Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 4 Issue 3 (2024)

Enhancing Fraud Detection and Risk Assessment in Financial Services Using Machine Learning and Predictive Analytics

Dr. RVS Praveen,
Director, Digital Engineering & Assurance, LTIMindtree Limited, M/s. Divija Commercial Properties THE SKYVIEW-Building No. 20, Raidurgam Village, Hyderabad, Telangana, India. praveen.rvs@gmail.com

Chandan Kumar,
Research Scholar, Department of Management, Magadh University, Bodh-Gaya, Bihar, India. ck5341176@gmail.com

Dr. E. Manigandan,
Associate Professor, Department of Information Technology, School of Business, Galgotias University, Greater Noida, Uttar Pradesh, India. e.manigandan@galgotiasuniversity.edu.in

Abbarapu Ashok,
Research Scholar, Department of Mathematics, VIT-AP University, India. ashokabbarapu@gmail.com

Dr Pushpa Rani,
Assistant Professor, Department of Management Sciences, Tecnia Institute of Advanced Studies, Delhi, India. pushpasangwan9@gmail.com

Subharun Pal,
PG Scholar, CEP, Indian Institute of Technology Patna, Bihar, India. subharunpal@gmail.com

Abstract –Financial fraud, which involves fraudulent practices to acquire financial gains, has recently become a major issue in businesses and organizations. It is inefficient, expensive, and time-consuming to discover fraudulent activities through manual verifications and inspections. The intelligent detection of fraudulent transactions is made possible by artificial intelligence through the evaluation of enormous amounts of financial data. Key components for ensuring operational integrity and limiting financial losses in the financial services business include fraud detection and risk assessment. Due to the increasing complexity of fraud schemes, traditional techniques of detection that depend on static rules and historical data are no longer adequate. In order to better detect fraud and evaluate risk in the financial services sector, this study explores the application of predictive analytics and machine learning (ML). Real-time data and adaptive algorithms are used to evaluate the performance of ML techniques such as supervised learning, unsupervised learning, and ensemble methods in detecting fraudulent actions. The results show a considerable improvement in detection accuracy and risk assessment over older methods. This paper also explores the possible obstacles of deploying these technologies, such as data privacy concerns, interpretability, and the need for ongoing model training.

***Keywords*—**Fraud Detection, Machine Learning, Threat Cycle, Risk Management.

I. Introduction

In recent years, numerous forms of financial fraud have become far more common and severe. Damage to private and public property is a direct outcome of these acts. Terrorists could utilize the stolen funds to finance their attacks, which puts national security at risk. Recognizing and tracking down instances of financial fraud is, thus, of the utmost importance. Financial transactions and trade networks are complex, making it difficult to detect fraud. Take the following definition of money laundering as an example: it is the practice of transferring funds or goods through transactions in order to mask their true origin [1]. The inclusion of intentionally false information on product quality, quantity, or cost is typical in money laundering invoices. All types of invoice fraud, including those involving quality, quantity, and price, are treated as minor exceptions when these figures are used to build our detection policy.

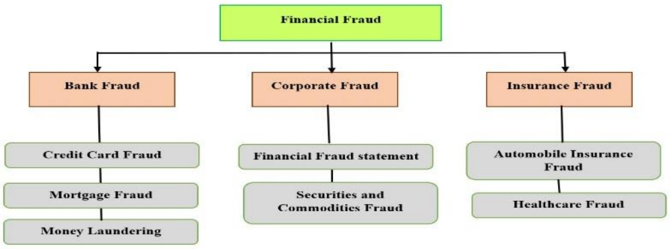

Fig. 1. Common Financial Fraud Categories

Money laundering and other more clandestine types of fraud do happen. Figure 1 displays typical financial fraud categories, as well as data mining and computational intelligence-based detection methods. There are a lot of different ways that money laundering might happen, such as buying and selling intangible assets, dealing with linked parties, or hiding cash while it's in transit. Among the many forms of internal financial fraud is the unauthorized acquisition of funds or assets belonging to another person [2]. An instance of external financial fraud would be providing stakeholders with an inaccurate picture of the company's financial status. Based on what we can see in the literature, most studies have concentrated on ways to identify external financial fraud. Surprisingly, not much is done to address the problem of internal financial fraud. Financial fraud identification is essential for preventing financial fraud and its catastrophic consequences. One of the primary goals of financial fraud detection (FFD) is to separate real financial data from fake data. That manner, the powerful can uncover the plots of the dishonest and foil them. The main objective of fraud detection is to maximize the quantity of correct predictions while keeping the number of inaccurate predictions within an acceptable range. The goal of achieving a high level of diagnostic confidence is to reduce the likelihood of fraudulent activity and false alarms. Both the financial industry and people's day-to-day lives are profoundly impacted by financial fraud. Expenditures on living, economy, and public trust can all take a knock when fraud happens [3]. The complexity of the situation renders traditional methods of fraud detection, such auditing, ineffective and prone to errors. The goal of financial fraud is to deceive people into parting with their money by means of dishonest means. Methods for financial fraud and data mining are distinct, and studies are being conducted to determine the most appropriate strategy for each case [4]. Modern technology, such as the internet and mobile computers, have increased the likelihood of financial fraud. The growing usage of credit cards has increased both spending and fraud. As fraudsters improve their methods, detection technologies must evolve to keep up. Misclassification of fraudulent and ordinary transactions is a common and costly problem in fraud detection. Data mining methods are good classifiers for identifying fraud because they can handle massive datasets and require little problem-specific knowledge.

II. Literature Survey

To commit financial fraud is to gain monetary gain by use of deceitful or unlawful means. Insurance, banking, taxes, and corporate finance are just a few of the many industries where financial fraud can take place [5]. Within the realm of supervised machine learning, support vector machines (SVMs) divide training data inputs into two categories by means of a maximum margin hyperplane [6]. Data mining, picture processing, and pattern recognition are just a few of the many fields that make use of the means clustering technique, which has been the subject of much research. Earlier research on fraud detection looked at how well K-means clustering might spot suspicious groups in data on monetary transactions. Making it easier to spot fraudulent conduct, the system can categorize data based on parameters such as transaction amount, frequency, and location. Support vector machines (SVMs) use labelled training sets for each class to identify fresh data points [7]. Using support vector machine methods to identify fraudulent transactions. A method that combines SVM and fusion Danger theory to identify fraud. When comparing their study to others, the researchers found that it was more efficient in terms of both time complexity and F-measure. use support vector machines to identify instances of medical billing system fraud. Quick and real-time detection of medical fraud is the goal of this study strategy [8]. The results of the experiments proved that the model was more effective than the alternatives. A more refined SVM for online credit card fraud detection. neural network algorithms for the detection of fraudulent charges on credit cards. By combining support vector machines (SVMs) with spike detection, the authors are able to get around the shortcomings of existing methods [9]. Previous approaches were surpassed by the suggested strategy. Supervised learning using support vector machines to differentiate between legitimate and fraudulent credit card transactions. Using support vector machines (SVMs), logic, and linear regression, the scientists improved the detection accuracy. With an emphasis on their concepts and uses in fraud detection, this article examines three significant adaptive algorithms: online learning, anomaly detection techniques, and reinforcement learning [10]. Reinforcement learning (RL) is a machine learning paradigm in which an agent learns to make decisions by interacting with its environment and receiving feedback. This strategy is particularly useful for optimizing fraud detection strategies over time[11]. In RL, feedback systems are essential for the learning process. The agent performs activities (such as detecting fraudulent transactions) and is rewarded or penalized based on the outcomes. Positive rewards stimulate correct identifications, whereas negative rewards discourage incorrect ones. Techniques such as Q-learning and Deep Q-Networks (DQNs) employ this input to update the agent's knowledge and build fraud detection tactics [12]. Optimizing detection approaches with RL requires balancing exploration (trying new strategies) and exploitation (applying established successful strategies). This balance is crucial for detecting emerging fraud tendencies while delivering reliable results [13]. RL algorithms can respond to changing patterns of fraudulent activity, making them particularly helpful in dynamic environments [14]. Neural networks, particularly deep learning models, excel at detecting anomalies because of their ability to capture complicated relationships in data. Auto encoders, a type of neural network, are commonly used for this task [15]. As they learn to compress and reconstruct input data, they become alert to anomalies that could suggest fraud, such as transactions with large reconstruction faults. Many well-known organizations in the banking, telecom, and consulting industries have used Classical Machine Learning (CML) fraud detection models, which are also a typical textbook exercise or capstone project [16]. Several CML models, such as Support Vector Machines, Random Forest Classifiers, Multivariate Logistic Regression, and Gradient Boosting Machines, as well as comparative studies across methods, or combining models in demonstrate high fidelity, robustness, ensembles, and ease of implementation [17]. Furthermore, researchers have applied many DL)approaches, provide information on auto encoders and RBM, and discuss graph neural networks [18]. Just two tedious steps—data cleaning and feature engineering—make up the model lifespan. A new form of machine learning called Quantum Computing (QC) was created by academics by combining the ideas of CML with QC. Using QC core units, qubits, and QML algorithms, the underlying concept is to surpass existing computing methods. Even though it's still young, QML has a lot of potential [19]. Improvements in computer power, hardware design, quality control, and quantum cloud technologies will lead to an increase in the use of QML for everyday tasks, similar to how large corporations are integrating CML into their operations. This system analyzes time series-based, highly imbalanced, multidimensional data from online transactions using QML in order to detect bogus records [20]. The CML models that used an improved SVM and quantum annealing solvers were compared to their method. This study delves into the difficulties of handling real-time transactional data and suggests that a quantum method would be more effective and relevant to other important commercial applications. Since the SOM model doesn't need any background knowledge, the proposed automation system can be enhanced in real time with fresh transaction data. Applying K-means clustering to credit card fraud characteristics in order to build a model for detecting fraudulent transactions [21]. The fraud detection problem is handled as a sequence classification challenge, with predictions based on long and short term memory (LSTM) [22]. Experimental results suggest that LSTM outperforms random forests in detecting credit card fraud. The Light Gradient Boosting Machine model's effectiveness in identifying credit card fraud to that of the Random Forest and Gradient Boosting Machine methods. In real datasets, the model achieved a higher total recall rate while also providing faster feedback.

III. METHODOLOGY

Financial institutions have recognized that using isolated security methods on individual delivery channels does not provide adequate protection against illegal account activity. Financial IT platforms are vulnerable to fraud due to security weaknesses that allow for large-scale monetary theft.

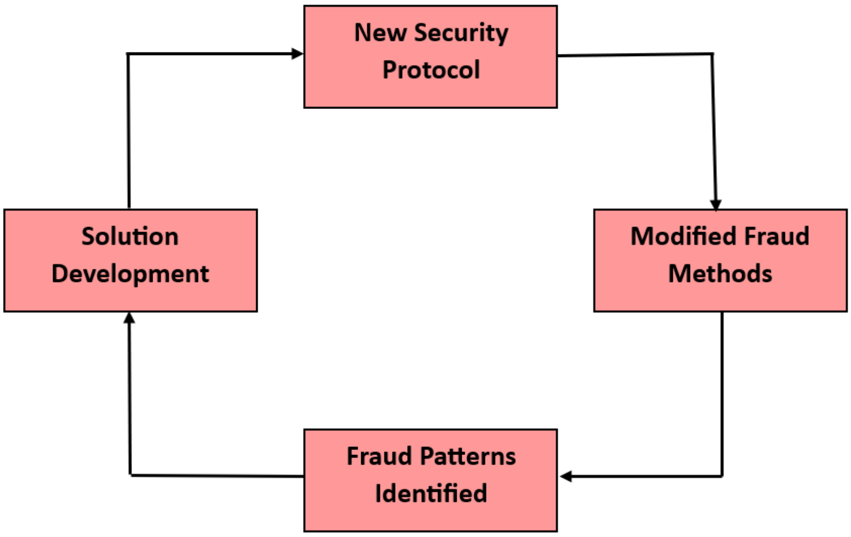

Fig. 2. Threat Cycle of Fraud

Malicious third parties can use weak authentication techniques including signatures, PINs, passwords, and Card Security Codes (CSCs) to conduct fraudulent financial transactions.

A. *Fraud Management*

Because scams can affect a financial institution's reputation, service quality, and bottom line, catching them early is a top priority. In order to combat the vulnerabilities in channelized authentication methods that fraudsters frequently take advantage of, numerous organizations are integrating "Swindle Management" with channelized security requirements. An all-encompassing fraud control architecture with multi-level security across all service networks is the end result of swindle management solutions' active screening of account activity data. To better serve their customers, financial institutions make use of modern technology including plastic credit/debit cards, ATMs, online banking, and mobile banking. The financial services within the business and data logic system levels are managed by allied network level servers, who direct received requests. For the purpose of verifying the legitimacy of users, network security policies and processes depend on the "what the user knows" and "what the user has" criteria [23]. Users are needed to submit necessary security information, including passwords, identification numbers, and personal details, in order to access banking services. As an additional layer of protection, they might verify their identity with a physical security token or smart card. Data mining allows Reactive Fraud Management to execute complex computations on recorded transaction data. Cases of fraud can be identified by comparing them to established patterns of fraud or by noting unusual behavior in relation to the documented history. By necessitating transactional data prior to utilizing data analysis tools, the "store now, query later" strategy can lengthen the time it takes to detect fraud. Keep waiting for the transaction to close and the related fiscal value to rise in order to get a preventative response. To predict the arrival of fresh data, reactive fraud management systems use labeled priming data sets. To keep up with emerging fraud risks posed by unlabeled transaction requests, models must be retrained on a regular basis. Large financial losses and undetected fraud instances happen because of delays caused by adding a large number of tagged swindle cases to the training data set.

B. *Types of Financial Fraud Detection*

Financial fraud can take numerous forms, which are briefly described here.

1) *Credit Card Fraud:*

The crime of credit card fraud occurs when an unauthorized third party uses a credit card to make a purchase [24]. When a physical card is lost or stolen, transactions can still be done remotely [25]. There are several techniques for obtaining the cardholder's information. Phishing is when a fraudster impersonates a finance official to gain access to a user's information. Swipers or skimmers can be used to read a user's card directly, or entire databases can be obtained by breaching the financial institution's network security or enlisting an accomplice [26]. Getting a replacement or fresh card for the user might be as easy as snooping on their mail. There has been a rise in organized crime related to credit card fraud due to the anonymity and accessibility of online methods [27]. Credit card fraud is typically detected by analyzing a customer's spending patterns and flagging transactions that deviate from the norm.

2) *Fiancial Fraud:*

Financial statements provide facts about a company's expenses, loans, income, and profits [28]. Management remarks on business performance and potential future difficulties may also be included [29]. For investors and potential borrowers, a company's financial statements are a window into its health and performance. Falsifying a company's financial statements to make it look wealthier is known as financial statement fraud or corporate fraud. There are a number of motivations that might lead to financial statement fraud, including the desire to artificially inflate success, reduce tax liabilities, or appease manipulative managers [30]. Financial statement fraud is challenging to detect due to a lack of understanding, infrequent occurrence, and the ability of educated industry professionals to conceal their deception.

3) *Insurance Fraud:*

Anyone in the insurance chain is capable of committing fraud at any point. False insurance claims, whether caused by inflated injuries or losses or entirely made up, constitute insurance claims fraud. Falsifying or intentionally provoking accidents that result in excessive repair and injury expenses is one example of automobile insurance fraud. People commit crop insurance fraud when they exaggerate the amount they will lose because of things like falling agricultural prices or natural disasters [31]. Insurance fraud can involve excessive billing, multiple claims, bribes to brokers, and "upcoding" of items.

4) *Mortagage Fraud::*

Financial fraud can take several forms, one of which is mortgage fraud, which is forging documents pertaining to real estate or mortgages. The practice of exaggerating a property's value in order to secure a loan is widespread.

5) *Money Laundering:*

Criminals engage in money laundering when they want to launder the proceeds of their illicit activity into more reputable companies. This masks their illicit activity by making it appear as though the funds are coming from legitimate sources. Because it gives criminals access to financial resources, money laundering is bad.

C. Machine Learning Methods for Financial Fraud Detection

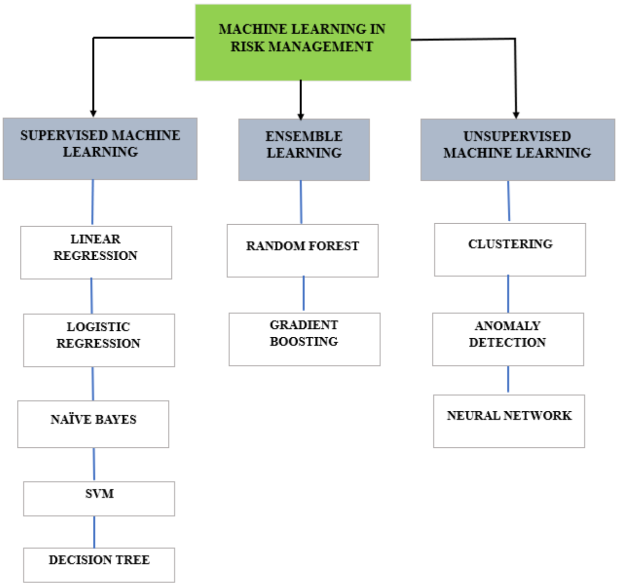

Fig. 3. Machine Learning in Risk Management

1) Supervised Learning

As part of its training procedure, the algorithm takes an input-output data set and uses it to build a mathematical model. Textual examples are used to teach algorithms, with input and output predetermined. The input set and the intended outputs are both obtained through the usage of an algorithm in this study. The Algorithm checks its real findings against the correct results to find out what happened. The model is then suitably updated after that. Classification, regression, prediction, and gradient boosting are supervised learning methods that estimate values using patterns.This interpretation is usually used in contexts where past data is used to forecast future occurrences. The two main functions of supervised learning are regression and classification. Decision trees, Naïve Bayes, and close neighbors are all instances of supervised machine learning. Supervised learning has a general foundation in classification. By studying real-world workflow examples, trainees can better understand and quantify how different classes behave when plotting vectors. The goal of inductive machine learning is to build a generalizable classifier by studying rules in previously encountered cases or by the execution of multiple tasks simultaneously.

a) Linear Regression

This algorithm for learning is the simplest. Predictive analysis is another possible application of this statistical tool. Continuous and quantitative factors like age, price, and sales are used to develop predictions. The method reveals a straight line between the dependent and independent variables in linear regression. It reveals the relationship between the values of the independent and dependent variables, showing how the latter changes.

b) Logistic Regression

Known as supervised learning, this method is utilized by most algorithms. To predict a collection of categorical dependent variables, this technique employs a set of independent variables. It analyzes the performance of a categorical dependent variable. Therefore, a value that can be categorized is necessary. Given a probability value between zero and one, it provides alternatives to yes/no and other binary choices. Except for that one detail, this method is very similar to Linear Regression. You can use Logistic Regression for classification problems and Linear Regression for regression concerns. Using this method, you can categorize observations based on many data sets and determine which factors are most important. Instead of fitting a regression line, we use an S-shaped logistic function that can take on two possible maximum values: 0 and 1. The logistic function's curve represents the probability of an event, such the color of an apple or a ball. Given its versatility, this approach ranks high among learning algorithms. It excels at both continuous and discrete dataset classification and prediction.

c) Naïve Bayes

This algorithm is designed for supervised learning. Its basis is Bayes' theorem. This is used for classification-related computing tasks. It can be used to classify texts in a multi-level dataset. Because of its simplicity and effectiveness, the Naive Bayes Classifier has become a go-to tool for training machine learning models. One such approach is a probability-based classifier, which relies on an object's likelihood to produce conclusions. Classifying articles, doing sentiment analysis, and detecting spam are just a few of its many applications.

d) SVM

This kind of supervised learning is useful for solving problems related to regression and classification. Its primary utility in machine learning is as a classification dataset. This method aims to find the optimal partition for partitioning an n-dimensional space into cases so that additional data points can be classified more easily in the future. A hyperplane is the term used to describe this dividing plane. Using SVM, the most extreme locations that can be used for hyperplane creation are taken into consideration. Due to the fact that these outliers are referred to as support vectors, the method is known as Support Vector Machine.

e) Decision Tress

Even while this supervised learning method works well for classification problems as well as regression ones, regression is where it truly excels. This classifier makes use of a hierarchical framework. Each leaf node represents the final result, the branches reflect the decision-making mechanisms, and the core nodes represent the qualities of the dataset. Two kinds of nodes can be seen in a decision tree: the decision node and the leaf node. A leaf node with no further branches is the end result of a decision node with numerous branches making a choice. The rating is based on features of the dataset that was provided. This graphical depiction shows all possible solutions to an issue given a particular set of parameters. This grows outward from the root node in the same manner that a tree's structure does. A classification and regression tree (CART) is a tool for retrieving variables from a tree. A decision tree divides a tree into subtrees based on yes/no answers.

2) UnSupervised Learning

Building a model using only inputs is known as unsupervised learning. Rather of relying on tagged output, this learning method uses unlabeled data. Association rules and K-means are algorithms that fall under this category. Unsupervised learning can be seen in Figure 3. A plethora of algorithms are utilized in unsupervised learning. The algorithms that are most commonly used include: Techniques for acquiring models using latent variables. Neural networks, clustering, and detecting anomalies

a) Clustering

Sorting items into categories according to how similar they are is an integral part of the process. Pattern recognition, machine learning, image analysis, computer graphics, and data retrieval are just a few examples of the many data mining applications that rely on the painstaking investigation of mathematical data. A lot of labor, not an algorithm, is required of the group's study. It is feasible to acquire knowledge on cluster formation comprehensively by use of algorithms that approach the subject from diverse angles. Data space shows perfect clusters with little overlap and separation.

b) Anomaly Detection

Organizing objects causes them to be more tightly packed into one set or cluster than into other groupings. Pattern recognition, machine learning, image analysis, computer graphics, and data retrieval are just a few examples of the many data mining applications that rely on the painstaking investigation of mathematical data. A lot of labor, not an algorithm, is required of the group's study. A cluster and its components can be obtained by various methods.

c) Neural Networks

Computer systems that mimic the structure and function of the brain are known as neural networks. Without task-specific rules, these networks can do tasks based on occurrences. Take photos of cats as an example. They learn to recognize them by comparing them to ones that are labeled as "cat" or "no cat" and then applying that knowledge to other images. Cats, including their fur, tails, cheekbones, and other feline-like facial characteristics, are completely foreign to them. From the models they analyze, they generate visual characteristics. A network of linked nodes called artificial neurons, which stand in for the actual neurons in an animal's brain, is the basis of the ANN. Synapses in the brain's blood supply are just one example of a link that can communicate with other neurons. The artificial neural network (ANN) gathers and evaluates signals before sending them on to connected neurons. In artificial neural network (ANN) applications, the input sum to each neuron is used to determine its output, and the signal to edges is a real value. In general, the symmetry of neurons and edges is congruent with how they learn. A link's signal intensity can be adjusted by changing its weight. Neurons can only transmit signals when their interaction signals exceed a certain threshold. It is common practice to organize neurons in layers. The input is changed in multiple ways via multiple levels. A series of steps is required for the symptoms to progress from the input layer (the first of many layers) to the extraction layer (the last). Originally, the ANN method aimed to mimic human problem-solving abilities; but, as it progressed, its concentration shifted to individual tasks, leading it to stray from its biological roots. Computer vision, machine translation, speech recognition, social networking, medical diagnostics, and even human services like painting are just a few of the many areas that have found usage for ANNs.

3) Ensemble Learning

Ensemble learning is the process of combining multiple machine learning algorithms to achieve better results than each algorithm could achieve on its own. The predictions provided by each learner are integrated using a combination rule to produce a single, more accurate forecast, rather than relying on a single model. There are two main types of ensemble methods: sequential and parallel ensembles [32]. The parallel approaches use a combiner to combine the predictions of many base classifiers that have been trained independently. Bagging and the random forest algorithm, which is an extension of it, are common parallel ensemble techniques. In order to promote variety among the ensemble members, parallel ensemble algorithms make use of the parallel generation of base learners.

a) Random Forest

One of the ensemble models, random forest (RF), creates several prediction models and combines them to produce the final prediction model [33]. From the original data, the RF creates multiple bootstrap samples (training data), and only part of the independent variables are used to train decision trees in each bootstrap sample. After averaging or voting the bootstrap tree predictions, final predictions are then produced when additional data points are presented for independent variables. The RF regression problem uses averaging. Using the remaining data that are not included in the bootstrap samples, the OOB error is a metric that assesses the prediction ability of the bootstrap trees. By employing the ideal number of bootstrap samples where the OOB error is reduced, the predictive performance of the model can be enhanced.

b) Gradient Boosting

A machine learning method called "boosting" can turn a weak classifier into a powerful one. This kind of ensemble meta-algorithm is employed to lessen variance and bias. On the other hand, classifiers that get somewhat better results than random guessing are considered weak learners, and classifiers that achieve much better results called strong learners; it is on the latter that boosting ensemble techniques are founded. Regarding the query of whether a class full of poor students may produce a single exceptional student. A group of subpar learners could generate a single excellent learner. A major influence on statistics and machine learning, which resulted in the creation of various boosting algorithms, such as XGBoost [34] and AdaBoost. Boosting involves iteratively applying a base learning algorithm to changed input data. By adjusting the training set to incorporate misclassified cases from the previous round, boosting techniques teach a weak learner with the input data, compute predictions, and then train a new weak learner. The iterative learning process is carried out until a predetermined quantity of baseline learners is collected and combined. The goal of boosting is to reduce bias, not variation. Base learners with low variance and high bias, such decision stumps (a decision tree with one internal node), benefit from this. Because misclassified samples are given more weight, the base learner tends to focus on them. Extra weight is given to samples that the algorithm determines to be prejudiced against in order to rectify the bias in the underlying classifier. Overfitting, which can occur when boosting algorithms place too much emphasis on noisy samples, makes them unsuitable for learning from noisy data [35]. Despite this, boosting-based ensemble approaches are highly effective in applied machine learning.

D. Structure of Financial Fraud Detection

A large number of training and test datasets improves the model's accuracy. Figure 4 shows the channels and patterns used by fraudsters. From data sets, extract these patterns. Instructions for the detection system to tell legitimate purchases apart from fraudulent ones. In real-time, trained models can reject or hold transactions for further investigation in order to detect and prevent fraud.All businesses can benefit from a reduction in global fraud if they work together to share their fraud experiences.

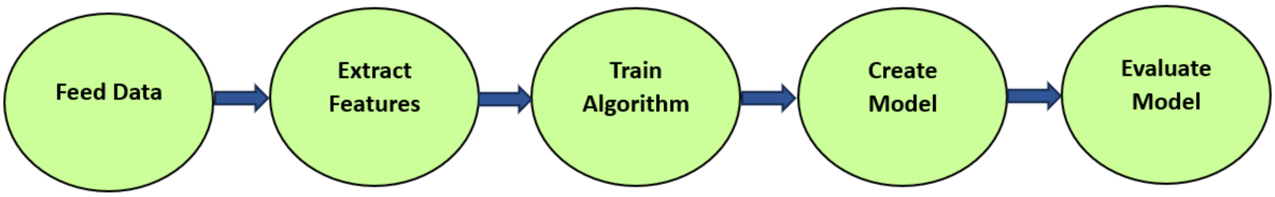

Fig. 4. Structure of Fraud Detection

Because they can be trained using test data to produce efficient and accurate results, Machine Learning Algorithms are good at detecting fraud. Monitoring, learning, detection, prevention, and continuous improvement are all essential components of a complete life cycle strategy for making decisions in real-time. Machine learning algorithms, when used properly, can drastically cut down on fraudulent transactions. Companies should not keep quiet about their fraudulent past if they want to remain ahead of scammers. Although there are systems in place to identify and notify companies of fraudulent transactions, they are siloed and do not facilitate learning across organizations to lessen the impact of fraud. A real-time platform that can automatically learn from previous incidents and alert global organizations is what the industry is aiming for. This is We provide a global model that makes use of AI and ML techniques. Collaborative fraud prevention initiatives cannot be successful without a central fraud management platform. The plan is to build an OS that will let businesses all across the globe communicate fraud trends, find fraudulent transactions before they happen, and secure their apps even more. Businesses should implement digital handshakes to improve communication. All parties concerned can agree on a uniform format for these entities to exchange the database. Since fraudulent transactions are now publicly available, the remaining organizations can take proactive security measures to prevent huge losses. A number of sectors can benefit from this centralized structure, including banking, telecommunications, the stock market, the internet, and social engineering. A dynamic, intelligence-driven approach to risk management is necessary for industries to prevent, detect, react to, and recover from cyber assaults.

IV. Results and Discussion

The performance of the classifiers was evaluated in this work using a variety of assessment metrics. The results of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) informed most of the measurements. Whereas "FP" denotes the anticipated number of valid transactions, "TP" denotes the anticipated number of fraudulent transactions. Amount of fraudulent transactions predicted as legitimate (TN) and number of legitimate transactions predicted as fraudulent (FN) are two different things. Metrics like accuracy, precision, recall, TPR, and FPR were used to evaluate classifier performance in the study. All of the assessment metrics are defined and presented in Table 1.

TABLE I. PERFORMANCE PREDICTION(%)

Metric	Formula and Description
True Positive Rate	$TRP = TP/(TP + FN)$
False Positive Rate	$FRP = FP/(FP + TN)$
Precision	$Precision = TP/(TP + FP)$
Recall	$Recall = TP/(TP + FN)$
F-Measure	$F - Measure = 2TP/(2TP + FP + FN)$
Accuracy	$Acc = (TP + TN)/(TP + TN + FP + FN)$

Using the IEEECIS dataset, Figure 5 displays the accuracy values of different methodologies applied to financial fraud detection. Greater values for accuracy indicate greater performance, since it measures how well each approach detects cases of fraud.

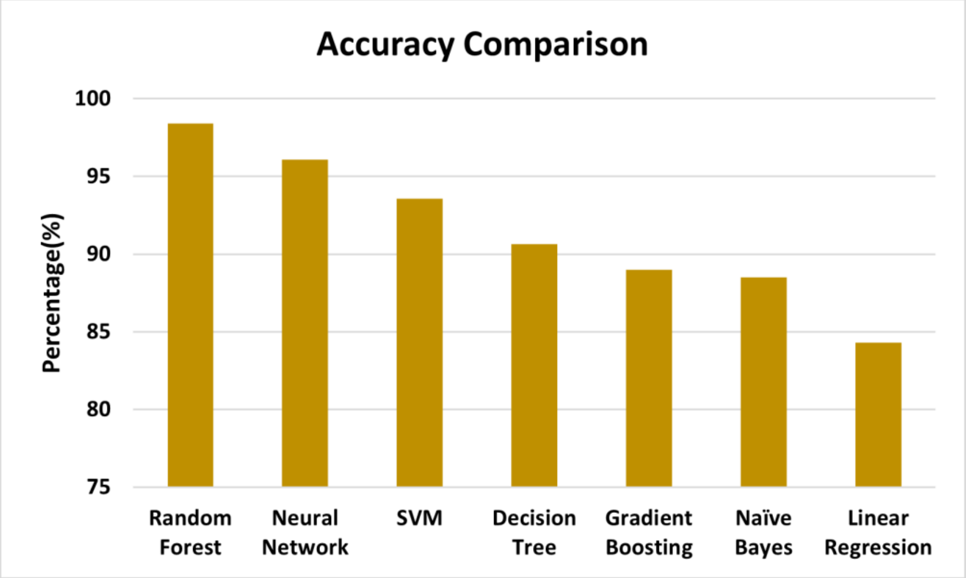

Fig. 5. Performance of Existing Methods

In conclusion, the Random Forest algorithm is the most effective strategy for detecting credit fraud in the specified dataset, with an accuracy of 0.98400 and an F1 score of 0.99194. The Random Forest technique was followed by the Neural Network, which achieved an accuracy of 0.97090 and F1 score of 0.98522. Additionally, we compared this finding to other references shown in Figure 5. The study found that three supervised machine learning techniques outperformed other algorithms: Random Forests, Neural Networks and SVM. Our findings and references support the performance evaluation of Random Forests and AdaBoost algorithms.

TABLE II. ACCURACY PREDICTION(%)

Metric	**Accuracy**
Random Forest	98.40
Neural Network	96.09
SVM	93.58
Decision Tree	90.63
Gradient boosting	89.00
Naïve Bayes	88.49
Linear Regression	84.30

Table II provides a comparison of the accuracy performance of various machine learning models. The most dependable models for this specific task are Random Forest and Neural Networks, which significantly surpass the others; in contrast, less effective models include Linear Regression and Naïve Bayes.

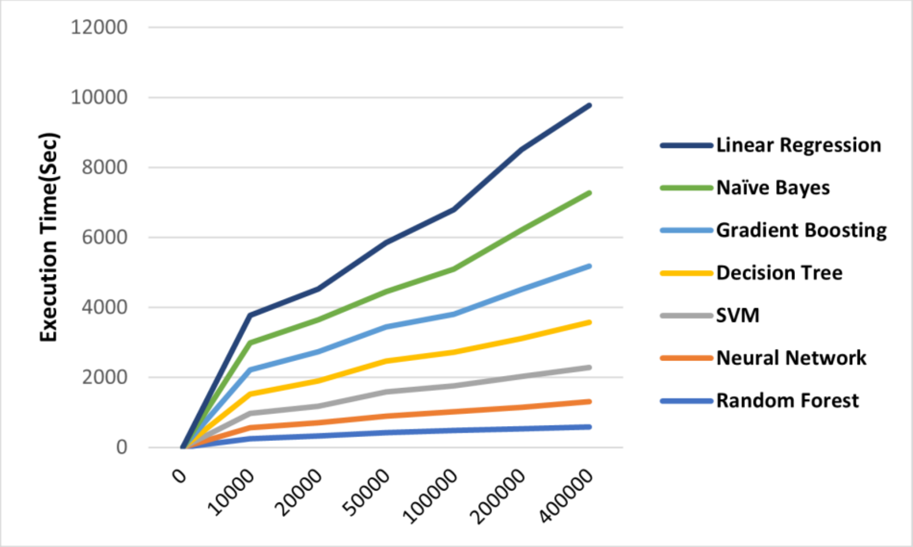

Fig. 6. Comparison of Existing and Proposed models' Computational Complexity

Furthermore, we simulated all datasets to assess the computational complexity of our model; Figure 6 displays the average execution times that resulted from this. The recommended model, Random Forest, has the quickest execution time, proving that it can efficiently handle high- and low-dimensional data and transactions. The model's performance has been fine-tuned, and it can now identify fraudulent users and transactions considerably faster, increasing its use in real-time fraud detection scenarios.

V. CONCLUSION AND FUTURE DIRECTIONS

One of the most important parts of running a contemporary bank is detecting fraud.Every method showed some promise in detecting different forms of financial fraud, even though they varied in how well they did it. Computational systems, such as support vector machines and neural networks, are effective against fraudsters' evolving strategies because they can adapt to new techniques. There are still a lot of unanswered questions about intelligent fraud detection. To completely comprehend certain forms of fraud and data mining techniques, additional research is required. Parameter tuning allows for computational fraud detection cost-benefit analysis and can enhance the performance of current systems. A more precise foundation for intelligent detection systems could be the result of additional research into the many forms of financial fraud.

REFERENCES

[1] P. Richhariya, B. Prashant, and K. Singh, “A Survey on Financial Fraud Detection Methodologies,” *Int. J. Comput. Appl.*, vol. 45, no. 22, pp. 975–8887, 2012.

[2] D. Huang, D. Mu, L. Yang, and X. Cai, “CoDetect: Financial Fraud Detection with Anomaly Feature Detection,” *IEEE Access*, vol. 6, pp. 19161–19174, 2018, doi: 10.1109/ACCESS.2018.2816564.

[3] J. West, M. Bhattacharya, and R. Islam, “Intelligent financial fraud detection practices: An investigation,” *Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng. LNICST*, vol. 153, pp. 186–203, 2015, doi: 10.1007/978-3-319-23802-9_16.

[4] Z. Huang, H. Zheng, C. Li, and C. Che, “Application of Machine Learning-Based K-means Clustering for Financial Fraud Detection,” *Acad. J. Sci. Technol.*, vol. 10, no. 1, pp. 33–39, 2024, doi: 10.54097/74414c90.

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

[6] N. M. Reddy, K. A. Sharada, D. Pilli, R. N. Paranthaman, K. S. Reddy, and A. Chauhan, "CNN-Bidirectional LSTM based Approach for Financial Fraud Detection and Prevention System," in *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, Jun. 2023, no. Icscss, pp. 541–546, doi: 10.1109/ICSCSS57650.2023.10169800.

[7] A. Mousa, "Detecting Financial Fraud Using Data Mining Techniques: A Decade Review from 2004 to 2015," *J. Data Sci.*, vol. 14, no. 3, pp. 553–570, 2022, doi: 10.6339/jds.201607_14(3).0010.

[8] Z. Zhang, X. Zhou, X. Zhang, L. Wang, and P. Wang, "A Model Based on Convolutional Neural Network for Online Transaction Fraud Detection," *Secur. Commun. Networks*, vol. 2018, 2018, doi: 10.1155/2018/5680264.

[9] T. Pourhabibi, K. L. Ong, B. H. Kam, and Y. L. Boo, "Fraud detection: A systematic literature review of graph-based anomaly detection approaches," *Decis. Support Syst.*, vol. 133, no. August 2019, p. 113303, 2020, doi: 10.1016/j.dss.2020.113303.

[10] A. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," *J. Netw. Comput. Appl.*, vol. 68, pp. 90–113, 2016, doi: 10.1016/j.jnca.2016.04.007.

[11] S. Obeng, T. V. Iyelolu, A. A. Akinsulire, and C. Idemudia, "Utilizing machine learning algorithms to prevent financial fraud and ensure transaction security," *World J. Adv. Res. Rev.*, 2024.

[12] R. Udayakumar, A. Joshi, S. S. Boomiga, and R. Sugumar, "Deep Fraud Net: A Deep Learning Approach for Cyber Security and Financial Fraud Detection and Classification," *J. Internet Serv. Inf. Secur.*, vol. 13, no. 4, pp. 138–157, 2023, doi: 10.58346/JISIS.2023.I4.010.

[13] Oluwatosin Ilori, Nelly Tochi Nwosu, and Henry Nwapali Ndidi Naiho, "Advanced data analytics in internal audits: A conceptual framework for comprehensive risk assessment and fraud detection," *Financ. Account. Res. J.*, vol. 6, no. 6, pp. 931–952, 2024, doi: 10.51594/farj.v6i6.1213.

[14] K. Koo, M. Park, and B. Yoon, "A Suspicious Financial Transaction Detection Model Using Autoencoder and Risk-Based Approach," *IEEE Access*, vol. 12, no. May, pp. 68926–68939, 2024, doi: 10.1109/ACCESS.2024.3399824.

[15] Ezekiel Onyekachukwu Udeh, Prisca Amajuoyi, Kudirat Bukola Adeusi, and Anwulika Ogechukwu Scott, "The role of big data in detecting and preventing financial fraud in digital transactions," *World J. Adv. Res. Rev.*, vol. 22, no. 2, pp. 1746–1760, 2024, doi: 10.30574/wjarr.2024.22.2.1575.

[16] M. Grossi *et al.*, "Mixed Quantum-Classical Method for Fraud Detection With Quantum Feature Selection," *IEEE Trans. Quantum Eng.*, vol. 3, no. August, pp. 1–12, 2022, doi: 10.1109/TQE.2022.3213474.

[17] N. K. Gyamfi and J. D. Abdulai, "Bank Fraud Detection Using Support Vector Machine," *2018 IEEE 9th Annu. Inf. Technol. Electron. Mob. Commun. Conf. IEMCON 2018*, no. November 2018, pp. 37–41, 2018, doi: 10.1109/IEMCON.2018.8614994.

[18] Y. Kumar, S. Saini, and R. Payal, "Comparative Analysis for Fraud Detection Using Logistic Regression, Random Forest and Support Vector Machine," *SSRN Electron. J.*, vol. 7, no. 4, pp. 726–731, 2021, doi: 10.2139/ssrn.3751339.

[19] O. Kyriienko and E. B. Magnusson, "Unsupervised quantum machine learning for fraud detection," pp. 1–10, 2022, [Online]. Available: http://arxiv.org/abs/2208.01203.

[20] C. Liu, Y. Chan, S. H. Alam Kazmi, and H. Fu, "Financial Fraud Detection Model: Based on Random Forest," *Int. J. Econ. Financ.*, vol. 7, no. 7, 2015, doi: 10.5539/ijef.v7n7p178.

[21] D. Yue, X. Wu, Y. Wang, Y. Li, and C. H. Chu, "A review of data mining-based financial fraud detection research," *2007 Int. Conf. Wirel. Commun. Netw. Mob. Comput. WiCOM 2007*, pp. 5519–5522, 2007, doi: 10.1109/WICOM.2007.1352.

[22] Z. Zhao and T. Bai, "Financial Fraud Detection and Prediction in Listed Companies Using SMOTE and Machine Learning Algorithms," *Entropy*, vol. 24, no. 8, pp. 1–17, 2022, doi: 10.3390/e24081157.

[23] C. Science, "Fraud Detection and Prevention Using Machine Learning Algorithms : A Review," 2021.

[24] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decis. Support Syst.*, vol. 50, no. 3, pp. 559–569, 2011, doi: 10.1016/j.dss.2010.08.006.

[25] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," *Decis. Support Syst.*, vol. 50, no. 3, pp. 602–613, 2011, doi: 10.1016/j.dss.2010.08.008.

[26] J. T. S. Quah and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence," *Expert Syst. Appl.*, vol. 35, no. 4, pp. 1721–1732, 2008, doi: 10.1016/j.eswa.2007.08.093.

[27] R. Jhangiani, D. Bein, and A. Verma, "Machine Learning Pipeline for Fraud Detection and Prevention in E-Commerce Transactions," *2019 IEEE 10th Annu. Ubiquitous Comput. Electron. Mob. Commun. Conf. UEMCON 2019*, pp. 0135–0140, 2019, doi: 10.1109/UEMCON47517.2019.8992993.

[28] P. Ravisankar, V. Ravi, G. Raghava Rao, and I. Bose, "Detection of financial statement fraud and feature selection using data mining techniques," *Decis. Support Syst.*, vol. 50, no. 2, pp. 491–500, 2011, doi: 10.1016/j.dss.2010.11.006.

[29] F. H. Glancy and S. B. Yadav, "A computational model for fi nancial reporting fraud detection," *Decis. Support Syst.*, vol. 50, no. 3, pp. 595–601, 2011, doi: 10.1016/j.dss.2010.08.010.

[30] W. Chen, M. Zhang, and K. Liu, “Machine Learning Techniques for Fraud Detection,” *Int. J. Electron. Commer.*, vol. 25, no. 1, pp. 56–78, 2021, [Online]. Available: https://doi.org/10.1080/10864415.2020.1848012.

[31] J. West and M. Bhattacharya, “Intelligent financial fraud detection: A comprehensive review,” *Comput. Secur.*, vol. 57, pp. 47–66, 2016, doi: 10.1016/j.cose.2015.09.005.

[32] I. D. Mienye and Y. Sun, “A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects,” *IEEE Access*, vol. 10, no. September, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.

[33] B. A. Tama and S. Lim, “Ensemble learning for intrusion detection systems: A systematic mapping study and cross-benchmark evaluation,” *Comput. Sci. Rev.*, vol. 39, p. 100357, 2021, doi: 10.1016/j.cosrev.2020.100357.

[34] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-Augu, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

[35] P. D. A and S. Homayouni, “Bagging and Boosting Ensemble Classifiers for Classification of Comparative Evaluation,” 2021.