A Financial Fraud Detection Model Using Artificial Intelligence

Khaleed Omair Alotaibi
College of Business, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia
komalotaibi@imamu.edu.sa (corresponding author)

Received: 30 October 2025 | Revised: 14 November 2025, 25 November 2025, and 27 November 2025 | Accepted: 28 November 2025

 DOI: https://doi.org/10.48084/etasr.15892

ABSTRACT

Finance, government, corporate, and consumer fraud all have broad effects on society, and as the use of new technologies increases, this problem is worsening. In the era of big data, traditional methods for detecting anomalies are slow, costly, inaccurate, inefficient, and impractical. There is a growing trend among financial institutions to adopt statistical and computational techniques in their operations. However, a comprehensive framework that integrates the entire AI lifecycle for adaptive fraud detection is still needed. This study presents, implements, and validates a five-phase Financial Fraud Detection Model (FFDM) framework. A prototype was developed and evaluated using a public dataset, comparing multiple machine learning models. The results show that the framework is operational and that an SVM model integrated into the FFDM achieved the best baseline performance for this task, demonstrating the model's practical feasibility.

Keywords-machine learning; fraud detection; artificial intelligence; design science

I. INTRODUCTION

Financial fraud poses a significant and escalating threat to global economic stability, corporate integrity, and consumer trust. The adoption of new digital technologies, while beneficial, has simultaneously expanded the attack surface for fraudulent activities, making them more sophisticated and more complex to detect. Traditional detection methods, such as manual auditing and rule-based systems, are increasingly inadequate, as they are often slow, costly, inefficient, and impractical for analyzing the vast volumes and complexity of modern financial data in real-time, leading to substantial economic losses and eroded confidence.

In response to these challenges, the financial industry has turned towards data mining and Artificial Intelligence (AI). AI, with its ability to imitate human cognitive functions, such as pattern recognition and logical inference, and to learn and improve over time, offers a significant paradigm shift. Machine Learning (ML) models can analyze large-scale transaction data to identify subtle and complex anomalies that would be impossible for human auditors or rigid rule-based systems to detect. Consequently, banks and financial institutions are actively investing in statistical and computational techniques to enhance their security frameworks.

Many existing AI-driven solutions focus on isolated components of the fraud detection pipeline. A significant gap remains in the development of a comprehensive end-to-end framework that seamlessly integrates the entire AI lifecycle from data ingestion and preprocessing to real-time decision-making and continuous model adaptation. Without this holistic approach, systems struggle to remain effective against rapidly evolving fraud tactics. This study focuses on this gap by designing, implementing, and evaluating a Financial Fraud Detection Model (FFDM) using AI. The operability of the proposed FFDM was demonstrated by building a prototype, testing it on a real-world dataset, benchmarking multiple AI models, and establishing a functional pipeline for real-time scoring and continuous learning. This end-to-end implementation provides a blueprint for practitioners and a validated artifact for researchers.

Many studies have shown that billions of dollars are lost or diverted from the banking and financial system each year due to negligence, omissions, and acts of commission. As technology advances, financial crimes and economic offenses have migrated online. To ensure the security of their business operations and assets, many organizations use AI-driven solutions and algorithms. According to [1], the likelihood that financial statements are manipulated can be analyzed using a model that examines related factors, such as the motives of those in power, the conditions that allow frauds to occur, and the attitude or values of people that can influence their likelihood of engaging in financial fraud. AI techniques can support both the mental capabilities of machines and humans [2], as they can handle large datasets and demonstrate impressive results [3]. Many cyberattacks focus on denial-of-service, infrastructure, and data protection issues, with 70% of banking and capital markets CEOs seeing cybersecurity threats as a critical challenge [4]. Security breaches in financial institutions are more than 300 times as many as in other service sectors.

The advancement of sophisticated technology has facilitated the more effective identification of intricate

fraudulent activities, thus improving digital performance within the e-commerce and banking sectors [5]. SVM-based methods reduce the time to fraud detection by minimizing the need for human analysis [6]. Furthermore, leveraging AI technologies for money laundering detection and prevention provides several non-financial advantages. The study in [7] focused on papers that demonstrated detection efficiency using existing financial datasets, outlining methods, algorithms, and validation techniques, and examining advantages and limitations. The study in [8] discussed how regulators, internet companies, and financial institutions are working together to develop a robust ecosystem to keep up with fraud schemes that are evolving at a rapid pace. According to [9], AI and ML have the potential to assist the U.S. financial industry in identifying and stopping money laundering and fraud.

In [10], multi-domain fraud detection methods for enterprise settings were described, demonstrating the advantages that modern learning methods can provide for pattern recognition and anomaly identification. In [11], the effectiveness of advanced ML techniques in detecting fraud in the banking industry was demonstrated, evaluating models against LightGBM, XGBoost, CatBoost, vote classifiers, and neural networks on a comprehensive dataset of banking transactions. In [12], the accuracy, sensitivity, specificity, precision, F-measure, and AUC of Logistic Regression (LR), Naive Bayes (NB), and K-Nearest Neighbor (KNN) algorithms were determined using three proportions of datasets, with skewed datasets undersampled randomly. Several digital forensic methods have been proposed to detect and investigate fraudulent activities within organizations. The use of digital forensics can be a potent tool in helping organizations identify potential fraud schemes, uncover financial misconduct, and gather legally sound evidence. Numerous works have been proposed to detect various forms of cyber-fraud, data breaches for monetary gain, and other digital risks [13, 14].

II. TYPES OF FINANCIAL FRAUD

There are many types of financial fraud [15], and some of the most significant are the following.

A. Credit Card Fraud

Unauthorized use of a credit card for fraudulent transactions does not require the user's knowledge [16, 17]. Several methods are used to obtain cardholder information. Fraudsters impersonate finance officials to trick users into revealing their details during phishing, use swipers and skimmers to read credit card information directly from ATMs or POS devices, or access whole databases by breaching the financial institution's network security or collaborating with a company insider [17, 18]. The user's card might also be intercepted in the mail containing a new or replacement card [18]. These remote methods have increased credit card fraud due to their anonymity and ease of access [17]. Credit card fraud can be detected by analyzing a customer's normal spending patterns and flagging transactions that differ significantly.

B. Securities and Commodities Fraud

Financial institutions sometimes employ fraudulent investment strategies to fool investors into investing in companies based on false or misleading information. This category includes pyramid schemes, Ponzi schemes, hedge fund fraud, foreign exchange fraud, and embezzlement [16].

C. Financial Statement Fraud

A company issues different types of financial statements to provide information on expenses, loans, income, and profits. As a result, these reports may also include management comments on the company's performance and future risks [19, 20]. It is helpful to have these financial statements available, as they offer a comprehensive overview of the company's status and can be used to measure success, influence stock prices, and determine loan eligibility [19]. There are instances of financial statement fraud, also known as corporate fraud, where these reports are manipulated to give a falsely improved impression of a business's profitability. Such fraud occurs to boost stock performance, lower tax burdens, or overstate results under management's influence. Due to limited industry knowledge, the rarity of financial statement fraud, and the fact that it is carried out by insiders skilled in deception, detecting it is often challenging [21].

D. Insurance Fraud

Insurance fraud can occur at any stage of the insurance process and by anyone involved. False claims can include exaggerated losses, injuries, or completely fake events. Automobile insurance fraud involves fabricating accidents or causing crashes to inflate repair and injury costs. In addition, some frauds occur on a larger scale, such as in the crop insurance industry, where the claimant overstates losses due to falling crop prices or natural disasters. In addition to paying excessive insurance premiums, submitting duplicate claims, giving rebates to brokers, or upcoding items, there are other types of insurance fraud [16].

E. Mortgage Fraud

Fraudulent manipulation of mortgage documents and property is a specific type of financial fraud that has been documented numerous times. There are several ways in which a lender can misrepresent the value of a property to influence financial institutions to lend on it [16].

F. Money Laundering

A money laundering scheme involves criminals infiltrating proceeds from illicit enterprises into legitimate companies to hide the source of the funds. The purpose of this tactic is to conceal the origin of the money, give it the appearance of legitimacy, and make it difficult to track the money trail of these criminals. As a result of money laundering, criminals can invest in their own businesses and become influential in the economy [16].

III. METHODOLOGY AND IMPLEMENTATION

This study employs the Design Science Research Methodology (DSRM) to develop and evaluate the proposed FFDM. The process consisted of two primary stages: Development and Implementation & Validation.

A. Development of the FFDM Framework

Based on a synthesis of the literature and identified gaps, the proposed FFDM comprises five integrated phases, as shown in Figure 1.

1) Data Ingestion and Aggregation

Raw data is gathered from various sources. These resources are collected in a central pool that includes third-party services, transaction logs, and customer profiles. This data aims to train the AI model to learn about the different patterns of malicious activities, generate hypotheses, and make decisions.

2) Data Preprocessing and Feature Engineering

Raw field data is cleaned and transformed into predictive signals through a mathematical algorithm. Detecting anomalies allows for the identification of suspicious activities that deviate from typical behavior patterns. In real-time, this solution protects the system from incidents that could lead to substantial financial losses, data breaches, and other malicious events that can also cause a notable decline in productivity.

3) Core AI/ML Engine

A key feature of the FFDM is its ability to localize fraud, achieved through a combination of supervised ML to detect known fraudulent samples and unsupervised ML to identify new malicious activities. In contrast, DL models establish the necessary connections for comprehensive detection.

4) Real-Time Scoring and Decision Engine

The FFDM can quickly determine whether malicious samples should be removed from the system based on the information they contain. In addition to real-time analysis, sample collection, AI model reasoning, and the generator model's likelihoods, it includes several other activities. Following this, a decision is made based on the activities that have occurred to this point. Depending on the fraud score, the transaction will be blocked if it exceeds 0.9; if it exceeds 0.7, it will trigger a challenge, such as two-factor authentication; and if it is below 0.7, it will be accepted, indicating that no fraud has been detected.

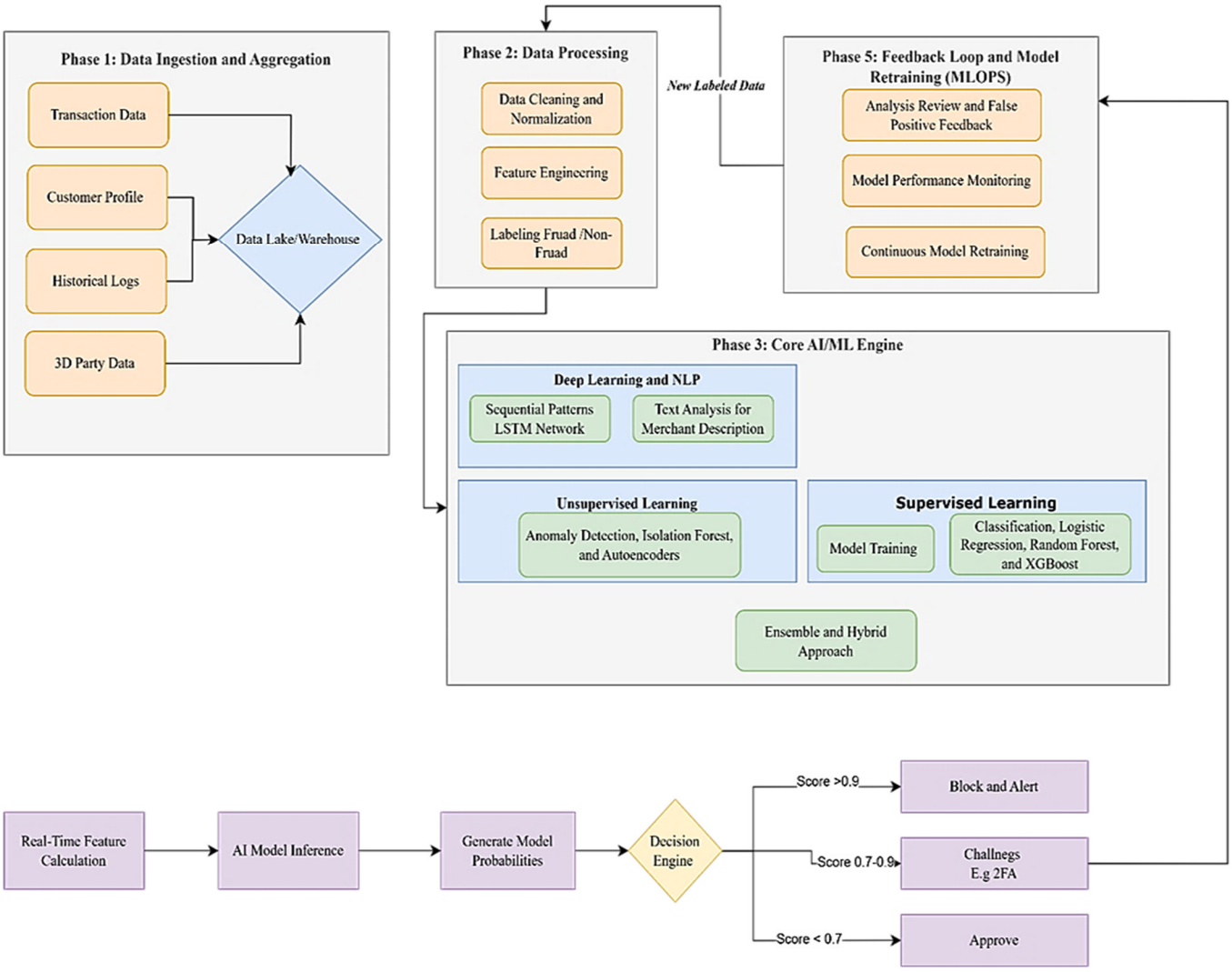

Fig. 1. The proposed FFDM using AI.

5) Feedback Loop and Model Retraining (MLOPS)

The individual's expert response and the execution metrics of the approach are continually fed back into the system. As a result, the models are retrained and fine-tuned to develop an adaptive system capable of addressing emerging illegal policies as they occur.

B. Implementation and Validation of the FFDM

To validate the FFDM's practicality and performance, a working prototype was implemented and evaluated on a financial transaction dataset. The goal was to test the operability of the entire five-phase pipeline, not just the performance of a single algorithm.

1) Dataset

The publicly available Credit Card Fraud Detection dataset [22] was selected for evaluation because it is a recognized benchmark that presents a realistic challenge with severe class imbalance (0.172% fraud) [23-25]. To reduce computational resources for this initial validation, a representative sample of 10,000 transactions was used, preserving the original fraud ratio. The data was split into 70/30 train/test sets, stratifying to maintain the fraud ratio in both subsets, a crucial step for reliably evaluating model performance on imbalanced data.

2) Experimental Setup

The system was implemented using Python with standard libraries (e.g., Scikit-learn, Pandas).

3) Implementation of FFDM Phases

- Phases 1 & 2: The dataset is ingested into the pipeline. The features were already preprocessed via PCA in the original dataset. A quality data check confirmed that no missing values were present. While the PCA features were provided, additional feature engineering was performed to derive eight final features for model training.
- Phase 3 - Core AI/ML Models and Training: Four algorithms were trained and evaluated to select the best model for the real-time engine: LR as a simple baseline, RF for its robustness with imbalanced data, SVM for its effectiveness in high-dimensional spaces, and an unsupervised Isolation Forest (IS) to test anomaly detection without labels. These algorithms were selected to evaluate both supervised and unsupervised approaches under the same framework. The models were trained on the preprocessed training set.
- Phase 4 - Real-Time Scoring Engine: The best-performing model (SVM, based on AUC score) was deployed into a simulated real-time decision engine. Decision thresholds (Block > 0.9, Challenge > 0.7, Accept < 0.7) were defined based on a standard industry practice that balances fraud prevention with customer experience. The logic was applied to the model's output scores to demonstrate how the framework would function in a live environment.
- Phase 5 - Feedback Loop: The MLOps feedback loop was integrated within the proposed framework, designing the data flow for continuous retraining. For this initial validation, continuous retraining was not performed; however, the structure is in place to incorporate new transaction data and investigator feedback to adapt the model to emerging fraud patterns in future work.

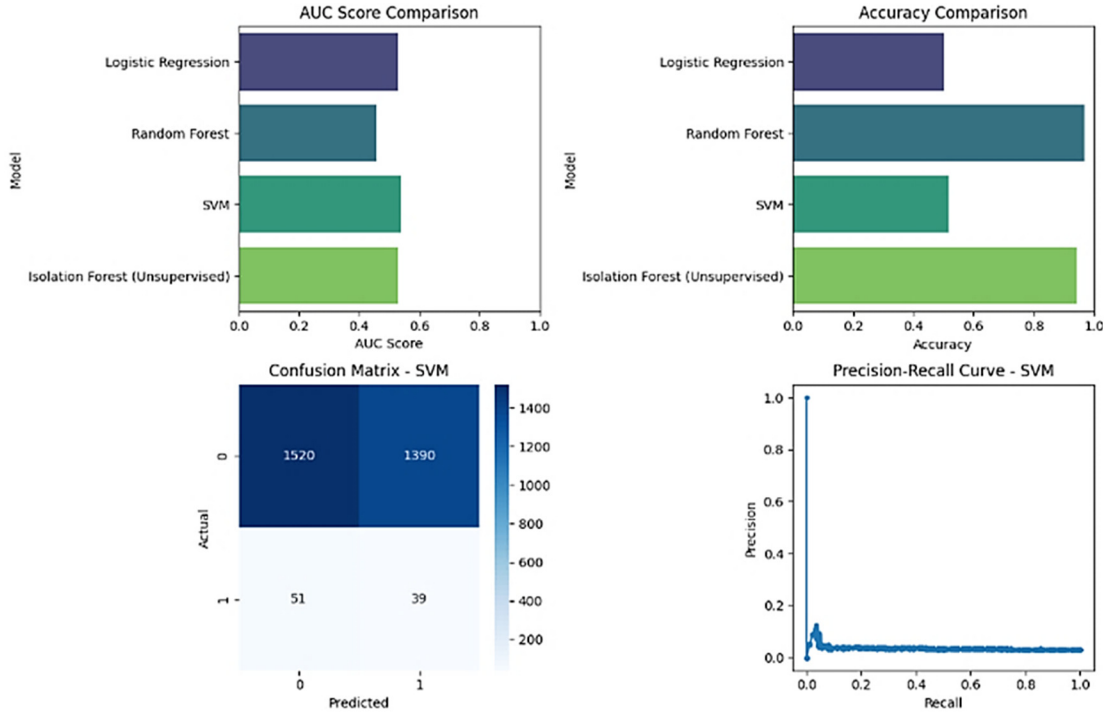

Fig. 2. (a) AUC score comparison, (b) Confusion matrix for SVM, (c) Accuracy comparison, (d) Precision-Recall curve for SVM.

IV. DISCUSSION

The implementation of the FFDM prototype yielded several key findings regarding both model performance and framework functionality.

1) Model Performance and Selection

Among the models tested, SVM demonstrated the highest discriminative power, although modest, with an AUC of 0.538. Consequently, it was selected as the core of the real-time decision engine. The model has a particularly notable precision (0.97) on legitimate transactions, which is critical for minimizing false positives and avoiding customer disruption. However, the recall of 0.43 for fraudulent transactions underscores the well-known challenge of detecting fraud in highly imbalanced datasets, highlighting an area for future improvement of the framework, such as experimenting with cost-sensitive learning or advanced sampling techniques.

B. Framework Operability

An important outcome of this study lies in the demonstration that the proposed FFDM is practically implementable, presenting an end-to-end pipeline, from raw data ingestion through to a functioning real-time scoring engine, proving the feasibility of integrating these often siloed phases into a cohesive, automated system.

V. CONCLUSION

The effects of fraud on society are broad and span finance, government, corporations, and consumers. In recent years, this problem has become increasingly severe due to advances in technology. Anomalies are often difficult to detect using traditional methods due to their slowness, high cost, inefficiency, and impracticality. Statistics and computation are becoming increasingly important to financial institutions. This study used AI to develop an FFDM in the following phases: Data Input, Data Preprocessing, Feature Engineering, Core AI/ML Engine, Real-Time Scoring and Decision Engine, and Feedback Loop and Model Retraining.

Despite promising results, this study has several limitations. Implementation and validation were conducted on a synthesized dataset with a specific fraud ratio, which may not fully capture the complexity and evolving patterns of real-world financial data. Furthermore, the computational resources required for the continuous retraining (MLOps) phase in a large-scale production environment were not extensively evaluated. Finally, the model's performance against coordinated, multi-vector fraud attacks remains an area for further testing. Future work will validate the effectiveness of FFDM development in real-world scenarios, such as actual transactions or datasets, and examine further adjustments to improve performance.

ACKNOWLEDGMENT

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2502).

REFERENCES

[1] J. K. Loebbecke and J. Willingham, "Review of SEC accounting and auditing enforcement releases," Unpublished Working Paper, 1998.

[2] S. Hasham, S. Joshi, and D. Mikkelsen, "Financial crime and fraud in the age of cybersecurity," McKinsey & Company, vol. 2019, 2019.

[3] O. Kaya, "Artificial intelligence in banking: A lever for profitability with limited implementation to date," Deutsche Bank Research, 2019.

[4] A. Vieira and A. Sehgal, "How Banks Can Better Serve Their Customers Through Artificial Techniques," in *Digital Marketplaces Unleashed*, C. Linnhoff-Popien, R. Schneider, and M. Zaddach, Eds. Berlin, Heidelberg: Springer, 2018, pp. 311–326.

[5] A. B. Malali and S. Gopalakrishnan, "Application of artificial intelligence and its powered technologies in the indian banking and financial industry: An overview," *IOSR Journal of Humanity and Social Sciences*, vol. 25, no. 4, pp. 55–60, 2020.

[6] G. Kumar, C. B. Muckley, L. Pham, and D. Ryan, "Can alert models for fraud protect the elderly clients of a financial institution?," *The European Journal of Finance*, vol. 25, no. 17, pp. 1683–1707, Nov. 2019, https://doi.org/10.1080/1351847X.2018.1552603.

[7] D. Choi and K. Lee, "An Artificial Intelligence Approach to Financial Fraud Detection under IoT Environment: A Survey and Implementation," *Security and Communication Networks*, vol. 2018, no. 1, 2018, Art. no. 5483472, https://doi.org/10.1155/2018/5483472.

[8] M. Narender and A. J. Anand, "Artificial Intelligence in Financial Fraud Detection," in *Handbook of AI-Driven Threat Detection and Prevention*, CRC Press, 2025.

[9] V. Boateng, E. K. Amoako, O. Ajay, and T. K. Adukpo, "Harnessing Artificial Intelligence for combating money laundering and fraud in the U.S. financial industry: A comprehensive analysis," *Finance & Accounting Research Journal*, vol. 7, no. 1, pp. 37–49, Feb. 2025, https://doi.org/10.51594/farj.v7i1.1814.

[10] M. M. Ismail and M. A. Haq, "Enhancing Enterprise Financial Fraud Detection Using Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 14854–14861, Aug. 2024, https://doi.org/10.48084/etasr.7437.

[11] U. Detthamrong, W. Chansanam, T. Boongoen, and N. Iam-On, "Enhancing Fraud Detection in Banking using Advanced Machine Learning Techniques," *International Journal of Economics and Financial Issues*, vol. 14, no. 5, pp. 177–184, Sept. 2024, https://doi.org/10.32479/ijefi.16613.

[12] F. Itoo, M. Singh, and S. Singh, "Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection," *International Journal of Information Technology*, vol. 13, no. 4, pp. 1503–1511, Aug. 2021, https://doi.org/10.1007/s41870-020-00430-y.

[13] A. Alshammari, "Detection and Investigation Model for the Hard Disk Drive Attacks using FTK Imager." *International Journal of Advanced Computer Science and Applications*, vol. 14, no.7, 2023

[14] F. Ullah, C. M. Pun, O. Kaiwartya, A. S. Sadiq, J. Lloret, and M. Ali, "HIDE-Healthcare IoT Data Trust ManagEment: Attribute centric intelligent privacy approach," *Future Generation Computer Systems*, vol. 148, pp. 326–341, Nov. 2023, https://doi.org/10.1016/j.future.2023.05.008.

[15] J. West and M. Bhattacharya, "Intelligent financial fraud detection: A comprehensive review," *Computers & Security*, vol. 57, pp. 47–66, Mar. 2016, https://doi.org/10.1016/j.cose.2015.09.005.

[16] 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," *Decision Support Systems*, vol. 50, no. 3, pp. 559–569, Feb. 2011, https://doi.org/10.1016/j.dss.2010.08.006.

[17] 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, Feb. 2011, https://doi.org/10.1016/j.dss.2010.08.008.

[18] J. T. S. Quah and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence," *Expert Systems with Applications*,

vol. 35, no. 4, pp. 1721–1732, Nov. 2008, https://doi.org/10.1016/j.eswa.2007.08.093.

[19] P. Ravisankar, V. Ravi, G. R. Rao, and I. Bose, "Detection of financial statement fraud and feature selection using data mining techniques," *Decision Support Systems*, vol. 50, no. 2, pp. 491–500, Jan. 2011, https://doi.org/10.1016/j.dss.2010.11.006.

[20] F. H. Glancy and S. B. Yadav, "A computational model for financial reporting fraud detection," *Decision Support Systems*, vol. 50, no. 3, pp. 595–601, Feb. 2011, https://doi.org/10.1016/j.dss.2010.08.010.

[21] S. Maes, K. Tuyls, B. Vanschoenwinkel, and B. Manderick, "Credit card fraud detection using Bayesian and neural networks," in *Proceedings of the 1st International Naiso Congress on Neuro Fuzzy Technologies*, 2002, Art. no. 270.

[22] "Credit Card Fraud Detection." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud.

[23] A. D. Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating Probability with Undersampling for Unbalanced Classification," in *2015 IEEE Symposium Series on Computational Intelligence*, Cape Town, South Africa, Sept. 2015, pp. 159–166, https://doi.org/10.1109/SSCI.2015.33.

[24] A. Dal Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," *Expert Systems with Applications*, vol. 41, no. 10, pp. 4915–4928, Aug. 2014, https://doi.org/10.1016/j.eswa.2014.02.026.

[25] B. Lebichot, G. M. Paldino, W. Siblini, L. He-Guelton, F. Oblé, and G. Bontempi, "Incremental learning strategies for credit cards fraud detection," *International Journal of Data Science and Analytics*, vol. 12, no. 2, pp. 165–174, Aug. 2021, https://doi.org/10.1007/s41060-021-00258-0.