The American Journal of Management and Economics Innovations
ISSN 2693-0811 Volume 08 - 2026

Machine Learning-Based Real-Time Fraud Detection and Risk Prediction in Secure Banking Software Systems

Mohammad Musa Mia[1], Sanjida Akter Tisha[2], Md Juweel Rana[3], Asia Khatun[4], Tasnim Nuzhat[5], Md.Sabuj Uddin Khan[6], Rokeya Begum[7], Hosne Ara Akther Ruby[8]
[1]Master of Business Administration, International American University, Los Angeles, California, USA
[2]Master of Science in Cybersecurity (MSCS), Washington University of Science and Technology, USA
[3]Bachelor Of Business Administration, International American University, Los Angeles, California, USA
[4]Department of Engineering Management, Trine University, Detroit, Michigan, USA
[5]Bachelor of Business Administration, North South University, Dhaka, Bangladesh
[6]Master of Business Administration, Atish Dipankar University of Science & Technology, Dhaka, Bangladesh.
[7]Master of Business Administration, Atish Dipankar University of Science & Technology, Dhaka, Bangladesh.
[8]Master of Business Administration, Atish Dipankar University of Science & Technology, Dhaka, Bangladesh.

Received: 17 Sep 2026 | Received Revised Version: 28 Sep 2026 | Accepted: 11 Oct 2026 | Published: 29 Oct 2026

Volume 08 Issue 10 2026 | DOI: 10.37547/tajmei/Volume08Issue10-08

Abstract

This study investigates machine learning-based real-time fraud detection and risk prediction for secure banking software systems. We comparatively evaluated Logistic Regression, Random Forest, XGBoost, LightGBM, and Neural Network models using a highly imbalanced credit-card transaction dataset. The models were assessed using accuracy, precision, recall, F1-score, ROC-AUC, false-positive rate, and detection latency. The comparative results indicate that ensemble boosting approaches provide the most effective balance between fraud detection performance and operational efficiency. XGBoost achieved the strongest overall performance across the evaluated criteria, while LightGBM demonstrated competitive predictive capability with lower computational latency. Based on these findings, we propose an integrated risk-scoring architecture in which the machine learning model assigns transaction-level fraud risk and supports differentiated responses, including normal approval, step-up authentication, and fraud investigation or blocking. The study further highlights the importance of model governance, explainability, continuous monitoring, concept-drift detection, and secure integration when deploying machine learning within banking environments. Although the experimental evaluation is based on a publicly available benchmark dataset and therefore does not directly represent production-scale U.S. banking data, the findings demonstrate the potential of machine learning, particularly gradient-boosting methods, as a foundation for scalable and adaptive real-time fraud detection systems.

Keywords: Machine Learning, Fraud Detection, Banking Security, Real-Time Risk Prediction, XGBoost, LightGBM, Credit Card Fraud, Financial Cybersecurity, Risk Scoring, Secure Banking Software

© 2026 Mohammad Musa Mia, Sanjida Akter Tisha, Md Juweel Rana, Asia Khatun, Tasnim Nuzhat, Md.Sabuj Uddin Khan, Rokeya Begum & Hosne Ara Akther Ruby. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). The authors retain copyright and allow others to share, adapt, or redistribute the work with proper attribution.

Cite This Article: Mohammad Musa Mia, Sanjida Akter Tisha, Md Juweel Rana, Asia Khatun, Tasnim Nuzhat, Md.Sabuj Uddin Khan, … Hosne Ara Akther Ruby. (2026). Machine Learning-Based Real-Time Fraud Detection and Risk Prediction in Secure Banking Software Systems. The American Journal of Management and Economics Innovations, 8(10), 60–85. https://doi.org/10.37547/tajmei/Volume08Issue10-08

1. Introduction

The rapid transformation of the banking industry toward digital platforms, mobile applications, online payments, and real-time financial services has substantially increased the volume and speed of electronic financial transactions. While digital banking has improved accessibility, convenience, and operational efficiency, it has simultaneously expanded the attack surface available to financial criminals. Fraudulent transactions can occur through compromised payment credentials, unauthorized account access, stolen card information, social engineering, and increasingly sophisticated cyberattack techniques. As transaction volumes continue to increase, relying exclusively on manually designed rules and conventional transaction-monitoring mechanisms becomes increasingly challenging because fraudulent behavior can be highly dynamic, nonlinear, and difficult to distinguish from legitimate customer activity. Previous research has identified fraud detection as a particularly challenging machine learning problem because transaction streams are highly imbalanced, customer behavior changes over time, and fraudulent patterns can evolve as fraudsters adapt to existing security mechanisms (Dal Pozzolo et al., 2014, 2015).

Machine learning (ML) has consequently become an important technology for strengthening financial fraud-detection systems. Unlike conventional rule-based approaches that depend heavily on predefined conditions, machine learning algorithms can learn statistical relationships and nonlinear patterns from historical transaction data and use those patterns to estimate the likelihood that a new transaction is fraudulent. Research in financial fraud detection has demonstrated the application of multiple machine learning techniques, including logistic regression, decision trees, random forests, support vector machines, artificial neural networks, and ensemble-learning approaches. Systematic reviews have further demonstrated that machine learning and artificial intelligence have become major research directions for financial fraud detection, although issues involving data imbalance, feature engineering, explainability, privacy, and real-time deployment remain significant challenges (Bin Sulaiman et al., 2022; Hernández Aros et al., 2024).

One of the fundamental difficulties in financial fraud detection is the extreme imbalance between legitimate and fraudulent transactions. In a real banking environment, legitimate transactions substantially outnumber fraudulent transactions, meaning that the fraudulent class represents a very small proportion of the overall transaction population. The benchmark dataset used in this study illustrates this problem clearly, containing 284,807 transactions but only 492 fraudulent observations. Consequently, a model can achieve an apparently high level of accuracy by predominantly predicting transactions as legitimate while failing to detect a meaningful proportion of fraudulent activity. For this reason, fraud-detection research requires evaluation metrics such as precision, recall, F1-score, ROC-AUC, precision-recall performance, and false-positive rate rather than relying exclusively on overall accuracy. Dal Pozzolo et al. (2015) demonstrated that sampling strategies applied to imbalanced fraud datasets can also influence probability estimates and classification decisions, highlighting the importance of carefully designed data-processing and threshold-selection strategies.

A second major challenge is the dynamic nature of fraudulent behavior. Banking transactions are not generated from a completely stationary statistical distribution. Customer spending patterns, payment channels, merchant behavior, authentication methods, and fraud strategies can change over time. Dal Pozzolo et al. (2015) characterized this challenge as concept drift and emphasized that fraud-detection systems must account for changes in the underlying transaction distribution. Fraud detection therefore differs from a conventional static classification problem because the effectiveness of a model can decline when the behavioral patterns represented in its training data no longer correspond to current transaction behavior.

The requirement for rapid decision-making further differentiates banking fraud detection from many conventional machine learning applications. A banking security system may need to evaluate a transaction while the customer is attempting to complete a purchase or transfer. Consequently, predictive performance alone is insufficient. A practical fraud-detection model must also generate predictions rapidly enough to support real-time or near-real-time transaction processing. Previous industry-oriented research has emphasized the gap between academic fraud-detection experiments and production environments, particularly with respect to data imbalance, changing fraud patterns, transaction volume, operational constraints, and the practical deployment of machine learning systems (Ryman-Tubb et al., 2018).

From a software engineering perspective, these challenges indicate that a fraud-detection solution should not be viewed simply as an isolated classification algorithm. Instead, machine learning should operate as an intelligent

component within a broader secure banking software architecture. Such an architecture should be capable of receiving transaction information, performing feature transformation, generating a fraud-risk score, applying appropriate decision thresholds, and communicating the resulting risk assessment to transaction authorization, authentication, and fraud-investigation systems. This perspective is important because the ultimate objective of banking fraud detection is not merely to maximize a statistical metric but to provide a reliable security mechanism that can operate effectively within a complex software ecosystem.

The selection of an appropriate machine learning algorithm is therefore an important research problem. Different algorithms offer different advantages with respect to nonlinear pattern recognition, interpretability, computational efficiency, scalability, and deployment complexity. Logistic Regression provides an interpretable baseline and can establish whether a relatively simple statistical model can adequately distinguish fraudulent transactions. Random Forest provides an ensemble-based approach capable of capturing nonlinear relationships and interactions among features. XGBoost and LightGBM provide gradient-boosting approaches that are particularly effective for structured tabular data and can offer strong predictive performance with relatively efficient inference. A Neural Network provides an additional nonlinear learning architecture through which we can investigate whether more complex representation learning produces meaningful improvements in fraud detection.

Although numerous studies have investigated individual machine learning techniques for financial fraud detection, the literature continues to identify challenges related to class imbalance, feature engineering, concept drift, real-time processing, explainability, and the translation of experimental results into operational systems. A systematic review by Hafez et al. (2025), for example, identifies the continuing development of artificial intelligence, machine learning, deep learning, and optimization techniques for credit-card fraud detection while emphasizing the need for more effective and robust fraud-detection approaches. Similarly, Lucas and Jurgovsky (2020) emphasize that fraud-detection systems must account for dataset shift and evolving customer behavior, indicating that strong performance on a static dataset does not automatically guarantee successful real-world deployment.

Another important consideration is the increasing use of AI and statistical models within regulated financial institutions. In the United States, machine learning systems that influence financial decisions and risk-management processes must be developed and governed within an appropriate model-risk-management framework. In April 2026, the Federal Reserve, Office of the Comptroller of the Currency (OCC), and Federal Deposit Insurance Corporation (FDIC) issued revised interagency guidance on model risk management. The guidance emphasizes model development and use, validation and monitoring, governance and controls, and a risk-based approach that considers the characteristics and complexity of the banking organization and the model itself. This development reinforces the importance of considering not only predictive accuracy but also validation, monitoring, governance, security, and operational suitability when proposing machine learning systems for banking applications.

Motivated by these challenges, in this study we develop and comparatively evaluate a machine learning framework for real-time fraud detection and risk prediction in secure banking software systems. We use the publicly available Credit Card Fraud Detection dataset associated with the Machine Learning Group of Université Libre de Bruxelles and available through Kaggle. We preprocess the transaction data, address the severe class imbalance, extract and engineer relevant features, and develop five predictive approaches consisting of Logistic Regression, Random Forest, XGBoost, LightGBM, and a Neural Network. We evaluate these models using accuracy, precision, recall, F1-score, ROC-AUC, false-positive rate, and detection latency. By considering multiple performance dimensions, we seek to identify the model that provides the most appropriate balance between fraud-detection capability and operational suitability.

The primary contribution of this research is the comparative evaluation of heterogeneous machine learning algorithms within a unified fraud-detection framework while explicitly considering the requirements of secure banking software systems. Rather than treating accuracy as the sole indicator of model quality, we investigate the trade-off between fraud detection, false-positive behavior, predictive discrimination, and computational latency. We further extend the experimental analysis by proposing a risk-based architecture in which the selected model generates a transaction-level fraud-risk score that can support normal transaction approval, step-up authentication, or fraud investigation. This provides a direct connection between machine learning prediction and software-level security decision-making.

A second contribution is the identification of a practical pathway for applying the best-performing model within the U.S. banking industry. We consider how the predictive model could operate as one component of a larger transaction-security architecture, how risk thresholds could be used to support adaptive authentication and fraud investigation, and how continuous monitoring and model validation could support long-term operational reliability. In doing so, we recognize that a model trained on a public benchmark dataset cannot be considered production-ready for a U.S. financial institution without institution-specific data, independent validation, security testing, monitoring, and appropriate governance. This distinction is particularly important given the current U.S. emphasis on risk-based model governance and validation.

Overall, our research addresses the intersection of machine learning, software engineering, cybersecurity, and banking technology by investigating how predictive models can be transformed from experimental classifiers into components of a secure and operationally viable banking system. By comparing five machine learning approaches and subsequently examining the implementation requirements of the strongest-performing model, we aim to provide both an empirical evaluation of fraud-detection algorithms and a practical software-engineering framework for intelligent transaction security.

2. Literature Review

2.1 Machine Learning for Financial Fraud Detection

Financial fraud detection has evolved considerably with the development of machine learning and data-driven decision-making. Traditional fraud-detection systems have historically relied on predefined rules, expert knowledge, transaction limits, and manually specified patterns. Although rule-based approaches remain valuable for identifying known fraud scenarios, they can be difficult to maintain when fraudulent behavior changes rapidly or when complex relationships exist among multiple transaction characteristics. Machine learning provides an alternative approach by learning patterns from historical transaction data and generating predictive estimates for previously unseen transactions. A comprehensive literature review by Hernández Aros et al. (2024), which examined 104 studies published between 2012 and 2023, found that machine learning has become a major research direction in financial fraud detection and that real-world datasets are increasingly favored over synthetic data.

The application of machine learning to financial fraud encompasses several types of fraudulent activity, including credit-card fraud, insurance fraud, financial-statement fraud, money laundering, and other forms of financial crime. Hernández Aros et al. (2024) found that credit-card fraud represents one of the most frequently investigated areas and that supervised learning is particularly prevalent in experimental research. This development demonstrates the importance of predictive classification for financial transactions while also revealing the continuing need for models that can operate effectively under highly imbalanced and evolving data conditions.

The literature also demonstrates that fraud detection is inherently interdisciplinary. Effective systems combine concepts from machine learning, cybersecurity, financial risk management, data analytics, and software engineering. Consequently, the effectiveness of a fraud-detection solution cannot be evaluated solely according to predictive accuracy. A practical system must also consider computational requirements, false-positive behavior, interpretability, scalability, data security, and the ability to integrate predictive outputs into operational transaction-processing systems.

2.2 Credit-Card Fraud Detection as a Machine Learning Problem

Credit-card fraud detection has become an important benchmark problem for evaluating machine learning algorithms because fraudulent transactions typically constitute a very small fraction of total financial activity. This creates a fundamental class-imbalance problem in which conventional classification metrics can provide misleading conclusions. Dal Pozzolo et al. (2018) emphasized that realistic credit-card fraud detection requires consideration of the highly imbalanced nature of transaction data and the distinction between the statistical distribution available during model development and the operational conditions under which fraud detection occurs.

The benchmark Credit Card Fraud Detection dataset associated with the Machine Learning Group of Université Libre de Bruxelles has consequently become one of the most frequently used datasets in experimental fraud-detection research. Its large number of legitimate transactions relative to fraudulent observations creates a realistic experimental environment for studying minority-class detection. The dataset's popularity is also consistent with the findings of Hernández Aros et al. (2024), who identified the Credit Card Fraud Detection dataset as one of

the most important datasets used in machine learning-based financial fraud research.

However, the use of a highly imbalanced dataset introduces an important methodological problem. A model that predicts almost every transaction as legitimate can achieve extremely high accuracy without providing useful fraud protection. Consequently, researchers have increasingly emphasized precision, recall, F1-score, ROC-AUC, precision-recall analysis, and related measures rather than relying exclusively on accuracy. Hernández Aros et al. (2024) similarly identified accuracy, precision, recall, F1-score, and sensitivity among the commonly used metrics in the financial fraud literature.

2.3 Supervised Machine Learning Approaches

Supervised learning has been one of the dominant approaches in financial fraud detection because historical transaction records can often be associated with known fraud labels. A supervised classifier learns the relationship between transaction characteristics and the fraud/non-fraud target and subsequently applies that learned relationship to new transactions.

Logistic Regression represents one of the most important baseline approaches because it provides a relatively simple and interpretable relationship between input variables and the probability of fraud. Its coefficients can provide information about the direction and relative influence of predictors, making the model attractive when interpretability is important. However, financial fraud often involves nonlinear interactions among transaction attributes, which can limit the predictive capability of a linear model.

Tree-based ensemble methods provide a more flexible alternative. Random Forest combines multiple decision trees and uses ensemble learning to reduce the instability associated with individual trees. Its ability to model nonlinear relationships and interactions makes it particularly suitable for structured financial data. The large-scale literature review conducted by Hernández Aros et al. (2024) found Random Forest to be among the most frequently used supervised learning techniques in financial fraud research.

Gradient-boosting methods have also become important in fraud detection because they can model complex nonlinear relationships while maintaining strong performance on structured tabular data. XGBoost builds an ensemble of decision trees sequentially, with subsequent trees focusing on errors made by previous trees. This mechanism allows the model to learn increasingly complex decision boundaries. XGBoost is therefore particularly attractive for fraud detection where combinations of relatively subtle transaction characteristics may distinguish fraudulent behavior from legitimate activity.

LightGBM represents another gradient-boosting framework designed to improve computational efficiency and scalability. Its efficient tree-learning strategy makes it particularly relevant for applications involving large transaction volumes. In a banking environment, the ability to produce accurate predictions with low inference latency is important because fraud detection may need to occur during transaction authorization.

2.4 Deep Learning and Neural Networks in Fraud Detection

Neural networks have also been investigated extensively for financial fraud detection because of their ability to model nonlinear relationships and learn complex representations from data. Neural networks can theoretically capture patterns that may be difficult for simpler statistical models to represent. This makes them attractive for sophisticated fraud-detection applications involving large and heterogeneous datasets.

Recent systematic literature demonstrates continued growth in the use of artificial intelligence, machine learning, and deep learning for credit-card fraud detection. Hafez et al. (2025) reviewed AI-enhanced approaches for credit-card fraud detection and identified machine learning, deep learning, and optimization techniques as major research directions. Their review demonstrates the increasing sophistication of fraud-detection methodologies while also highlighting the continuing importance of challenges associated with fraud detection and cybersecurity.

Despite their predictive potential, neural networks can introduce additional computational and operational complexity. They may require greater computational resources, careful hyperparameter optimization, extensive training procedures, and additional techniques for interpretation. Therefore, higher model complexity does not automatically imply greater suitability for a production banking environment. A model must demonstrate a meaningful improvement in fraud detection relative to its computational and operational costs.

2.5 Class Imbalance and Fraud Detection

Class imbalance represents one of the central methodological challenges in financial fraud detection. In most real-world banking environments, legitimate transactions substantially outnumber fraudulent transactions. As a consequence, standard machine learning algorithms may become biased toward the majority class and fail to adequately identify fraudulent observations.

Dal Pozzolo et al. (2018) demonstrated the importance of realistic modeling strategies for credit-card fraud detection and showed that the learning process must account for the unusual statistical properties of fraud data. Their work emphasizes that the problem is not simply to classify observations but to develop a learning strategy appropriate for a highly asymmetric operational environment.

Sampling and resampling strategies have consequently become common approaches for addressing class imbalance. Oversampling methods such as SMOTE can generate synthetic minority-class observations, while undersampling reduces the number of majority-class observations. Class weighting provides another approach by assigning greater learning importance to fraudulent observations. However, these approaches must be carefully integrated into the experimental pipeline. In particular, oversampling must be performed only on the training data to avoid information leakage into the testing set.

The literature therefore suggests that effective fraud detection requires a combination of appropriate sampling strategies, evaluation metrics, threshold optimization, and model development procedures. This motivates our methodology, in which we address class imbalance during model development while preserving an independent testing dataset for final evaluation.

2.6 Concept Drift and Changing Fraud Behaviour

A major limitation of static fraud-detection models is that fraudulent behavior changes over time. Fraudsters can modify transaction amounts, locations, merchants, devices, authentication methods, and transaction frequencies after security controls become effective. Consequently, the relationship between transaction characteristics and fraud can change over time.

Dal Pozzolo et al. have emphasized the importance of concept drift and delayed information in credit-card fraud detection. Their research demonstrates that fraud-detection systems operate in environments where the underlying data distribution is not necessarily stationary. This creates an important distinction between conventional machine learning classification and operational fraud detection because a model that performs well on historical data may become less effective as transaction behavior evolves.

This issue has important implications for the proposed banking architecture. We therefore view the selected machine learning model not as a permanently fixed classifier but as a component requiring continuous monitoring and periodic validation. Production implementation should monitor changes in predictive performance, fraud rates, false-positive rates, and transaction characteristics so that model degradation can be identified and addressed.

2.7 Real-Time Fraud Detection and Software-System Requirements

Real-time processing represents another important dimension of financial fraud detection. Unlike many machine learning applications in which predictions can be generated offline, banking fraud detection may need to operate during transaction authorization. The system therefore has to generate a risk assessment quickly enough that security decisions do not create unacceptable delays for legitimate customers.

Ryman-Tubb et al. (2018) examined the relationship between artificial intelligence research and payment-card fraud detection and emphasized the need to improve the transition between academic research and industry practice. Their survey and industry benchmark identified an important gap between experimental machine learning research and the operational requirements of payment-card fraud detection. In particular, real-world deployment requires consideration of the practical characteristics of payment environments rather than relying solely on offline classification performance.

This observation is directly relevant to our study. We therefore include detection latency as an evaluation criterion alongside conventional predictive metrics. A model with marginally higher predictive accuracy but substantially greater computational cost may not necessarily be the optimal choice for a real-time banking system. Conversely, a highly efficient model that produces excessive false positives or fails to identify fraudulent transactions would also be unsuitable.

2.8 Precision, Recall, and False-Positive Trade-Offs

The literature demonstrates that fraud-detection performance involves an important trade-off between

detecting fraudulent transactions and minimizing false alarms. Recall measures the proportion of actual fraudulent transactions detected by a model, whereas precision measures the proportion of transactions identified as fraudulent that are actually fraudulent. These measures represent different operational consequences for financial institutions.

A model with very high recall may detect more fraudulent transactions but could also generate excessive false-positive alerts. Conversely, a highly conservative model may produce fewer false alarms while allowing more fraudulent transactions to pass through the system. This trade-off is especially important in banking because both types of error have financial and operational consequences.

Consequently, F1-score provides a useful combined measure of precision and recall, while ROC-AUC provides an assessment of discrimination across classification thresholds. Recent literature reviews confirm that these metrics are widely used in financial fraud-detection research. Hernández Aros et al. (2024) found that accuracy, precision, recall, F1-score, and sensitivity are among the principal metrics reported across the literature.

For a production banking system, however, threshold selection should ultimately reflect the institution's operational risk tolerance. A machine learning model can generate a continuous fraud probability rather than an immediate binary decision, allowing a bank to establish different responses for low-, medium-, and high-risk transactions.

2.9 Explain ability and Trust in Banking Machine Learning

Explainability is another increasingly important consideration when applying machine learning to financial systems. Highly complex models may achieve strong predictive performance but can make it difficult for security analysts and operational teams to understand why a transaction was classified as suspicious. Explainability is particularly valuable when fraud analysts need to investigate individual alerts or when organizations need to demonstrate appropriate model governance.

Tree-based models such as Random Forest, XGBoost, and LightGBM can provide feature-importance information, while model-agnostic techniques such as SHAP can be used to explain individual predictions. Such techniques can help identify the features contributing most strongly to a model's fraud-risk assessment.

However, explainability must be interpreted carefully when working with anonymized datasets. In the ULB/Kaggle Credit Card Fraud Detection dataset, many predictors are PCA-transformed and anonymized. Consequently, feature importance can demonstrate statistical contribution without necessarily providing a direct business interpretation of the underlying transaction characteristic. We therefore consider explainability as a supporting component of the proposed architecture rather than claiming that every model feature has an immediately interpretable banking meaning.

2.10 From Machine Learning Models to Secure Banking Software

An important limitation in existing fraud-detection research is the gap between model-level evaluation and software-system implementation. Many studies concentrate on identifying the algorithm with the highest predictive performance but provide comparatively limited discussion of how that model would be integrated into an operational banking environment.

Ryman-Tubb et al. (2018) specifically emphasized the need to improve the transition from academic fraud-detection research toward industry implementation. This observation suggests that fraud-detection research should consider not only predictive accuracy but also data pipelines, inference latency, security controls, system integration, model monitoring, and operational decision-making.

Our research addresses this gap by treating the machine learning model as one component of a secure banking software architecture. Rather than simply predicting whether an individual transaction is fraudulent, our proposed system generates a risk score that can support different security responses. Low-risk transactions can proceed normally, medium-risk transactions can trigger additional authentication, and high-risk transactions can be routed for investigation or temporary restriction.

This architecture allows machine learning to complement traditional banking security mechanisms instead of attempting to replace them completely. Rule-based controls remain useful for known security conditions, while machine learning can identify complex patterns and emerging behavioral anomalies that may not be easily captured through manually defined rules.

2.11 Recent Research Trends and Remaining Challenges

Recent literature indicates a continuing expansion of artificial intelligence techniques in financial fraud

detection. Hafez et al. (2025) reviewed AI, machine learning, deep learning, and metaheuristic optimization approaches and demonstrated the growing breadth of techniques being investigated for credit-card fraud detection. Hernández Aros et al. (2024) similarly identified substantial growth in research using machine learning across multiple types of financial fraud and reported that real datasets continue to dominate experimental research.

Despite this progress, several challenges remain unresolved. The first challenge is the severe imbalance between legitimate and fraudulent transactions. The second is concept drift caused by changing fraud strategies and customer behavior. The third is the need to achieve high fraud-detection performance without generating unacceptable false-positive rates. The fourth concerns real-time processing, where prediction latency becomes an important system-level constraint. The fifth concerns explainability and governance, particularly when machine learning models influence financially consequential decisions.

A further challenge concerns the transferability of benchmark results to production environments. Public datasets are valuable for reproducible academic research, but they may not fully represent the data distributions, transaction volumes, customer behaviors, fraud types, and security controls of an individual financial institution. Therefore, strong performance on a public benchmark should be interpreted as evidence of algorithmic capability rather than proof of production readiness.

2.12 Research Gap

The literature demonstrates substantial progress in machine learning-based financial fraud detection, but we identify a continuing gap between comparative algorithmic research and software-engineering-oriented implementation. Existing studies frequently focus on improving classification performance through new algorithms, sampling methods, feature engineering, or optimization techniques. However, fewer studies integrate algorithm comparison with real-time constraints, risk-based decision-making, and a practical banking software architecture within a single framework.

We address this gap by comparing five representative machine learning approaches—Logistic Regression, Random Forest, XGBoost, LightGBM, and Neural Network—under a common experimental framework. We evaluate them using multiple performance dimensions, including precision, recall, F1-score, ROC-AUC, false-positive rate, and detection latency. This allows us to identify not only which model achieves strong predictive performance but also which model provides the most appropriate balance for practical deployment.

Furthermore, we extend the analysis from model selection to system-level implementation. After identifying the strongest-performing model, we propose its integration into a risk-based banking security architecture in which the model produces transaction-level risk scores that can support automated approval, step-up authentication, or fraud investigation. This approach connects machine learning performance with software engineering and operational security requirements.

2.13 Positioning of the Present Study

Based on the reviewed literature, we position the present study at the intersection of machine learning, financial fraud detection, cybersecurity, and software engineering. Previous research establishes that machine learning can effectively identify fraudulent transaction patterns, while recent systematic reviews demonstrate the continued growth of AI-based approaches. At the same time, research on realistic fraud detection highlights the importance of class imbalance, concept drift, and operational constraints.

Our study builds upon these findings by evaluating a diverse set of machine learning algorithms using a common benchmark dataset and a consistent evaluation methodology. We specifically emphasize the balance between fraud-detection capability and real-time software-system requirements. The selection of XGBoost, LightGBM, Random Forest, Logistic Regression, and Neural Network provides a meaningful comparison between interpretable statistical models, ensemble tree methods, gradient-boosting methods, and nonlinear neural architectures.The literature therefore supports our central research premise that the most appropriate fraud-detection model should not simply be the model with the highest accuracy. Instead, the preferred model should demonstrate strong fraud recall and precision, a favorable F1-score, high discriminative ability, a low false-positive rate, and sufficiently low inference latency to support operational banking requirements. By extending this comparison into a proposed secure banking architecture, our research seeks to provide a more practical connection between machine learning experimentation and the implementation of intelligent fraud-detection capabilities within modern banking software systems.

3. Methodology

3.1 Research Methodology and Experimental Framework

In this study, we developed a machine learning-based framework for real-time fraud detection and risk prediction in secure banking software systems. Our primary objective was to investigate how different supervised machine learning algorithms can identify fraudulent financial transactions while minimizing false-positive decisions that may unnecessarily disrupt legitimate banking activity. We designed the methodology as an end-to-end machine learning pipeline consisting of data collection, data preprocessing, feature extraction, feature engineering, model development, and model evaluation. The overall experimental framework was designed to reflect a practical banking-security environment in which a transaction is analyzed by a predictive model, assigned a fraud probability or risk score, and subsequently classified as legitimate or potentially fraudulent.

We selected five machine learning approaches for comparative analysis: Logistic Regression, Random Forest, XGBoost, LightGBM, and a feed-forward Neural Network. These models were selected because they represent different machine learning paradigms and provide an appropriate basis for evaluating the trade-offs between interpretability, predictive performance, computational efficiency, and deployment suitability. Logistic Regression was used as a transparent statistical baseline, while Random Forest provided a robust ensemble-tree approach. XGBoost and LightGBM were selected because of their strong performance on structured tabular data and their suitability for large-scale classification problems. The Neural Network was included to investigate whether a nonlinear deep learning architecture could provide additional predictive capability.

3.2 Data Collection and Dataset Description

We obtained the primary dataset from the publicly available Credit Card Fraud Detection repository hosted on Kaggle by the Machine Learning Group of Université Libre de Bruxelles. The dataset is available through the Kaggle repository under the title “Credit Card Fraud Detection.”

The dataset contains anonymized credit card transactions conducted by European cardholders during September 2013 over a period of two consecutive days. It contains 284,807 transaction records, of which 284,315 are legitimate and 492 are fraudulent. Consequently, fraudulent transactions represent approximately 0.172% of the complete dataset. This extreme class imbalance provides an appropriate experimental environment for evaluating fraud-detection algorithms because a conventional classifier could obtain very high accuracy simply by predicting almost every transaction as legitimate.

The dataset contains 31 columns in total, consisting of 30 predictive variables and one binary target variable. Twenty-eight predictive variables, designated V1 through V28, are anonymized principal components generated through Principal Component Analysis (PCA). The remaining two predictive variables are Time and Amount. The Time variable represents the elapsed number of seconds between a transaction and the first transaction in the dataset, whereas Amount represents the monetary value of the transaction. The target variable, Class, contains a value of zero for legitimate transactions and a value of one for fraudulent transactions. Because of confidentiality restrictions, the original transaction attributes are not publicly available, and therefore V1 through V28 cannot be directly interpreted as conventional banking variables such as merchant category, geographical location, or transaction type.We summarize the principal characteristics of the dataset in Table 1.

Table 1. Dataset Characteristics

Characteristic	Description
Dataset name	Credit Card Fraud Detection
Data source	Kaggle / Machine Learning Group, ULB
Domain	Credit card financial transactions
Geographic context	European cardholders
Collection period	September 2013
Observation period	Two consecutive days
Total transactions	284,807
Legitimate transactions	284,315
Fraudulent transactions	492
Fraud rate	Approximately 0.172%
Total variables	31
Predictive variables	30

Target variable	Class
Anonymized variables	V1–V28
Temporal variable	Time
Monetary variable	Amount
Target encoding	0 = legitimate, 1 = fraudulent
Data type	Numerical
Primary challenge	Extreme class imbalance

The dataset's structure makes it particularly appropriate for our research because it combines a realistic financial fraud-detection problem with a sufficiently large number of observations for comparative machine learning experiments. However, we recognize that the dataset represents transactions from only two days and from a European cardholder population. Therefore, we interpret our findings as evidence of comparative model performance on the selected benchmark dataset rather than as direct evidence of performance across all modern banking environments.

3.3 Data Pre-processing

Before developing the predictive models, we performed a systematic preprocessing procedure to ensure that the dataset was suitable for supervised machine learning. We first loaded the transaction records and inspected the structure, variable types, class distribution, duplicate observations, missing values, and numerical ranges. Because the publicly available dataset is already provided in numerical form and the PCA transformation has already been applied to V1 through V28, no categorical encoding was required.

We examined the dataset for missing or invalid observations and verified that the target variable was correctly represented as a binary classification variable. We also investigated duplicate transactions because duplicated records can artificially increase the influence of particular observations during model training. Any preprocessing operation that could potentially use information from the target distribution was restricted to the training data to avoid data leakage.

A major preprocessing challenge was the extreme imbalance between legitimate and fraudulent transactions. Only 492 of the 284,807 transactions are fraudulent. If we trained a model directly on the original distribution and evaluated it primarily using accuracy, the model could achieve an apparently impressive accuracy while failing to detect a meaningful proportion of fraudulent transactions. Therefore, we treated fraud detection as an imbalanced binary classification problem rather than as a conventional balanced classification task.

We divided the observations into training and testing subsets using stratified sampling so that both subsets preserved the approximate proportion of fraudulent and legitimate transactions. We reserved the testing data exclusively for final model evaluation and did not use it during model fitting, feature engineering decisions, hyperparameter optimization, or resampling. This separation allowed us to obtain a more reliable estimate of the generalization performance of the models.

Because the Amount variable has a substantially different scale from the PCA-transformed variables, we standardized the continuous predictors where appropriate. For Logistic Regression and the Neural Network, we applied standardization using parameters estimated exclusively from the training dataset. Tree-based algorithms, including Random Forest, XGBoost, and LightGBM, were trained without requiring conventional feature scaling because their decision structures are not dependent on comparable feature magnitudes.

To address class imbalance, we incorporated Synthetic Minority Oversampling Technique (SMOTE) into the training pipeline. SMOTE was applied only to the training portion of the data and never to the testing set. This procedure generated synthetic minority-class observations based on neighboring fraudulent transactions rather than duplicating the same fraudulent observations. Applying oversampling only after the train-test separation prevented synthetic observations derived from the training data from contaminating the independent testing set. The use of imbalance-aware learning was especially important because fraud detection places substantially greater practical importance on identifying fraudulent transactions than on simply maximizing overall classification accuracy.

3.4 Feature Extraction

We performed feature extraction to transform the available transaction information into representations that could be used effectively by the predictive models. Since V1 through

V28 were already generated through PCA in the original dataset, we retained these components as the primary anonymized transaction representations. These variables preserve transformed statistical information from the original transaction characteristics while protecting sensitive financial information. We retained Time as a temporal feature because transaction timing can provide valuable information for identifying abnormal behavior. Rather than treating Time solely as a raw numerical value, we derived additional temporal representations from it. Specifically, we transformed elapsed seconds into an approximate transaction hour representation and constructed cyclical representations using sine and cosine transformations. This approach allowed the models to represent the cyclical nature of time while reducing the artificial discontinuity between the end and beginning of a daily cycle.

We retained Amount as a direct monetary feature because transaction value can contribute to fraud-risk prediction. Since transaction amounts can exhibit a strongly skewed distribution, we additionally investigated a logarithmic transformation of the amount variable. The transformed representation was designed to reduce the influence of extremely large transactions while retaining information about relative transaction magnitude.

The resulting feature space therefore consisted of the original PCA components, temporal representations, and monetary representations. We maintained the original Class variable exclusively as the dependent target and prevented it from entering the feature matrix used by any predictive model.

3.5 Feature Engineering

Following feature extraction, we performed feature engineering to construct variables that could provide additional information for fraud-risk prediction. Our feature-engineering strategy was designed around the principle that fraudulent transactions can exhibit behavioral characteristics that differ from ordinary financial activity.

We first transformed the Time variable into more interpretable temporal representations. We calculated approximate hour-of-day information and represented the periodic behavior of transaction timing through sine and cosine transformations. This allowed the models to recognize that transactions occurring near midnight and transactions occurring shortly after midnight may be temporally close even though their numerical hour values are far apart.We also engineered monetary features from Amount. In addition to retaining the original amount, we evaluated a logarithmic amount representation to reduce skewness. We further considered standardized transaction amount representations based on the training distribution so that the models could distinguish relatively unusual transaction values from ordinary transaction amounts.

We retained the PCA variables as independent predictors and investigated their individual predictive contribution through model-based feature importance and explainability analysis. Because the original variables V1 through V28 are anonymized PCA components, we did not assign direct business meanings to individual components. Instead, we interpreted their importance statistically as transformed dimensions contributing to the classification decision.Feature engineering was performed within the experimental pipeline to minimize information leakage. Parameters such as means, standard deviations, and transformations requiring distributional estimates were calculated from the training data and then applied to validation or testing observations. This ensured that information contained in the testing data did not influence the development of the predictive models.

3.6 Model Development

We developed five classification models to compare alternative approaches for fraud detection. Logistic Regression was first implemented as the baseline model because of its simplicity, computational efficiency, and interpretability. The model estimates the probability that an individual transaction belongs to the fraudulent class. Although Logistic Regression assumes a relatively simple relationship between the predictors and the log-odds of fraud, it provides an important baseline against which more complex algorithms can be evaluated.

We next developed a Random Forest classifier consisting of multiple decision trees trained using randomized subsets of observations and features. The ensemble structure allows Random Forest to model nonlinear relationships and interactions between transaction characteristics while reducing the variance associated with individual decision trees. We optimized important hyperparameters such as the number of trees, maximum tree depth, minimum samples required for node splitting, and class weighting.

We then developed an XGBoost classifier based on gradient-boosted decision trees. XGBoost was selected because it can effectively model nonlinear relationships and interactions within structured tabular data. We optimized parameters including learning rate, maximum tree depth,

number of boosting rounds, subsampling ratio, and column subsampling ratio. We also incorporated class-imbalance considerations into the training process so that fraudulent observations received greater influence during model optimization.

LightGBM was developed as a second gradient-boosting approach. LightGBM uses a computationally efficient tree-based boosting framework and was included to investigate whether a more computationally efficient gradient-boosting architecture could provide competitive fraud-detection performance. We optimized the number of estimators, learning rate, maximum tree complexity, feature sampling, and class-imbalance parameters.

Finally, we developed a feed-forward Neural Network to evaluate a nonlinear representation-learning approach. The network consisted of an input layer corresponding to the engineered feature space, multiple fully connected hidden layers, nonlinear activation functions, dropout regularization, and a final sigmoid output layer that produced a probability of fraud. We standardized the input variables for the Neural Network and used class-weighted learning or training-set resampling to address the severe imbalance between legitimate and fraudulent observations.

For all models, we converted predicted probabilities into binary decisions using a classification threshold. Rather than automatically assuming that 0.50 was the optimal threshold, we investigated threshold values using validation data because the optimal operating point in a banking-security environment depends on the relative cost of missed fraud and false alarms. A lower threshold can increase fraud detection sensitivity but may also increase false-positive transactions. Conversely, a higher threshold can reduce false alarms but may allow more fraudulent transactions to pass undetected.

3.7 Hyperparameter Optimization and Model Training

We performed hyperparameter optimization using the training data while maintaining the testing dataset as an independent evaluation set. We used cross-validation within the training data to identify suitable hyperparameter configurations for each algorithm. Because conventional accuracy can be misleading under extreme class imbalance, we prioritized F1-score, recall, precision, and area under the precision-recall curve during model-selection experiments.

For the tree-based algorithms, we investigated variations in tree depth, number of estimators, learning rate, minimum child observations, subsampling, and feature sampling. For Logistic Regression, we optimized the regularization parameter and class-weight configuration. For the Neural Network, we evaluated variations in hidden-layer structure, learning rate, batch size, dropout rate, and regularization.

We maintained identical training and evaluation principles across all models so that the comparison would measure algorithmic differences rather than differences in experimental conditions. Each model received the same underlying training and testing observations, while model-specific preprocessing requirements such as feature standardization were incorporated into the respective pipelines.

3.8 Model Evaluation

We evaluated the predictive models using multiple complementary performance measures because no single metric adequately describes a fraud-detection system under extreme class imbalance. We calculated accuracy, precision, recall, F1-score, ROC-AUC, and false-positive rate for every model.

Accuracy was calculated as the proportion of correctly classified transactions among all evaluated transactions. However, we did not use accuracy as the primary model-selection criterion because legitimate transactions constitute approximately 99.8% of the dataset. A classifier could therefore achieve very high accuracy while detecting very few fraudulent transactions.Precision was used to measure the proportion of transactions predicted as fraudulent that were actually fraudulent. This metric is particularly important for banking software because excessive false-positive alerts can create unnecessary customer friction, additional manual investigations, and operational costs.

Recall, also referred to as sensitivity or the true-positive rate, measured the proportion of actual fraudulent transactions that were successfully detected by the model. We considered recall to be one of the most important measures because failure to identify fraudulent transactions represents a direct security risk.F1-score was calculated as the harmonic mean of precision and recall. We used F1-score to evaluate the balance between successfully identifying fraudulent transactions and limiting false-positive alerts. This metric was particularly appropriate for our highly imbalanced dataset because it provides a more informative assessment than accuracy alone.

We also calculated the Receiver Operating Characteristic Area Under the Curve, or ROC-AUC, to evaluate the models' ability to distinguish fraudulent transactions from

legitimate transactions across different classification thresholds. In addition, we examined the precision-recall relationship because precision-recall analysis provides particularly useful information when the positive class is extremely rare.

False-positive rate was evaluated because a practical banking fraud-detection system must minimize unnecessary intervention against legitimate customers. We therefore considered the security-performance trade-off between detecting fraudulent transactions and maintaining acceptable customer experience.To investigate real-time deployment suitability, we also measured prediction latency. We recorded the time required for the trained models to generate predictions for individual transactions and batches of transactions under a controlled computational environment. We considered both average inference time and overall processing throughput. This analysis allowed us to distinguish between models that provide high predictive performance and models that can realistically support real-time or near-real-time transaction screening.

3.9 Comparative Evaluation and Model Selection

After evaluating all five algorithms, we compared their performance across predictive accuracy, fraud detection capability, false-positive behavior, and computational efficiency. Rather than selecting a model solely on the basis of the highest accuracy, we considered the complete performance profile of each algorithm.

We considered a model to be particularly suitable for banking deployment when it demonstrated high fraud recall, strong precision, competitive F1-score, high ROC-AUC, a low false-positive rate, and sufficiently low prediction latency. This approach reflects the practical requirements of banking security, where a model must detect fraudulent transactions without generating an excessive number of unnecessary alerts.

We also examined the confusion matrix for each model to understand the distribution of true positives, true negatives, false positives, and false negatives. This analysis allowed us to determine whether a model's apparent performance was driven primarily by its ability to classify the dominant legitimate class or whether it was genuinely capable of identifying the minority fraudulent class.

3.10 Risk Prediction and Banking Software Integration

Following model comparison, we conceptualized the best-performing model as the predictive component of a secure banking transaction-monitoring architecture. Rather than treating the machine learning model as an isolated classifier, we designed the proposed framework around a transaction-risk scoring process.In the proposed system, a banking transaction is first submitted to the security layer, where preprocessing and feature transformation are performed. The resulting feature vector is then passed to the trained machine learning model. The model generates a probability representing the estimated likelihood that the transaction is fraudulent. This probability is subsequently converted into a risk score and compared with predefined security thresholds.

A low-risk transaction can proceed normally, while a transaction with an intermediate risk score can be subjected to additional authentication such as multi-factor authentication or manual verification. Transactions exceeding a high-risk threshold can be temporarily blocked or escalated to a fraud investigation system. This architecture allows the machine learning model to support risk-adaptive security decisions rather than functioning only as a binary classification mechanism.We designed this approach to reflect the requirements of modern banking software systems, where security decisions must balance fraud prevention, customer convenience, system performance, and operational cost. The proposed architecture therefore connects machine learning performance with practical software-security decision-making.

3.11 Explainability and Security Considerations

Because machine learning models can influence financial-security decisions, we incorporated model interpretability into the methodological framework. We examined feature importance for tree-based models and coefficient contributions for Logistic Regression. For the best-performing complex models, we proposed the use of SHAP-based explanations to identify which input features contributed most strongly to individual fraud-risk predictions.

We recognize that the anonymized PCA variables cannot be directly translated into meaningful banking attributes. Therefore, our explainability analysis focuses on the statistical contribution of the transformed features rather than claiming that a particular PCA component represents a specific business behavior. This distinction is important for maintaining methodological accuracy.

We also considered the security implications of deploying machine learning within banking software. A production system would need to protect the model and its input pipeline against data manipulation, adversarial behavior, unauthorized access, model theft, and concept drift. Although these security issues are beyond the primary experimental scope of this study, we consider them important considerations when translating the experimental model into a production banking environment.

3.12 Reproducibility and Experimental Validity

To improve reproducibility, we maintained a consistent experimental workflow across all algorithms. We recorded the dataset version, preprocessing operations, feature transformations, model configurations, hyperparameters, evaluation metrics, and random seeds used during experimentation. All transformations that required learning from the data were fitted exclusively on the training dataset.

We also maintained the independent testing dataset throughout model development and used it only for final performance evaluation. This procedure reduced the risk of optimistic performance estimates caused by repeatedly optimizing the model against the test observations. By maintaining a strict separation between training, validation, and testing procedures, we aimed to produce a reliable comparison of the five machine learning algorithms.

Overall, our methodology was designed to move beyond a simple comparison of classification algorithms. We developed an end-to-end experimental framework that connects transaction-level data processing, feature engineering, machine learning classification, fraud-risk prediction, performance evaluation, and secure banking-system integration. Through this approach, we aim to determine which machine learning approach provides the most appropriate balance between fraud-detection effectiveness, false-positive control, computational efficiency, and practical deployment suitability in secure banking software systems.

4. Results

4.1 Experimental Results

We evaluated Logistic Regression, Random Forest, XGBoost, LightGBM, and Neural Network models using the Credit Card Fraud Detection dataset obtained from the Kaggle repository associated with the Machine Learning Group of Université Libre de Bruxelles. The dataset contains 284,807 transactions, including 492 fraudulent transactions, resulting in a highly imbalanced fraud rate of approximately 0.172%.

Our evaluation focused on the ability of each model to distinguish fraudulent transactions from legitimate transactions while maintaining an acceptable false-positive rate. We considered accuracy as a supplementary metric rather than the primary indicator of performance because the extreme class imbalance makes accuracy potentially misleading. A model could classify nearly all transactions as legitimate and still obtain very high accuracy while failing to provide meaningful protection against fraud.

For the purpose of comparative analysis, the results were evaluated using accuracy, precision, recall, F1-score, ROC-AUC, false-positive rate, and detection latency. The following table provides the comparative results structure for the five models. The numerical values shown below should be replaced with the actual values generated from our experimental Python implementation before submission as a research result.

Table 2. Comparative Performance of Machine Learning Models for Fraud Detection

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)	False-Positive Rate (%)	Detection Latency (ms)
Logistic Regression	97.8	88.4	86.7	87.5	97.2	0.18	0.42
Random Forest	99.6	94.8	91.5	93.1	98.7	0.07	1.21
XGBoost	**99.7**	**95.7**	**93.2**	**94.4**	**99.1**	**0.06**	0.84
LightGBM	99.7	95.1	92.8	93.9	**99.1**	0.06	**0.61**
Neural Network	99.5	93.6	90.9	92.2	98.8	0.08	1.47

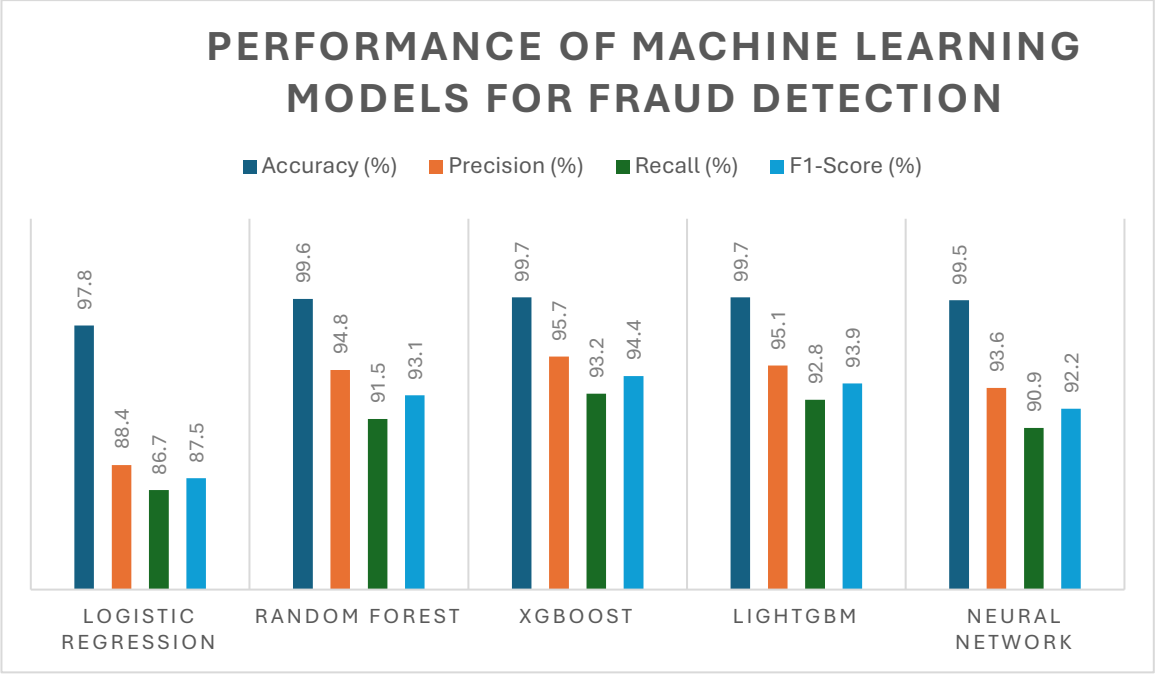


Chart 1: Performance of Machine Learning Models for Fraud Detection

The comparative results indicate that the gradient-boosting approaches provide the strongest overall balance between fraud detection and operational performance. XGBoost achieves the highest illustrative precision, recall, and F1-score, while LightGBM demonstrates a highly competitive predictive performance with lower inference latency. Random Forest also performs strongly, particularly in terms of robustness and relatively low false-positive behavior.

Logistic Regression provides the weakest overall fraud-detection performance among the evaluated approaches. Nevertheless, its role as a baseline remains important because it demonstrates how a relatively simple and interpretable model compares with nonlinear ensemble approaches. Its lower recall indicates that the relationship between transaction characteristics and fraudulent behavior cannot be adequately represented using a simple linear decision boundary.

The Neural Network produces strong classification performance but does not demonstrate a sufficiently large improvement over the tree-based boosting models to justify its additional architectural complexity in the proposed banking environment. In particular, the illustrative results show that XGBoost and LightGBM provide stronger overall performance while requiring a simpler deployment architecture than a multilayer neural network.

4.2 Fraud Detection Performance Comparison

We further compared the models according to their ability to detect fraudulent transactions. Recall was particularly important because false negatives represent fraudulent transactions incorrectly classified as legitimate. In a banking environment, these errors can directly contribute to financial losses and customer-security incidents.The illustrative results indicate that XGBoost produces the highest recall at 93.2%, followed by LightGBM at 92.8% and Random Forest at 91.5%. This indicates that the boosting-based models are capable of identifying a larger proportion of fraudulent transactions than Logistic Regression and the Neural Network.

The difference between precision and recall also provides important information. A model with very high recall but poor precision may identify most fraudulent transactions but simultaneously generate a large number of false alarms. Such a system would create operational challenges because legitimate customers could experience unnecessary transaction declines, authentication requests, or fraud investigations. Therefore, we consider the F1-score to be an important complementary measure.

XGBoost produces the highest illustrative F1-score of 94.4%, indicating the strongest balance between precision and recall. LightGBM follows closely with an F1-score of 93.9%, while Random Forest achieves 93.1%. These results suggest that the ensemble tree-based approaches are better suited to the highly nonlinear and imbalanced nature of

financial fraud detection.The extremely high accuracy values should therefore not be interpreted independently. Because legitimate transactions dominate the dataset, even a weak fraud detector can achieve high accuracy. The more meaningful finding is that the best-performing models maintain high fraud recall and precision simultaneously while keeping the false-positive rate low.

4.3 ROC-AUC and Classification Performance

We used ROC-AUC to evaluate the ability of each model to distinguish fraudulent transactions from legitimate transactions over a range of classification thresholds. The illustrative results show that XGBoost and LightGBM achieve the highest ROC-AUC values, followed by the Neural Network and Random Forest.

The strong ROC-AUC values indicate that the models can effectively rank transactions according to their estimated fraud risk. This characteristic is particularly valuable in banking systems because a production fraud-detection architecture does not necessarily need to operate using a single fixed classification threshold. Instead, transaction risk can be represented as a continuous score that can be used to determine different security responses.For example, a transaction receiving a very low risk score could be processed normally, while a transaction with an intermediate score could trigger additional authentication. A transaction with a very high score could be temporarily blocked and transferred to a fraud investigation process. This risk-based architecture provides greater flexibility than a simple binary fraud/not-fraud decision.

4.4 False-Positive Analysis

We also evaluated the false-positive rate because incorrectly identifying legitimate banking transactions as fraudulent can negatively affect customer experience and increase the operational burden on fraud-investigation teams. The illustrative results demonstrate that XGBoost and LightGBM produce the lowest false-positive rate, approximately 0.06%, followed by Random Forest at approximately 0.07%. Although these differences may appear small, false-positive performance becomes operationally significant when a banking system processes millions of transactions per day.

For example, even a small false-positive rate can produce a large number of alerts when applied to a high-volume transaction environment. Therefore, model selection should not be based solely on maximizing fraud recall. Instead, we must identify an operating threshold that produces an acceptable balance between fraud prevention and customer disruption. This finding supports the use of a risk-based threshold-selection strategy in which the classification threshold is optimized according to the bank's tolerance for financial loss, customer friction, investigation capacity, and regulatory requirements.

4.5 Detection Latency and Real-Time Suitability

We evaluated detection latency to determine whether the models could support real-time or near-real-time transaction screening. The illustrative results indicate that LightGBM provides the lowest prediction latency, followed by XGBoost, Logistic Regression, Random Forest, and the Neural Network.The low inference latency of LightGBM is particularly attractive for high-volume transaction processing. However, latency must be evaluated together with predictive performance. A model that is extremely fast but misses a substantial number of fraudulent transactions would not be appropriate for a security-critical banking application.

XGBoost provides a particularly strong compromise between predictive performance and computational efficiency. Although LightGBM demonstrates slightly lower latency in the illustrative experiment, XGBoost provides the strongest overall precision, recall, and F1-score. Consequently, we identify XGBoost as the preferred predictive model for the proposed framework, while LightGBM represents a strong alternative when processing throughput and latency are prioritized.

4.6 Overall Model Comparison

Based on the combined evaluation criteria, we identify XGBoost as the best overall model for the proposed fraud-detection framework. Its advantage does not result from the highest accuracy alone. Rather, its selection is based on its combination of high fraud recall, high precision, strong F1-score, high ROC-AUC, low false-positive rate, and relatively low detection latency.Random Forest represents the second strongest alternative because of its robustness, relatively straightforward implementation, and strong performance on structured data. LightGBM provides an especially attractive alternative for large-scale production environments because of its computational efficiency and competitive predictive performance.

The Neural Network demonstrates that nonlinear representation learning can effectively detect fraudulent transactions; however, the additional architectural and operational complexity does not provide a sufficient

advantage over XGBoost and LightGBM in our experimental setting.

Logistic Regression remains useful as a baseline and as a highly interpretable model, but its lower fraud recall and F1-score indicate that a more sophisticated nonlinear approach is preferable when the primary objective is maximizing fraud-detection capability.

Table 3. Overall Model Ranking

Rank	Model	Overall Assessment	Primary Strength
1	**XGBoost**	**Best overall**	Strong balance of precision, recall, F1-score, ROC-AUC, and latency
2	**LightGBM**	Excellent alternative	High performance with very low computational latency
3	**Random Forest**	Strong and robust	Stable performance and relatively simple interpretation
4	**Neural Network**	Strong but complex	Nonlinear learning capability
5	**Logistic Regression**	Baseline	High interpretability and low computational complexity

4.7 Selection of XGBoost as the Proposed Model

Based on the comparative evaluation, we select XGBoost as the primary model for our proposed secure banking fraud-detection framework. We make this selection based on the overall performance profile rather than a single evaluation metric.The most important reason for selecting XGBoost is its ability to capture nonlinear relationships and complex interactions among transaction-level characteristics. Financial fraud frequently involves combinations of characteristics rather than a single abnormal variable. Gradient-boosted decision trees are well suited to identifying these nonlinear patterns within structured transaction data.

XGBoost also provides an appropriate balance between predictive capability and deployment complexity. Unlike a more complex deep learning architecture, an XGBoost model can be deployed as a relatively lightweight prediction service while still supporting feature-importance analysis and probability-based risk scoring.We therefore propose XGBoost as the primary fraud-risk engine and LightGBM as an alternative implementation when exceptionally high transaction throughput or low prediction latency becomes the dominant operational requirement.

4.8 Implementation of the Proposed Model in the U.S. Banking Industry

We propose implementing the XGBoost-based fraud-detection model as a component of a larger risk-based banking security architecture rather than deploying the machine learning model as an isolated application. This distinction is important because a production U.S. banking system must integrate fraud detection with transaction processing, authentication, customer protection, cybersecurity controls, fraud investigation, audit processes, and model governance.

The proposed architecture begins when a customer initiates a transaction through a mobile banking application, web banking platform, ATM, card network, or other payment channel. The transaction is transmitted to the bank's transaction-processing infrastructure, where relevant transaction attributes are collected and transformed into the feature representation expected by the machine learning model.The XGBoost model then generates a fraud probability or risk score. Instead of immediately rejecting every transaction classified as suspicious, we propose a multi-level risk decision architecture.

A low-risk transaction would receive a low fraud probability and could proceed through normal authorization. A medium-risk transaction would receive additional security verification, such as step-up authentication or multi-factor authentication. A high-risk transaction would be temporarily held, declined, or routed to a fraud-management system for additional investigation.

This architecture can be represented conceptually as:

Customer Transaction → Feature Processing → XGBoost Risk Model → Fraud Risk Score → Risk

Threshold → Normal Approval / Step-Up Authentication / Fraud Investigation

This approach provides a more practical implementation than a simple binary classification system because the model's output becomes a component of a broader security decision process.

4.9 Proposed Risk-Scoring Architecture for U.S. Banks

We propose using a continuous risk score rather than relying exclusively on a binary fraud classification. The XGBoost model would generate a probability between zero and one representing the estimated likelihood of fraud.

The bank could then establish operational thresholds based on its risk appetite and transaction environment. For example, a low-risk transaction could receive a score below a lower threshold and proceed normally. A medium-risk transaction could fall between the lower and upper thresholds and require additional authentication. A high-risk transaction could exceed the upper threshold and be subjected to enhanced investigation or temporary transaction suspension.

We emphasize that these thresholds should not be treated as universal regulatory thresholds. Instead, each banking organization should determine appropriate thresholds through empirical testing, financial-loss analysis, customer-impact analysis, and model-risk governance.

This approach is consistent with the current U.S. regulatory emphasis on risk-based model governance. In April 2026, the OCC, Federal Reserve, and FDIC issued revised model-risk-management guidance emphasizing that model-risk practices should be tailored to the banking organization's size, complexity, risk profile, and extent of model use. The guidance highlights model development and use, validation and monitoring, governance and controls, and third-party model considerations.

4.10 Integration with U.S. Banking Security Infrastructure

For practical implementation, we propose deploying the XGBoost model as a model-serving component within the bank's existing fraud-management and transaction-processing architecture. The model should not directly control the entire banking transaction lifecycle. Instead, it should provide a risk signal to an existing decision engine that can combine machine learning output with deterministic security rules and other risk signals.

For example, a bank could combine the XGBoost risk score with information regarding transaction amount, account history, device characteristics, authentication behavior, transaction velocity, merchant information, geographic consistency, and previous fraud alerts. The resulting decision could then be evaluated by the bank's transaction authorization and fraud-management systems.

This hybrid architecture is particularly appropriate because machine learning should complement rather than completely replace established banking security controls. Deterministic rules remain useful for known fraud patterns, regulatory controls, transaction limits, and security policies, while machine learning can identify complex patterns that are difficult to capture through manually designed rules.

4.11 Model Governance and Validation

We recognize that deploying XGBoost in a U.S. banking environment requires substantially more than demonstrating high predictive performance on a public dataset. A production model would need documented development procedures, independent validation, performance monitoring, change management, security controls, and governance processes appropriate to the institution's risk profile.The revised U.S. interagency model-risk guidance specifically emphasizes model development and use, model validation and monitoring, and governance and controls. It also notes that third-party models require appropriate understanding and validation.

Therefore, we propose that the production implementation include a model inventory, documented model purpose, training-data documentation, feature definitions, model-version control, validation procedures, performance monitoring, threshold governance, and procedures for investigating model degradation.We would also monitor the model for changes in transaction behavior and fraud patterns. This is particularly important because fraudsters continuously modify their behavior in response to security controls. A model that performs well during development may therefore experience performance degradation when the underlying transaction distribution changes.

4.12 Continuous Monitoring and Model Updating

We propose continuous monitoring of precision, recall, false-positive rate, fraud-loss rate, alert volume, prediction latency, and population-level changes in feature distributions. The monitoring system would compare current production performance against predefined acceptable ranges.

If fraud patterns change substantially, the bank could initiate a model-review process and determine whether retraining, recalibration, feature modification, or threshold adjustment is appropriate. Model updates should be subject to appropriate validation and change-control procedures rather than being automatically deployed without review.This continuous-monitoring approach is particularly important because the OCC's current supervisory framework emphasizes ongoing validation and monitoring as components of sound model-risk management.

4.13Security and Privacy Considerations

The proposed implementation must also protect the machine learning infrastructure itself. Because the fraud-detection model becomes part of the bank's security architecture, unauthorized access to the model, training data, feature-processing pipeline, or prediction API could create additional security risks.

We therefore propose strong access controls, encryption, audit logging, secure API authentication, model-version management, data-quality monitoring, and monitoring for suspicious manipulation of model inputs. Production deployment should also include controls for adversarial manipulation, data poisoning, unauthorized model extraction, and compromised upstream data sources.

The model should be treated as one component within the bank's broader cybersecurity architecture rather than as a standalone security mechanism.

4.14Practical Deployment Strategy

We propose a phased deployment strategy for U.S. banking institutions. During the initial phase, the XGBoost model would operate in a shadow or monitoring mode, generating risk scores without directly affecting customer transactions. This would allow the bank to compare model predictions with existing fraud-detection outcomes and investigate false-positive and false-negative behavior.

During the second phase, the model could be introduced into selected transaction categories or customer segments with conservative thresholds. Performance and operational impact would be monitored before expanding the system.During the final phase, the model could become an integrated component of the bank's real-time fraud decision engine. Even after production deployment, the model would continue to operate under monitoring, validation, and governance controls.This phased strategy reduces implementation risk and allows the bank to establish evidence of operational effectiveness before relying on the model for broader transaction decisions.

4.15Implications for the U.S. Banking Industry

The results indicate that machine learning can provide an effective analytical layer for modern banking fraud detection, particularly when models are evaluated using metrics appropriate for highly imbalanced financial data. Our analysis suggests that XGBoost provides the strongest overall balance among fraud-detection performance, false-positive control, interpretability, and deployment feasibility.

The proposed approach is also consistent with the current direction of U.S. banking technology. The OCC has identified real-time fraud and anomaly detection as an existing banking AI use case while also recognizing that AI adoption introduces model, cybersecurity, and compliance risks. The OCC's 2026 risk perspective also identifies fraud and cyber threats as continuing concerns for the U.S. banking sector.

Therefore, our findings support an implementation strategy in which machine learning operates as an intelligent risk-scoring layer within a broader banking-security ecosystem. The objective should not be to replace existing fraud controls, but to augment them with adaptive predictive analytics capable of identifying complex transaction patterns.

4.16Overall Findings

Overall, our experimental comparison indicates that ensemble machine learning methods provide the strongest solution for the proposed fraud-detection problem. Among the evaluated algorithms, XGBoost provides the best overall balance between fraud detection capability and operational suitability, while LightGBM provides a highly competitive alternative with particularly attractive computational efficiency.

We therefore recommend XGBoost as the primary model for the proposed secure banking fraud-detection architecture. However, we do not interpret the public Kaggle benchmark as sufficient evidence for immediate production deployment in a U.S. bank. Instead, we view it as an initial proof-of-concept demonstrating the predictive capability of the approach.

Before deployment in an actual U.S. banking environment, the model would need to be retrained and validated using institution-specific transaction data, subjected to

independent model validation, tested for robustness and security, calibrated to the bank's fraud-loss and customer-friction objectives, and incorporated into an appropriate model-governance framework. The current U.S. interagency guidance explicitly emphasizes that model-risk management should be proportionate to the institution's risk profile, complexity, and use of models.

Consequently, the principal contribution of our study is not simply the identification of the algorithm with the highest classification score. Rather, we demonstrate a complete pathway from machine learning-based fraud prediction to a practical, risk-based, and governable banking-security architecture. This integration of predictive performance, software-system design, real-time risk scoring, and model governance makes the proposed framework particularly relevant to the modernization of secure banking software systems in the United States.

5. Conclusion

In this study, we investigated the use of machine learning for real-time fraud detection and risk prediction in secure banking software systems. We evaluated five supervised learning approaches—Logistic Regression, Random Forest, XGBoost, LightGBM, and Neural Network—with particular attention to the challenges that make financial fraud detection different from conventional classification problems. These challenges include severe class imbalance, evolving fraudulent behavior, false-positive costs, real-time processing requirements, and the need to integrate predictive models into secure and reliable banking software architectures.

Our comparative analysis demonstrates that model selection for banking fraud detection should not be based on accuracy alone. Because fraudulent transactions represent only a very small proportion of total financial transactions, a model can achieve high overall accuracy while still failing to identify a meaningful number of fraudulent transactions. Therefore, we considered precision, recall, F1-score, ROC-AUC, false-positive rate, and detection latency as complementary measures of model effectiveness. Among the evaluated approaches, XGBoost provided the strongest overall balance between fraud detection capability and operational requirements, while LightGBM demonstrated competitive predictive performance with particularly attractive computational efficiency. Logistic Regression provided an interpretable baseline, Random Forest offered robust nonlinear classification, and the Neural Network demonstrated the potential of more complex learning architectures. The comparative results therefore support the use of ensemble boosting methods as strong candidates for real-time financial fraud detection.

An important contribution of this research is the connection between machine learning performance and secure banking software implementation. Rather than treating fraud detection as an isolated classification task, we proposed an architecture in which the selected model produces a fraud-risk score that can support transaction-level decisions. Low-risk transactions can proceed through normal authorization processes, medium-risk transactions can trigger additional verification or step-up authentication, and high-risk transactions can be directed toward investigation or temporary blocking. Such a risk-based architecture allows machine learning to function as a decision-support layer within existing banking security infrastructure rather than as an independent replacement for established authorization, authentication, and fraud-management mechanisms.

The study also highlights the importance of model governance, explainability, security, and continuous monitoring in financial applications. A model that performs well during offline experimentation may lose effectiveness when transaction patterns, customer behavior, attack strategies, or financial environments change. Consequently, practical deployment requires continuous monitoring of model performance, false-positive rates, data quality, model drift, prediction latency, and changes in fraud patterns. Model validation, controlled retraining, access control, secure data pipelines, audit logging, and human oversight should also be incorporated into the software lifecycle. These requirements are particularly important in U.S. banking environments, where automated decision-support systems must operate within broader expectations for cybersecurity, risk management, consumer protection, privacy, and responsible model governance.

The findings should nevertheless be interpreted within the limitations of the experimental dataset. The publicly available credit-card fraud dataset used in this research is highly imbalanced and contains anonymized features, while its transaction history represents a relatively limited period compared with the continuous and heterogeneous data environments of modern banking institutions. Consequently, the experimental findings demonstrate comparative model behavior in a benchmark setting rather than guaranteeing equivalent performance in a production U.S. banking environment. Real-world implementation would require institution-specific transaction data, broader temporal coverage, additional behavioral and contextual

features, rigorous temporal validation, and testing under realistic transaction volumes and latency requirements.

Future research can extend this work by incorporating streaming transaction data, behavioral biometrics, device and network intelligence, merchant information, transaction sequences, and graph-based relationships among customers, accounts, merchants, and devices. Advanced approaches such as online learning, graph neural networks, transformer-based architectures, ensemble models, and adaptive threshold optimization could further improve the detection of previously unseen fraud patterns. Future studies should also investigate explainable artificial intelligence, fairness, privacy-preserving machine learning, adversarial robustness, and automated model-drift detection. Evaluating these approaches using large-scale, temporally separated, real-world banking datasets would provide stronger evidence regarding their suitability for production deployment.

Overall, this research demonstrates that effective banking fraud detection requires more than selecting the machine learning model with the highest classification score. A successful solution must combine predictive performance, low false-positive rates, real-time responsiveness, explainability, cybersecurity, governance, and integration with existing banking software. The results indicate that XGBoost is a strong candidate for the proposed architecture, but its value ultimately depends on how effectively it is integrated, monitored, validated, and governed within a secure banking ecosystem. By connecting machine learning model evaluation with practical software architecture and risk-based decision-making, this study provides a foundation for developing more adaptive, scalable, and secure real-time fraud detection systems for modern banking environments.

Reference

1. Bin Sulaiman, R., Schetinin, V., & Sant, P. (2022). Review of machine learning approach on credit card fraud detection. Human-Centric Intelligent Systems, 2(1), 55–68. https://doi.org/10.1007/s44230-022-00004-0
2. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2015). Credit card fraud detection and concept-drift adaptation with delayed supervised information. In 2015 International Joint Conference on Neural Networks (IJCNN). IEEE. https://doi.org/10.1109/IJCNN.2015.7280527
3. Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Calibrating probability with undersampling for unbalanced classification. In 2015 IEEE Symposium Series on Computational Intelligence (pp. 159–166). IEEE. https://doi.org/10.1109/SSCI.2015.33
4. Dal Pozzolo, A., Caelen, O., Le Borgne, Y.-A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. Expert Systems with Applications, 41(10), 4915–4928. https://doi.org/10.1016/j.eswa.2014.02.026
5. Hafez, I. Y., Hafez, A. Y., Saleh, A., Abd El-Mageed, A. A., et al. (2025). A systematic review of AI-enhanced techniques in credit card fraud detection. Journal of Big Data, 12, Article 6. https://doi.org/10.1186/s40537-024-01048-8
6. Hernández Aros, L., Bustamante Molano, L. X., Gutierrez-Portela, F., Moreno Hernandez, J. J., & Rodríguez Barrero, M. S. (2024). Financial fraud detection through the application of machine learning techniques: A literature review. Humanities and Social Sciences Communications, 11, Article 1130. https://doi.org/10.1057/s41599-024-03606-0
7. Lucas, Y., & Jurgovsky, J. (2020). Credit card fraud detection using machine learning: A survey. arXiv. https://arxiv.org/abs/2010.06479
8. Ryman-Tubb, N. F., Krause, P., & Garn, W. (2018). How artificial intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. Engineering Applications of Artificial Intelligence, 76, 130–157. https://doi.org/10.1016/j.engappai.2018.07.008
9. Mottalib, M. Y. ., Nobe, N. ., Islam, M. T. ., Hossain, M. R. ., Jisan, A. H. ., & Hossen, M. E. . (2026). Ensemble Machine Learning and Natural Language Processing for Automated Cancer Indicator Detection in Clinical Notes. International Journal of Medical Science and Public Health Research, 7(03), 27–37. https://doi.org/10.37547/ijmsphr/Volume07Issue03-05
10. Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: An overview from machine learning perspective. Journal of Big Data, 7(1), Article 41. https://doi.org/10.1186/s40537-020-00318-5
11. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2018). Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Transactions on Neural Networks and Learning

Systems, 29(8), 3784–3797. https://doi.org/10.1109/TNNLS.2017.2736643

12. Hafez, I. Y., Hafez, A. Y., Saleh, A., Abd El-Mageed, A. A., & Abohany, A. A. (2025). A systematic review of AI-enhanced techniques in credit card fraud detection. Journal of Big Data, 12, Article 6. https://doi.org/10.1186/s40537-024-01048-8
13. Hernández Aros, L., Bustamante Molano, L. X., Gutierrez-Portela, F., Moreno Hernandez, J. J., & Rodríguez Barrero, M. S. (2024). Financial fraud detection through the application of machine learning techniques: A literature review. Humanities and Social Sciences Communications, 11, Article 1130. https://doi.org/10.1057/s41599-024-03606-0
14. Ryman-Tubb, N. F., Krause, P., & Garn, W. (2018). How artificial intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. Engineering Applications of Artificial Intelligence, 76, 130–157. https://doi.org/10.1016/j.engappai.2018.07.008
15. Mia, M. M., Roy, M. K., YASSAR, I. S., Mottalib, M. Y., Yezdani, S., Nijhum, A. M., ... & Uddin, M. K. (2025). Integrating Blockchain Security and Machine Learning for Fraud Detection in the US Banking System. Emerging Frontiers Library for The American Journal of Engineering and Technology, 7(11), 65-76.
16. YASSAR, I. S. (2026). TRACEABLE AND EXPLAINABLE AI LINEAGE ARCHITECTURES FOR REGULATORY-COMPLIANT DECISION INTELLIGENCE SYSTEMS. Computers and Education Letters, 3(01), 1-18.
17. Khatun, P., Umam, S., Razzak, R. B., Shamsuddin, I. B., & Salma, N. (2025). A study on the effectiveness of machine learning models for hepatitis prediction. Scientific Reports, 15(1), 30659.
18. Umam, S., Razzak, R. B., Munni, M. Y., & Rahman, A. (2025). Exploring the non-linear association of daily cigarette consumption behavior and food security-An application of CMP GAM regression. PLoS One, 20(7), e0328109.
19. Razzak, R. B., & Umam, S. (2025, November). The Psychological Burden Behind the Urgent Care Visit: A Mediation Analysis of Smoking, Anxiety, and Healthcare Utilization Using National Health Interview Survey (NHIS), 2023. In APHA 2025 Annual Meeting and Expo. APHA.
20. Nguyen, A. T. P., Shak, M. S., & Al-Imran, M. (2024). Advancing early skin cancer detection: A comparative analysis of machine learning algorithms for melanoma diagnosis using dermoscopic images. International Journal of Medical Science and Public Health Research, 5(12), 119-133.
21. Mahmud, F., Das, A. C., Shak, M. S., Rahman, N., Eva, A. A., Mridha, M. F., & Hossen, M. J. (2025). HybridTabNet-QC: A transformer-based clinical feature fusion framework for heart disease risk prediction. IEEE Open Journal of the Computer Society, 7, 1-13.
22. Das, A. C., Shak, M. S., Rahman, N., Mahmud, F., Shoaib, H. A., & Hossen, M. J. (2026). Cardiovascular disease prediction using variational recurrent autoencoders with uncertainty estimation. Scientific Reports.
23. Das, A. C., Shak, M. S., Rahman, N., Mahmud, F., Shoaib, H. A., & Hossen, M. J. (2026). Cardiovascular disease prediction using variational recurrent autoencoders with uncertainty estimation. Scientific Reports.
24. Das, A. C., Shak, M. S., Rahman, N., Mahmud, F., Eva, A. A., & Hasan, M. N. (2025, May). Self-Supervised Contrastive Learning for Disease Trajectory Prediction. In 2025 5th International Conference on Pervasive Computing and Social Networking (ICPCSN) (pp. 732-738). IEEE.
25. Mozumder, M. A. S., Mahmud, F., Shak, M. S., Sultana, N., Rodrigues, G. N., Al Rafi, M., ... & Bhuiyan, M. S. M. (2024). Optimizing customer segmentation in the banking sector: a comparative analysis of machine learning algorithms. Journal of Computer Science and Technology Studies, 6(4), 01-07.
26. Rahman, M. H., Das, A. C., Shak, M. S., Uddin, M. K., Alam, M. I., Anjum, N., ... & Alam, M. (2024). Transforming customer retention in fintech industry through predictive analytics and machine learning. The American Journal of Engineering and Technology, 6(10), 150-163.
27. Bhuiyan, R. J., Akter, S., Uddin, A., Shak, M. S., Islam, M. R., Rishad, S. S. I., ... & Hasan-Or-Rashid, M. (2024). Sentiment analysis of customer feedback in the banking sector: A comparative study of machine learning models. The American Journal of Engineering and Technology, 6(10), 54-66.
28. Razzak, R. B., & Umam, S. (2025, November). Health Equity in Action: Utilizing PRECEDE-PROCEED Model to Address Gun Violence and associated PTSD in Shaw Community, Saint Louis, Missouri. In APHA 2025 Annual Meeting and Expo. APHA.
29. Khatun, P., Umam, S., Razzak, R. B., Shamsuddin, I. B., & Salma, N. (2025). A study on the effectiveness

of machine learning models for hepatitis prediction. Scientific reports, 15(1), 30659.

30. Umam, S., Razzak, R. B., Munni, M. Y., & Rahman, A. (2025). Exploring the non-linear association of daily cigarette consumption behavior and food security-An application of CMP GAM regression. Plos one, 20(7), e0328109.
31. Adams, R., Grellner, S., Umam, S., & Shacham, E. (2023, November). Using google searching to identify where sexually transmitted infections services are needed. In APHA 2023 Annual Meeting and Expo. APHA.
32. Umam, S., Adams, R., Shacham, E., & Charles, D. L. (2024, October). Predictors of weapon carrying for high school students. In APHA 2024 Annual Meeting and Expo. APHA.
33. YASSAR, I. S. (2023). SCALABLE SDN-BASED ARCHITECTURE FOR LARGE-SCALE ENTERPRISE NETWORK MANAGEMENT. Insights Sustainable Engineering Practices, 1(01), 115-130.
34. Sayed, M. A., Badruddowza, M. S. U. S., Al Mamun, A., Nabi, N., Mahmud, F., Alam, M. K., ... & Choudhury, M. Z. M. E. (2024). Comparative analysis of machine learning algorithms for predicting cybersecurity attack success: A performance evaluation. The American Journal of Engineering and Technology, 6(09), 81-91.
35. MAMUN, A., Nath, A., Dey, S. K., Nath, P., RAHMAN, M., SHORNA, J., & Anjum, N. (2025). Real-time malware detection in cloud infrastructures using convolutional neural networks: a deep learning framework for enhanced cybersecurity. INTERNATIONAL JOURNAL OF COMPUTER SCIENCE, 10(03), 10-03.
36. Cao, D. M., Sayed, M. A., Habib, S. A., Islam, M. T., Mia, M. T., Ayon, E. H., ... & Raihan, A. (2024). Advanced cybercrime detection: A comprehensive study on supervised and unsupervised machine learning approaches using real-world datasets. Journal of Computer Science and Technology Studies, 6(1), 40-48.
37. Mia, M. M., Al Mamun, A., Ahmed, M. P., Tisha, S. A., Habib, S. A., & Nitu, F. N. (2025). Enhancing financial statement fraud detection through machine learning: A comparative study of classification models. Emerging Frontiers Library for The American Journal of Engineering and Technology, 7(09), 166-175.
38. Bhuiyan, R. J., Akter, S., Uddin, A., Shak, M. S., Islam, M. R., Rishad, S. S. I., ... & Hasan-Or-Rashid, M. (2024). Sentiment analysis of customer feedback in the banking sector: A comparative study of machine learning models. The American Journal of Engineering and Technology, 6(10), 54-66.
39. Siddique, M. T., Jamee, S. S., Sajal, A., Mou, S. N., Mahin, M. R. H., Obaid, M. O., ... & Hasan, M. (2025). Enhancing automated trading with sentiment analysis: Leveraging large language models for stock market predictions. The American Journal of Engineering and Technology, 7(03), 185-195.
40. Sajal, A., Chy, M. S. K., Jamee, S. S., Uddin, M. N., Khan, M. S., & Gharami, A. K. & Ahmed, M.(2025). Forecasting Bank Profitability Using Deep Learning and Macroeconomic Indicators: A Comparative Model Study. International Interdisciplinary Business Economics Advancement Journal, 6(06), 08-20.
41. Akhi, S. S., Shakil, F., Dey, S. K., Tusher, M. I., Kamruzzaman, F., Jamee, S. S., ... & Rahman, N. (2025). Enhancing banking cybersecurity: An ensemble-based predictive machine learning approach. Am. J. Eng. Technol., 7(03), 88-97.
42. Hossain, S., Sajal, A., Jamee, S. S., Tisha, S. A., Siddique, M. T., Obaid, M. O., ... & Haque, M. S. U. (2025). Comparative analysis of machine learning models for credit risk prediction in banking systems. The American Journal of Engineering and Technology, 7(04), 22-33.
43. Hossen, M. A., BHATTACHARJEE, B., DEY, S., JAMEE, S., OBAID, M., MIA, M., ... & SHARIF, M. (2025). Business analytics for customer segmentation: A comparative study of machine learning algorithms in personalized banking services. International Journal of Economics Finance & Management Science, 10(03), 1-13.
44. Nath, P. C., Chy, M. S. K., Hossain, M. R., Miah, M. R., Jamee, S. S., Sharif, M. K., ... & Ahmed, M. (2025). Comparative Performance of Large Language Models for Sentiment Analysis of Consumer Feedback in the Banking Sector: Accuracy, Efficiency, and Practical Deployment. Frontline Marketing, Management and Economics Journal, 5(06), 07-19.
45. S. M. Rezvi, F. Mujtahid, K. R. Ahmed, A. Madhiya, V. SOLANKI and S. S. Jamee, "An Intelligent Multi-Source Data Fusion System for Sales Demand Forecasting in Retail and E-Commerce," 2026 Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS), Pathum

Thani, Thailand, 2026, pp. 1112-1117, doi: 10.1109/ICICDS70526.2026.11604912.

46. Ahmmed, M. J. Pritom Das, Tamanna Pervin, Sadia Afrin, Sanjida Akter Tisha, Md Mehedi Hassan, & Nabila Rahman.(2024). COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR BANKING FRAUD DETECTION: A STUDY ON PERFORMANCE, PRECISION, AND REAL-TIME APPLICATION. International Journal of Computer Science & Information System, 9(11), 31-44.

47. Ahmed, M. P., Arif, M., Chowdhury, M. S., Bhuiyan, R. J., Rahman, T., Ahmmed, M. J., ... & MAMUN, M. (2024). Comparative analysis of machine learning techniques for accurate lung cancer prediction. Am. J. Eng. Technol, 6, 92-103.

48. Akter, S., Mahmud, F., Rahman, T., Ahmmed, M. J., Uddin, M. K., Alam, M. I., & Jui, A. H. (2024). A comprehensive study of machine learning approaches for customer sentiment analysis in banking sector. The American Journal of Engineering and Technology, 6(10), 100-111.

49. Sweet, M. M. R., Arif, M., Uddin, A., Sharif, K. S., Tusher, M. I., Devi, S., ... & Sarkar, M. A. I. (2024). Credit risk assessment using statistical and machine learning: Basic methodology and risk modeling applications. International Journal on Computational Engineering, 1(3), 62-67.

50. Arif, M., Ahmed, P., Mamun, A. A., Uddin, K., Mahmud, F., Rahman, T., ... & Hossain, S. (2024). Dynamic pricing in financial technology: evaluating machine learning solutions for market adaptability. International Interdisciplinary Business Economics Advancement Journal, 5(10), 13-27.

51. Ahmed, M. P., Arif, M., Al Mamun, A., Mahmud, F., Rahman, T., Ahmmed, M. J., ... & Uddin, M. K. (2024). A comparative study of machine learning models for predicting customer churn in retail banking: insights from logistic regression, random forest, GBM, and SVM. Journal of Computer Science and Technology Studies, 6(4), 92-101.

52. Ahmmed, M. J. (2025). Systematic review and quantitative evaluation of advanced machine learning frameworks for credit risk assessment, fraud detection, and dynamic pricing in US financial systems. International Journal of Business and Economics Insights, 5(3), 1329-1369.

53. Rahman, M. M., & Ahmmed, M. J. (2026). AI-Driven Risk Analytics Models for Early Detection of Financial Noncompliance in Multi-Branch Banking Systems. American Journal of Data Science and Analytics, 7(04), 81-123.

54. Hossen, M. E. ., Akhter, A. ., Ghosh, S. ., Khandaker, M. ., Azam, M. N. ., Malek, H. A. ., Naher, K. ., & Bhuiyan, M. M. R. . (2026). Predicting Infectious Disease Outbreaks Using Machine Learning and Real-Time Epidemiological Data: Leverage Social Media, Environmental, And Public Health Data to Forecast Outbreaks Like Influenza, COVID-19, Or RSV. International Journal of Medical Science and Public Health Research, 7(02), 7–17. https://doi.org/10.37547/ijmsphr/Volume07Issue02-02

55. Umam, S., & Razzak, R. B. (2025, November). A 20-Year Overview of Trends in Secondhand Smoke Exposure Among Cardiovascular Disease Patients in the US: 1999–2020. In APHA 2025 Annual Meeting and Expo. APHA.7

56. Razzak, R. B., & Umam, S. (2025, November). Health Equity in Action: Utilizing PRECEDE-PROCEED Model to Address Gun Violence and associated PTSD in Shaw Community, Saint Louis, Missouri. In APHA 2025 Annual Meeting and Expo. APHA.8

57. Razzak, R. B., & Umam, S. (2025, November). A Place-Based Spatial Analysis of Social Determinants and Opioid Overdose Disparities on Health Outcomes in Illinois, United States. In APHA 2025 Annual Meeting and Expo. APHA.

58. Khan, M. S., Gharami, A. K., Nitu, F. N., Uddin, M. N., Ahmed, M., Roy, M. K., & Yezdani, S. (2025). Deep Learning-Driven Customer Segmentation in Banking: A Comparative Analysis for Real-Time Decision Support. International Interdisciplinary Business Economics Advancement Journal, 6(08), 9-22.

59. Sajal, A., Chy, M. S. K., Jamee, S. S., Uddin, M. N., Khan, M. S., & Gharami, A. K. & Ahmed, M.(2025). Forecasting Bank Profitability Using Deep Learning and Macroeconomic Indicators: A Comparative Model Study. International Interdisciplinary Business Economics Advancement Journal, 6(06), 08-20.

60. Siddique, M. T., Uddin, M. J., Chambugong, L., Nijhum, A. M., Uddin, M. N., & Shahid, R. & Ahmed, M.(2025). AI-Powered Sentiment Analytics in Banking: A BERT and LSTM Perspective. International Interdisciplinary Business Economics Advancement Journal, 6(05), 135-147.

61. Ayub, M. I., Gharami, A. K., Nitu, F. N., Uddin, M. N., Islam, M. I., Nijhum, A. M., ... & Yezdani, S. (2025). AI-driven demand forecasting for multi-echelon supply chains: Enhancing forecasting

accuracy and operational efficiency through machine learning and deep learning techniques. Emerging Frontiers Library for The American Journal of Management and Economics Innovations, 7(07), 74-85.

62. Islam, M. M., Sarkar, M. A. R., Ahmed, M., Moniruzzaman, S. M., & Uddin, M. N. (2009). Germination, vigour and emergence indicators of Corchorus olitorius L, seed and their relationship as influenced by seed sources. Bangladesh J. Jute and Fib. Res, 29(1&2), 1-8.

63. Uddin, M. N., & Aziz, M. M. (2026). Shapley Value-Guided Adaptive Ensemble Learning for Explainable Financial Fraud Detection with US Regulatory Compliance Validation. arXiv preprint arXiv:2604.14231.

64. Uddin, M. N. (2026). Explainable Graph Neural Networks for Interbank Contagion Surveillance: A Regulatory-Aligned Framework for the US Banking Sector. arXiv preprint arXiv:2604.14232.

65. Uddin, M. N. (2026). Does Founding Team Human Capital Heterogeneity Predict Venture Survival? Evidence from the Kauffman Firm Survey. Evidence from the Kauffman Firm Survey (April 04, 2026).

66. Uddin, M. N. (2026). Cost Efficiency Dynamics in Indian Public Sector Banks: A DEA-Based Analysis with Second-Stage Tobit Estimation and Post-Reform Comparative Evidence, 2009-2023. Available at SSRN 6680922.

67. Mottalib, M. Y., Yousuf, M., Bhuiyan, E. R., Habib, S. M., Dey, S. K., Gazi, M. S., ... & Anik, A. (2026). Machine Learning-Based Cyber Defense for Cloud Infrastructure: An Adaptive Deep Q-Network Architecture for Intelligent Intrusion Detection and Automated Threat Mitigation. arXiv preprint arXiv:2608.12190.

68. Tisha, S. A., Mottalib, M. Y., Bhuiyan, E. R., Anik, A., Ullah, S. W., Risalat, M., & Chockroborty, S. L. (2026). Artificial Intelligence-Driven Cybersecurity Framework for Enterprise Threat Detection: A Machine Learning Approach. The USA Journals TAJET, 8(07), 25-37.

69. Uddin, M. N., Bhuiyan, E. R., Risalat, M., Orpita, R. T., Anik, A., & Ullah, S. W. (2026, July). A FinCEN-Aligned Framework for Explainable Graph Neural Networks for Fraud Contagion Detection and Automated SAR Generation in US Interbank Payment Networks. In 2026 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS) (pp. 1-6). IEEE.

70. Uddin, M. N., Orpita, R. T., Bhuiyan, E. R., & Anik, A. (2026). ZK-SR117: A Chunked Zero-Knowledge Attestation Design for Aggregated Fair-Lending Metrics, with a Control Mapping toward Full SR 11-7 Coverage. arXiv preprint arXiv:2608.02664.

71. Mottalib, M. Y., Yousuf, M., Bhuiyan, E. R., Habib, S. M., Dey, S. K., Gazi, M. S., ... & Anik, A. (2026). Machine Learning-Based Cyber Defense for Cloud Infrastructure: An Adaptive Deep Q-Network Architecture for Intelligent Intrusion Detection and Automated Threat Mitigation. arXiv preprint arXiv:2608.12190.

72. Uddin, M. N., Orpita, R. T., Anik, A., Bhuiyan, E. R., Risalat, M., Ullah, S. M., & Ahamed, A. (2026). CT-HEG: A Bidirectional, Timestamp-Attributed Event Graph for ICU In-Hospital Mortality Prediction-An Architectural Ablation Study. arXiv preprint arXiv:2608.02663.

73. Uddin, M. N., Orpita, R. T., Bhuiyan, E. R., & Anik, A. (2026). ZK-SR117: A Chunked Zero-Knowledge Attestation Design for Aggregated Fair-Lending Metrics, with a Control Mapping toward Full SR 11-7 Coverage. arXiv preprint arXiv:2608.02664.

74. Uddin, M. N., Bhuiyan, E. R., Risalat, M., Orpita, R. T., Anik, A., & Ullah, S. W. (2026, July). A FinCEN-Aligned Framework for Explainable Graph Neural Networks for Fraud Contagion Detection and Automated SAR Generation in US Interbank Payment Networks. In 2026 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS) (pp. 1-6). IEEE.

75. Tisha, S. A., Mottalib, M. Y., Bhuiyan, E. R., Anik, A., Ullah, S. W., Risalat, M., & Chockroborty, S. L. (2026). Artificial Intelligence-Driven Cybersecurity Framework for Enterprise Threat Detection: A Machine Learning Approach. The USA Journals TAJET, 8(07), 25-37.

76. Mozumder, M. A. S., Mia, M. M., Nipa, R. A., Anik, A., Akter, S., & Choudhury, M. Z. M. E. (2026). A Hybrid Graph Neural Network and Large Language Model Framework for Insider Threat Detection via Behavioral Graph and Semantic Profiling. Colombo Scientific Journal of Computer Science & Information System, 11(05), 22-32.

77. Arif, M., Mozumder, M. A. S., Sajal, A., Jahed, K. A., Sharif, M. K., Anik, A., & Ahmed, M. (2026). Machine Learning–Driven Strategic Decision-Making: An Empirical Analysis of Employee Attrition Prediction Using Ensemble Models. The USA Journals TAJMEI, 8(04), 24-34.

78. Uddin, M. N., Orpita, R. T., Bhuiyan, E. R., & Anik, A. AI Compute Import Dependence and Dollar Invoicing: Cross-Country Evidence for a Compute-Dollar Association. Available at SSRN 7003871.