Financial Fraud Detection Analysis Using Machine Learning Models

Mehar Kapoor
mehar081bteceai24@igdtuw.ac.in
Indira Gandhi Delhi Technical
University for Women

Kanan Singhal
kanan091btcseai24@igdtuw.ac.in
Indira Gandhi Delhi Technical
University for Women

Disha Jain
disha036bteceai24@igdtuw.ac.in
Indira Gandhi Delhi Technical
University for Women

Tanvi Bisht
tanvi145btcsai22@igdtuw.ac.in
Indira Gandhi Delhi Technical
University for Women

Dr. Shweta Jindal
miss.shweta.singhal@gmail.com
Indira Gandhi Delhi Technical
University for Women

***Abstract*— Growth of the financial market and rising transaction techniques in the market have led to several frauds. A wide range of researched are employed in detecting financial frauds that can take place and analyse them using several models. In the current paper, the machine learning models that have been implemented are Logistic Regression, Decision Trees, Boost, AdaBoost and a 1D Convolutional Neural Network (CNN), to check their efficiency of fraud detection. The paper concludes AdaBoost has the highest accuracy (0.92) and other evaluation metrics (Precision: 0.59, Recall: 0.81 and F1 Score: 0.68). Thus, it can improve the detection efficiency significantly.**

(Keywords: Financial fraud, Logistic regression, Decision Tree, XGBoost, AdaBoost, SMOTE, Keras, Tensorflow, Deep Learning, Machine Learning, 1D Convolutional Neural Network)

I. Introduction

As the global market is expanding, so is the rate of transactions. With the expanding financial landscape and evolution of more arenas in the context of financial fraudulent activities, various new methods have come up which have led to huge losses to the companies across the world. Increasing rate of cybercrimes are a testimony to the increasing rate of financial frauds. These frauds can be a result of carelessness or lack of awareness or a sense of fraud in the corporate executives.

The impact of these can lead to substantial monetary losses for individuals, businesses and government. Several new technologies have come up including Blockchain [1] [2] and Machine Learning [3] [4, 5] [6] [7] models like LSTM techniques [8] , Logistic Regression [9] [10] and K-means clustering [11] which cater to minimizing the fraudulent activities taking place. These models analyze huge transactional data to recognize unusual patterns and irregularities, and then determine the financial fraud taking place in real-time.

The current paper aims to look for innovative solutions to enhance fraud detection capabilities and demonstrate models like [12] Logistic Regression, [13] Decision Trees, XGBoost, AdaBoost techniques. The findings provide actionable insights into strengthening fraud prevention frameworks and improving audit practices.

II. Literature Review

The research [13] analyzed several machine learning techniques being utilized for financial fraud detection, Support Vector Machine (SVM) [13] and Artificial Neural Network (ANN) are frequently used for this purpose. Further, the most common type of fraud is Credit card fraud as per the reviewed studies. The study highlights the assessment rubrics, including [14] accuracy, precision, recall, and F1 score. Imbalanced datasets which are more non-fraudulent than fraudulent transactions pose a challenge, often addressed through oversampling techniques. The research poses a suggestion that in future research, unsupervised methods like clustering for anomaly detection and the use of unstructured data to improve model accuracy should be explored. SVM and ANN are widely adopted for fraud detection, but future research should explore unsupervised learning and unstructured data to enhance detection accuracy.

The research analyzed models like Logistic Regression, SVM, Decision Trees [13], and Neural Networks for detecting financial statement fraud. Based on past reports, various machine learning techniques, with findings indicating that machine learning techniques, particularly SVM and neural networks, often surpass traditional statistical methods in effectiveness. The model accuracy can be further enhanced by using key financial indicators and ratios.

A significant challenge posed by class imbalance (more non-fraudulent cases) is often addressed with oversampling techniques like SMOTE. Furthermore, machine learning shows potential for integration into potential for integration into automated fraud detection systems for greater efficiency and reliability. Machine learning models, especially SVM and Neural Networks, are more effective than statistical methods, and their performance improves further with key financial indicators and techniques like SMOTE to handle class imbalance.

The primary focus of the study [12] is to create a precise model to detect financial scams, specifically focusing on credit card transactions. With the rise of online financial transactions, financial fraud is becoming increasingly common. Traditional methods struggle with large datasets and evolving fraud techniques. The paper describes a fraud detection model working on LSTM [15] which is a type of recurrent neural network (RNN) known for handling time series data along with taking care of the long-term dependencies. Coached on a dataset of [16] credit card transactions, the dataset contained over 280,000 transactions with a small fraction labeled as fraud. The data was pre-processed by normalizing values and separating it into training and [17] testing sets. LSTM cells allows the model to optimize memory usage an store essential information over a period of time, this makes them suitable in recognizing the hidden patterns and detecting fraudulent patterns. This model when compared to traditional models, gives a higher accuracy of 99.95%, proving to be more efficient when working with complex and large datasets, explaining its ability to easily adapt to new fraud patterns.

This study [18] focuses on the combined use Machine Learning and Blockchain to detect manipulations or scams in financial transactions even more effectively. The major problem lies in the complexity of financial fraud hence making it difficult to be detected by conventional methods and static rules. Machine learning helps address this by usage of different types of algorithms to make accurate predictions based on historical data. Furthermore, Blockchain technology creates a secure, unchangeable record of transactions, which helps prevent tampering enabling transparency, as all parties can access the same transaction history. The combined approach includes aspects of both the technologies wherein Machine Learning detects suspicious activity in real-time, while Blockchain securely records these transactions. Smart contracts on Blockchain can automatically enforce rules, making the fraud detection process faster and even more reliable. The data used for detecting fraud includes transaction records, account details, and customer profiles. Integration of Blockchain with the machine learning models provides a secure system to cater to the privacy concerns and security risks. However, the current research offers significant challenges including those of computational powers, scalability issues and working on complex datasets carrying large transaction volumes.

This research [15] studies the use of the Fraud Diamond Theory. This theory includes elements like pressure, opportunity, rationalization and capability in [15] order to detect fraud patterns in data related to property and real estate.

The basis of the research is to detect how these four elements contribute to such activities. The study evaluates that opportunity and capability [15] play a key role in enabling financial frauds. Meanwhile, emphasizing on the need for ethical government frameworks to prevent such actions and hold the people involved in these actions responsible.

This research analyzed [19] the impact of Pentagon fraud theory [19] on real estate, property, and building contractors. Beneish M-Score Model [19] was used to analyse the effect of every feature of this theory. This model uses eight different financial ratios to check a company's records in terms of committing frauds in their transactional statements. The feasibility of the model was monitored through a Goodness and Fit test [19]. These outcomes show the estimation power of the model [15] to predict the probability of the model accuracy of 68.3% [20] wherein 36.5% is tagged as not fraud and 86% of fraud has been predicted by the model. Based on the results, it was concluded that pressure and [17] arrogance have relevance to financial fraud detection whereas opportunity, rationalism, and competence have null effect. The study revealed that pressure and arrogance significantly influence fraud occurrence, while opportunity, rationalism, and competence showed no substantial effect on fraud likelihood.

It proposed an analytical study [21] of several machine learning algorithms, including CATBoost, AdaBoost, LightGBM, and (CNN) for credit card fraud detection. This research examined the importance of various methods for data balancing, including SMOTE which notably improvised the results of the models. The CNN model stood out by achieving the highest ROS-AUC score of 95.25. This depicts its capability in detecting frauds by optimizing ensemble techniques using synthetic oversampling to handle class imbalances.

This study [22] focuses on detection of network attacks using algorithms like XGBoost, including various pre-processing steps and feature extraction. This method efficiently classifies cyberattacks on datasets such as DARPA and KDDcenev, giving a 100% accuracy in several scenarios. This method outperforms other traditional methods, demonstrating its high dimensionality even in non- linearly anomaly scenarios.

This paper [23] explored predictive abilities of three machine learning models including LSTM, Random Forest Regression, Lasso Regression. The main objective was to predict short term stock prices in the Indian Pharmaceutical market and the gas sectors. This was done using the historical data obtained from the laboratories of Dr. Reddy and Gujarat Gas Ltd. Amongst the three models, Random forest gave the best results, MAPE values include 0.6 and 1.47 for both the companies respectively. The research highlights how the ensemble techniques are successful and efficient in time series forecasting as well as fraud detections, even in conditions of temporal dependencies and non-linearities.

The study emphasizes [24] on a comparative analysis between the basic and ensemble machine learning models like Random Forest, XGBoost, LightBGM, AdaBoost and Stacking. The study analysed this through prediction of loan approvals in the financial sector. The results depicted ensemble methods to consistently outperform the traditional methods, wherein Random Forest achieved the highest accuracy and ROC-AUC scores of 88% and 87% respectively

The research analyses [25] and detects malicious nodes in IoT networks using a machine learning based framework. For this purpose, SensorNetGuard dataset was used and four supervised ML models were used, including Random Forest, Decision Trees, [26] Support Vector Machine and K-Nearest Neighbors. As a result of the study, Random Forest was found to achieve the highest accuracy of 99.99%. This result depicts the predictive capability of Random Forest and its efficiency in anomaly detection meanwhile handling class imbalances.

III. Algorithms Used

A. Logistic Regression- it is a probabilistic model used for two-class classification problems, where the output [5] has only 2 possible outcomes. It uses sigmoid function to map values in the range of 0 to 1 which represents probabilities. It is widely used for its simplicity, interpretability and efficiency for linear separable data, but it may struggle with complex relationships.

B. Decision Trees- [27] [28] This algorithm is widely used for problems based on classification and regression. It functions and makes decisions following a tree-like structure, wherein every node is like an assessment of a feature, each branch [26] is the aftermath of the testing process and every leaf node represents the output prediction. Decision trees [17] [29] are effective with non- linear and complex data making it a good fit for financial fraud detection and other pattern recognition problems.

C. XGBoost- This algorithm is a advanced and scalable machine learning technique used by scientists for supervised learning tasks including regression and classification. Uses basics of Decision trees but adds up the concept of gradient boosting. It optimizes the process by adding features like regularization to reduce overfitting, handling missing values internally and employing parallel processing for faster computation. It is capable of handling large datasets, imbalanced classes and complex non-linear relationships.

D. AdaBoost- [28] A strong ensemble learning algorithm used for classifying tasks. It aims at improving accuracy of weak classifiers by combining multiple simple models into a single stronger predictive model. It minimizes errors, making it robust to overfitting. It can be sensitive to noisy data and outliers. Due to its interpretability and solid theoretical foundation in boosting it is widely used for two-class and multi-class classifying problems.

E. 1D CNN [17] - A deep learning model that works in one dimension to find hidden patterns amongst sequences. It puts small filters in form of layers over the input data to look for patterns and learn important features from it, without any manual intervention. It is a suitable model for classification tasks keeping in mind its ability and adaptability.

IV. Experimental Setup

Hardware: Intel Xeon CPU with 2 vCPUs (virtual CPUs) and 13GB of RAM
Software: Python 3.10, Pandas, NumPy, Matplotlib, Seaborn, Scikit-Learn
Operating System: Ubuntu 22

V. Methodology

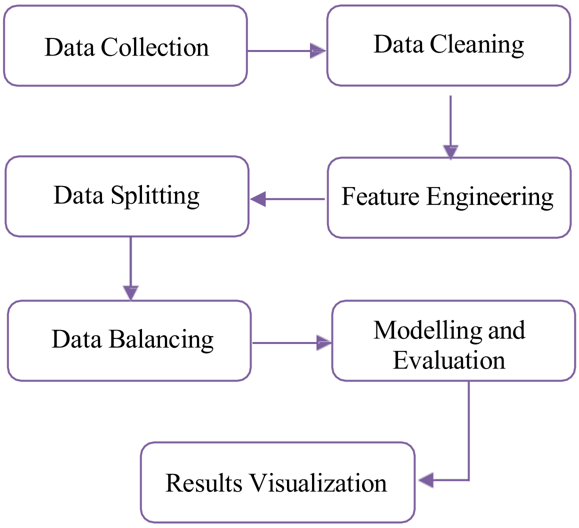

Fig 1. Flowchart of the process

The Fig 1. lists out the flow of the methodological process used in the paper. Furthermore, the dataset utilized in this study is titled "Fraud_Detection_FrostBoost". This was synthetically synthesized to depict real world fraud detection scenarios. It consists of 1,000 records and 20 features, along with a binary target variable indicating the nature of the transactions as fraud or not fraud. The dataset was particularly designed with a 90:10 class imbalance ratio, representing the typical real-world distribution in which fraud cases are rare but critical. The synthetic generation process was inspired by characteristics of real transaction logs, particularly those found in publicly available datasets such as the [29] Kaggle Credit Card Fraud dataset (ULB). Although no real data was used for privacy and compliance reasons.

To address class imbalancing, SMOTE (Synthetic Minority Over-sampling Technique) was utilized and applied on the training set during model development. SMOTE [29] generates new instances of the minority class (fraud) by interpolating between existing ones, allowing classifiers to better learn the patterns of rare events. This process plays a crucial role in boosting recall without compromising the generalization of the models.

i. Feature Engineering

New features are created to evaluate relationships between existing ones or contextual information. The gap feature is created which measures the time difference between Timestamp1 and LastLogin. This feature can reflect the user behavior patterns.

Encoding Categorical Features, Machine learning models cannot directly process categorical data. Encoding transforms categorical data into numeric representations. In this code, Label Encoding is used to allot different integers to categories, beneficial for ordinal data.

ii. Data Splitting

Breaking the acquired data into 2 sets for training and testing is critical for assessing generalization abilities of the model. Training is needed to fit the model, while testing data estimates its performance on hidden data. 80% of the dataset is allotted to the training set (X_train, Y_train). 20% is assigned to the testing set (X_test, Y_test) [17].

This is a common split ratio, balancing the need for sufficient training data and a reliable test set size for evaluation. random_state=42 ensures reproducibility by fixing the random seed for splitting.

iii. Data Balancing

SMOTE (Synthetic Minority Oversampling Technique): It creates artificial samples for the minority class by incorporating available minority samples and their nearest neighbors.

This avoids simple duplication and introduces slight variations, making the dataset more balanced.

If one class (e.g., fraud) is underrepresented, the model may become partial toward the majority class (e.g., non-fraud). This can result in poor recall for the minority class, which is often critical in problems like fraud detection. SMOTE ensures [20] that the training data contains sufficient examples from both classes, allowing the model to learn patterns for both.

iv. Modelling and Evaluation

a. Logistic Regression:

A linear algorithm for two-class classification. [30] It estimates the probability of a class using the cost function. Logistic regression is inherently designed for two class classification tasks, making it suitable for scenarios like fraud detection or identifying suspicious activity.
Data Preparation: To mimic actual fraud detection situations, a synthetic dataset of 1000 samples, 20 characteristics, and a 90:10 class imbalance ratio was developed.
Multiple performance metrics are calculated to evaluate the model.

b. Decision Trees:

Decision trees [20] are used because they are well-suited for the binary classification task at hand. They can comprehend complex, non-linear relationships present in the data, handle both categorical and numerical features without extensive pre-processing, and are robust to outliers.

Their interpretability allows for clear visualization of decision-making paths, which is critical in applications where understanding the rationale behind predictions is important.

Data Preparation: To mimic actual fraud detection situations, a synthetic dataset of 1000 samples, 20 characteristics, and a 90:10 class imbalance ratio was developed.

c. XGBoost Algorithm

The XGBoost algorithm was used in financial fraud detection through performance optimization and the correction of unbalanced data. The approach used was:

1. Data Preparation: To mimic actual fraud detection situations, a synthetic dataset of 1000 samples, 20 characteristics, and a 90:10 class imbalance ratio was developed.
2. Data Splitting: To preserve class distribution, the dataset was splitted into training and testing sets (80:20) [20] using stratified sampling.
3. Handling Class Imbalance: To create a balanced training dataset, SMOTE was utilised to oversample the minority class (fraud cases).
4. Model Configuration and Training: Training data was converted to DMatrix format, and key hyperparameters were set. Objective was set as binary: logistic, evaluation metric was logloss, learning rate was 0.1, maximum depth as 5 and subsample and. Column sampling as 0.8. The optimal number of boosting rounds was determined through 3-fold cross- validation with early stopping.
5. Threshold Customization: A custom probability threshold of 0.4 was applied to classify test set predictions, balancing precision and recall.
6. Evaluation Metrics: The model achieved 92% accuracy, with [20] precision of 59%, recall [17] of 76%, and an F1-score of 67%, effectively identifying fraud while managing the compromise between precision and recall. This approach highlights the effectiveness of XGBoost in handling imbalanced datasets and detecting fraudulent transactions.

d. AdaBoost Algorithm:

1. Data Preparation: A synthetic dataset with 1000 samples, 20 features, and a 90:10 class imbalance (fraud cases) was created.
2. Data Splitting: The dataset was divided 80:20 into training and testing sets implementing stratified random sampling.
3. Class Imbalance Handling [20]: SMOTE was applied to oversample the minority class (fraud cases) in the training set.
4. Model Configuration and Tuning: An AdaBoostClassifier with a DecisionTreeClassifier base estimator (max_depth=1) was used. Hyperparameters like number of estimators, learning rate, and max depth were upgraded using grid search with 3-fold cross-validation. The most suited model was selected based on the F1-score.
5. Model Evaluation: The algorithm was evaluated using Accuracy, Precision, Recall, F1-Score, ROC- AUC, and ROC curve on the test set.
6. Results Visualization: Performance was compared with Logistic Regression, Decision Tree, and XGBoost using evaluation metrics.

e. 1D Convolutional Neural Network:

The 1D CNN [31] algorithm is used in fraud detection to capture hidden insights or patterns from input data by forming layers. The following approach was utilized:

1. Data Preparation: A data which was synthetic in nature and consisted of 1000 samples, 20 features or characteristics in a 90:10 class imbalance ratio were used and worked upon.
2. Data Splitting: The input data was split into training and testing data by stratified sampling method
3. Handling Class Imbalance: In order to ensure a balanced and unbiased input data, SMOTE was used to handle the class imbalance by resampling the minority classes.
4. Model Architecture: It consisted of three layers: Convolutional, Droupout and Dense. Optimization techniques like Adam and cross entropy function were used to enhance the model and learning rate and early stopping was inculcated to avoid overfitting. The model was further tuned with multiple epochs and multiple layers.
5. Evaluation metrics: The model gave a high accuracy of 92%, folloed by a high precision score of 77%. But the resulting recall and f1 score dropped down to 33.3% and 46.7%. This highlights the potential of 1D CNN in feature extraction and finding hidden patterns.

V. RESULT

The Result utilizes evaluation metrics like accuracy, precision, recall and f1 score to compare these predictions made by the algorithms.

Models	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.605	0.083	0.428	0.00
Decision Trees	0.83	0	0	0
XGBoost	0.92	0.59	0.76	0.67
AdaBoost	0.92	0.59	0.81	0.68
1D CNN	0.92	0.77	0.33	0.46

Table 1. Result Analysis Table

The table 1 shows the comparable result parameters used. It depicts the result and helps us to easily compare them. It can be seen that as the model used was enhanced, the results improvised.

The figure 2 depicts the comparison between the four machine learning models. The graph depicts how the four parameters increased from logistic regression to XGBoost.

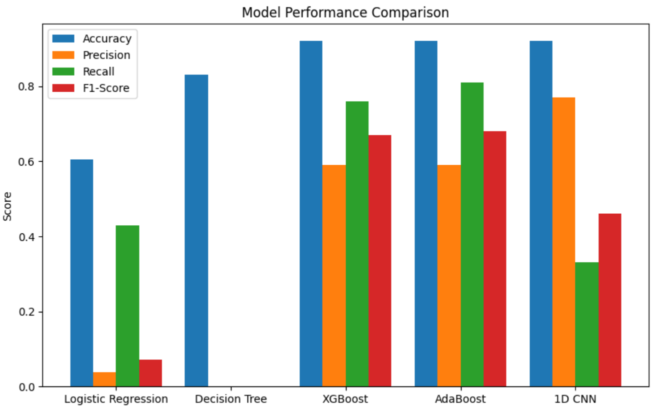

Figure 2: Model Performance Comparison

Afterwards, XGBoost and AdaBoost had the same accuracy and precision but AdaBoost had a higher recall and F1 Score making it better model than XGBoost.

This was followed by the deep learning model, 1D CNN which even though had a high accuracy and precision score as compared to the rest of the models but had a very low recall and f1score resulting in a high number of false positives despite of hyperparameter tuning and optimization. Hence the best possible output was achieved by using AdaBoost that stands as the best model.

VI. CONCLUSION

The investigation deals with how the financial fraudulent activities can be classified using various machine learning models. The process required analyzing a dataset, pre-processing it by ensuring there are no null values. Further it involved setting up the training and testing set-ups for the algorithms, which had to be trained later on.

The research used four models and them compared them to give an analysis on the best model for fraud identification based on four parameters: accuracy, precision, recall and F1 score

Unlike other studies that focus on a single model, the current paper evaluates multiple models—Logistic Regression, Decision Trees, XGBoost, and AdaBoost—providing a broader perspective on financial fraud detection. Prior research based on fraud detection emphasizes models like SVM and ANN, which have limitations with imbalanced datasets. The current study integrates SMOTE for balancing the dataset, improving recall, and F1-score.

The result depicts that AdaBoost was the best predictive model for financial fraud detection with the accuracy of 0.92, precision of 0.59, recall of 0.81 and F1 score 0.68. Even though AdaBoost had similar accuracy and precision to that of XGBoost, it is still considered best because higher recall of AdaBoost indicates that it was able to identify more positive fraud cases, leaving less number of undetected cases.

Further a higher F1 score of AdaBoost is primarily due to its hard-to-classify instances. AdaBoost assigns higher weights to misclassified data, which gives it an edge in prediction over XGBoost. Thus, AdaBoost is the best model for financial fraud detection, out of the four models tested.

VII. References

[1] L. Liu, W.-T. Tsai, M. Z. A. Bhuiyan, H. Peng and M. Liu, ""Blockchain-enabled fraud discovery through abnormal smart contract detection on Ethereum."," in *Future Generation Computer Systems 128*, 2022.

[2] S. Tatineni, " "Enhancing Fraud Detection in Financial Transactions using Machine Learning and Blockchain."," in *International Journal of Information Technology and Management Information Systems (IJITMIS) 11, no. 1*, 2020.

[3] A. Oza, ""Fraud detection using machine learning."," in *Transfer 528812, no. 4097*, 2018.

[4] Ashtiani, M. N. and B. Raahemi, ""Intelligent fraud detection in financial statements using machine learning and data mining: a systematic literature review."," 2021.

[5] K. Bah1, A. W. Jallow2, A. N. Bah3 and M. Touray, "Developing a Machine Learning Algorithm for Improved Management of Congestive Heart Failure Patients in the Emergency Department," in *Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei 11031, Taiwan*, 2023.

[6] R. T. Potla, ""AI in fraud detection: leveraging real- time machine learning for financial security."," in *Journal of Artificial Intelligence Research and Applications 3, no. 2*, 2023.

[7] E. Pan, ""Machine Learning in Financial Transaction Fraud Detection and Prevention."," in *Business and Management Research 5*, 2024.

[8] Benchaji, Ibtissam, S. Douzi and B. E. Ouahidi, ""Credit card fraud detection model based on LSTM recurrent neural networks."," in *Journal of Advances in Information Technology 12, no. 2*, 2021.

[9] Talenti, Luca, M. Luck, A. Yartseva, N. Argy, S. Houzé and C. Damon, ""L1 logistic regression as a feature selection step for training stable classification trees for the prediction of severity criteria in imported malaria," in *arXiv preprint*, 2015.

[10] Arsov, Nino, M. Pavlovski and L. Kocarev, ""Stability of decision trees and logistic regression."," in *arXiv preprint arXiv:1903.00816*, 2019.

[11] Huang, Zengyi, H. Zheng, C. Li and C. Che, ""Application of machine learning-based k-means clustering for financial fraud detection."," in *Academic Journal of Science and Technology 10, no. 1*, 2024.

[12] Alghofaili, Yara, A. Albattah and a. M. A. Rassam, ""A financial fraud detection model based on LSTM deep learning technique."," in *Journal of Applied Security Research 15, no. 4*, 2020.

[13] R. Rao and V. Mandhala, "Unveiling financial fraud: A comprehensive review of machine learning and data mining techniques.," in *Ingénierie des Systèmes d'Information, Vol. 29, No. 6*, 2024.

[14] J. P. Singh, M. P. Singh, A. K. Singh, S. Mukhopadhyay, J. K. Mandal and P. Dutta, "Computational Intelligence in Communications and Business Analytics," in *6th International Conference, CICBA 2024*, Patna, 2024.

[15] N. D. Rahmatika, M. D. Kartikasari, D. Indriasih, I. A. Sari and A. Mulia., ""Detection of Fraudulent Financial Statement; Can Perspective of Fraud Diamond Theory be applied to Property, Real Estate, and Building Construction Companies in Indonesia?."," in *European Journal of Business and Management Research 4, no. 6*, 2019.

[16] D. Ramadhani, I. Rachmawati, ,. D. N. Cahyaningsih, H. Ayuningtias, A. Gunawan and D. (. Dennyra, "Acceleration of Digital Innovation & Technology towards Society 5.0," in *Proceedings of the International Conference on Sustainable Collaboration in Business, Information and Innovation*, indonesia, 2022.

[17] R. J. Mohan, B. R. K. Raju, V. C. Sekhar and T. V. K. Prasad, "Algorithms in Advanced Artificial Intelligence," in *CRC Press*, 2025.

[18] Mehbodniya, Abolfazl, I. Alam, S. Pande, R. Neware, K. P. Rane, M. Shabaz and M. V. Madhavan, ""[Retracted] Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques."," in *Security and Communication Networks 2021, no. 1*, 2021.

[19] A. M. Rahmatika, D. A. Sari, R. P. Nugroho, &. Wijayanti and L. P., "Application of Fraud Pentagon Theory in Detecting Financial Statement Fraud: Evidence from Indonesia's Property and Construction Sector.," in *Journal of Financial Crime Studies*, 2021.

[20] R. J. Mohan, B. R. K. Raju, V. C. Sekhar and T. V. K. P. Prasad, "Algorithms in Advanced Artificial Intelligence," London, 2025.

[21] S. Singh, B. Hazel, P. Asthana, S. Barnwal and G. Srivastava, "Deceptive Plastic: Unveiling the World of Financial Fraud," *Procedia Computer Science,* pp. 416-423, 2025.

[22] S. Huang, J. Cao, Y. Yang, X. Du, J. Yuan and K. Yu, "Research on intelligent detection of network attack based on XGBoost," *Procedia Computer Science,* pp. 1130-1136, 2025.

[23] P. K. Sarangi, E. Singh, A. Kaushal, B. Sharma, M. Dutta and V. P. Dubey, "Analysing the fluctuations in commodity prices and forecasting the future directions using machine learning

techniques," *Procedia Computer Science, 259,* pp. 1827-1836, 2025.

[24] L. Hota, P. K. Jain and A. Kumar, "A comparative performance assessment for prediction of loan approval in financial sector," *Procedia Computer Science,* pp. 298-307, 2025.

[25] V. K. Pandey, S. Prakash, S. K. Jha, T. Yang and R. S. Rathore, "An efficient and robust framework for IoT security using machine learning techniques," *Procedia Computer Science,* pp. 118-124, 2025.

[26] E. A. Kozyrev, E. A. Ermakov, A. S. Boiko, I. A. Mednova, E. G. Kornetova, N. A. Bokhan and S. A. Ivanova, "Building predictive models for schizophrenia diagnosis with peripheral inflammatory biomarkers," 2023.

[27] Anyaebosi, 2. A Chioma1, Onyedikachi4, G. O. S. E. Favour, Rosita5, E. Chinyere, Egbe6, P. Njock, Adam7, F. Muhammad, Atisu8, J. C, Khan9, A. Mahmud, Doe10 and Francis, "An Intelligent Fraud Detection model for Oil and Gas Financial Statements Using Machine Learning & Big Data Mining," in *Global Scientific Journal*, 2024.

[28] B. Unhelkar, H. M. Pandey, A. P. Agrawal and A. Choudhary, "Advances and Applications of Artificial Intelligence & Machine Learning," in *Proceedings of ICAAAIML 2022*, 2023.

[29] H. Ren, Y. Zheng, C. Li, F. Jing, Q. Wang, Z. Luo and W. Cheng, "Using Machine Learning to Predict Cognitive Decline in Older Adults From the Chinese Longitudinal Healthy Longevity Survey: Model Development and Validation Study," in *JMIR aging*, 2025.

[30] H. L. Gururaj, F. Flammini, R. Kumar and N. S. & Prema, "Recent Trends in Healthcare Innovation," in *Proceedings of the Annual International Conference on Recent Trends in Healthcare Innovation (AICRTHI 2024)*, London, 2025.

[31] H. L. Gururaj, F. Flammini, S. Srividhya, M. L. Chayadevi and S. Selvam, "Computer Science Engineering: Proceedings of the 1st International Conference on Computing and Intelligent Information Systems," Banglore, 2024.

[32] Anyaebosi, 2. A Chioma1, G. O. S. E. F. Onyedikachi4, E. C. Rosita5, P. N. Egbe6, F. M. Adam7, J. C. Atisu8, A. M. Khan9 and F. Doe10, "An Intelligent Fraud Detection model for Oil and Gas Financial Statements Using Machine Learning & Big Data Mining," in *Global Scientific Journal*, 2024.

[33] R. Al-Haddad, F. Sahwan, A. Aboalmakarem, G. Latif and Y. M. Alufaisan, "Email text analysis for fraud detection through machine learning techniques," in *3rd Smart Cities Symposium*, 2020.

[34] A. Ali, S. A. Razak, S. H. Othman, T. A. E. Eisa, A. Al-Dhaqm, M. Nasser, T. Elhassan, H. Elshafie and A. Saif, ""Financial fraud detection based on machine learning: a systematic literature review."," in *Applied Sciences 12, no. 19*, 2022.

[35] J. L. Perols and A. C. Dzuranin, ""A picture is worth a thousand words: Using interactive data visualization to assess fraud risk."," in *Journal of Forensic Accounting Research 6, no. 1*, 2021.