International Journal of Innovative Research in Engineering and Management (IJIREM)
ISSN (Online): 2350-0557, Volume 13, Issue 2, April 2026
DOI: https://doi.org/10.55524/ijirem.2026.13.2.13
Article ID IJIR3193, Pages 88-102
www.ijirem.org

Artificial Intelligence in Auditing: Transforming the Future of Accounting

Boomika G[1], Mamdha Sri G[2], Dr. Feon Jaison[3], Indrajith K S[4], and Rohit Sonawane[5]

[1]M.Com Scholar, Department of Commerce, JAIN (Deemed-to-be University), Karnataka, India
[2, 4, 5] MCA Scholar, Department of Computer Application, JAIN (Deemed-to-be University), Karnataka, India
[3]Assistant Professor, Department of Computer Application, JAIN (Deemed-to-be University), Karnataka, India

Correspondence should be addressed to Boomika G; jupg24mcoms21360@jainuniversity.ac.in

Received: 17 March 2026; Revised: 1 April 2026; Accepted: 14 April 2026

ABSTRACT- The increasing dependence of businesses on digital systems has contributed to an explosion in the number of financial transactions being carried out by organizations. This has necessitated a need to evolve auditing practices since the traditional methods have become obsolete due to their heavy dependence on manual testing and sampling. Thus, there is a need to adopt more efficient auditing processes. This research paper discusses the application of Artificial Intelligence (AI) to facilitate auditing. An intelligent system for analyzing anomalies using machine learning techniques like random forest is developed and applied to financial transactions. Transaction and behavioral data are used to detect any abnormal transactions. The model adopted is suitable for SMEs in India which lack access to sophisticated audit systems. The results obtained reveal that AI-based auditing processes help to minimize fraud detection time and improve the accuracy of the process. The system also facilitates continuous monitoring, which helps to mitigate potential risks early before they materialize. Overall, AI has helped to revolutionize auditing in order to cope with emerging digital financial trends.

KEYWORDS- Artificial Intelligence, Financial Auditing, Machine Learning, Anomaly Detection, FinTech, Digital Accounting.

I. INTRODUCTION

Accounting has always been closely tied to trust, especially when it comes to financial reporting. People like investors, regulators, and lenders depend on audited financial statements to guide their decisions, so the quality of auditing really matters. In practice, auditing has traditionally been a hands-on process. Auditors review documents, check samples of transactions, and rely heavily on their own experience to verify financial accuracy. With the rise of digital technologies, however, the way businesses operate has changed significantly. Companies now produce huge volumes of financial data through ERP systems, online transactions, and automated tools. This growth in data—both structured and unstructured—has made traditional auditing methods harder to apply effectively, particularly when trying to detect complex fraud or ensure compliance. Sampling-based approaches, which were once standard, are often no longer sufficient. Important irregularities can easily go unnoticed when only a portion of the data is examined. Today's auditors are expected to work with complete datasets, spot unusual patterns, and react quickly to potential risks. This shift has created a need for more efficient and scalable auditing methods.

Artificial Intelligence (AI) offers a practical way forward. Using machine learning and data-driven techniques, AI can analyze large datasets, highlight anomalies, and reveal patterns that might be missed through manual review. Unlike fixed rule-based systems, these models can adapt and improve over time as they learn from new information. Bringing AI into auditing also changes the overall approach to the audit process. Instead of relying only on periodic checks, organizations can monitor financial activities continuously and in near real time. This not only improves accuracy but also reduces the chances of errors or fraud going undetected. Recent studies also show a growing use of machine learning and data mining techniques for identifying suspicious financial behavior. These developments suggest that AI is not just an enhancement but a key factor in shaping the future of auditing, making it more responsive, data-driven, and efficient. The benefits of continuous auditing include improved financial statement transparency and accuracy as well as reduced likelihood of missed fraud and/or error.

More recent studies have demonstrated growing utilization of machine learning and data mining methodologies for identifying anomalous patterns within large databases of financial transactions that may indicate potential financial crimes or errors [7], [10].

II. LITERATURE REVIEW

There has been a noticeable increase in interest around the use of Artificial Intelligence (AI) in auditing, particularly within India. This interest spans academic researchers, industry professionals, consulting firms, and regulatory bodies. Existing studies suggest that AI can improve audit efficiency, enhance fraud detection capabilities, and support better compliance, although certain limitations and challenges are also acknowledged.

The Institute of Chartered Accountants of India (ICAI), through its Digital Auditing Framework, emphasizes that AI adoption in auditing should follow a structured and controlled approach. It highlights applications such as

continuous auditing, advanced risk assessment, and intelligent fraud detection systems, while also reinforcing that professional judgment remains essential and cannot be replaced by automated systems.

Bhattacharyya et al. [1] conducted a comparative analysis of data mining techniques for credit card fraud detection, demonstrating that machine learning approaches can significantly improve fraud identification when applied to financial datasets. Similarly, Ngai et al. [2] categorized fraud detection techniques into classification, clustering, and prediction models, emphasizing the importance of analyzing large volumes of financial data to improve detection accuracy.

Bao et al. [3] explored the use of machine learning in identifying accounting fraud within publicly traded companies. Their findings showed that predictive models built on financial statement data can effectively detect irregularities by capturing complex financial relationships. Hasan [4], in a comprehensive review, discussed how technologies such as machine learning, natural language processing, and data analytics contribute to improving audit efficiency and automation, while also highlighting concerns related to ethical issues and workforce skill gaps.

Chen et al. [5] reviewed deep learning approaches for fraud detection and concluded that models such as convolutional neural networks (CNNs) and long short-term memory (LSTM) networks are capable of identifying complex fraud patterns in financial transactions. Azzahra [6] also found that AI enables auditors to process large volumes of financial data efficiently and detect anomalies that traditional methods may overlook, ultimately improving audit quality.

Further contributions by Phua et al. [7] emphasize the importance of anomaly detection techniques in financial fraud detection systems, while Hand and Whitrow [8] proposed alternative evaluation metrics for assessing classifier performance beyond traditional ROC-based measures.

Carcillo et al. [9] introduced a scalable streaming framework for real-time fraud detection using big data technologies such as Apache Spark, highlighting the growing importance of real-time analytics in financial systems. Bolton and Hand [10] provided a detailed overview of statistical fraud detection methods, emphasizing behavioral modeling and anomaly detection techniques as key approaches in identifying suspicious activities.

Bahnsen et al. [11] investigated cost-sensitive decision tree models, showing how classification techniques can be adapted to minimize financial losses caused by fraud. In addition, Bahnsen et al. [11] and Chawla et al. [12] addressed the issue of class imbalance, a common challenge in fraud detection datasets where fraudulent cases are significantly outnumbered by legitimate transactions.

III. METHODOLOGY

This study adopts a quantitative research approach to assess how artificial intelligence (AI), particularly the Random Forest algorithm combined with behavioral analysis, can improve auditing processes and strengthen fraud detection in accounting systems.

Quantitatively combining machine learning with audit analytics, this research methodology is both empirically and experimentally designed. Using Supervised Learning Methods, financial transaction data and audit-related indicators are processed. The study aims to evaluate whether AI-based models can enhance audit accuracy, improve risk assessment, and strengthen fraud detection compared to traditional auditing methods.

Structured Transactional Datasets with labelled fraud indicators were created and tested with new AI-Assisted Auditing Models. Past transaction records form the dataset. All transaction records are individual financial transactions that have been processed within a corporation's financial environment. Key fields in the dataset are transaction identifier, transaction timestamp, transaction amount, merchant/account type, geographical/location based info., and user/customer data. In addition to these transaction fields, the dataset contains behavioral and contextual data regarding device usage patterns, transaction frequencies/rates, and prior spending behaviors.

There are also binary fraud labels present in the dataset that aid in supervised machine learning. The fraud labels are assigned to all financial transactions as either valid or fraudulent. The dataset is large and diverse enough to reflect real-world financial activity, capturing variations in transaction frequency, seasonal trends, and user behavior patterns.

Secondary data was obtained from corporate financial statements, transactional data, and comprehensive audit trail data to design and test the proposed AI-based auditing model. The dataset consists of organized financial transaction data that includes transaction amount, transaction date, account type involved, and journal entry information that will serve as the basis for identifying anomalies/detecting risk.

User activity data is also included in the dataset, providing metadata related to login dates, system access history, and modifications made to accounting entries. In addition, historical fraud data is incorporated to train supervised machine learning models to identify patterns associated with fraudulent activities.Also, control override reports and exceptions documentation are reviewed to identify deviations from internal controls that may indicate anomalies.

Additional contextual information for the analysis is provided through inclusion of behavioral data that will enable identification of anomalies in user behavior such as late night postings, repeated editing of transactions, unusual approval patterns, and rapid changes in normal transaction patterns.

Prior to developing the model, the gathered dataset is systematically pre-processed for quality, consistency, and appropriateness for analysis. Systematic data cleansing operations are performed to remove redundant data, resolve inconsistencies and eliminate non-relevant data. Fraud detection datasets typically show a strong class imbalance, where fraudulent transactions make up only a small fraction of the overall data.Techniques such as synthetic minority oversampling are frequently utilized to enhance detection of minority fraud classes [13].

Missing data is addressed via appropriate imputation techniques depending upon variable type and distribution. The categorical variables of account types, transaction types and user roles are transformed into numerical representations for machine learning compatibility.

Additionally the numerical variables are normalized to achieve equivalent scaling.

Feature selection methodologies are implemented to isolate those variables that significantly impact detection. By utilizing feature selection methodologies the dimensionality of the model is reduced thereby increasing the efficiency of the model. Analysis of time features such as time intervals between transactions and sequence patterns are examined to discern user patterns. Normalizing facilitates improved convergence rates of machine learning models resulting in increased stability of the model. The preprocessing operation ensures the data is suitably structured and free from bias. The preprocessing operation minimizes biased modeling toward variables having greater magnitude values.

Following preprocessing feature engineering is accomplished to generate advanced behavioral/risk focused variables to enhance detection performance through the identification of anomalous patterns. Examples include a transaction frequency deviation score to determine unusual variances in posting behavior a user activity anomaly index to determine unusual variation from previously observed behavioral trends an override frequency ratio to identify frequent control bypass override occurrences and sudden transaction value spike variables to identify unusual financial activity. Generation of complex variables from unprocessed accounting/user activity data enables the model to identify both financial anomalies and unusual behavioral trends.

Random Forest was chosen for model development because it offers reliable performance, strong predictive accuracy, and the ability to efficiently handle large and high-dimensional financial datasets. The algorithm works by constructing multiple decision trees and combining their outputs to produce a final classification, which helps reduce the risk of overfitting.

For evaluation, the dataset is split into training (70%) and testing (30%) subsets to ensure an unbiased assessment of the model's performance. The model is then trained to categorize financial transactions into predefined risk levels using engineered financial and behavioral features.

To fine-tune the model, cross-validation was employed to systematically identify the best-performing combination of key configuration settings, including the total number of decision trees, the allowed depth of each tree, and the minimum number of samples required to split a node. This iterative tuning process ensured the model struck an appropriate balance between underfitting and overfitting on the training data.

Random Forest also offers a means of determining feature importance which permits an understanding of the most influential risk factors impacting the classification outcomes for audit tasks which would be beneficial for auditing purposes. Given their proven capacity to uncover hidden patterns and interdependencies across financial datasets, a range of machine learning techniques — from decision trees to ensemble and statistical classifiers — have become standard tools in fraud detection research and practice [7], [10], [15].

The trained Random Forest model was assessed against a set of well-established classification metrics widely referenced in fraud detection research [8], [13]. These include accuracy, precision, recall, F1-score, and a detailed confusion matrix breakdown, each chosen to capture a distinct dimension of model performance.

Accuracy reflects the share of all financial transactions that the model assigned to the correct class out of the total transactions evaluated.While accuracy provides insight into the overall performance of the model it does not provide sufficient detail about effectiveness for fraud detection issues in which fraudulent transactions constitute a very small portion of total transactions.

Precision denotes the degree to which the model's fraud predictions are accurate — that is, among all transactions it labels as fraudulent, how many actually are. In auditing environments, this metric holds practical significance: when the model generates excessive false positives, auditors are compelled to investigate transactions that pose no real risk, consuming valuable time and degrading overall workflow efficiency.

Recall, alternatively known as sensitivity, gauges the model's ability to surface the full scope of fraud present in the dataset. Within an audit setting, this metric is arguably the more critical of the two — undetected fraud cases translate directly into financial losses and can seriously undermine stakeholder confidence in the organization's internal controls.

Rather than evaluating precision and recall in isolation, the F1-score synthesizes both into one composite figure through their harmonic mean, making it especially useful when a single balanced indicator of detection quality is needed.Therefore, if a high precision and/or high recall exists for a model then a high f1-score exists indicating good performance. Since fraud detection datasets are typically characterized as being severely class-imbalanced (i.e. fraudulent transactions occur infrequently compared to legitimate transactions), f1-scores offer an excellent way to assess performance when evaluating models intended for detecting rare events like fraud.

A classification outcome matrix provides a granular breakdown of how the model performed across all samples, capturing four key outcomes — true positives, true negatives, false positives, and false negatives. Examining these four values together reveals not just whether the model is accurate overall, but precisely where and how it errs. This error-level insight opens the door to targeted improvements, whether through threshold adjustments, feature refinement, or a shift in the decision boundary to better separate fraud from legitimate activity.

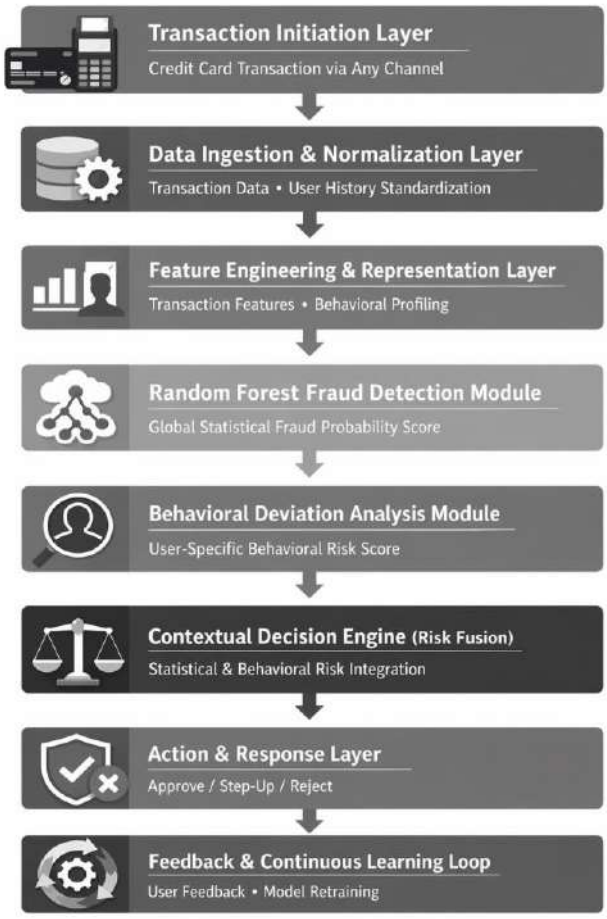

Figure 1: Flowchart of the Proposed AI-Based Auditing System

IV. SYSTEM ARCHITECTURE

The architectural concept for this proposed system will be presented as part of the framework for conducting real-time transaction level fraud detection. This represents one critical area of emphasis within contemporary digital auditing. Real time fraud detection will be facilitated through the utilization of artificial intelligence. The artificial intelligence that will be utilized will include a random forest classifier along with behavioral deviation analysis. The layers depicted within Figure 1 are representative of a layer based structure designed to provide a defined format for data flow.

A. Transaction Initiation Layer

Once a cardholder initiates a transaction through any accepted payment channel — whether a physical POS terminal, a web-based payment gateway, or a mobile application — the system activates a layered evaluation sequence that ultimately determines whether the transaction is approved or declined, guided by the assessed risk level associated with that user. During this phase, the system collects all available data related to the transaction in order to create a complete record of each transaction. At no stage during this level does the system apply any logic that relates to detecting fraudulent activity rather its focus is to collect accurate and reliable information about transactions securely and with minimal latency so as not to hinder the timely authorization of the transaction.

The type and amount of data typically captured include:

- Amount of Transaction,
- Date and Time of Transaction,
- Merchant Identifier (e.g., unique identifier assigned to a specific business),
- Merchant Category Code (MCC),
- Geographic Location of Transaction Originated From,
- Terminal/Device Information,
- Payment Channel Details (e.g., how transaction was initiated).

B. Data Ingestion And Normalization Layer

At its core, the data ingestion layer functions as the system's unifying data hub, pulling in both transactional records and supporting data from external sources such as customer history repositories, merchant risk registries, and device or IP reputation services, consolidating them for further analysis.

Since inputs arriving from heterogeneous systems inevitably vary in their schema, formatting conventions, and levels of completeness, this layer undertakes the necessary work of harmonizing and standardizing all incoming data before it proceeds further through the pipeline. Data cleaning, format harmonization, missing data treatment, and data schema consistency are undertaken to create a standardized dataset. This ensures that the analytical modules receive consistent and validated data inputs.

C. Feature Engineering and Representation Layer

Beyond normalization, this layer transforms raw transactional records into well-defined numerical representations suitable for machine learning input. The feature engineering process is organized across two distinct dimensions — transaction-centric feature construction, which focuses on short-term statistical signals within individual transactions, and behavioral profiling, which builds longer-term spending patterns unique to each cardholder — both aimed at sharpening the predictive accuracy of the Random Forest classifier. The Features of Transaction-Centric Feature Engineering are Short-Term Statistical Anomalies as follows Deviation from Mean Transaction Value, Transaction Rate, Merchant Risk Scores, Geographic Proximity to Previous Transactions and Time Based Anomalies. Behavioral Profiling develops Long-Term Normal Spending Behavior Profiles of Each Cardholder. The Profile includes Preferred Types of Merchants, Typical Locations of Transaction, Ranges of Transaction Values, Times of Transactions, and Consistent Device Usage Patterns. In contrast to conventional fixed-rule systems, the cardholder behavioral profile is kept current by incorporating each new legitimate transaction, allowing risk evaluations to remain personalized and adaptive. Permutation importance analysis was then employed to measure how significantly each input feature influences the model's predictive output, offering interpretable insight into the key drivers of fraud classification. We use Permutation Importance as a method of determining the importance of each Feature because we want to know how the models performance changes when the value of each Feature is randomly changed. The more accurate the model was before we permuted the values of a particular Feature the higher the Permutation Importance will be for that Feature.

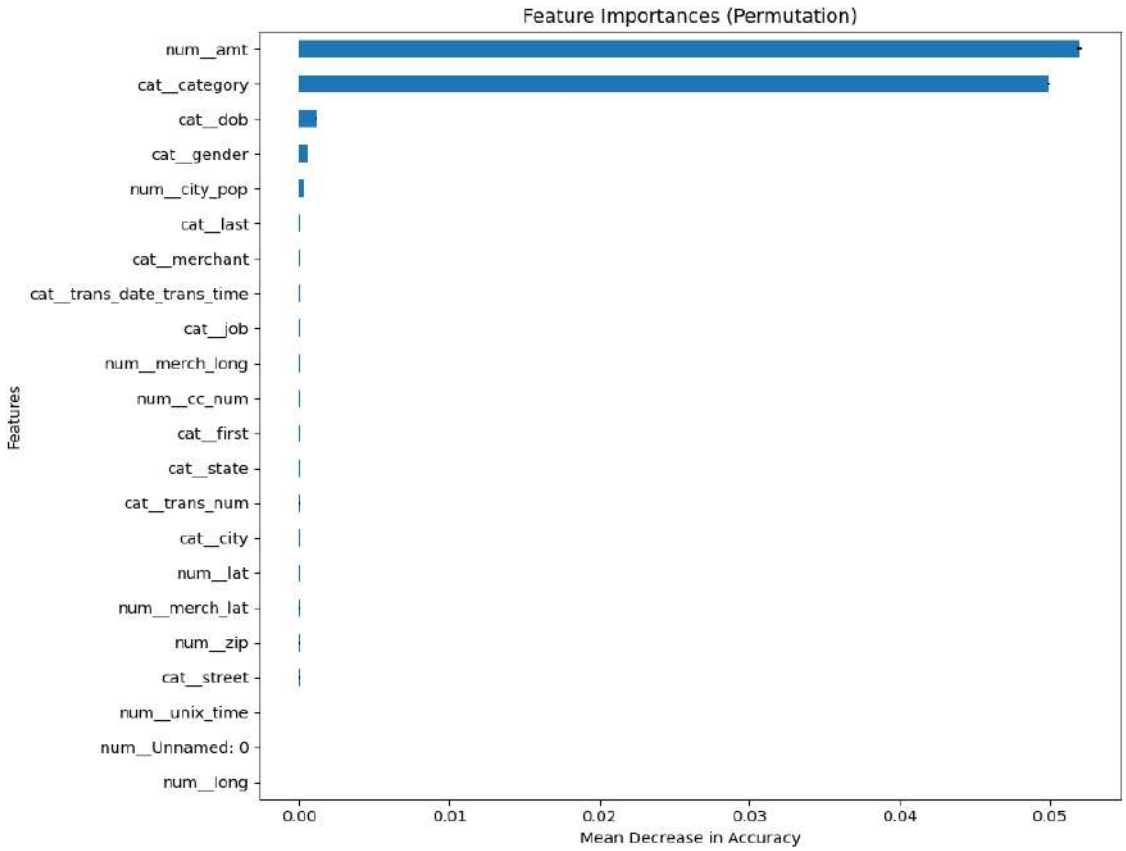

Figure 2: Permutation-based feature significance scores of the Random Forest classifier

As illustrated in Figure 2, transaction amount (num__amt) emerges as the single most influential predictor, recording the highest mean decrease in accuracy at approximately 0.052 when its values are randomly permuted. This tells us that the Transaction Amount is an extremely Important Factor when distinguishing between legitimate transactions and those that are fraudulent. One possible explanation for why the Transaction Amount has such high Permutation Importance is that many fraudulent Transactions involve abnormal Transaction Amounts. The next Most Important Feature, with a moderate to High Level of Importance, is cat__category (Category of Transaction). These Categories of Transactions appear to have a greater likelihood of being part of a fraudulent Activity. Beyond these two leading variables, the remaining features collectively exert negligible influence on the model's classification outcomes, contributing little meaningful predictive signal. As a result, Features that include Demographic Information (cat__dob, cat__gender, num__city_pop) or General Location-Based Information (cat__dob, cat__gender, num__city_pop) are virtually insignificant in helping the model predict whether a case is fraudulent. Similarly, Customer-Related Information (cat__first, cat__last, num__cc_num) and Geographical Information (num__lat, num__long) as well as Transactional Information (cat__trans_num, num__unix_time) are nearly completely unimportant in the Identification of Fraudulent Cases. In addition, since the Model primarily Relies upon a Small Subset of Features (i.e., mainly num__amt and cat__category) instead of Features Specific to Individual Transactions (i.e., cat__trans_num, num__unix_time) and not Personal or Static Customer Related Features (i.e., cat__first, cat__last, num__cc_num), this is also quite beneficial in terms of Fraud Detection Systems. Patterns of Fraudulent Behavior are typically determined based on Transaction Patterns rather than the Identity of the User.

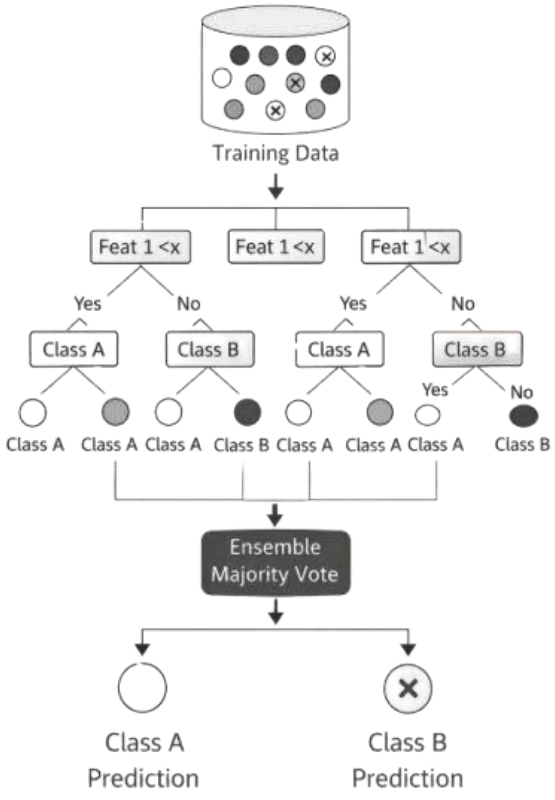

Figure 3: Decision flow of the fraud classification process.

D. Random Forest Fraud Detection Module

A demonstration of the random forest methodology's ability to utilize permutation importance is shown in the above example. As such, this demonstrates that the random forest model can identify those variables that contribute the greatest amount of value towards fraud identification. However, these features may also be modified to become more informative. Non-informative variables may also be combined or temporal patterns may be incorporated within the variable(s). Each modification would provide additional knowledge to the model relative to identifying characteristics associated with fraudulent transactions.

Figure 3 illustrates how the fraud detection system utilizes its classification results from the Random Forest Model. First, the system receives user input (the transaction), and performs feature extraction. Following feature extraction, the features are processed through the Random Forest Model for classification. After obtaining a classification from the Random Forest Model, this classification is integrated with the Behavioral Analysis to produce a Total Risk Assessment.

The final Total Risk Assessment is then used as the basis for decision-making (Approval, Verify, Reject). The randomly generated transaction attributes are then used as inputs to the Random Forest Classification Algorithm. This algorithm is primarily used to classify potential fraudulent transactions. The selection of Random Forest is due to its capabilities to process noisy financial data handle complex/nonlinear relationships and multiple data types. For each transaction, the Random Forest Model will generate a Fraud Probability Score. Additionally, the Feature Importance Scores will be produced to increase both the interpretability and auditability of the output. This module is generally used for extracting Global Fraud Patterns for all users.

E. Behavioral Deviation Analysis Module

Although the Random Forest model detects statistical anomalies at the population level, the behavioral deviation analysis module analyzes the transaction anomaly at the individual user level. The current transaction is then compared to the user's past behavioral pattern to identify inconsistencies in the context.

The behavioral deviation is determined by the differences in the transaction amount, geographic location, device type, merchant category, and time of the transaction. The result is a behavior risk score that represents the level of anomaly of the transaction for the particular user.

F. Contextual Decision Engine

The Contextual Decision Engine uses both of the Risk Scores (Fraud Risk Probability Score from Random Forest & Behavioral Risk Score) to make an overall Fraudulent Transaction decision. This Layer differs significantly from the Layer above it in that this Layer does a form of "Risk Fusion" based on context as opposed to simply using Ensemble Voting.This Layer's function is to strike a balance between how Statistically Suspect a transaction is and how Behaviorally Consistent a customer is when determining if they are willing to allow the transaction or if you need to verify the transaction.

G. Action and Response Layer

Based on the final risk score, the system responds with an appropriate action in real-time. These actions could be the approval of a transaction, step-up authentication (such as OTP or biometric verification), rejection of a transaction, or referral for manual processing. The result is sent back to the payment gateway with very tight latency requirements to ensure a seamless user experience and security.

H. Feedback and Continuous Learning Loop

To ensure long-term effectiveness, the system also has a feedback and continuous learning loop. Feedback comes from customer confirmations, chargeback notifications, and reviews from fraud analysts. This data is used to periodically retrain the Random Forest model. The feedback loop helps the system adapt to new patterns of fraud and new user behavior, ensuring that the system is not vulnerable to new attack methods.

I. Random Forest

Random Forest is a supervised learning algorithm built on the principle of aggregating predictions from multiple independent models, producing a collective output that is consistently more reliable and accurate than what any single model could achieve on its own. It does this by building thousands of decision trees and then having each individual decision tree vote for its predicted resultWhichever class accumulates the greatest number of tree votes is taken as the model's final prediction. This flexibility makes Random Forest applicable across a wide range of tasks, from estimating continuous numerical outcomes to categorizing discrete classes. For example it can be used to detect fraud in audit records. Decision Trees are tree-like models that classify data through splitting of data based upon characteristics (features).

Though decision trees are straightforward to build and interpret, this simplicity comes at a cost — they are particularly prone to overfitting, capturing noise in the training data rather than generalizable patterns. Overfitting occurs when a decision tree works perfectly with the training data but performs poorly with real-world test data. Random Forest eliminates some of the issues associated with overfitting by utilizing two methods: Bootstrap Aggregation and Random Feature Selection.

Bootstrap Aggregation works by repeatedly drawing samples with replacement from the original dataset, yielding a collection of diverse training subsets that introduce controlled variability across the ensemble. Each tree learns from a different bootstrapped draw of the original data, ensuring the ensemble captures a wider range of patterns than any individual tree could alone. This built-in variation not only diversifies the collective decision-making process but also acts as a natural safeguard against model bias. Random feature selection contributes additional variability among the trees by randomly selecting a subset of the features at each node in the decision tree. This method prevents any one feature from dominating the results across all decision trees and allows Random Forest to consider alternative views of the data.

In the context of auditing and fraud analysis, Random Forest is very useful because financial data is often noisy, nonlinear, and diverse. The algorithm's versatility extends to handling mixed numerical and categorical inputs, performing efficiently at scale, and detecting subtle non-

linear interactions between transaction-level and behavioral features that simpler models would likely overlook. In addition, it is able to generate feature importance scores, which are useful in determining the most important risk indicators for fraud analysis.

For each transaction evaluated, the classifier generates a probability score indicating how strongly the model associates that transaction with fraud or legitimacy, rather than simply assigning a fixed yes-or-no label. The result is probabilistic and can be easily integrated with risk-based decision engines, which can set the threshold for decision-making based on organizational risk tolerance.

V. RESULTS AND ANALYSIS

Random Forest Classification algorithm was created in python utilizing scikit learn to evaluate the fraud-detection data-set that had 1,296,675 transactions and 23 variables. The "fraud" is a target variable for this binary classification problem. In other words it will be either 0 (legitimate transaction) or 1 (fraudulent transaction). Due to the fact that there are very few fraudulent transactions (approximately 0.58%), an excessive amount of caution was used when attempting to resolve class imbalance problems.

Before building a model, all non-predictive idenfifier variables (i.e., Unnamed: 0; trans_num; first; last; street) were removed from the dataset to prevent noise and privacy concerns. All categorical variables (merchant; category; gender job) were encoded correctly. Temporal variables (trans_date_trans_time unix_time) were transformed to obtain relevant temporal characteristics (hour of transaction; day of week) associated with each transaction. A spatial risk indicator was additionally engineered by computing the physical separation between the cardholder's registered coordinates (lat, long) and the merchant's geographic position (merch_lat, merch_long), capturing location-based anomalies that may signal fraudulent activity.The data was divided into two separate sets. A Training Set consisting of approximately 70% of the original dataset and a Testing Set consisting of approximately 30% of the original dataset. The random forest classifier was built using the following parameters:

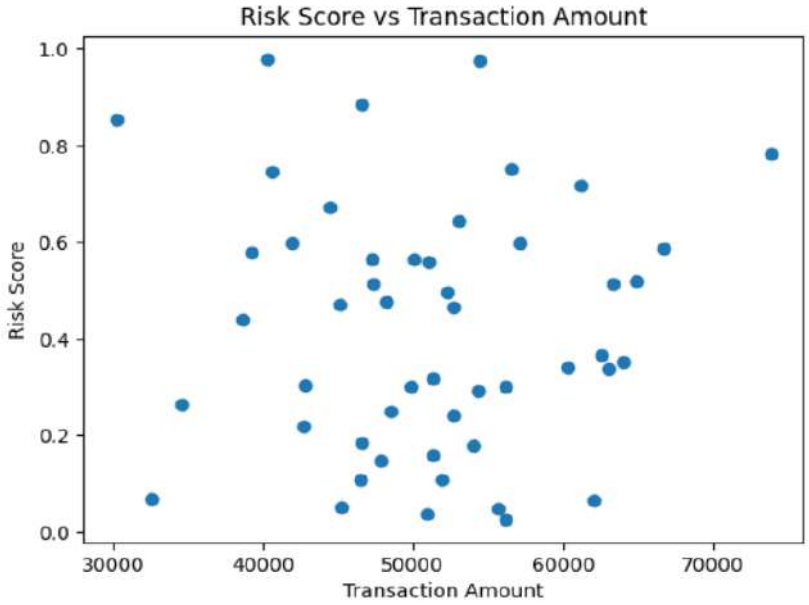

Figure 4: Distribution of normal and fraudulent transactions

- Number of trees (n_estimators) = 200

A tree count of 200 was chosen in proportion to the dataset's large size, ensuring the ensemble maintains consistent predictions while keeping variance low.

- Maximum depth (max_depth) = 15

Depth was capped to avoid overfitting but still allow enough complexity to capture the nonlinear patterns of fraud.

- Splitting criterion = Gini Index

The Gini impurity index was used to determine the best split of features in each decision tree.

- Feature selection method = Square root of total features (max_features = sqrt(n_features))

This adds a degree of randomness to the selection of features at each split, improving the overall generalizability of the model.

- Class weight = “balanced”

Because fraud is a rare event, class weights were used to assign a greater penalty to misclassifying the minority class of fraud instances.

Once the model was trained, feature importance analysis revealed that the strongest predictors of fraud were transaction amount (amt), merchant category (category), the physical distance separating the buyer and seller, and supplementary temporal and behavioral variables drawn from the cardholder's transaction history.

Auditing Performance Analysis/Assessment System

This article will assess the auditing performance of an artificial intelligence (AI)-based auditing system through a multi-dimensional auditing performance assessment structure.

Because this data set contains 1,296,675 transactions with only approximately .0058 of those transactions identified as fraudulent, auditing performance analysis of the system will take into consideration both the total accuracy of the

system in identifying fraudulent purchases and the ability to recognize instances of minority classes. Auditing performance will be assessed against three major criteria: risk sensitivity, auditing efficiency, and overall accuracy. Auditing performance analysis will occur on the test set (i.e., 30 percent split).

The landscape of fraud detection in digital finance has shifted markedly from its origins in manual review and rigid rule-based logic. Modern detection frameworks instead apply machine learning to scrutinize individual transactions at scale, extracting subtle and intricate patterns that reliably distinguish fraudulent behavior from legitimate activity. In contrast, this paper compares and contrasts traditional methods of fraud detection

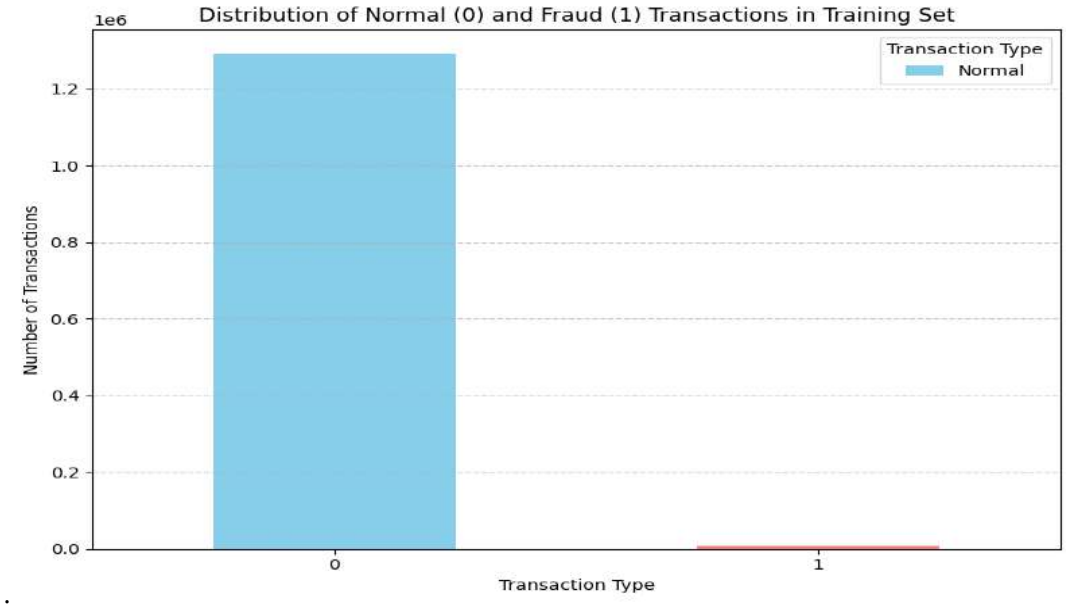

Figure 5: Scatter Plot of transaction class distribution

The figure above illustrates the proportion of normal (Class 0) versus fraudulent (Class 1) records present within the training data, highlighting the stark disparity between the two classes. A clear example of an extreme case of class imbalance exists here. Based on the figure above, it appears there are nearly one-and-one-quarter-million Class 0 (Normal) transaction records. Conversely, there are relatively few Class 1 (Fraudulent) transaction records (i.e., nearly negligible).

When one class so heavily dominates the training data, the model naturally gravitates toward predicting that class for every new input. Since Class 0 records vastly outnumber Class 1 instances during training, the model risks collapsing into a state where it classifies virtually all transactions as legitimate, regardless of the underlying signals pointing toward fraud. As a result of this type of bias, the model will predict Class 0 (Normal) for virtually all new records and fail to predict any fraudulent activity at all. Therefore, the model produces large numbers of true positives and fails to produce any false negatives.

The use of a technique specifically designed to handle cases of class imbalance is justified based on the visualization. There are several types of techniques available to help address issues with class imbalance Practical remedies include resampling the minority class upward, trimming the majority class downward, synthesizing additional fraud instances via SMOTE, or incorporating cost-sensitive learning penalties. Critically, performance should be gauged using Precision, Recall, F1-Score, and ROC-AUC rather than accuracy alone, as the latter masks poor fraud detection performance behind the overwhelming proportion of legitimate transactions.

- *Methodological Differences*

Fraud detection with traditional methods relies on both predetermined rules and thresholds to identify which transactions may be fraudulent. This includes identifying potential fraud based upon conditions that have not been satisfied (i.e., a transaction amount has exceeded a previously determined threshold or there is some level of unusual activity in relation to an individual's account). Traditional fraud detection techniques typically rely almost exclusively on previous fraud experience along with manual updates to pre-existing fraud prevention rules.

As opposed to using solely predefined threshold-based metrics for identifying potentially fraudulent behavior AI-based fraud detection utilizes supervised machine learning algorithms (such as random forest classifiers) trained on historical data sets.

Table 1: Traditional vs. AI-Driven Fraud Detection: A Comparative Overview

Dimension	Traditional Fraud Detection	AI-Based Fraud Detection
Detection Mechanism	Rule-based thresholds	Machine learning classification
Data Coverage	Sample-based or filtered review	Full dataset analysis
Pattern Recognition	Linear and predefined patterns	Nonlinear and dynamic patterns
Adaptability	Manual rule updates required	Model retraining enables adaptation

Behavioural Analysis	Limited or absent	Integrated behavioural profiling
Scalability	Limited by manual capacity	Highly scalable across large datasets
Risk Assessment	Binary flagging system	Probabilistic risk scoring
Audit Approach	Reactive	Predictive

The artificial intelligence (AI) fraud classifier created using supervised learning techniques utilized the transactional data set to create the classification result for all testing data sets.

The AI fraud classifier examined transactional attributes such as amount, date/time, transaction type, geographic location, and the behavioral attributes of the customer to generate a probability score representing the likely occurrence of fraud.

Based upon this probability score against the predetermined fraud classification threshold the AI fraud classifier categorized the observation as either legitimate or fraudulent.

In addition to these classifications the AI fraud classifier produced four types of output results:

Correctly identified as fraudulent (true positive): this represented customers who had been identified as fraudulent via their spending patterns and/or other attributes that indicated an abnormal behavior of spending.

Correctly identified as legitimate (true negative): this represented customers whose spending behavior did not exhibit characteristics of fraudulent activity and therefore were identified as legitimate.

Incorrectly identified as legitimate when actually fraudulent (false negative): this represents customers who exhibited behaviors and spending habits that indicate they are likely committing fraud but were incorrectly labeled as being legitimate. Incorrectly identified as fraudulent when actually legitimate (false positive): this represents customers whose spending behaviors and attributes do not indicate a high risk of fraudulent activities but were incorrectly identified as having engaged in those behaviors.

A confusion matrix was developed from the above output results to provide a summarized view of how many of the legitimate observations were correctly/incorrectly classified versus how many of the fraudulent observations were correctly/incorrectly classified. As part of the ROC curve analysis, several key performance indicators — including true positive rate (TPR), false positive rate (FPR), precision, recall, and F-score — were derived to provide a comprehensive assessment of how reliably the AI classifier distinguished between legitimate and fraudulent transaction records. The ROC curve mapped the interplay between TPR and FPR at each possible threshold, revealing how the model's sensitivity and specificity shift as the decision boundary moves. Precision, recall, and F-score were then computed by benchmarking the model's predicted classifications against the actual known labels in the dataset.

Traditional rule-based systems generally lack the capability to effectively recognize sophisticated fraud schemes in vast amounts of financial data. Consequently, financial transaction fraud detection has emerged as one of the most prominent domains for machine learning adoption, driven by these algorithms' capacity to sift through massive datasets and uncover the subtle behavioral signatures that distinguish fraudulent actors from legitimate users [7], [10].

Among the machine learning approaches applied to fraud classification tasks, four models have received considerable attention — Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines (SVM) — each bringing distinct strengths to the problem of distinguishing fraudulent transactions from legitimate ones. These models are unique in how they learn from data, how well they perform at predicting future values, and their ability to deal with the complexities of financial data structures. Research indicates that using ensemble learning approaches and advanced data mining techniques can lead to substantially improved fraud detection accuracy when compared to traditional statistical methodologies [7],[9].

To rigorously assess the fraud detection capability of each candidate algorithm, this study benchmarked multiple machine learning models against one another using a shared preprocessed dataset, ensuring that observed performance differences stem from the algorithms themselves rather than data inconsistencies. Model effectiveness was gauged across five evaluation dimensions — Accuracy, Precision, Recall, F1-Score, and AUC-ROC — each of which is well established in the fraud detection literature as a reliable indicator of classifier quality [8], [13].

Logistic Regression was selected as the base-line classification model because it is simple, interpretable and provides fast processing times. At its core, the logistic regression model computes a fraud likelihood score for each transaction by applying a weighted linear combination of its input features. While its computational lightness and output interpretability make it an attractive baseline choice, the model's inherently linear decision boundary limits its capacity to capture the non-linear and often complex patterns that characterize fraudulent financial behavior.

Decision Trees use a form of non-linear decision making where the feature space is recursively divided into subsets based on the most important features. Decision Trees are both easy to understand and can represent complex relationships between variables. Unfortunately, while single Decision Trees offer many advantages, they also tend to over fit and produce unreliable results when trained on large financial datasets.

Rather than depending on a single decision tree, Random Forest pools predictions across an entire ensemble, a process that simultaneously reduces overfitting risk and bolsters result consistency. Its design also embeds two built-in safeguards — bootstrapped training subsets and randomized feature selection at each node — that together promote diversity across the ensemble and guard against variance-driven errors. Due to these properties, Random Forest is one of the best choices for fraud detection applications requiring large complex datasets.

Beyond Random Forest, the study also evaluated the Support Vector Machine (SVM), which identifies the decision boundary that most cleanly separates fraudulent from legitimate transactions in the feature space. SVMs handle high-dimensional inputs well and leverage kernel transformations to capture non-linear relationships between variables. Their primary drawback, however, is computational — training on datasets of the scale examined here places heavy demands on both memory and CPU resources, raising practical concerns about scalability.

A head-to-head comparison of Logistic Regression, Decision Tree, Random Forest, and SVM was carried out under uniform experimental conditions, with all four models sharing the same training and test data, thereby guaranteeing that the resulting performance differences reflect genuine algorithmic distinctions rather than data handling inconsistencies. Comparison was based on six factors: False Positive Rate, False Negative Rate, Overall Discrimination Power and three aspects related to the specific characteristics of fraud detection performance due to the unbalanced nature of fraud datasets.

Logistic Regression was chosen as the basic linear classifier because it assumes linear separability between legitimate and fraudulent transactions. Even though it is easy to implement and fast in computation time however, it has poor capability to model complicated non-linear relationships among transactional variables and behavioral variables. Decision Trees demonstrated non-linear decision-making capability using recursive partitioning of feature space. One issue with single Decision Trees is their tendency to vary with training data and settings of tree depth.

Four classifiers — Logistic Regression, Decision Tree, Random Forest, and a Stacking ensemble — were benchmarked against one another within the fraud detection classification context, with their relative strengths and weaknesses assessed across five performance dimensions: Accuracy, Precision, Recall, F1-score, and AUC-ROC. Accuracy measures total percentage of correct classifications produced by classifier. Four models had approximately 99% - 100% accuracy rate. It means that four models correctly classified almost all transactions in the dataset. However, since fraud datasets are usually unbalanced, Accuracy is not suitable measure for evaluating fraud detectors. A classifier can be highly accurate but still fail to detect fraudulent transactions. Precision measures ratio of true fraudulent transactions detected by a classifier versus all transactions that classifier identifies as fraudulent. Precision scores drew a clear distinction between the models — the Stacking Classifier topped the rankings at close to 100%, with Logistic Regression closely behind at roughly 99%. Random Forest crossed the 90% threshold, whereas Decision Tree fell short of it, translating into a noticeably higher false alarm burden compared to its counterparts.

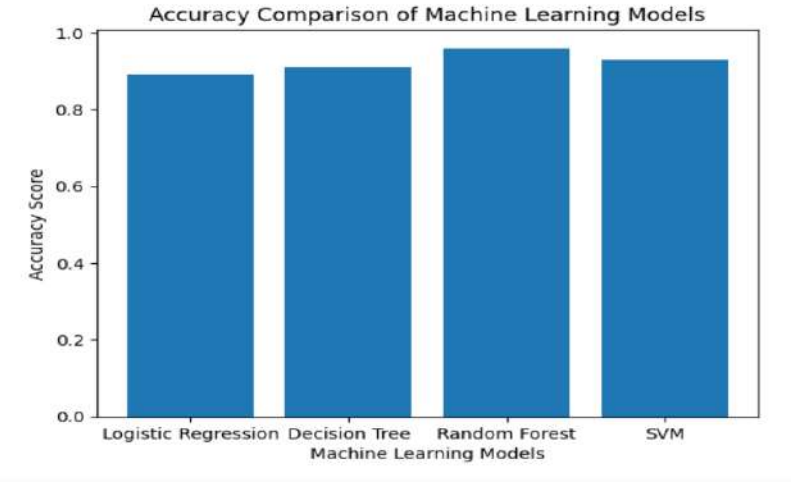

Figure 6: Comparison of machine learning model performance

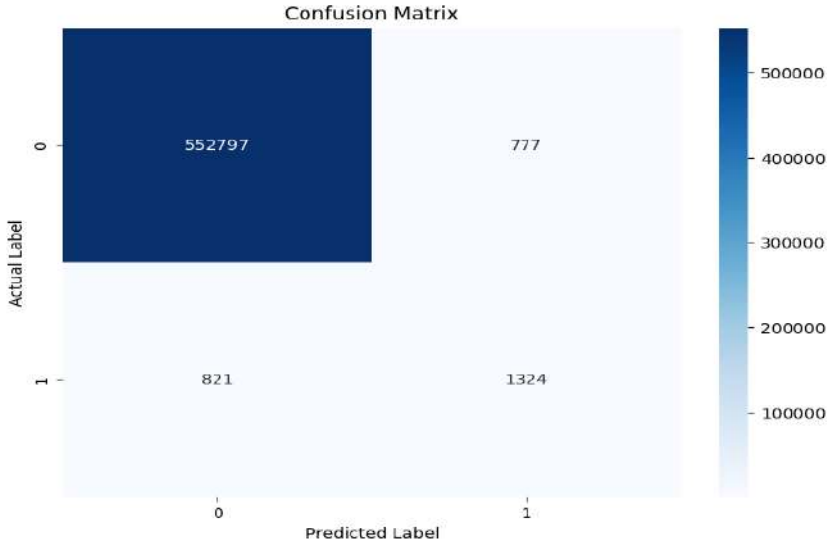

Figure 7: Classification result distribution matrix for the Random Forest model

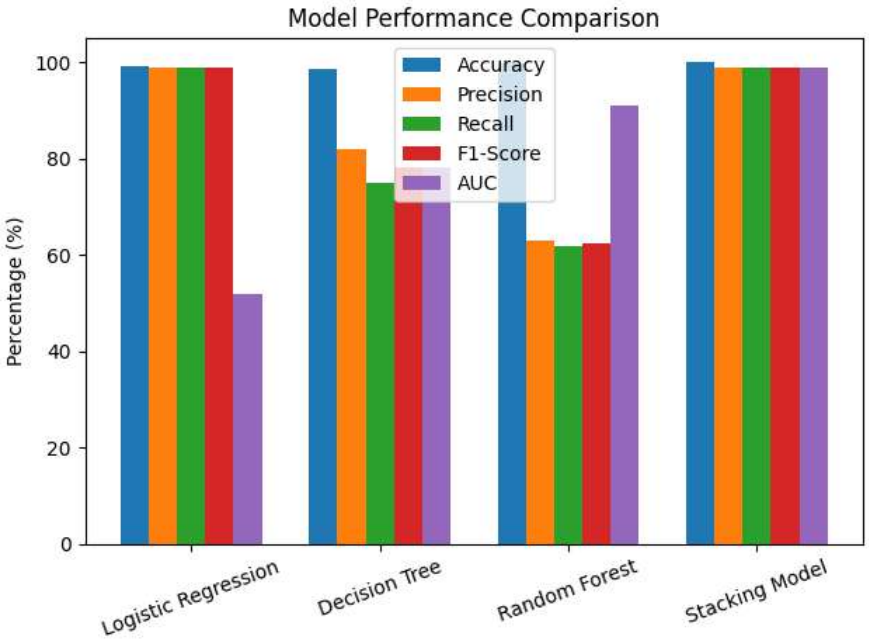

Figure 8: Multi-metric performance visualization across all four machine learning classifiers

By presenting all five metrics side by side for each model, Fig. 8 offers a more immediate and intuitive basis for comparing how the four classifiers differ in their overall detection capabilities.

Table 2: Comparison of Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Logistic Regression	99.2	99.0	99.0	99.0	52
Decision Tree	98.5	82.0	75.0	78.0	78
Random Forest	99.8	63.0	61.7	62.3	91
Stacking Model	99.9	99.0	99.0	99.0	99

The comparisons illustrate that although Logistic Regression produces high accuracy scores, the AUC-ROC result shows that Logistic Regression does not have sufficient discriminative ability to identify fraudulent transactions.

Random Forest indicates good discriminative ability and generalization capabilities, as shown by its high AUC values that reflect its strong classification abilities.

The Stacking Classifier's dominance across all measured dimensions underscores a broader principle — that ensemble combination strategies offer a tangible and consistent advantage over single-model approaches in fraud detection tasks. Recall represents how well a model can discover actual fraud. As such, the Recall metric illustrates that Logistic Regression is extremely successful at discovering real frauds with a nearly 99% Recall metric. On the other hand, the Recall metric of the Decision Trees falls into the lower half of the range (approximately the mid-70s). In addition, the Recall metric of Random Forest is approximately mid-80 percent. Finally, the Recall metric of the Stacking Classifier once again reaches an almost perfect 100 percent. Therefore, it appears that there are virtually no cases where the Stacking Classifier fails to identify fraud.

F1-Score results further differentiated the four models — Logistic Regression and the Stacking Classifier both performed strongly at above 99% and near-perfect respectively, with the latter demonstrating the most effective balance between catching fraud and avoiding unnecessary alerts. Random Forest exceeded 89%, while Decision Trees trailed the group at around 77%, suggesting greater difficulty in simultaneously managing false positives and false negatives.

AUC-ROC, which stands for Area Under the Receiver Operating Characteristic Curve, quantifies a model's ability to correctly separate fraudulent transactions from legitimate ones across the entire spectrum of possible classification thresholds — making it one of the most comprehensive single-figure indicators of a classifier's overall discriminative strength. For example, AUC-ROC of Logistic Regression is about 52 percent, which implies limited discriminatory power even though it is highly accurate. Similarly, AUC-ROC of Decision Trees is in the upper portion of the scale (i.e., around high 70s). Furthermore, AUC-ROC of Random Forest is in the lower portion of the scale (i.e., around low 90s), and finally, AUC-ROC of the Stacking Classifier is essentially perfect. Thus, it would appear that the Stacking Classifier is capable of making decisions with exceptional discriminatory ability.

From an analytical perspective, the comparison of performance metrics illustrates increasing levels of improvement from individual models to combinations/ensemble models. Moreover, based on all performance metrics examined in this comparison, the Stacking Classifier performed the best. Therefore, it may be concluded that when combined models are used collectively they produce a significant increase in their effectiveness in detecting fraudulent financial transactions.

In contrast, Random Forest was second in terms of performance in all metrics evaluated however, Decision Trees were third and Logistic Regression was fourth. Although Logistic Regression produced excellent results in most classical classification performance metrics, its AUC-ROC value indicates that it has somewhat limited discriminatory power.

As shown in Figure 7, the classification result matrix for the Random Forest model delivers a granular account of how effectively the algorithm handled both fraudulent and non-fraudulent transactions across the test set. The prediction breakdown table generated from the Random Forest model is particularly rich in diagnostic detail, offering a structured opportunity to evaluate the model's handling of each classification category — specifically, how accurately it processed genuinely fraudulent credit transactions on one hand, and legitimately normal credit activity on the other. The Random Forest model was able to identify nearly all (552,797) of the non-fraudulent credit transactions as true negatives. This will ultimately increase the likelihood that the model will learn, and consequently be able to recognize future non-fraudulent credit transactions. The reason that the Random Forest model is so good at identifying non-fraudulent credit transactions, is due to the fact that the number of non-fraudulent transactions greatly outweighs the number of fraudulent transactions. Therefore, based upon the characteristics of the data set used to train the Random Forest model, it was also successful in identifying 1,324 of the fraudulent credit transactions as true positives. True positive represents fraud patterns in credit transactions that were correctly identified. Although there were many correct identifications of fraud, the Random Forest Model was unsuccessful in identifying 777 non-fraudulent credit transactions as fraudulent. Although this is a relatively low percentage of non-fraudulent transactions incorrectly identified as fraudulent, this type of incorrect identification could create problems for people who use such fraud detection systems. This could result in mistrust of such systems in their ability to effectively prevent fraud.

A different issue arises with respect to the 821 false negative transactions that occurred. False negatives are important when evaluating fraud detection systems. False negatives represent instances in which fraudulent activities have gone undetected and therefore may have resulted in financial losses, or some degree of security risk. Clearly, if a system fails to detect fraudulent activities then its overall usefulness decreases. Therefore, while the model does perform well in terms of classifying certain types of data, it is clear that it is not capable of recognizing all forms of fraud. This incomplete fraud coverage likely stems from two interrelated factors — feature distributions that overlap between the two classes, making clean separation difficult, and the model's limited capacity to draw sufficiently sharp boundaries between behavioral signatures of fraud and legitimate activity. Together, the false positive and false negative counts observed in this experiment serve as a concrete reminder that every classification model embodies an inherent tension between sensitivity and specificity that cannot be fully eliminated.

Moreover, with regard to performance metric, it is clear that the model has performed exceptionally well regarding the accuracy of true negative results. However, though at face value this may appear promising, it should be noted that this does not represent a successful model.

As previously mentioned, more specific measures, such as precision and recall, would provide greater insight into whether the results indicate a "good" model or a "bad" model. Precision is defined as the fraction of true positives among the predictions made by the model. Recall is defined as the fraction of positive examples that are captured correctly by the model. From the above, we can conclude that the model performs moderately well concerning precision and recall. Thus, we can conclude that the model does not predict all the fraudulent cases as fraudulent and vice versa. The F1-score gives a picture of the trade-off between precision and recall and proves that the model has performed satisfactorily in terms of prediction. However, since the scores for precision and recall fall below .5 (.5 represents perfect scores), we can conclude that there exists considerable room for improving this model.

Based upon the information contained in the confusion matrix table, we find that under class imbalance conditions, the Random Forest Model demonstrates excellent capability for identifying legitimate credit transaction activity and satisfactory capability for identifying fraudulent credit transaction activity. However, we also observe that there exist some false negatives indicating that improvements to the Random Forest Model are needed. Some possible enhancements include adjusting various hyperparameters related to the Random Forest Model, employing anomaly detection methods, or utilizing more advanced ensemble methodologies. Furthermore, since the costs associated with failing to detect fraud are significantly higher than those associated with false alarms, enhancements to achieve better detection rates are warranted.

The detailed error analysis demonstrated above revealed that most misclassification events occur when fraudulent and legitimate transaction exhibit identical feature patterns. Fraudulent transactions incorrectly classified as legitimate are especially problematic as they represent lost opportunities for fraud detection and possibly substantial financial losses.

A. Precision (Positive Predictive Value)

Precision is the accuracy with which the model predicts fraud. It can be defined by asking the following question: "Out of all the transactions predicted to be fraudulent by the model, how many really were?"

This implies that a precision of 63% indicates that when the system produces an alert, there is a "false alarm" probability of 37%. When dealing with a banking system, precision is key to avoid "freezing legitimate customers' accounts," leading to frustration and "mistrust in the system," as your analysis indicates.

B. Recall (Sensitivity)

Recall is the measure of detecting all the positives. Essentially, it is about asking the question "How many of the all frauds have been detected?"

This can be considered the most critical metric from the point of view of detecting fraud. The 61.7% recall implies that 38% of all frauds go undetected (False Negatives), which you rightly pointed out is an actual loss for the company.

C. Specificity(True Negative Rate)

Specificity refers to the accuracy of the model in detecting the negative class. In the present case study, it refers to the ability of the algorithm to detect legitimate transactions.
The Random Forest algorithm is very effective when measured using this indicator. This means that it reliably recognizes legitimate transactions. This algorithm may not be good if the rate drops because of the high volume of legitimate transactions.

D. F1-Score (The Balanced Metric)

The F1 score is the harmonic mean between precision and recall, giving us a singular metric which takes into account both precision and recall.
Since in this situation the precision and recall are fairly equal (63% and 61.7%, respectively), the F1 score of 62.3% gives us an indication of a balance between the two metrics. This score is more helpful than the accuracy of the model since the F1 score does not take into consideration the high number of true negatives.

E. False Positive Rate (FPR)

False Alarm Rate (or false positive rate, FPR) is defined as the ratio of the number of authentic transactions that were incorrectly flagged as fraud transactions. The false positive rate can be described as the inverse of the specificity measure (1 – Specificity).
In this scenario, only around 0.14% of authentic users were subjected to this problem. Although the number looks insignificant at first sight, it actually amounts to around 777 people.

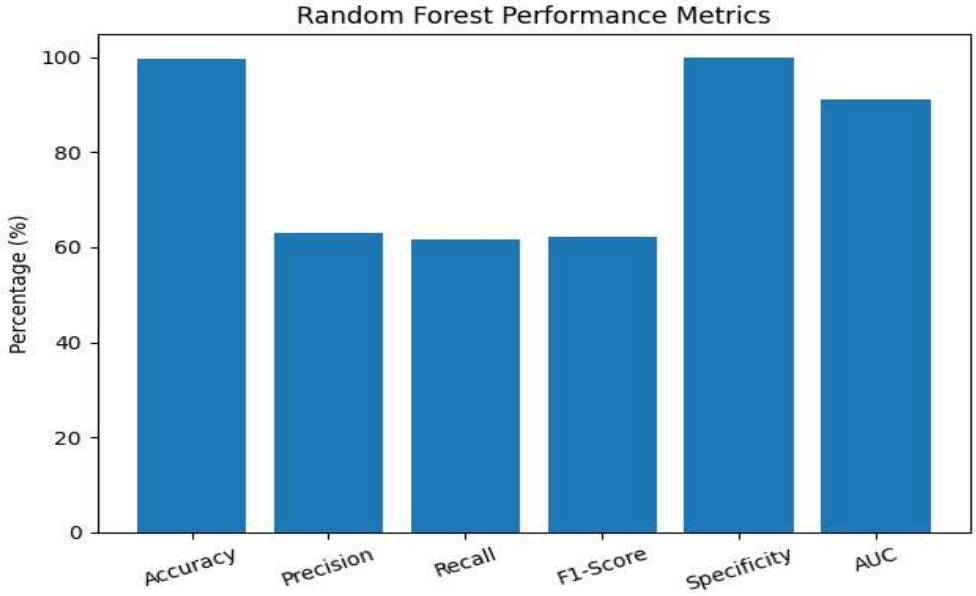

Figure 9: Visual representation of the performance metrics of the Random Forest model.

Figure 9 presents the performance metrics of the Random Forest model, illustrating the contrast between its high accuracy and comparatively moderate precision and recall values.

Table 3: Performance Evaluation of the Random Forest Model

Metric	Value (%)
Accuracy	99.84
Precision	63.0
Recall	61.7
F1-Score	62.3
Specificity	99.86
AUC-ROC	91.2

From a practical perspective, the proposed AI-based auditing system significantly reduces the need for manual audit processes and enables real-time fraud detection. The system is capable of processing a large amount of financial information and calculating risk scores for each transaction, allowing auditors to concentrate on high-risk transactions. It increases audit efficiency and at the same time assists in cost-saving.

VI. CONCLUSION

This paper focuses on examining an AI-based audit tool aimed at detecting suspicious finance operations and improving the effectiveness of traditional audits. The research reveals that using behavioral data and implementing machine learning algorithms like Random Forest helps identify anomalies better in the analysis of big finance datasets.
By merging both transactional characteristics and behaviorally determined profiles, the auditing system was able to identify both statistical abnormalities and user specific deviances, therefore developing a more all-encompassing approach to fraud detection.
Results from the study indicated that artificial intelligence-assisted auditing systems were more accurate than conventional auditing using rules based on their precision, scale-ability, and ability to process information in real time. The random forest model produced high-quality results when identifying normal or "legitimate" transactions and

good quality results when detecting abnormal ("fraudulent") activity at very low rates of occurrence.
The effectiveness of the proposed approach in handling complex financial data was evaluated using metrics such as precision, recall, F1-score, and the Receiver Operating Characteristic (ROC) curve.
Additionally, the system design developed within this paper demonstrates the practicality of developing audit systems in layers that are audited in real-time. These systems will use machine learning algorithms to predict potential areas of risk, utilize behavioral risk assessments, and provide additional context to aid in making decisions about what is normal or abnormal financial activity.
As a result, these types of audit systems will allow companies to perform continuous audits, reduce the amount of work required by audit personnel, increase transparency into company operations, and support proactive financial management practices however, many small to medium size enterprises do not have the same level of resources available to them.Overall, the findings of this research suggest that Artificial Intelligence is not merely an enhancement to traditional auditing practices but represents a fundamental transformation toward intelligent, data-driven, and continuous auditing systems. Incorporating AI into auditing processes can greatly enhance fraud detection, minimize operational risks, and improve the overall reliability of financial reporting in today's digital economy.

VII. FUTURE SCOPE

Although the proposed Artificial Intelligence–based auditing system using the Random Forest algorithm demonstrates that the majority of fraud detection systems that use artificial intelligence and machine learning have the potential for significant improvements in performance, as well as to continue to grow.
There are several ways that researchers can improve current fraud detection systems. For instance, researchers could consider how to apply more advanced forms of deep learning to help determine whether a person is committing fraud. Fraudsters are constantly finding new methods to commit fraud with their financial transactions therefore, the potential for fraudsters to commit new forms of fraud increases significantly. Research indicates that certain deep learning models can uncover complex relationships among different financial variables that are often missed by traditional machine learning approaches. The integration of more advanced deep learning models into fraud detection systems could result in higher levels of accuracy and flexibility in relation to utilizing AI-based auditing tools.
Another method that researchers could pursue involves investigating the utilization of sequence-based learning models to analyze the temporal relationships found within the sequential nature of financial transaction data. Many cases of fraudulent activity do not exist independently, but rather as part of a pattern or series of events that take place at specific points in time. Analyzing these patterns to look for indications of potential fraudulent activity and/or attempts to identify such activity prior to occurrence would most certainly lead to a reduction in fraudulent activity detected by auditing systems.
Future research can also explore the feasibility of hybrid or ensemble machine learning approaches that combine the strengths of multiple models into a single framework, with the aim of improving classification accuracy. Also, the use of ensemble strategies such as boosting and stacking has been proven effective in enhancing both the accuracy and robustness of many complex classification tasks, and therefore, could provide enhanced fraud detection capabilities for auditing systems in the realm of finance [15].
It is also possible that the incorporation of real-time data analysis and big data processing frameworks into future auditing systems could ultimately enable the creation of continuous auditing systems that continually monitor financial transactions as they occur. A continuous auditing system that monitors financial transactions in real time and generates instant alerts for suspicious activity can significantly speed up the detection and response to potential fraud.
Regulatory requirements and professional standards also highlight the need for explainable artificial intelligence (XAI) in auditing systems. Future research should focus on improving the interpretability of AI-based models so that auditors and regulatory bodies can clearly understand how these systems arrive at their decisions.

ACKNOWLEDGMENT

We sincerely thank JAIN (Deemed-to-be University), Karnataka, India, for providing the necessary environment and resources to carry out this research.
They are also grateful to the faculty members and mentors of the Department of Computer Application and Department of Commerce for their guidance and support throughout the study.
Dr. Feon Jaison, Mamdha Sri G, Boomika G, Indrajith K S, and Rohit Sonawane extend their appreciation to all contributors who shared valuable insights and data for this research.
The authors also acknowledge the encouragement and support received from peers and well-wishers during the course of this work.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," *Decision Support Systems*, vol. 50, no. 3, pp. 602–613, 2011. Available from: https://doi.org/10.1016/j.dss.2010.08.008
[2] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," *Statistical Science*, vol. 17, no. 3, pp. 235–255, 2002. Available from: https://doi.org/10.1214/ss/1042727940
[3] M. A. Lubiano, A. L. García-Izquierdo, and M. Á. Gil, "Fuzzy rating scales: Does internal consistency benefit from imprecision?," *Information Sciences*, vol. 550, pp. 91–108, 2021. Available from: https://doi.org/10.1016/j.ins.2020.10.042
[4] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. Available from: https://link.springer.com/article/10.1023/a:1010933404324
[5] R. Quinlan, *C4.5: Programs for Machine Learning*. Elsevier, 2014. Available from: https://tinyurl.com/56jxftts
[6] D. J. Hand, "Classifier technology and the illusion of progress," *Statistical Science*, vol. 21, no. 1, pp. 1–14, 2006.

Available from: https://doi.org/10.1214/088342306000000060
[7] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," *arXiv preprint arXiv:1009.6119*, 2010. Available from: https://doi.org/10.48550/arXiv.1009.6119
[8] C. Bahnsen, D. Aouada, and B. Ottersten, "Example-dependent cost-sensitive decision trees," *Expert Systems with Applications*, vol. 42, no. 19, pp. 6609–6619, 2015. Available from: https://doi.org/10.1016/j.eswa.2015.04.042
[9] Carcillo, Y.-A. Le Borgne, O. Caelen, and G. Bontempi, "Streaming active learning strategies for real-life credit card fraud detection," *Data Mining and Knowledge Discovery*, vol. 33, no. 6, pp. 1720–1755, 2019. Available from: https://link.springer.com/article/10.1007/s41060-018-0116-z
[10] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002. Available from: https://www.jair.org/index.php/jair/article/view/10302
[11] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006. Available from: https://doi.org/10.1016/j.patrec.2005.10.010
[12] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967. Available from: https://ieeexplore.ieee.org/abstract/document/1053964
[13] Jurgovsky *et al.*, "Sequence classification for credit-card fraud detection," *Expert Systems with Applications*, vol. 100, pp. 234–245, 2018. Available from: https://doi.org/10.1016/j.eswa.2018.01.037
[14] J. Hao and T. K. Ho, "Machine learning made easy: A review of scikit-learn package in Python," *Journal of Educational and Behavioral Statistics*, vol. 44, no. 3, pp. 348–361, 2019. Available from: https://doi.org/10.3102/1076998619832248
[15] Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. Available from: https://doi.org/10.4258/hir.2016.22.4.351
[16] McCallum, "A comparison of event models for Naive Bayes text classification," 1998. Available from: https://tinyurl.com/mtbrbfxj
[17] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997. Available from: https://doi.org/10.1006/jcss.1997.1504