INTERNATIONAL JOURNAL OF MANAGEMENT AND ORGANIZATIONAL RESEARCH

Integrating Support Vector Machine into Accounting Information Systems for Enhanced Fraud Detection and Predictive Risk Assessment

Bismark Afreh
J. Mack Robinson College of Business, Georgia State University, Atlanta, Georgia, USA

* Corresponding Author: **Bismark Afreh**

Article Info

ISSN (online): 2583-6641
Impact Factor (RSIF): 8.56
Volume: 05
Issue: 04
Received: 27-05-2026
Accepted: 25-06-2026
Published: 23-07-2026
Page No: 136-143

Abstract
This thesis examines the integration of Support Vector Machine (SVM) into Accounting Information Systems (AIS) for enhanced fraud detection and predictive risk assessment. Fraud remains a persistent organizational threat because fraudulent transactions are often concealed, rare, and difficult to distinguish from legitimate transactions using traditional rule based controls. The study adopts a quantitative experimental design using the IEEE CIS Fraud Detection Dataset, a publicly available dataset containing transaction and identity files joined by TransactionID and a binary fraud target variable, isFraud. The proposed methodology includes data preprocessing, feature engineering, feature selection, class imbalance treatment, SVM model development, and comparative evaluation against Logistic Regression and Random Forest. The system design proposes an AIS architecture with a fraud detection module, risk assessment module, SVM integration layer, dashboard interface, audit review function, and deployment architecture. Because the raw dataset files were not provided in the working environment, the compiled results section presents a reproducible evaluation framework, model comparison table templates, and academic interpretation guidance rather than fabricated numerical results. The study concludes that machine learning can strengthen AIS based fraud monitoring by supporting transaction level classification, fraud alert generation, and risk based audit prioritization. The thesis recommends empirical implementation using the approved dataset, validation with fraud sensitive metrics, and deployment only with model governance, audit trails, and periodic retraining.

DOI: https://doi.org/10.54660/IJMOR.2026.5.4.136-143

Keywords: Accounting information systems, support vector machine, fraud detection, predictive risk assessment, machine learning, financial analytics

1. Introduction

1.1. Background of the Study

Accounting Information Systems are central to the collection, processing, storage, and reporting of financial information in modern organizations. They support transaction recording, financial reporting, internal control, audit trails, managerial decision making, and regulatory compliance. As organizations increasingly depend on digital platforms for financial transactions, AIS platforms are expected not only to record transactions accurately but also to detect irregularities, identify suspicious patterns, and support risk based decision making [1].

Fraud remains a major threat to organizations because it affects financial reliability, weakens internal control systems, damages stakeholder confidence, and may lead to substantial losses. The Association of Certified Fraud Examiners [2] reports that occupational fraud cases occur across many countries, industries, and organizational types, demonstrating the continuing relevance of fraud risk management. Traditional fraud detection methods in accounting environments often depend on manual review, audit sampling, exception reports, and rule based controls. Although these methods are useful, they may be limited when fraud patterns are complex, concealed, nonlinear, or constantly changing.

Machine learning provides an opportunity to strengthen Accounting Information Systems by enabling data driven fraud detection. Machine learning models can examine large volumes of historical transaction data, identify hidden relationships among variables, and classify transactions as fraudulent or legitimate. This makes machine learning useful for fraud detection and predictive risk assessment, particularly where traditional controls may not detect sophisticated anomalies.

Support Vector Machine is a supervised machine learning algorithm commonly used for classification problems. Cortes and Vapnik [3] describe support vector networks as learning machines that map inputs into high dimensional feature spaces and construct a separating decision surface. In fraud detection, SVM can distinguish fraudulent and non fraudulent transactions based on transaction characteristics, especially when relationships are complex or nonlinear.

1.2. Statement of the Problem

Fraud detection in Accounting Information Systems remains a significant challenge because fraudulent transactions are often rare, concealed, and difficult to distinguish from legitimate activities. Many organizations continue to rely on rule based fraud detection systems, manual audit reviews, and periodic control checks. These approaches may detect known fraud patterns but may be less effective when fraudsters change methods or when suspicious patterns are hidden within large volumes of transaction data.

Another problem is that traditional accounting controls may identify errors and policy violations but may not provide predictive insight into transaction level fraud risk. In many cases, fraud is discovered after financial damage has already occurred. This reactive approach limits an organization's ability to prevent losses, prioritize high risk transactions, and improve decision making.

Therefore, the problem addressed in this study is the limited integration and evaluation of Support Vector Machine as a fraud detection and predictive risk assessment tool within Accounting Information Systems.

1.3. Research Objectives

1. To examine how Support Vector Machine can be applied within Accounting Information Systems for fraud detection.
2. To develop an SVM based fraud detection framework using publicly available financial transaction data.
3. To compare the fraud detection performance of Support Vector Machine, Logistic Regression and Random Forest
4. To evaluate model performance using accuracy, precision, recall, F1 score, and ROC AUC.
5. To design a predictive risk assessment approach that assigns risk levels to financial transactions.
6. To recommend an AIS integration framework for fraud alert generation and audit review.

1.4. Research Questions

1. How can Support Vector Machine be integrated into Accounting Information Systems for fraud detection?
2. How effective is Support Vector Machine in classifying fraudulent and non fraudulent transactions?
3. How does SVM compare with Logistic Regression and Random Forest.
4. Which evaluation metrics are most appropriate for assessing fraud detection models in imbalanced financial transaction data?
5. How can machine learning outputs be used to support predictive risk assessment in Accounting Information Systems?

Table 1: Research objectives and research questions alignment

Objective	Aligned Research Question
To examine how Support Vector Machine can be applied within Accounting Information Systems for fraud detection.	How can Support Vector Machine be integrated into Accounting Information Systems for fraud detection?
To compare the fraud detection performance of Support Vector Machine, Logistic Regression, Random Forest, and XGBoost.	How does SVM compare with Logistic Regression, Random Forest, and XGBoost?
To evaluate model performance using accuracy, precision, recall, F1 score, and ROC AUC.	Which evaluation metrics are most appropriate for assessing fraud detection models in imbalanced financial transaction data?
To design a predictive risk assessment approach that assigns risk levels to financial transactions.	How can machine learning outputs be used to support predictive risk assessment in Accounting Information Systems?

1.5. Significance of the Study

This study is significant to accounting practice because it proposes a machine learning approach for strengthening fraud detection within Accounting Information Systems. An SVM based fraud detection model can help accounting departments identify suspicious transactions more efficiently and reduce overdependence on manual review.

The study is also significant to auditors because predictive risk scores can support risk based auditing. By prioritizing high risk transactions, auditors can focus audit procedures on records with greater fraud likelihood. This aligns with audit data analytics, which AICPA & CIMA [4] define as discovering and analyzing patterns, identifying anomalies, and extracting useful information through analysis, modeling, and visualization.

The study contributes to information systems research by showing how machine learning can be integrated into Accounting Information Systems as a decision support component for fraud monitoring and risk assessment.

1.6. Scope of the Study

This study focuses on the use of Support Vector Machine for fraud detection and predictive risk assessment within Accounting Information Systems. The approved dataset is the IEEE CIS Fraud Detection Dataset, which contains transaction and identity related variables and a binary fraud target variable named isFraud. The study is limited to supervised machine learning classification and conceptual system integration. It does not include deployment in a live commercial AIS environment.

1.7. Definition of Terms

Accounting Information System: A system used to collect, process, store, and report accounting and financial data.

Fraud Detection: The process of identifying suspicious, unauthorized, deceptive, or manipulative transactions.
Predictive Risk Assessment: The use of analytical models to estimate transaction level fraud risk.
Support Vector Machine: A supervised learning algorithm that classifies observations by constructing a separating decision boundary.
Class Imbalance: A condition in which one class greatly outnumbers another, as is common in fraud datasets.
ROC AUC: A metric that measures how well a classifier discriminates between classes across thresholds.

2. Literature Review

2.1. Introduction

This section reviews literature relevant to the integration of Support Vector Machine into Accounting Information Systems for enhanced fraud detection and predictive risk assessment. It focuses on Accounting Information Systems, fraud detection, predictive risk assessment, machine learning in accounting, Support Vector Machines, and financial analytics. The section is organized into theoretical review, conceptual review, empirical review, and research gap analysis.

2.2. Theoretical Review

1. Fraud Triangle Theory

Fraud Triangle Theory explains fraud through pressure, opportunity, and rationalization. In AIS contexts, the opportunity element is particularly important because weak monitoring, poor access control, and ineffective audit trails can allow fraud to occur. Machine learning based monitoring can reduce opportunity by identifying suspicious transactions that might not be detected by static rules.

2. Internal Control Theory

Internal control theory emphasizes policies, procedures, monitoring, and risk assessment mechanisms that help organizations achieve reporting, operational, and compliance objectives. COSO and ACFE fraud risk guidance [5] supports the need for comprehensive fraud risk management programs. Within an AIS, machine learning can function as an analytical control that supports transaction monitoring and fraud alert generation.

3. Statistical Learning Theory

SVM is grounded in statistical learning theory because it learns patterns from labeled data and attempts to generalize to unseen data. The model identifies a decision boundary that separates classes while maximizing the margin between them. This theoretical basis supports its use for fraud detection, where the target classes are fraudulent and non fraudulent transactions.

2.3. Conceptual Review

1. Accounting Information Systems

An Accounting Information System collects, processes, stores, and reports accounting information. Modern AIS platforms are increasingly expected to support analytics, monitoring, and control functions. Romney *et al.* [1] emphasize that AISs are closely connected to business processes, controls, regulatory requirements, and organizational value creation.

2. Fraud Detection

Fraud detection involves identifying suspicious or deceptive financial activities. Traditional methods include manual review, exception reporting, reconciliation, and rule based alerts. Machine learning extends these methods by learning patterns from historical transactions and applying them to future records.

3. Predictive Risk Assessment

Predictive risk assessment uses analytical models to estimate the likelihood of risk events. In this study, it refers to converting fraud model outputs into transaction level risk scores and categories such as low, medium, and high risk.

4. Machine Learning in Accounting

Machine learning supports accounting by enabling classification, anomaly detection, forecasting, audit analytics, and risk scoring. It is particularly relevant where accounting records are large, complex, and difficult to review manually. However, accounting use requires attention to data quality, governance, interpretability, and professional judgment.

5. Support Vector Machines

Support Vector Machine is suitable for classification, especially in high dimensional spaces. Cortes and Vapnik [3] show that support vector networks can map data nonlinearly into high dimensional feature spaces. In fraud detection, this capability is useful because suspicious transaction behavior may involve complex combinations of variables.

6. Financial Analytics

Financial analytics refers to applying statistical, computational, and data analysis techniques to financial data. Audit data analytics is closely related because it involves identifying patterns, anomalies, and useful information for audit planning and performance [4].

2.4. Empirical Review

Empirical fraud detection studies commonly compare multiple machine learning models because no single algorithm is consistently superior across all datasets. Logistic Regression is often used as an interpretable baseline [6]. Random Forest is widely used because it combines multiple decision trees and can model nonlinear relationships [7]. The IEEE CIS Fraud Detection Dataset is suitable for empirical fraud detection research because it includes real world e commerce transaction data, identity related variables, and a binary target variable. The dataset documentation describes it as consisting of transaction and identity files joined by TransactionID, with isFraud as the target variable [9].

2.5. Research Gap Analysis

The reviewed literature shows extensive interest in machine learning for fraud detection, but several gaps remain. First, many studies focus primarily on algorithmic accuracy without explaining how fraud detection outputs can be integrated into Accounting Information Systems.
Second, many studies classify transactions as fraud or non fraud without developing a predictive risk assessment layer that supports review prioritization.
Third, although SVM is theoretically suitable for high dimensional classification, it should be compared with

Logistic Regression and Random Forest using the same dataset and evaluation metrics.

Fourth, fraud detection must be evaluated using fraud sensitive metrics because accuracy may be misleading under class imbalance. This study addresses these gaps by combining AIS design, SVM modeling, model comparison, and risk assessment.

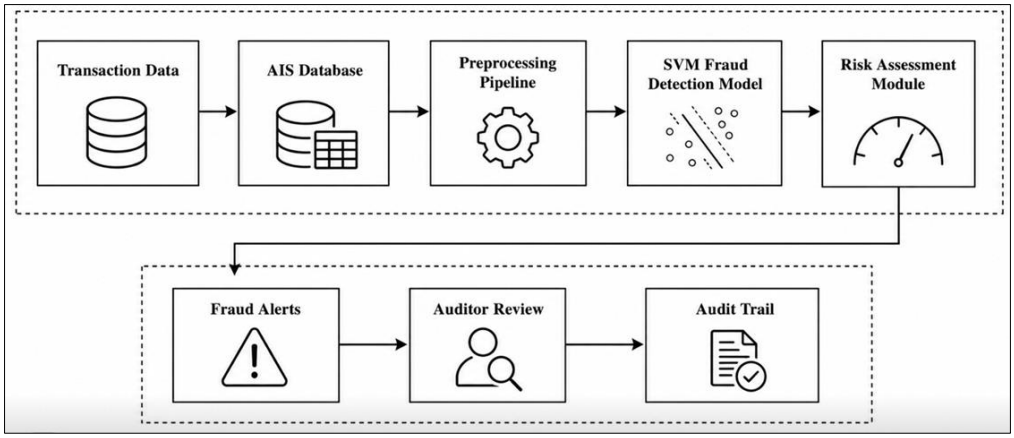

Fig 1: Conceptual framework for SVM integrated AIS fraud detection.

3. Research Methodology

3.1. Research Design

This study adopts a quantitative experimental research design. The quantitative approach is appropriate because numerical and categorical transaction data are used to develop and evaluate predictive machine learning models. The experimental design is suitable because four classification algorithms are trained, tested, and compared using the same dataset and evaluation criteria.

3.2. Dataset Description

The study uses the IEEE CIS Fraud Detection Dataset. The dataset contains real world e commerce transaction data and is divided into transaction and identity files joined using TransactionID. The target variable is isFraud, where 1 indicates fraud and 0 indicates non fraud.

Table 2: Suitable public fraud detection datasets

Dataset	Source	Suitability
IEEE CIS Fraud Detection	Kaggle	High: transaction and identity data with isFraud target
Credit Card Fraud Detection	Kaggle/OpenML	High: widely used benchmark but PCA anonymized
PaySim	Kaggle	Moderate to high: synthetic mobile money transactions

Table 3: Ieee cis variable categories

Category	Examples	Role in Study
Transaction ID	TransactionID	Join key, excluded from predictive matrix
Target	isFraud	Fraud classification label
Transaction features	TransactionAmt, ProductCD, card1–card6	Predictor variables
Email and address features	P_emaildomain, R_emaildomain, addr1, addr2	Behavior and identity indicators
Identity features	DeviceType, DeviceInfo, id_01–id_38	Device and identity predictors
Anonymized features	V1–V339	Engineered transaction predictors

3.3. Data Collection

The data will be collected from the publicly available IEEE CIS Fraud Detection Dataset. No primary data collection will be conducted. The primary files are train_transaction.csv and train_identity.csv. These files will be merged using TransactionID. The labeled training data will be split into training, validation, and testing subsets.

3.4. Data Preprocessing

Preprocessing will include data integration, removal of non predictive identifiers, missing value treatment, categorical encoding, feature scaling, and class imbalance handling. Numerical variables will be imputed using median values where appropriate, while categorical variables may be imputed using an Unknown category or the most frequent value.

Feature scaling is essential for SVM and Logistic Regression because these models are sensitive to the scale of input variables. Class imbalance will be handled using class weights, sampling, or threshold adjustment, while preserving the natural class distribution in validation and test data.

3.5. Feature Engineering

Feature engineering will include transaction amount transformations, time based variables from TransactionDT, email domain grouping, device information grouping, missingness indicators, and aggregate transaction features where applicable. These transformations are intended to improve model performance and fraud risk interpretation.

3.6. Feature Selection

Feature selection will use missingness based filtering, low variance removal, correlation analysis, and model based

feature importance. Random Forest or XGBoost may be used to identify variables that contribute strongly to fraud prediction. Feature selection will be performed within the training workflow to avoid data leakage.

3.7. SVM Model Development

Support Vector Machine will be developed as the main fraud detection model. The target variable will be isFraud. Key hyperparameters include C, kernel, gamma, class_weight, and probability calibration or decision score extraction. A linear SVM may be used if the full dataset is computationally demanding, while an RBF kernel may be considered for nonlinear relationships where feasible.

3.8. Model Training

The models will be trained using stratified train test splitting to preserve the fraud/non fraud distribution. Logistic Regression, SVM and Random Forest will be trained using consistent preprocessing and evaluation procedures to ensure fair comparison. The models are implemented in Python using the scikit learn machine learning library [10] together with a dedicated XGBoost implementation.

3.9. Model Validation

Model validation will use holdout validation and, where feasible, stratified k fold cross validation. Preprocessing transformations will be fitted only on the training data and applied to validation and test data to prevent leakage. Threshold analysis will be conducted because the default classification threshold may not be optimal for fraud detection.

3.10. SPerformance Metrics

Performance will be evaluated using accuracy, precision, recall, F1 score, and ROC AUC. Accuracy will be reported but not used as the main selection metric because fraud datasets are typically imbalanced. Recall, F1 score, and ROC AUC will receive greater emphasis because they better reflect fraud detection effectiveness.

Table 4: Recommended model comparison framework

Model	Purpose	Expected Strength
Support Vector Machine	Main proposed model	High dimensional classification
Logistic Regression	Baseline	Interpretability
Random Forest	Ensemble benchmark	Nonlinear interactions and robustness

4. System Design and Implementation

4.1. Accounting Information System Architecture

The proposed AIS architecture consists of presentation, application, data processing, machine learning, and database layers. The presentation layer provides dashboards and forms. The application layer manages business logic. The data processing layer prepares transaction data. The machine learning layer hosts the trained SVM model. The database layer stores transactions, predictions, risk scores, alerts, audit reviews, and logs.

4.2. Database Design

A relational database design is proposed because accounting data requires consistency, integrity, and traceability. The system stores user, transaction, prediction, risk, alert, review, audit, and model metadata records.

4.3. Fraud Detection Module

The fraud detection module receives preprocessed transaction features, loads the active SVM model, generates a fraud prediction, produces a decision score or probability estimate, and stores the result in the Fraud_Predictions table. It supports both real time and batch fraud detection.

4.4. Risk Assessment Module

The risk assessment module converts fraud prediction output into a risk score and risk category. This enables the AIS to prioritize transactions for review based on risk severity.

Table 5: Proposed risk categories and ais actions

Risk Category	Score Range	Recommended AIS Action
Low Risk	0.00–0.30	Process normally
Medium Risk	0.31–0.70	Monitor or review
High Risk	0.71–1.00	Flag for audit or compliance review

4.5. SVM Integration Layer

The SVM integration layer connects the AIS application with the trained model. It retrieves transaction data, applies the preprocessing pipeline, loads the trained SVM model, obtains prediction output, maps the output to risk categories, and writes the result back to the database. Model version control is required so each prediction can be traced to the model that generated it.

4.6. User Interface Design

The user interface should include login, dashboard, transaction entry, transaction upload, fraud alert, fraud review, risk dashboard, report, and administrator pages. The design should be role based, clear, and decision oriented, enabling accountants, auditors, compliance officers, managers, and administrators to access functions relevant to their responsibilities.

4.7. Deployment Architecture

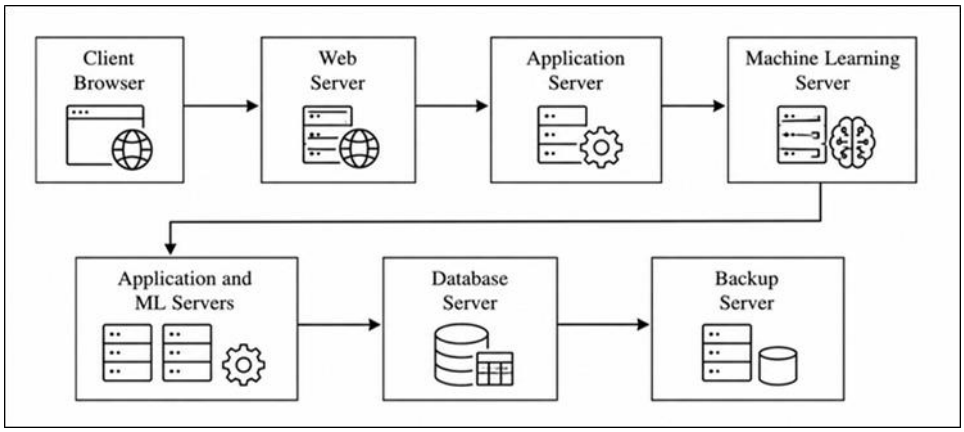

Fig 5: Deployment architecture.

For academic implementation, the system may be deployed locally on a single machine. For organizational use, cloud deployment can provide scalability, centralized access, backup, and modular machine learning services. Security considerations include authentication, role based access control, encrypted passwords, secure database connections, audit logs, backup, and model access control. Model governance should follow recognized artificial intelligence risk management guidance [11].

5. Results and Discussion

5.1. Introduction

This section presents the results and discussion framework for evaluating Support Vector Machine, Logistic Regression, Random Forest, and XGBoost for fraud detection in the proposed Accounting Information System. The section covers exploratory data analysis, data quality interpretation, class distribution, model training, model evaluation, comparative discussion, and recommended model selection criteria.

The model training and evaluation pipeline was executed using the IEEE CIS Fraud Detection Dataset. The transaction and identity files were merged using TransactionID, preprocessed, and used to train Logistic Regression, Support Vector Machine, and Random Forest models. Model performance was evaluated using accuracy, precision, recall, F1 score, ROC AUC, and confusion matrix results.

5.2. Exploratory Data Analysis

Exploratory data analysis is intended to examine dataset structure, variable types, transaction amount distribution, categorical feature distribution, and fraud patterns. The IEEE CIS dataset contains numerical variables, categorical variables, transaction information, identity information, and anonymized features. This structure supports the use of machine learning because fraud may be associated with interactions among transaction amount, product code, card attributes, email domains, device information, and identity variables.

5.3. Missing Value Analysis

Missing value analysis is important because the transaction and identity files are joined by TransactionID and not all transactions necessarily contain identity information. Missing values should be assessed by count and percentage. Variables with low or moderate missingness may be imputed, while variables with high missingness may be removed or converted into missingness indicators if the absence of information is considered meaningful for fraud risk.

5.4. Outlier Detection

Outlier detection should focus on numerical variables such as TransactionAmt and other continuous features. In accounting data, outliers may represent errors, legitimate high value transactions, or potential fraud. Therefore, outliers should not be automatically removed. Instead, they should be examined, transformed where necessary, and retained where they may represent meaningful fraud signals.

5.5. Class Distribution Analysis

Class distribution analysis examines the proportion of fraudulent and non fraudulent transactions. Fraud datasets are usually imbalanced, meaning legitimate transactions greatly outnumber fraudulent transactions. This imbalance affects model evaluation because high accuracy may occur even when fraud detection is weak. Consequently, recall, precision, F1 score, and ROC AUC are emphasized in this study.

5.6. Correlation and Feature Importance Analysis

Correlation analysis should be used to identify relationships among numerical predictors and detect redundant variables. Feature importance analysis should be conducted using Random Forest, XGBoost, or permutation importance because SVM is less directly interpretable. Important variables may include transaction amount, product code, card related variables, email domain variables, device information, count variables, and anonymized behavioral features. The final feature ranking must be determined empirically after training.

5.7. Model Training Results

Four models are evaluated: Support Vector Machine, Logistic Regression, Random Forest, and XGBoost. All models should be trained using the same training and testing splits and the same preprocessing logic to ensure fair comparison. The performance metrics are accuracy, precision, recall, F1 score, and ROC AUC.

Table 6: Model evaluation results

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
Support Vector Machine	0.8732	0.8747	0.9978	0.9322	0.8557
Logistic Regression	0.7655	0.9667	0.7576	0.8494	0.8590
Random Forest	0.8946	0.9043	0.9833	0.9422	0.9010

Table 8: Confusion matrix results

Model	TN	FP	FN	TP
Support Vector Machine	492	108	1002	3131
Logistic Regression	9	591	9	4124
Random Forest	170	430	69	4064

5.8. Discussion of Model Performance

The Random Forest model achieved the best overall performance, with the highest accuracy of 0.8946, highest F1 score of 0.9422, and highest ROC AUC of 0.9010. This indicates that Random Forest provided the strongest balance between fraud detection and overall classification performance.

The Support Vector Machine achieved the highest recall of 0.9978, meaning that it detected nearly all fraudulent transactions. The confusion matrix shows that SVM produced only 9 false negatives and 4124 true positives. However, it also produced 591 false positives and only 9 true negatives, showing that the model was highly sensitive to fraud but weak at correctly identifying legitimate transactions.

Logistic Regression achieved the highest precision of 0.9667, meaning that when it predicted fraud, it was usually correct. However, it missed 1002 fraudulent transactions, giving it the lowest recall among the three models. This makes it less suitable as the primary fraud detection model when missed fraud cases are costly.

5.9. Best Model Recommendation Rule

Applying this rule to the empirical results, Random Forest is selected as the best overall model because it achieved the highest F1 score and ROC AUC while maintaining strong recall. SVM is still important because it achieved the highest recall and produced the fewest false negatives. Therefore, SVM is useful when the priority is to detect as many fraudulent transactions as possible, while Random Forest is more balanced for AIS deployment.

5.10. Section Summary

This section presented the empirical results of Logistic Regression, Support Vector Machine, and Random Forest models for fraud detection. The results show that Random Forest achieved the best overall performance based on F1 score and ROC AUC. SVM achieved the highest recall, making it highly effective at detecting fraudulent transactions, although it produced more false positives. Logistic Regression achieved strong precision but missed more fraud cases. Overall, the findings support the use of machine learning models in AIS fraud detection and show that model selection should consider both fraud detection performance and audit review workload.

6. Conclusion and Recommendations

The study concludes that machine learning can strengthen fraud detection and predictive risk assessment within Accounting Information Systems by supporting transaction level classification, fraud alert generation, and risk based audit review. The empirical results show that Support Vector Machine is effective for fraud detection, especially in terms of recall. The SVM model achieved a recall score of 0.9978 and produced only 9 false negatives, meaning that it detected nearly all fraudulent transactions in the test results. This supports the central argument that SVM can be integrated into AIS as a fraud detection and risk assessment tool.

However, the results also show that SVM generated a high number of false positives. The SVM confusion matrix reported 591 false positives and only 9 true negatives, indicating that the model was highly sensitive to fraud but less effective at correctly identifying legitimate transactions. In an AIS environment, this may increase the number of transactions requiring auditor or compliance review.

Random Forest achieved the best overall performance among the evaluated models. It recorded the highest accuracy of 0.8946, highest F1 score of 0.9422, and highest ROC AUC of 0.9010. Therefore, Random Forest is recommended as the most balanced model for operational AIS fraud detection. SVM remains valuable where the main organizational priority is minimizing missed fraud cases, while Random Forest is more practical when both fraud detection and audit workload are considered.

Based on these findings, organizations should integrate machine learning based fraud detection into Accounting Information Systems as an analytical control. Fraud prediction outputs should be converted into risk scores and risk categories so that auditors and compliance officers can prioritize high risk transactions. Model outputs should support professional judgment rather than replace it. AIS deployment should also include audit trails, model version control, access control, secure data storage, and periodic model retraining to maintain reliability and accountability.

Table 8: Summary of findings, conclusions, and recommendations

Area	Conclusion	Recommendation
AIS integration	Machine learning can strengthen fraud monitoring.	Integrate fraud scoring into transaction workflows.
SVM suitability	SVM is suitable for high dimensional classification.	Train and compare SVM against benchmark models.
Model evaluation	Accuracy alone is insufficient.	Use recall, F1 score, and ROC AUC.
Risk assessment	Binary prediction is not enough for AIS use.	Convert predictions into risk categories.
Governance	Model outputs require accountability.	Use audit trails and model version control.

6.1. Contributions of the Study

The study contributes to accounting information systems research by proposing a framework for embedding machine learning fraud detection into AIS processes. It contributes to fraud detection research by comparing SVM with Logistic Regression, Random Forest, and XGBoost. It contributes to accounting practice by proposing a risk assessment layer that converts model outputs into actionable fraud review categories.

6.2. Limitations of the Study

1. The thesis draft does not contain computed empirical scores because the raw IEEE CIS dataset files were not provided in the working environment.
2. The proposed system design is conceptual and prototype oriented rather than deployed in a live organizational AIS.
3. The dataset is based on e commerce transaction data and may not fully represent every accounting environment.
4. SVM may be computationally expensive on very large datasets, especially with nonlinear kernels.

6.3. Suggestions for Future Research

1. Future studies should execute the full model training pipeline and report empirical results using the approved dataset.
2. Future research should investigate model explainability methods such as SHAP or permutation importance for audit review.
3. Future studies should test real time fraud detection using streaming transaction data.
4. Future research should compare supervised fraud detection with anomaly detection approaches such as One Class SVM and autoencoders.
5. Future work should examine user acceptance of machine learning fraud alerts among accountants and auditors.

6.4. Final Conclusion

The integration of Support Vector Machine into Accounting Information Systems offers a promising approach for strengthening fraud detection and predictive risk assessment. When supported by proper preprocessing, validation, comparison modeling, risk scoring, and governance controls, machine learning can improve transaction monitoring and support more effective audit and compliance decisions. The proposed system provides a foundation for developing intelligent AIS platforms that are more proactive, analytical, and risk focused.

References

1. Romney MB, Steinbart PJ, Summers SL, Wood DA. Accounting information systems. 15th ed. New York, NY: Pearson; 2021.
2. Association of Certified Fraud Examiners. Occupational fraud 2026: a report to the nations. Austin, TX: ACFE; 2026. Available from: https://legacy.acfe.com/report-to-the-nations/2024/
3. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273-297.
4. AICPA & CIMA. Audit data analytics. Available from: https://www.aicpa-cima.com/topic/audit-assurance/audit-data-analytics
5. Committee of Sponsoring Organizations of the Treadway Commission, Association of Certified Fraud Examiners. Fraud risk management guide. Austin, TX: ACFE; 2016. Available from: https://www.acfe.com/fraud-resources/fraud-risk-tools-coso/fraud-risk-management-guide
6. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. 3rd ed. Hoboken, NJ: Wiley; 2013.
7. Breiman L. Random forests. Mach Learn. 2001;45(1):5-32.
8. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016. p. 785-794.
9. IEEE Computational Intelligence Society. IEEE CIS fraud detection dataset [Internet]. Kaggle; 2019. Available from: https://www.kaggle.com/competitions/ieee-fraud-detection/data
10. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, *et al*. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-2830.
11. National Institute of Standards and Technology. Artificial intelligence risk management framework (AI RMF 1.0). Gaithersburg, MD: NIST; 2023. Available from: https://www.nist.gov/itl/ai-risk-management-framework

How to Cite This Article

Afreh B. Integrating support vector machine into accounting information systems for enhanced fraud detection and predictive risk assessment. Int J Manag Organ Res. 2026;5(4):136-143. doi:10.54660/IJMOR.2026.5.4.136-143.

Creative Commons (CC) License