Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

INTERNAL FRAUD DETECTION IN FINANCIAL INSTITUTIONS USING MACHINE LEARNING TECHNIQUES

JOSUE CORREA[1], SOLANGE ROJAS[1]
[1]Faculty of Engineering and Architecture, Universidad César Vallejo, Piura, Perú
E-mail: [1] {jcorreaqu, sorojasc}@ucvvirtual.edu.pe

ABSTRACT

This research aimed to implement a machine learning model aligned with Sustainable Development Goal No. 12: Responsible Consumption and Production, which focuses on strengthening access to information and technological knowledge. The objective of this study was to develop a machine learning model to enhance the detection of internal fraud within a financial institution located in Sullana, Piura, Perú. This was an applied research study, employing a quantitative approach, a pre-experimental design, and a descriptive scope. The population consisted of 6000 operational records from the financial institution and 25 employees from the business department. Observation guides and questionnaires were used as data collection instruments during the execution of the project. The results showed a 37% reduction in unusual behaviors, a 50-minute decrease in the average detection time, a 27.5% increase in accuracy, and a 96% rate of high employee perception. These findings indicate that the machine learning model proved to be an effective tool for fostering a culture of transparency, thereby reducing operational risks and reinforcing trust in the local financial system.

Keywords: *Risk Management, Financial Institutions, Artificial Intelligence, Data Processing, Automation*

1. INTRODUCTION

At present, internal fraud in financial institutions represents one of the main challenges to operational sustainability and trust within the global financial system. It is estimated that magnitude of the problem and its direct impact on the economic stability of organizations [1]. Microfinance institutions are particularly vulnerable due to their organizational structure, which is based on trust and close relationships with micro-entrepreneurs. This exposes them to significant risks of data manipulation, omission of records, or misappropriation of payments [2].

The digital transformation has brought a substantial increase in the volume and complexity of financial data. Likewise, security gaps and opportunities for unusual behaviors within institutions have also grown. Vargas [3] argues that adopting digital tools is essential to reduce costs and expand financial access; however, without improvement in internal controls, these advances may create new scenarios of vulnerability.

Small and medium-sized financial entities, lacking intelligent monitoring systems, face limitations in detecting and preventing internal anomalies that impact their operations in a timely manner.

The case study developed in Sullana, Piura, Perú, clearly reflects this issue. Microfinance institutions play a key role in financial inclusion and local development; however, they lack advanced technological tools to detect anomalous staff behavior, such as data alteration, access outside working hours, privilege abuse, or unauthorized credit approvals. Although these events may appear isolated, they distort accounting records, undermine institutional credibility, and erode client trust in the services provided.

Therefore, this research proposes integrating machine learning model into internal control systems as an effective and scalable response to these risks. The main difference between traditional methods and machine learning models lies in their ability to analyze large volumes of data—such as transactions—identify patterns, and generate real-time alerts, thereby increasing the precision, speed, and predictive capacity of financial institutions [4].

Despite the adoption of digital technologies, most small and medium-sized microfinance institutions still rely on traditional internal control mechanisms, such as rule-based monitoring, periodic audits, and manual transaction reviews. These approaches are largely reactive and depend on static thresholds or predefined rules, which limits their ability to adapt to evolving behavioral patterns. Consequently,

detections are often delayed, false-positive rates increase, and operational workloads grow significantly. In addition, manual supervision does not scale effectively with the growing volume and complexity of financial data, making it difficult to consistently identify subtle or previously unseen anomalous behaviors.

To address these limitations, the proposed framework adopts a data-driven predictive approach that learns behavioral patterns from operational logs and credit-process variables, enabling the detection of anomalous activities beyond static rules. It preserves deployment realism through temporally consistent training and validation to prevent information leakage and supports decision-making via probabilistic risk scores and configurable alert thresholds, balancing detection sensitivity with review workload.

Moreover, the application of these technologies not only aims to optimize processes but also to strengthen institutional transparency and ethics, contributing to the achievement of Sustainable Development Goal No. 12, which promotes responsible and sustainable practices in the consumption and production of financial services [5].

From a social and organizational perspective, the application of machine learning models manifests within the technological dimension, representing progress toward and ethical and proactive institutional culture focused on fraud prevention, transparency and trust between employees and clients. The practical relevance of this study lies in its direct application within a Peruvian financial institution. Developing a machine learning model trained with real operational data allows for technological adaptation to local market characteristics and enables the evaluation of its impact in terms of operational efficiency, reduction in detection time, and alert accuracy. This approach seeks to optimize control processes, improve credit risk management, and ensure institutional continuity in the face of potential internal irregularities.

Several studies have demonstrated the efficiency of machine learning in detecting financial fraud, both in traditional banking environments and in microfinance institutions. Recent advances in artificial intelligence facilitate the processing of large volumes of transactions and the identification of anomalous behaviors that conventional systems fail to detect [6].

2. RELATED WORKS

A literature review began with Supriyadi [7], who developed a predictive model for an Indonesian microfinance institution using the XGBoost algorithm to detect fraud committed by employees. The study adopted two perspectives—technological and psychosocial—and successfully identified behavioral patterns linked to workplace pressure high stress levels, and financial habits. Similarly, Boubker [8] designed a predictive model to detect fraud based on financial transactions in a Moroccan bank. Using a Gradient Boosting model combined with SMOTE, the study achieved highly promising results in identifying fraudulent transactions, even though only 1.8% of all transactions corresponded to actual fraud cases.

The study by Mediana and Sandari [9] analyzed the application of artificial intelligence in internal banking audits, focusing on its contribution to fraud detection and prevention. The results showed that machine learning algorithms and neural networks can process large volumes of data in real time. They reported that 85% of participants perceived an improvement in operational efficiency, enhancing transparency and responsiveness to financial anomalies.

Likewise, Marulanda et al. [10] indicated that one of the main causes of institutional failure in countries such as Nicaragua, Bolivia, and Honduras was the presence of internal fraud arising from weak controls, poor organizational culture, and the absence of preventive technological systems. The study emphasized the need to strengthen staff training, foster a culture of internal control and adopt technological tools that allow for the timely detection of behavioral changes among employees.

In a similar vein, Garcia [11] developed in Chile a predictive model using unsupervised learning techniques such as Local Outlier Factor and Components Analysis, successfully identifying unusual behaviors and employees requiring closer supervision—even with limited data. This demonstrated that effective tools for fraud prevention and institutional protection can be built using small datasets.

In the Peruvian context, Llanos [12] proposed a scoring based on patterns of unauthorized operations, capable of generating real-time alerts. The model detected potential losses exceeding S/745,943.95 and reduced response time to suspicious events by 30%.

Addressing such a complex issue requires more than the use of data or technology; it is essential to understand the underlying causes of these behaviors. The theoretical frameworks presented in this research help analyze both the human factors that lead to fraudulent actions and the technical methods that facilitate detection and prevention.

From a theoretical standpoint, this study is based on Jensen's Fraud Triangle Model [13], which posits that fraud arises from three human factors: pressure, opportunity, and rationalization. This framework allows for an understanding of not only the fraudulent act itself but also its emotional and organizational context. Additionally, Agency Theory, as explained by Guggengerber et al. [14], describes the relationship in which a principal delegates decision-making authority to an agent, directly influencing perceptions of efficacy and reliability within organizations.

The study also incorporates Vapnik's Statistical Learning Theory [15], which provides the mathematical foundation enabling algorithms to learn from historical data. This theory is critical for building predictive models capable of recognizing behavioral patterns and accurately identifying early signs of potential internal fraud.

Based on both theorical and empirical analyses of internal fraud in financial institutions, the present research aims to develop a machine learning model to enhance the detection of internal fraud in the credit administration of a financial institution in Sullana, Piura, Perú. To achieve this goal, the following specific objectives were established: (a) identify behavioral patterns associated with internal fraud in credit administration; (b) analyze the average time required to detect internal fraud using traditional methods versus a machine learning model; (c) determine the accuracy of fraud detection before and after model implementation; and (d) assess the perception of business area employees before and after the deployment of the machine learning model.

In line with the stated objectives, the general hypothesis proposed that the machine learning model would improve internal fraud detection in the credit administration of a financial institution. The following specific hypotheses were formulated: (a) behavioral patterns associated with internal fraud in credit administration include access to the premises outside working hours, data alteration, and privilege abuse; (b) the average detection time for internal fraud will differ by 30% after implementing the machine learning model; (c) the accuracy of internal fraud detection will increase by 25% following implementation; and (d) 70% of employees in the business area will report a high level of perception following the deployment of the machine learning model.

3. MATERIALS AND METHODS

The purpose of this study is to develop and evaluate a machine learning model for the detection of internal fraud within the credit administration processes of a financial institution in Sullana, Piura, Perú. The methodological process involves several structured stages including data acquisition, preprocessing, model training, evaluation, and validation with operational and perceptual indicators.

3.1 Type and Design of Research

This is an applied study with a quantitative approach and a pre-experimental design, consisting of a pretest and posttest with a single group. The research aimed to assess the impact of the proposed machine learning model on internal fraud detection efficiency and employee perception.

Ethical considerations were observed to ensure data confidentiality and compliance with institutional policies. All operational data were anonymized before analysis, and access was restricted to authorized personnel.

3.2 Data Collection

The dataset consisted of 6,000 operational records from the institution's internal credit management system, covering variables such as access logs, transaction approvals, and system modifications. Data were collected through direct observation and internal monitoring systems between August and September 2025.

The variables were selected in coordination with the institution's risk management unit to reflect operational behaviors potentially associated with internal fraud, such as unauthorized access, unusual approval timing, or data modification frequency.

3.3 Ethical Considerations, Consent and Data Anonymization

This research was conducted under ethical principles aimed at ensuring the protection of the individuals involved, the confidentiality of the information, and the responsible use of data. To strengthen methodological transparency and enhance confidence in the results, this section describes the data collection process, consent procedures, and anonymization strategies applied throughout the study.

First, the data used in this research were obtained from operational records generated during the routine use of the financial system by the institution's employees. The organization formally authorized the use of this information for applied research and internal control improvement purposes, under policies of confidentiality, legality, and restricted access limited to the research team.

Prior to analysis, the data were anonymized by removing direct personal identifies and replacing internal employee and client codes with non-reversible surrogate identifiers. Additionally, the principle of data minimization was applied, retaining only the variables strictly necessary for modeling and evaluation. This approach significantly reduced the risk of re-identification.

Regarding the employee perception instrument (questionnaire), participation was voluntary and based on informed consent. Participants were informed about the objectives of the study, that their participation would not affect their employment conditions, and that their responses would be treated confidentially. No personally identifiable information was collected, and results are reported exclusively in aggregated form.

3.4 Data Preprocessing

It was implemented through a preprocessing pipeline designed to preserve temporal consistency and prevent data leakage. Data were extracted from MySQL while excluding payments recorded after each installment's due date and chronologically ordering all historical installments. Type conversion, zero-imputation, and the generation of 40 numerical features related to loan progression, payment ratios, client repayment history, and temporal metrics were applied. The target variable was defined using business rules based on delinquency severity, payment level, and historical behavior. Outliers detected using the IQR method were logged without removal, and the dataset was split sequentially (70% training, 10% validation, 20% testing) to ensure future-oriented evaluation.

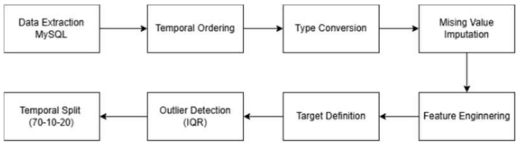

Figure 1. Data preparation workflow

3.5 Model Development

The model was developed using Random Forest due to its strong performance on tabular data and in class-imbalanced scenarios. A *Random Forest Classifier* was employed with an optimized base configuration (300 trees and a maximum depth of 15). Hyperparameter optimization was conducted through *RandomizedSearchCV* with 50 iterations and temporal cross-validation, exploring variations in the number of trees, depth, minimum samples for splits and leaves, and feature-selection strategies, using ROC-AUC as the primary optimization metric.

Model validation was carried out at three levels: temporal cross-validation, a separate validation set, and a final test set. The main evaluation metrics were PR-AUC, MCC, F2-score, and Balanced Accuracy, which are well suited for imbalanced classification problems. Interpretability was addressed through Gini-based feature importance, highlighting variables related to prior delinquency, historical repayment performance, and payment ratios.

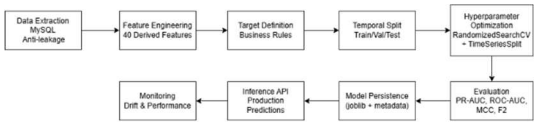

Figure 2. Machine learning model development pipeline

3.6 Algorithm and Implementation Details

The proposed risk detection engine was implemented as a Random Forest classifier trained within a reproducible and leakage-aware pipeline. A total of 6,000 operational records were initially extracted from the institution's internal systems. After preprocessing, data cleaning, and feature-consistency validation, 5,000 events with complete and reliable information were retained for modeling.

The positive class prevalence was 30.26% (1,513/5,000). It is important to note that this label corresponds to the institution's operational definition of suspicious or fraud-related behavior according to internal rules, rather than confirmed fraud cases.

To preserve chronological consistency and avoid look-ahead bias, the dataset was split sequentially using a temporal strategy (70%/10%/20%): 3,500 samples for training, 500 for validation, and 1,000 for testing.

Table 1: Random forest configuration

Parameter	Value	Purpose
n_estimators	300	Number of trees
max_depth	15	Maximum tree depth
min_samples_split	20	Minium samples

		required to split a node
min_samples_leaf	10	Minimum samples per leaf
max_features	sqrt	Feature subset per split
class_weight	balanced	Handles class imbalance
bootstrap	true	Enables bagging
max_samples	0.8	Subsampling ratio for each tree
oob_score	true	Out-of-bag internal validation
random_state	42	Reproducibility
n_jobs	-1	Parallel training

Hyperparameters were optimized using RandomizedSearchCV with 50 iterations and temporal cross-validation (TimeSeriesSplit with 5 folds and no shuffling). ROC-AUC was used as the primary optimization metric. The explored search space included variations in the number of trees, tree depth, minimum samples per split and leaf, feature selection strategy, bootstrap sampling, and subsampling ratio.

To prevent temporal leakage, all predictors were computed strictly using information available up to the prediction cut-off time. Class imbalance was mitigated through balanced class weighting. Additionally, model artifacts, preprocessing configurations, random seeds, and the operational decision threshold (τ) were versioned to ensure traceability, reproducibility, and drift-aware updates.

3.7 Evaluation Metrics

The evaluation focused on metrics suitable for imbalanced classification. The confusion matrix allowed the identification of critical errors, particularly false negatives. The main metrics included precision, recall, F1-score, F2-score, and balanced accuracy. The robust metrics PR-AUC, ROC-AUC and MCC provided a discriminative and balanced assessment of model performance. The ROC and Precision-Recall curves were used to determine optimal thresholds based on Youden's index and F1 maximization, respectively [16].

3.8 Technologies and Components

The system relies on a set of technologies that keep the architecture flexible and allow it to integrate properly with the predictive model. The frontend, built with React, provides a dynamic interface for end users. The backend, developed in Laravel/PHP, offers REST services that handle business logic, JWT auth, and access to the MySQL database, which stores the operational records used by both the transactional system and the machine learning pipeline. The predictive component operates as an independent Python microservice that loads the trained Random Forest model and processes backend requests to generate risk predictions.

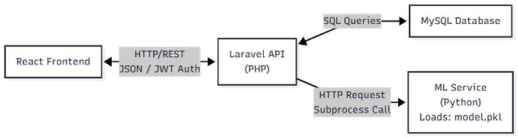

Figure 3. System architecture

3.9 Calculation of Operational Indicators

For the first three organizational indicators—anomaly frequency, detection time, and staff perception—structured observation guides were used during the pretest and posttest stages. Each indicator was computed using standardized formulas, as described below.

(a) Frequency of Unusual Behaviors

The frequency of internal anomalous behaviors recorded through Observation Guide No. 01 was calculated as a percentage based on the number of events detected during the observation period. The indicator was computed using Equation (1):

$$F(\%) = \frac{C}{N} \; x \; 100$$

F (%): percentage frequency of unusual behaviors

C = number of unusual behaviors detected

N = total number of observations conducted

This formula allowed quantifying the proportion of irregular behaviors present during the pretest stage.

(b) Average Fraud Detection Time

The average time required to detect internal fraud events, as recorded through Observation Guide No. 02, was computed based on the duration elapsed between the start and end of each detection process. The indicator was calculated using Equation (2):

$$TP = \frac{\sum_{i=1}^{n}(HF_i - HI_i)}{n}$$

TP = average fraud detection time (in minutes)

HF_i = end time of fraud detection for case i

HI_i = start time of fraud detection for case i

n = total number of detection cases observed

This formula enabled determining the mean detection time during the pretest stage and served as the basis for comparing detection efficiency before and after the implementation of the machine learning model.

(c) Precision in Fraud Detection

The precision of the fraud detection process, measured through Observation Guide No. 03, was calculated based on the proportion of correctly identified fraud cases relative to all positive detections. The indicator was computed using Equation (3):

$$Precision\ (\%) = \frac{TP}{TP + FP} x\ 100\%$$

TP = true positives, corresponding to real fraud events correctly identified

FP = false positives, corresponding to non-fraudulent cases incorrectly flagged as fraud

This formula allowed assessing the accuracy of the inferences made during the pretest stage and served as the basis for comparing the model's performance before and after the implementation of the machine learning system.

3.9.1 Staff Perception of System Effectiveness

Staff perception regarding the effectiveness of the fraud detection system was assessed using a structured Likert-scale questionnaire administered during the pre-test and post-test stages. The instrument consisted of ten items (P1–P10) that measured employees' evaluations of system reliability, usefulness, operational clarity, and perceived contribution to internal control. Each item was rated on a five-point Likert scale, where 1 = strongly disagree and 5 = strongly agree.

For each participant, the scores across the ten items were averaged to obtain an overall perception rating. The resulting means were then classified into qualitative categories according to their numerical value: Low (1.0–2.50), Moderate (2.51–3.50), and High (3.51–5.0). These categorizations allowed determining the general level of acceptance and perceived effectiveness of the system after the implementation of the machine learning model. The results of the pre-test and post-test evaluations are presented in Tables 2 and 3.

Table 2: Pre-test perception levels

Perception Category	Range	Number of Participants	Percentage
Low perception	1.00 – 2.50	10	40%
Moderate perception	2.51 – 3.50	14	56%
High perception	3.51 – 5.00	1	4%
	TOTAL	**25**	**100%**

Table 3: Post-test perception levels

Perception Category	Range	Number of Participants	Percentage
Low perception	1.00 – 2.50	0	0%
Moderate perception	2.51 – 3.50	1	4%
High perception	3.51 – 5.00	24	96%
	TOTAL	**25**	**100%**

3.9.2 Validation and Perception Analysis

To complement the quantitative analysis, a perception survey was applied to 25 credit and business staff members. The instrument demonstrated good validity (Aiken's V = 0.85) and reliability (Cronbach's α = 0.869). Descriptive and inferential statistics were used to analyze the impact of the model.

This dual validation approach allowed for both technical verification of the model's performance and assessment of its perceived organizational impact.

3.10 Statistical Analysis

To determine whether the data followed a normal distribution, the Shapiro–Wilk test was applied to both pretest and posttest values for each indicator. Depending on the normality results, either parametric (Student's t-test) or non-parametric (Wilcoxon signed-rank test) methods were used to assess significant differences between pre- and post-intervention stages. The significance threshold was set at α = 0.05.

3.10.1 Indicator 1: Frequency of Unusual Behaviors

The Shapiro–Wilk test showed p-values greater than 0.05 (Pre = 0.850; Post = 0.272), indicating normal distribution. Therefore, the Student's t-test was applied ($p = 0.016 < 0.05$). The null hypothesis was rejected, confirming a significant reduction in unusual behaviors after model implementation.

Table 4: Normality and significance tests for frequency of unusual behaviors

Shapiro-Wilk	
Pre-test	0.850
Post-test	0.272
Student's T	
0.016	

3.10.2 Indicator 2: Average Detection Time of Internal Frauds

The Shapiro–Wilk test indicated non-normality (Pre = 0.007; Post = 0.011). Consequently, the Wilcoxon signed-rank test was used ($p = 0.012 < 0.05$), rejecting the null hypothesis and confirming a significant improvement in detection speed after implementation.

Table 5: Normality and significance tests for average fraud detection time

Shapiro-Wilk	
Pre-test	0.007
Post-test	0.011
Wilcoxon	
0.012	

3.10.3 Indicator 3: Precision of Inferences in Fraud Detection

Normality was not met (Pre = 0.000; Post = 0.000), so the Wilcoxon test was applied ($p = 0.046 < 0.05$). The null hypothesis was rejected, indicating a significant increase in detection accuracy with the model.

Table 6: Normality and significance tests for fraud detection precision

Shapiro-Wilk	
Pre-test	0.000
Post-test	0.000
Wilcoxon	
0.046	

3.10.4 Indicator 4: Perception of System Effectiveness

The Shapiro–Wilk results (Pre = 0.035; Post = 0.000) confirmed non-normality. The Wilcoxon signed-rank test ($p = 0.000 < 0.05$) showed a significant improvement in perception, with participants rating the model as more effective post-implementation.

Table 7: Normality and significance tests for perception of system effectiveness

Shapiro-Wilk	
Pre-test	0.035
Post-test	0.000
Wilcoxon	
0.000	

These results statistically validate the impact of the machine learning model, confirming significant differences between pretest and posttest stages across all four indicators. All statistical analyses were performed using SPSS version 26 and Microsoft Excel, applying a 95% confidence level.

4. RESULTS

The proposed machine learning model for internal fraud detection was implemented and tested using 6,000 operational records from the credit administration system of a financial institution in Sullana, Piura, Perú. The analysis compared pretest and posttest results to assess changes in behavioral patterns, detection efficiency, model accuracy, and employee perception following the deployment of the model.

4.1 Detection of Internal Anomalous Behavior

During the pretest stage (September 8–12, 2025), eight cases of unusual internal behaviors were recorded, including unauthorized data alterations, privilege abuse, and after-hours access. Specifically, 62.5% of cases corresponded to privilege abuse, 37.5% to data alteration, 25% to after-hours access, and 12.5% to unauthorized approvals.

After implementing the machine learning model (September 24–October 2, 2025), the frequency of such behaviors decreased by 37%, indicating the system's effectiveness in identifying and limiting risky actions. The model successfully detected anomalies based on behavioral indicators and access logs, automatically flagging irregular activities in real-time.

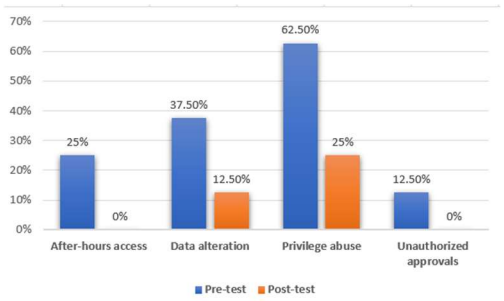

Figure 4: Comparison Between Pre-test and Post-test Results

4.2 Fraud Detection Efficiency

The model significantly reduced the average fraud detection time from 78.44 minutes in the pretest to 28.77 minutes in the posttest, reflecting a 63% improvement in detection speed. This result confirms the model's operational effectiveness, enabling faster and more accurate identification of unusual internal behaviors within the business area.

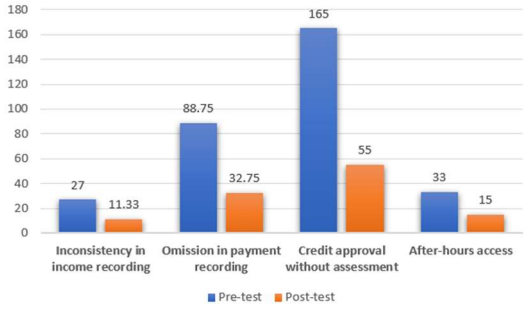

Figure 5: Comparison Between Pre-test and Post-test Results

4.3 Detection Accuracy and Operational Outcomes

The accuracy of fraud detection improved notably after implementing the machine learning model. During the pretest, the traditional method achieved 60% true positives and 40% false positives, reflecting a high rate of incorrect alerts due to static rules and limited contextual analysis.

In the posttest stage, the model reached 87.50% true positives and 12.50% false positives, demonstrating a significant increase in detection reliability and a marked reduction in false alerts, confirming the model's effectiveness in identifying real fraud cases.

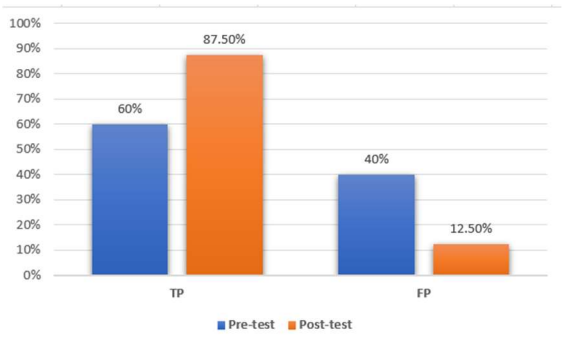

Figure 6: Comparison Between Pre-test and Post-test Results

4.4 Organizational Impact and Employee Perception

To evaluate human and organizational impact, a perception survey was applied to 25 employees in the business and credit areas. Results showed that 96% of participants perceived the implementation of the model positively, highlighting greater transparency and faster anomaly detection in daily operations.

According to respondents, the system enhanced internal control mechanisms, reduced the opportunity for manipulation, and improved trust in the credit administration process. These findings indicate that the model's integration did not only strengthen technical detection but also reinforced ethical conduct and accountability among staff.

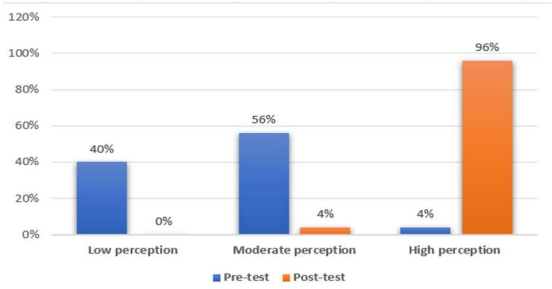

Figure 7: Comparison Between Pre-test and Post-test Results

5. DISCUSSION

5.1 Effectiveness of the Machine Learning Model

The implementation of the machine learning model was associated with improved internal fraud monitoring within the financial institution. The system efficiently identified unusual behavioral patterns such as unauthorized access, data manipulation, privilege abuse, and irregular credit approvals. Compared with the pretest phase, the frequency of anomalous behaviors decreased by

37%, and the average detection time was reduced from to approximately 50 minutes. These findings confirm that automated pattern recognition effectively supports early intervention and strengthens internal control mechanisms.

Furthermore, the post-implementation precision reached 87.5%, compared to 60% in the pretest of limited data volume and imbalance between fraudulent and normal records, which are common in small financial entities. Despite this limitation, the model's capacity to flag suspicious activities in near real time provided a valuable analytical advantage for the institution.

5.2 Fraud Behavior Trends

Before the model's implementation, the most frequent anomalies were privilege abuse and data alteration, accounting for 62.5% and 37.5% of detected cases, respectively. After the deployment, a notable decline in both behaviors was observed, suggesting that predictive monitoring contributed to deterring internal misconduct. These patterns align with prior findings by Boubker [8], who also highlighted privilege misuse as a key indicator of internal fraud in financial environments.

In addition, new detection cases such as omission of payment records and credit approval without proper evaluation were identified during the posttest stage. This indicates that the model not only detected previously known risks but also uncovered new forms of operational inconsistency that were not systematically monitored before.

5.3 Feature and Detection Insights

Feature analysis revealed that the most influential variables in fraud detection were login time deviations, frequency of data modification, and authorization anomalies. The integration of these behavioral indicators allowed the model to identify subtle deviations that manual reviews often overlook. Collaborators' perception also improved notably — 96% rated the system as useful or very useful — reflecting a positive attitude toward data-driven decision-making and technological transparency in internal auditing.

This improvement in perception suggests a gradual cultural shift toward preventive control and accountability, consistent with the results reported by Mediana and Sandari [9], who found that artificial intelligence tools foster confidence and operational integrity in banking environments.

5.4 Limitations and Future Work

The primary limitation of this study lies in the reduced dataset, composed of approximately 6,000 operational records. While the results are encouraging, expanding the data scope to include multiple branches or additional time periods could increase model robustness and accuracy. Another limitation concerns the relatively low precision rate, which could be improved through ensemble learning or resampling techniques such as SMOTE, as suggested by Boubker [8].

Future research should explore integrating explainable AI methods to improve interpretability of model predictions and strengthen trust among internal auditors. Extending the model's application to other financial institutions, such as credit unions or municipal savings banks, would also enable comparative validation under different organizational contexts.

5.5 Significance of the Study

The findings of this study highlight the practical importance of applying machine learning to internal fraud detection in small financial institutions. The implementation of the model not only enhanced operational control but also reduced detection time and improved decision accuracy. These outcomes demonstrate the potential of intelligent systems to strengthen transparency and accountability in credit management.

From a theoretical perspective, the results contribute to the growing body of research supporting the use of anomaly detection and behavioral analytics in financial environments. The evidence obtained validates that data-driven models can complement traditional audit mechanisms, aligning institutional management with modern digital governance standards.

5.6 Discussion and Implications of the Findings

The results presented in Sections 5.1 to 5.5 demonstrate that the implementation of a machine learning–based detection model generated a significant and positive impact on internal fraud control within the financial institution. The reduction in anomalous behaviors, improvement in detection

time, and increase in precision confirm that automated behavioral analysis constitutes an effective complement to traditional internal audit mechanisms.

These findings are consistent with the conclusions of Hernández et al. [17], who, in a comprehensive literature review, report that machine learning techniques outperform conventional rule-based systems in financial fraud detection due to their capacity to learn complex and evolving patterns. Similarly, Hilal et al. [1] emphasize that anomaly detection approaches are particularly effective for identifying rare and previously unseen fraudulent behaviors, which supports the observed decrease in anomalous internal activities after the model's implementation.

The improvement in predictive precision aligns with the results reported by Bello et al. [6], who found that advanced analytics and machine learning models significantly enhance classification accuracy in fraud detection tasks. Likewise, Dávila et al. [18] demonstrated that supervised learning algorithms achieve higher detection performance when trained on operational transaction data, reinforcing the effectiveness of the predictive strategy adopted in this study.

Regarding the prevention of internal fraud in microfinance and lending environments, the findings coincide with Supriyadi et al. [7], who showed that machine learning models such as XGBoost contribute substantially to strengthening internal controls in microlending business processes.

Beyond technical performance, the organizational impact identified in this study is supported by Mediana and Sandari [9], who concluded that the implementation of artificial intelligence in internal audit functions improves operational efficiency and strengthens trust in control mechanisms.

From a theoretical perspective, the results are compatible with Vapnik's [15] statistical learning theory, which states that well-trained supervised models can generalize effectively to unseen data, explaining the model's improved classification capability. Furthermore, Dietterich [19] emphasizes that ensemble and advanced learning methods often yield superior performance compared to single classifiers, which is consistent with the improvements reported in this research.

Unlike many prior studies that focus primarily on algorithmic accuracy, this research integrates technical indicators (precision, detection time, anomaly reduction) with organizational indicators (staff perception), thereby extending the existing body of knowledge by demonstrating the multidimensional impact of machine learning adoption on internal fraud control.

Overall, these findings suggest that future research should focus on expanding datasets, incorporating explainable artificial intelligence techniques as proposed by Hasan et al. [20], and evaluating long-term organizational effects in order to consolidate the applicability and scalability of predictive fraud detection systems across diverse financial contexts.

6. CONCLUSIONS

This study demonstrates that machine learning–based internal fraud detection is not only technically viable but also organizationally transformative for small financial institutions.

Beyond the technical findings, this study has broader implications for society and organizational governance. By demonstrating that machine learning models can effectively detect internal fraud in small financial institutions, this research contributes to strengthening transparency, accountability, and trust in the financial system. Early detection of anomalous internal behaviors not only protects institutional assets but also safeguards clients, promotes ethical practices, and reduces reputational and operational risks in vulnerable financial environments.

Furthermore, the proposed model presents potential applications beyond the scope of the current study. Similar approaches could be adapted for use in other financial organizations such as credit unions, municipal savings banks, and microfinance institutions facing comparable limitations in data volume and monitoring capacity. Additionally, the integration of behavioral analytics into internal audit processes may support compliance frameworks, risk management systems, and digital governance strategies in both public and private sectors.

Finally, this research encourages continued exploration and practical implementation of intelligent fraud detection systems. Future efforts should focus on expanding datasets, incorporating explainable artificial intelligence techniques, and evaluating long-term organizational impacts.

Financial institutions are encouraged to adopt data-driven monitoring tools as complementary mechanisms to traditional audits, fostering a proactive culture of prevention and continuous improvement in internal control systems.

REFERENCES:

[1] W. Hilal, S. A. Gadsden, y J. Yawney, «Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances», *Expert Syst. Appl.*, vol. 193, p. 116429, may 2022, doi: 10.1016/j.eswa.2021.116429.

[2] E. Waweru, S. M. Karume, y A. Kibet, «A Review of Human Vulnerabilities in Cyber Security: Challenges and Solutions for Microfinance Institutions», *J. Inf. Secur.*, vol. 16, n.º 1, Art. n.º 1, nov. 2024, doi: 10.4236/jis.2025.161006.

[3] A. H. Vargas Garcia, «La banca digital: Innovación tecnológica en la inclusión financiera en el Perú», *Ind. Data*, vol. 24, n.º 2, pp. 99-120, jul. 2021, doi: 10.15381/idata.v24i2.20351.

[4] F. L. Garcia Cabrera y J. L. Romero Baltazar, «Billetera digital y su impacto en la inclusión financiera de las bodegas de Lima Norte 2023», *Univ. Peru. Cienc. Apl. UPC*, dic. 2023, Accedido: 24 de mayo de 2025. [En línea]. Disponible en: https://repositorioacademico.upc.edu.pe/handle/10757/673062

[5] M. García Goldar, «Propuestas para garantizar modalidades de consumo y producción sostenibles (ODS 12)», *Rev. Fom. Soc.*, n.º Nº 299, pp. 91-114, may 2021, doi: 10.32418/rfs.2021.299.4582.

[6] O. A. Bello, A. Folorunso, J. Onwuchekwa, O. E. Ejiofor, F. Z. Budale, y M. N. Egwuonwu, «Analysing the Impact of Advanced Analytics on Fraud Detection: A Machine Learning Perspective», p. 104, 2023, doi: 10.37745/ejcsit.2013/vol11n6103126.

[7] H. Supriyadi, D. S. Priyarsono, N. A. Achsani, y T. Andati, «Preventing Internal Fraud in Microlending Business Processes with Machine Learning Models: Confirmatory Factor Analysis (CFA) and Extreme Gradient Boosting (XGBoost)», *Asia Proc. Soc. Sci.*, vol. 8, n.º 1, pp. 34-37, may 2021, doi: 10.31580/apss.v8i1.1943.

[8] M. B. Boubker, «Fraud Detection in Financial Transactions Using Machine Learning with Oversampling Techniques: A Case Study of a Moroccan Bank», *J. Electr. Syst.*, vol. 20, n.º 10s, pp. 6443-6448, jul. 2024.

[9] A. M. Mediana y T. E. Sandari, «Implementation of Artificial Intelligence in Fraud Detection and Prevention in Internal Audit: (Case Study in the Banking Sector)», *Int. J. Educ. Soc. Stud. Manag. IJESSM*, vol. 4, n.º 3, pp. 1230-1237, dic. 2024, doi: 10.52121/ijessm.v4i3.532.

[10] B. Marulanda, L. Fajury, M. Paredes, y F. Gomez, «Lo bueno de lo malo en Microfinanzas: Lecciones aprendidas de experiencias fallidas en América Latina», *IDB Publ.*, oct. 2010, doi: 10.18235/0009864.

[11] D. I. García Jurado, «Desarrollo de un modelo predictivo para detectar casos de fraude interno en una institución bancaria», 2016, Accedido: 24 de mayo de 2025. [En línea]. Disponible en: https://repositorio.uchile.cl/handle/2250/143695

[12] L. M. Llanos Chávez, «Propuesta de implementación de un sistema de alertas tempranas para minimizar fraudes en cuentas pasivas en una entidad financiera», 2021, Accedido: 24 de mayo de 2025. [En línea]. Disponible en: https://repositorio.usil.edu.pe/entities/publication/a1b31d76-e560-40b8-9793-9613694ef696

[13] M. C. Jensen y W. H. Meckling, «Theory of the firm: Managerial behavior, agency costs and ownership structure», *J. Financ. Econ.*, vol. 3, n.º 4, pp. 305-360, oct. 1976, doi: 10.1016/0304-405X(76)90026-X.

[14] T. Guggenberger, L. Lämmermann, N. Urbach, A. Walter, y P. Hofmann, «Task Delegation From AI to Humans : A Principal-Agent Perspective», Hyderabad, India, 2023. Accedido: 17 de octubre de 2025. [En línea]. Disponible en: https://eref.uni-bayreuth.de/id/eprint/87383/

[15] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY: Springer, 2000. doi: 10.1007/978-1-4757-3264-1.

[16] H. Silva-Marchan, J. P. Garces, R. S. Ancajima, y G. B. Rea, «Modelo predictivo para la detección precoz de infartos en la población del norte de Perú», *Int. J. Comput. Innov. Intell. Syst. AI*, vol. 1, n.º 1, pp. 62-77, jul. 2025, doi: 10.64439/cisai.v1i1.9.

[17] L. Hernandez Aros, L. X. Bustamante Molano, F. Gutierrez-Portela, J. J. Moreno Hernandez, y M. S. Rodríguez Barrero, «Financial fraud detection through the application of machine learning techniques: a literature review», *Humanit. Soc. Sci. Commun.*, vol. 11, n.º 1, pp.

1-22, sep. 2024, doi: 10.1057/s41599-024-03606-0.

[18] R. C. Dávila-Morán *et al.*, «Aplicación de Modelos de Aprendizaje Automático en la Detección de Fraudes en Transacciones Financieras», *Data Metadata*, vol. 2, n.º 109, ene. 2023, doi: 10.56294/dm2023109.

[19] T. G. Dietterich, «Ensemble Methods in Machine Learning», en *Multiple Classifier Systems*, Berlin, Heidelberg: Springer, 2000, pp. 1-15. doi: 10.1007/3-540-45014-9_1.

[20] M. R. Hasan, M. S. Gazi, y N. Gurung, «Explainable AI in Credit Card Fraud Detection: Interpretable Models and Transparent Decision-making for Enhanced Trust and Compliance in the USA», *J. Comput. Sci. Technol. Stud.*, vol. 6, n.º 2, Art. n.º 2, abr. 2024, doi: 10.32996/jcsts.2024.6.2.1.