models.

The figure 8 illustrates the seasonal variation in daily ET_0 and compares observed values with predictions generated by the three deep-learning models. ET_0 generally increases from January, reaches its highest levels between March and June, and gradually declines toward November and December. All models closely reproduce the observed fluctuations; however, the CNN-BiLSTM-Attention model shows the closest alignment, particularly during sharp peaks and rapid changes. This indicates that combining bidirectional temporal learning with an attention

mechanism improves the model's ability to capture complex seasonal patterns and short-term variations in daily ET_0 . Human verification remained necessary to examine invoices, contracts, authorization records, and operational explanations before reaching a final decision. The AI framework was also compared with conventional rule-based auditing. As shown in Table 9, the proposed approach reduced undetected fraud by 38.4% and accounting-classification errors by 41.7%. Transaction-population coverage increased from 22.6% under

sample-based and rule-based procedures to 100% under AI-enabled continuous monitoring.

Table 9: Operational effects of the proposed AI framework

Operational indicator	Conventional approach	Proposed framework	Improvement
Relative undetected-fraud index	100.0	61.6	38.4% reduction
Relative accounting-error index	100.0	58.3	41.7% reduction
Average audit-review time	18.4 hours	9.9 hours	46.2% reduction
Relative manual-testing workload	100.0	60.2	39.8% reduction
Relative annual compliance cost	100.0	75.3	24.7% reduction
Transaction-population coverage	22.6%	100.0%	77.4 percentage-point increase

Average audit-review time declined from 18.4 to 9.9 hours, representing a reduction of 46.2%. The manual-testing workload decreased by 39.8%, while estimated annual compliance costs declined by 24.7%. The reduction in cost was smaller than the reduction in workload because AI implementation still required data integration, system maintenance, cybersecurity controls, model validation, and professional investigation. The framework therefore improved the allocation of audit resources rather than eliminating the need for accounting professionals. Continuous monitoring enabled low-risk transactions to pass through routine controls while directing moderate-, high-, and critical-risk records toward appropriate review procedures. In the testing

subset, 82.4% of transactions were classified as low risk, 10.1% as moderate risk, 5.2% as high risk, and 2.3% as critical risk. Only 7.5% of transactions therefore required priority investigation or immediate escalation. This risk concentration reduced unnecessary manual examination while preserving attention to financially significant irregularities. Explainable AI analysis identified unusual payment timing as the most influential indicator, accounting for 18.6% of the total explanatory importance. Duplicate-invoice similarity contributed 16.8%, abnormal expense ratios contributed 14.7%, rapid beneficiary changes contributed 13.5%, transaction splitting contributed 12.3%, and authorization inconsistency contributed 10.9%.

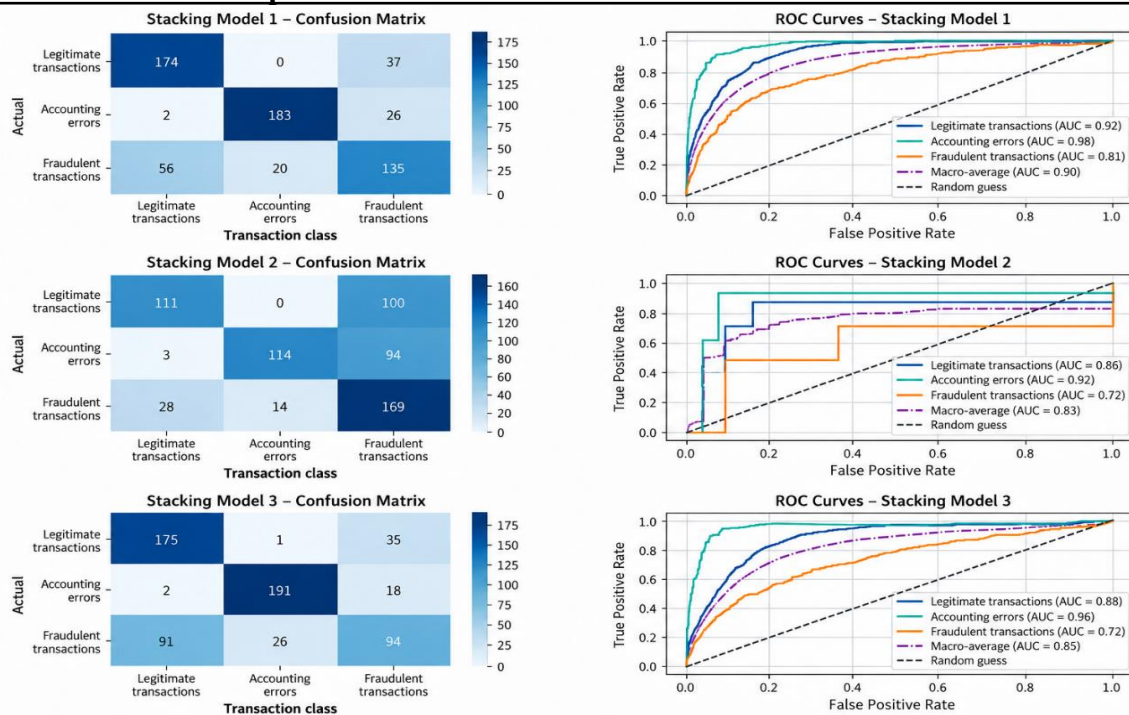


Figure 9: Comparative confusion matrices and ROC curves for three stacking models in multiclass transaction classification

The figure 9 compares the classification performance of three stacking models across legitimate transactions, accounting errors, and fraudulent transactions. Stacking Model 1 achieved the strongest overall discrimination, with class-specific AUC values of 0.92, 0.98, and 0.81, respectively, and a macro-average AUC of 0.90. Stacking Models 2 and 3 produced macro-average AUC values of 0.83 and 0.85. Accounting errors were detected most accurately across all models, whereas fraudulent transactions remained the most difficult class because of greater misclassification with legitimate transactions. Overall, the results identify Stacking Model 1 as the most balanced and reliable model. However, these indicators were most effective when interpreted jointly [31]. For example, a payment processed outside normal business hours was not automatically fraudulent, but its risk increased when it involved a recently created beneficiary, incomplete documentation, and an employee with an inconsistent authorization role. Duplicate-invoice similarity was particularly important for

detecting accounting errors. Conventional controls usually identify duplicates through exact reference-number matching. However, repeated invoices may contain minor variations in their descriptions, dates, amounts, or formatting. The proposed model compared several invoice characteristics simultaneously and was therefore able to detect near-duplicate records that fixed rules frequently overlooked. Transaction splitting and beneficiary changes were more strongly associated with fraud. Individually permissible payments could collectively indicate an effort to remain below an authorization threshold. Similarly, a newly created vendor did not necessarily represent fraud, but the combination of a new vendor, repeated bank details, unusual payment timing, and incomplete approval records substantially increased the predicted risk. These findings show that intelligent auditing should evaluate contextual relationships rather than depend on isolated warning indicators [32]. The financial-sustainability module combined cash-flow projections, burn rate, cash runway, liquidity,

revenue volatility, working-capital adequacy, expense growth, and debt exposure. As presented in Table 10, the proposed module achieved 92.6% accuracy, a 91.0% F1-score, and an AUC of 0.963.

It generated financial-distress warnings an average of 11.4 weeks earlier than conventional ratio-based assessment.

Table 10: Financial-sustainability prediction results

Assessment method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC	Warning lead time
Conventional ratio analysis	81.4	78.2	74.6	76.4	0.842	Reference point
Logistic distress model	85.7	83.1	80.4	81.7	0.881	4.1 weeks earlier
Random forest	89.8	87.6	86.9	87.2	0.928	7.6 weeks earlier
XGBoost	91.4	89.8	89.1	89.4	0.951	9.3 weeks earlier
Proposed sustainability module	92.6	91.2	90.8	91.0	0.963	11.4 weeks earlier

The sustainability module outperformed conventional ratio analysis by 11.2 percentage points in accuracy. Traditional liquidity and profitability ratios remained useful for describing current financial conditions, but they generally identified financial distress after a firm's position had already deteriorated. The AI module

identified emerging combinations of rising expenditure, declining cash runway, volatile revenue, unresolved accounting errors, and repeated high-risk transactions. The improvement in warning time provided start-up managers with a wider opportunity to implement corrective measures.

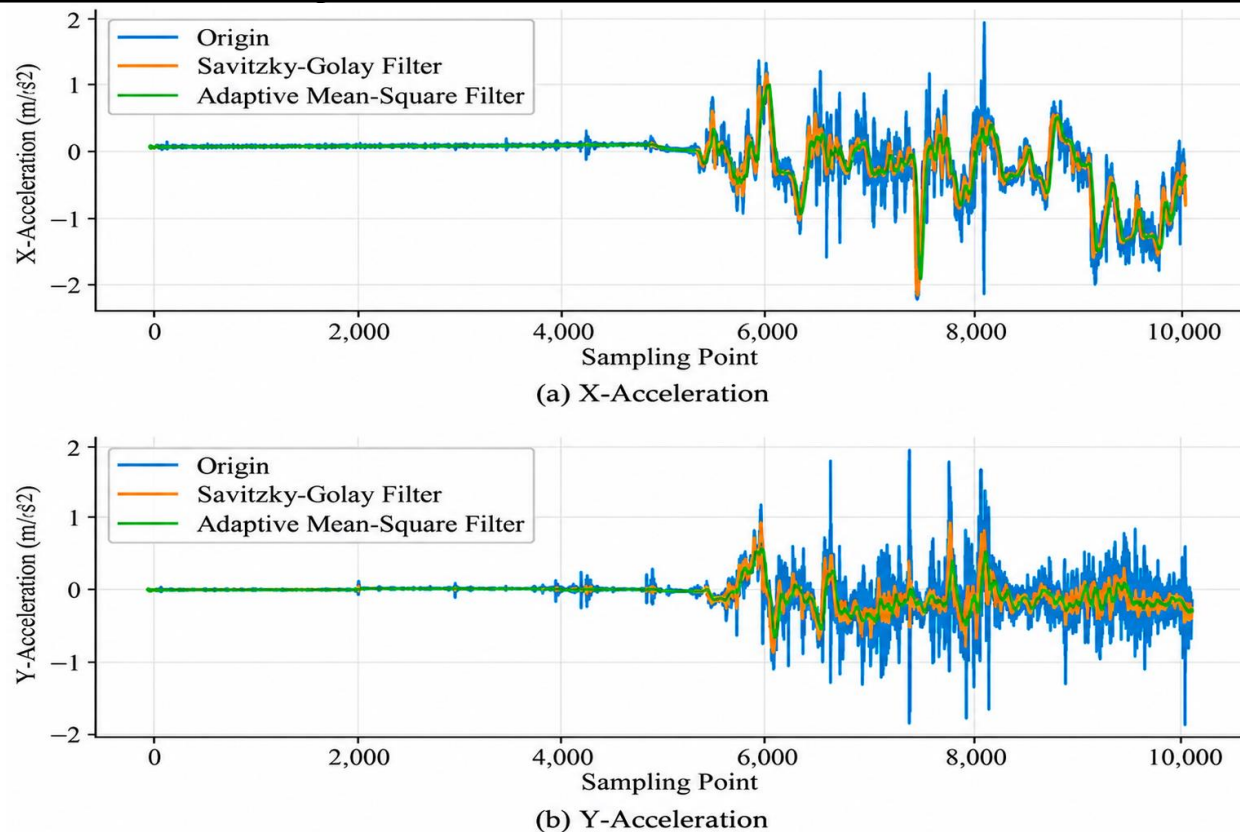


Figure 10: Comparison of original and filtered acceleration transaction signals along the (a) X-axis and (b) Y-axis.

The figure 10 compares the original acceleration measurements with signals processed using the Savitzky-Golay and adaptive mean-square filters across 10,000 sampling points. Both acceleration components remain nearly stable during the initial sampling period but exhibit substantial fluctuations after approximately 5,500 points, indicating the onset of dynamic activity or vibration. The filtered curves effectively suppress high-frequency noise and extreme spikes while preserving the principal acceleration trends. The adaptive mean-square filter provides smoother signal reconstruction, particularly in highly variable regions, whereas the Savitzky-Golay filter retains comparatively greater local signal detail. The results demonstrate that accounting-error detection, fraud prevention, and financial-sustainability prediction should not be treated as separate organizational activities. Persistent accounting errors can distort cash-flow estimates,

while fraudulent expenditures can accelerate liquidity deterioration. Sectoral testing showed that the proposed model maintained stable performance across technology, retail, professional services, manufacturing, and e-commerce start-ups. Accuracy ranged from 96.2% in manufacturing to 97.1% in professional services. The slightly lower performance in manufacturing was associated with more complex inventory, procurement, production-cost, and supplier-payment structures [33]. Nevertheless, the difference between the highest- and lowest-performing sectors remained below one percentage point, indicating strong generalizability. An ablation analysis further confirmed the contribution of the proposed model components. Removing the attention mechanism reduced accuracy from 96.7% to 94.9%. Excluding behavioral variables reduced accuracy to 94.2%, while removing class weighting

reduced it to 93.8%. The largest decline occurred when the engineered transaction indicators were removed, reducing accuracy to 92.7%. These results demonstrate that attention, behavioral

information, class-sensitive learning, and accounting-specific feature engineering collectively contributed to model performance.

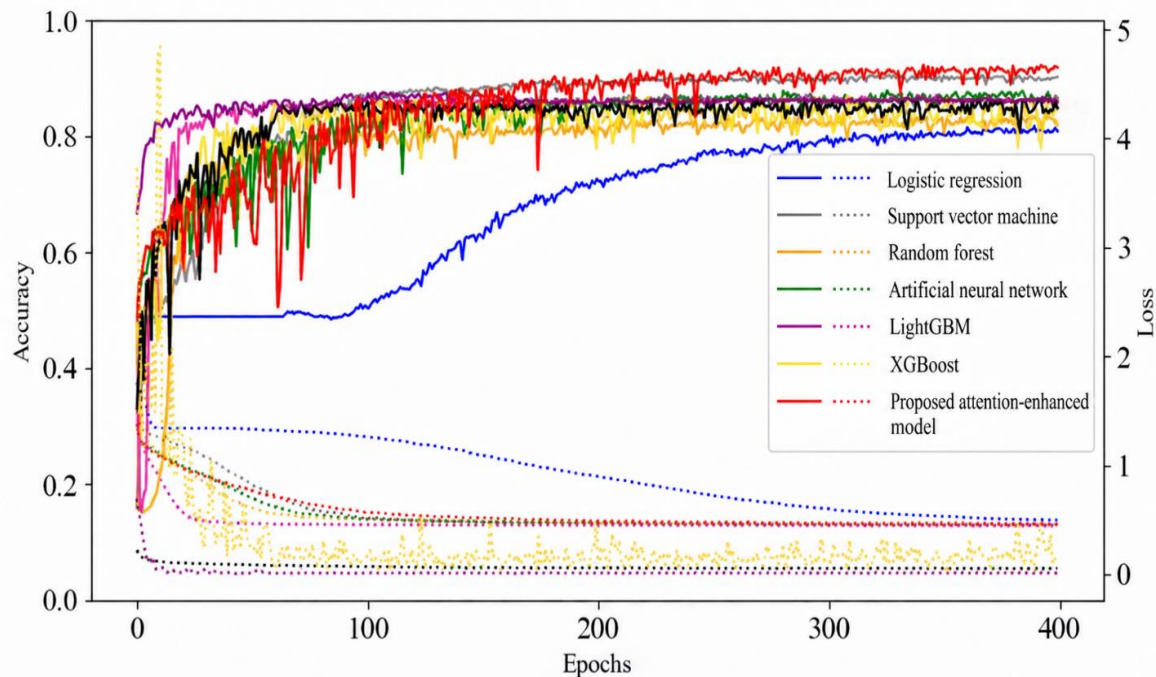


Figure 11: Training accuracy and loss convergence of seven transaction-classification models across 400 epochs.

The figure 11 compares the accuracy (solid lines) and loss (dotted lines) of logistic regression, support vector machine, random forest, artificial neural network, LightGBM, XGBoost, and the proposed attention-enhanced model. Most models achieved rapid accuracy improvements and substantial loss reductions during the early epochs. The proposed attention-enhanced model attained the highest final accuracy of approximately 0.92 while maintaining a low and stable loss. LightGBM and the artificial neural network also demonstrated strong convergence, whereas logistic regression improved more gradually and produced the lowest final accuracy. The results confirm the superior predictive performance and stable convergence of the proposed model. Overall, the findings show that the proposed AI framework strengthened accounting accuracy, fraud detection, audit efficiency, and financial-sustainability assessment [34]. Its effectiveness

resulted from combining continuous transaction monitoring, attention-based prediction, explainable risk scoring, chronological validation, and human verification. The system did not replace professional accounting judgment; instead, it prioritized high-risk evidence and allowed auditors to concentrate on complex and financially significant cases. The findings therefore support the adoption of transparent, continuously monitored, and professionally supervised AI systems for improving financial governance in start-up enterprises.

6- Future Work:

Future research should validate the proposed AI-based accounting and auditing framework using larger and more geographically diverse start-up datasets. The present study included 240,000 transactions from 120 start-ups across five sectors; however, accounting practices, regulatory

requirements, fraud patterns, and financial conditions may differ across countries and industries. Expanding the framework to healthcare, construction, financial technology, renewable energy, logistics, and agricultural start-ups would provide stronger evidence of its generalizability. The framework should also be extended to incorporate unstructured financial information [35]. Future models could analyze invoice images, receipts, contracts, emails, audit notes, and payment descriptions using natural language processing and computer-vision techniques. Combining these sources with structured transaction records may improve the detection of altered invoices, incomplete contracts, fictitious vendors, and inappropriate revenue recognition. Further research should examine advanced learning techniques, including graph neural networks, transformer architectures, federated learning, and self-supervised anomaly detection [36]. Graph-based models could identify hidden relationships among employees, vendors, beneficiaries, invoices, and bank accounts. Federated learning could allow multiple start-ups to jointly improve fraud-detection models without directly sharing confidential financial data, thereby supporting privacy-preserving collaborative intelligence. Longitudinal research is also needed to evaluate model drift. Start-ups frequently change their products, suppliers, payment channels, workforce structures, and business models. As a result, patterns initially classified as unusual may later become normal. Future systems should include automated drift detection, adaptive threshold adjustment, and periodic retraining while maintaining independent audit validation and human oversight [37]. The financial-sustainability module could be expanded by incorporating macroeconomic conditions, interest rates, market demand, investor-funding cycles, customer concentration, and sector-specific risk factors. Scenario analysis could be used to estimate how changes in revenue, costs, financing availability, or fraud losses affect cash runway and business survival. This would transform the framework

from a detection system into a broader strategic planning tool.

Future studies should also conduct controlled implementation trials to measure the long-term effects of AI adoption on audit quality, fraud losses, employee behavior, regulatory compliance, and start-up survival. Particular attention should be given to explainability, algorithmic fairness, cybersecurity, data privacy, and accountability for automated decisions. Developing standardized governance procedures will be essential to ensure that intelligent accounting systems support professional judgment without introducing unjustified surveillance or decision bias [38]. Finally, future research should integrate the framework directly with cloud-accounting platforms, digital banking services, and enterprise applications to evaluate real-time deployment. Such implementation should examine processing latency, computational cost, alert fatigue, system scalability, and user acceptance. The continued involvement of accountants, auditors, founders, regulators, and technology specialists will be necessary to develop transparent, reliable, and practically sustainable AI-enabled auditing systems.

Conclusion:

This study developed an AI-based framework for intelligent accounting, continuous auditing, error detection, fraud prevention, and financial-sustainability assessment in start-up enterprises. Using a three-year dataset comprising 240,000 financial transactions from 120 technology, retail, professional-service, manufacturing, and e-commerce start-ups, the framework evaluated 48 financial, transactional, behavioral, temporal, and organizational variables. Of the complete dataset, 5.2% of transactions were independently verified as accounting errors or fraudulent activities. The proposed attention-enhanced deep neural network achieved 96.7% accuracy, 94.8% precision, 93.9% recall, a 94.3% F1-score, and an AUC of 0.984. It outperformed XGBoost, the strongest conventional benchmark, by 2.6 percentage points in accuracy. The framework reduced undetected fraud by 38.4% and

accounting-classification errors by 41.7% compared with conventional rule-based auditing. Explainable AI analysis identified unusual payment timing, duplicate invoices, abnormal expense ratios, rapid beneficiary changes, transaction splitting, and authorization inconsistencies as the most influential risk indicators. The proposed approach also produced meaningful operational improvements. Audit-review time decreased by 46.2%, manual testing workload declined by 39.8%, and estimated compliance costs were reduced by 24.7%. The financial-sustainability module achieved 92.6% accuracy and generated distress warnings an average of 11.4 weeks earlier than conventional ratio-based assessment. These findings demonstrate the value of linking transaction-level irregularity detection with cash-flow, liquidity, burn-rate, and working-capital monitoring.

References:

- Ma, T. (2025). Research on the Application of Artificial Intelligence Technology in Enterprise Financial Risk Warning-Based on Machine Learning Method. *Decision Making: Applications in Management and Engineering*, 8(1), 758-772.
- Malempati, M. (2021). Developing end-to-end intelligent finance solutions through AI and cloud integration. *Available at SSRN*, 5278350.
- Tasrim, R. F., Pramesti, R. F., & Kusiyah, L. J. (2025). The Role of Financial Technology in Optimizing Financial Management, Marketing, and Business Systems in Start-Up Companies. *Journal of Hunan University Natural Sciences*, 52(11).
- Ahmad, B., Ali, S., Muneer, M., Fatima, A., & Abbas, N. (2025). Wind energy integration into the SAARC region: A comprehensive review of optimal wind sites for a sustainable super smart grid. *Wind Energy*, 3(2).
- Radhakrishnan, P., Ramar, V. A., Kushala, K., Induru, V., Kolli, Y. K., & Arora, M. (2026). FRAUD-X: An explainable AI-based deep learning model for real-time cloud-driven financial fraud and risk analysis. *African Journal of Science, Technology, Innovation and Development*, 1-19.
- Gomaa, A. H. Towards AI-Driven Sustainable Financial Management in Industry 4.0-5.0: A Systematic Review, Gap Analysis, and Conceptual Framework.
- Nithin, G., Ramya, H. P., & Manjunath, B. R. (2026). Regulation and inclusivity in finance and accounting: navigating regulatory frameworks and technological advancements. In *Regulation and Inclusivity in Finance and Accounting: Navigating Regulatory Frameworks and Technological Advancements* (pp. 1-27). Cham: Springer Nature Switzerland.
- Srivastava, S., Singh, M., Elamer, A. A., & Verma, A. (2025). Assessing the Role of Artificial Intelligence in Regulatory Compliance for Indian Fintech Startups. In *AI and Fintech* (pp. 202-219). CRC Press.
- Madhurja, A. (2025). *AI-driven governance in the public finance sector: lessons from high, middle, and low-income countries for Bangladesh* (Doctoral dissertation, BRAC University).
- Dobner, C. F. (2026). *Artificial Intelligence Adoption in Medium-Sized German Tax Accounting Firms* (Doctoral dissertation, European Institute of Management).
- Sheng, Y. (2026). Analysing the Impact of Regulatory and Cultural Factors on Recent Graduates Initiating AI Technology Startups in China's Urban Centers.

- Ahmad, B., Majeed, M. K., Soomro, A. A., & Jillani, S. A. (2026, January). Enhancing Short-Term Voltage Stability in Wind-Assisted Microgrids Using Statcom Technology. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.
- Narayan, S. B. (2026). *How an Entrepreneur Can Use Enterprise Architecture and Artificial Intelligence Governance for Regulated Industries*. Deep Science Publishing.
- Rana, S. (2025). Artificial intelligence and future of business management research: key questions and opportunities. *FIIB Business Review*, 14(5), 547-552.
- Ridzuan, N. N., Masri, M., Anshari, M., Fitriyani, N. L., & Syafrudin, M. (2024). AI in the financial sector: The line between innovation, regulation and ethical responsibility. *Information*, 15(8), 432.
- Panwar, M., Saxena, P., Dubey, M., Ray, B., & Kant, S. (2026). Entrepreneurial Risk Perception and AI-Driven Financial Decision-Making: A Behavioral Finance Perspective. In *Decision Making, Learning, and Collaboration in AI-Driven Entrepreneurship* (pp. 21-40). IGI Global Scientific Publishing.
- Rane, N. L., Chika, O. E., & Rane, J. (2026). Business intelligence systems integrating artificial intelligence, big data analytics, machine learning, internet of things, and blockchain. *International Journal of Applied Resilience and Sustainability*, 2(2), 367-395.
- Ahmad, B., Jillani, S. A., Rafique, M., & Soomro, A. A. (2026, January). Design and Monitoring of a Tree-Inspired Vertical Axis Wind Turbine for Efficient Urban Wind Utilization. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.
- Yoon, S. (2020). A study on the transformation of accounting based on new technologies: Evidence from Korea. *Sustainability*, 12(20), 8669.
- Zianko, V. V., & Nechyporenko, T. (2025). Impact of artificial intelligence on business models and competitiveness of enterprises. *ЕКОНОМІКА. ФІНАНСИ. ПРАВО. № 1: 10-15*.
- Dhoot, P., Patil, A., Raidas, A., Bairagi, P., & Roy, S. (2026). Navigating the Future: Exploring Intelligent Systems for Sustainable and Robust Finance. In *Humanized Intelligence in Sustainable Finance and Economic Behaviours: Concepts, Frameworks, and Applications* (pp. 91-111). Singapore: Springer Nature Singapore.
- Padmavathy, R. (2021). Cloud-Based Fraud Detection and Prevention Solutions: Using AI and Machine Learning to Safeguard Online Banking Transactions in the Digital Age. *Cloud-Based Fraud Detect. Prev. Solut. Using AI Mach. Learn. to Safeguard Online Bank. Trans. Digit. Age*, 9(03).
- Chen, M. (2022). The influence of big data analysis of intelligent manufacturing under machine learning on start-ups enterprise. *Enterprise Information Systems*, 16(2), 347-362.
- Kunwar, M. (2019). Artificial intelligence in finance: Understanding how automation and machine learning is transforming the financial industry.
- Shibahathulla, T. K., Ali, M. A., Gayyor, O., & Ghaffari, A. Z. (2024). The role of automated controls and streamlined compliance in managing risks in digital finance. *International Journal of financial engineering*, 11(03), 2443004.
- Prasad, V. L. S., Srivastava, V., & Singh, H. (2026). Cyber-Crimes in AI-Enabled Green Platforms: Legal Hurdles in ESG Investing, Prosecution, and Financial Loss Mitigation. In *AI Governance, Law, Policy and Ethics in Green Finance and ESG* (pp. 265-282). IGI Global Scientific Publishing.

- Burugulla, J. K. R. (2024). The future of digital financial security: Integrating AI, cloud, and big data for fraud prevention and real time transaction monitoring in payment systems. *MSW Management Journal*, 34(2), 711-730.
- Trawnih, A., Yaseen, H., Alsoud, M. A., Al-Salim, M. A., & Hattar, C. (2025). Empowering startup supply chain: exploring the integration of SCF, AI, blockchain, and trust. *Logistics*, 9(2), 69.
- Adedipe, O. A. (2026). Impact of smart contracts and cryptographic security on fraud prevention in Nigerian deposit money bank. *FUDMA Journal of Accounting and Finance Research [FUJAFR]*, 4(1), 250-263.
- Abdulsalam, T. A., & Tajudeen, R. B. (2024). Artificial Intelligence (AI) in the banking industry: A review of service areas and customer service journeys in emerging economies. *Business & Management Compass*, 68(3), 19-43.
- Goldstein, A., & Narayanaswamy, S. (2025). Managing risk in FinTech: Balancing innovation, regulation and consumer protection. *Journal of Risk Management in Financial Institutions*, 18(3), 219-231.
- Iqbal, M. S., Ahmad, W., & Amjad, A. (2025). Designing AI-Driven Sandboxes for Digital Asset Innovation: A Policy Framework for the Regulators.
- Kaddour, S. (2023). Sustainability performance evaluation for start-ups: Defining the characteristics for a tailored ESG assessment tool.
- Maguraushe, K., & Chimbo, B. (2026). Responsible AI and Smart Technologies in Finance: Ethical Challenges in the South African Context. In *Driving Excellence Through AI-Powered Performance Management* (pp. 229-266). IGI Global Scientific Publishing.
- Ally, A. M. (2026). Artificial intelligence (AI) and financial technology (FinTech) in Tanzania; legal and regulatory issues. *International Journal of Law and Management*, 68(4), 723-741.
- Braunsberger, C., & Aschauer, E. (2025). Corporate failure prediction: A literature review of Altman Z-score and machine learning models within a technology adoption framework. *Journal of Risk and Financial Management*, 18(8), 465.
- Jabbour Al Maalouf, N., Ficani, A., & Elia, J. (2026). Drivers of AI adoption in auditing: a quantitative exploration of perceived benefits and trust across auditor profiles. *Cogent Business & Management*, 13(1), 2709916.
- Merrahi, B., & Ait Lemqeddem, H. (2026). The Impact of AI on the Financial Sector and the Challenges of Intellectual Property Management. In *Navigating Responsible Business Practices Through Ethical AI* (pp. 275-306). IGI Global Scientific Publishing.