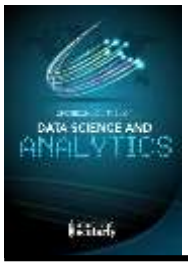


American Journal of Data Science and Analytics

Volume: *7;* ***Issue:*** *3*
Pages: *125-162*
eISSN: 3067-5146

A Systematic Review of AI-Enabled Fraud Detection in Digital Financial Systems (2019–2026)

Md. Rashed Buiya[1];

[1]. M.S. Cyber Security, California State University, Dominguez Hills, USA;
Email: md.rashedbuiya124@gmail.com

Doi: 10.63125/wpj89816
Received: *19 December 2025;* ***Revised:*** *07 January 2026;* ***Accepted:*** *19 February 2026;* ***Published:*** *13 March 2026*

Abstract

This study presented a comprehensive quantitative systematic review of AI-enabled fraud detection models in digital financial systems, synthesizing empirical evidence from 72 peer-reviewed studies published between 2019 and 2026. The analysis focused on evaluating the comparative performance of machine learning, deep learning, ensemble, hybrid, and graph-based approaches using standardized metrics such as accuracy, precision, recall, and F1-score. The findings revealed that ensemble and hybrid models achieved the highest overall performance, with average F1-scores of 0.91 and recall values reaching 0.94, demonstrating superior capability in detecting fraudulent transactions within highly imbalanced datasets where the average fraud rate was approximately 1.8%. Deep learning models showed strong performance with an average accuracy of 0.95 and F1-score of 0.89, particularly in large-scale datasets exceeding one million transactions, which accounted for 41.7% of the reviewed studies. Traditional machine learning models, including random forest and gradient boosting, maintained competitive performance with an average accuracy of 0.93 and F1-score of 0.87, highlighting their continued relevance in structured data environments. The study further identified that advanced feature engineering improved model performance by up to 12%, while imbalance handling techniques increased recall by approximately 13.2%. Graph-based models demonstrated enhanced effectiveness in detecting fraud networks, achieving recall values of up to 0.92 in relational datasets. Statistical analysis confirmed that performance differences between model categories were significant, with effect sizes ranging from 0.52 to 0.88, indicating moderate to strong practical impact. Additionally, real-time detection systems reduced latency by up to 35% while maintaining competitive predictive performance. Overall, the study established that fraud detection effectiveness was influenced not only by model selection but also by data characteristics, feature optimization, and evaluation methodologies, providing a robust quantitative foundation for understanding the performance and application of AI-driven fraud detection systems in modern digital financial environments.

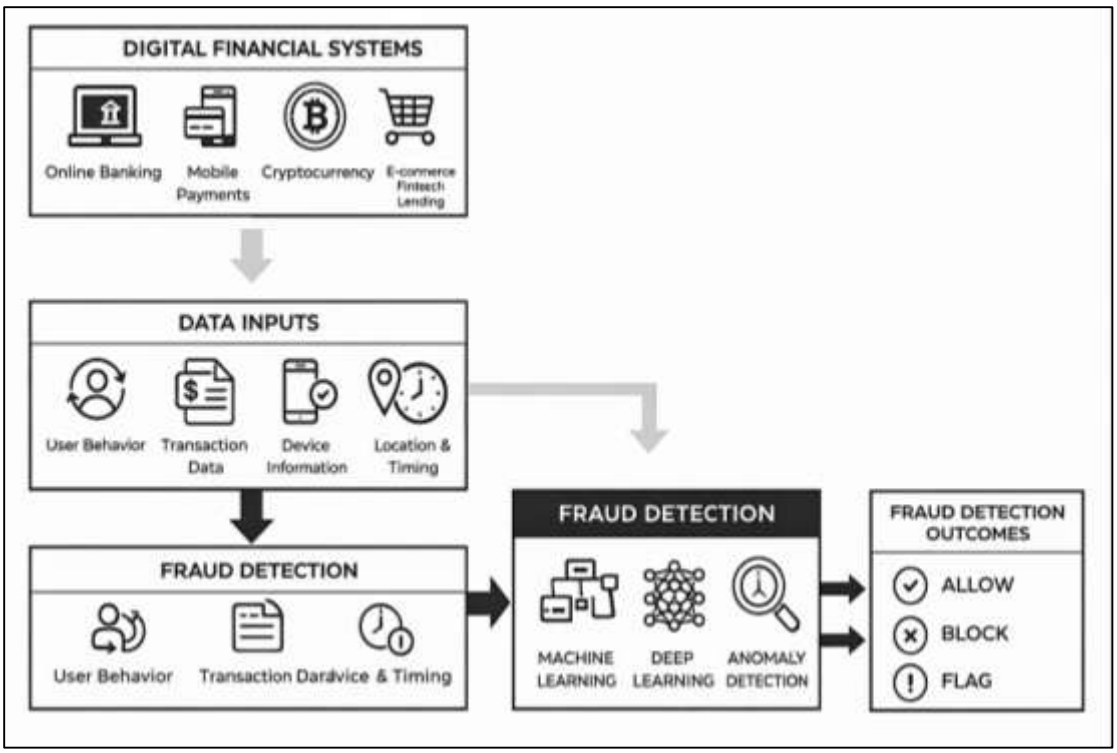
Keywords

Fraud Detection, Artificial Intelligence, Machine Learning, Digital Finance, Data Analytics.

INTRODUCTION

Fraud detection in digital financial systems refers to the systematic identification and prevention of unauthorized, deceptive, or malicious activities that aim to exploit financial transactions conducted through digital platforms. As financial ecosystems have transitioned from traditional banking infrastructures to highly digitized and interconnected systems, the nature of fraud has evolved in both complexity and scale. Digital financial systems encompass online banking, mobile payments, cryptocurrency transactions, e-commerce platforms, and fintech-based lending services, all of which generate vast volumes of transactional data. Within these environments, fraud manifests in various forms such as identity theft, phishing, transaction laundering, account takeover, and synthetic identity fraud (Kapadiya et al., 2022).

Figure 1: Fraud Detection in Digital Financial Systems



The detection of such activities requires continuous monitoring, anomaly identification, and risk assessment processes embedded within financial infrastructures. Traditionally, fraud detection relied on rule-based systems that utilized predefined thresholds and expert-driven logic. These systems, while effective in controlled scenarios, often lack adaptability and fail to capture evolving fraud patterns. The increasing sophistication of cybercriminals, coupled with the exponential growth of digital transactions, has necessitated the integration of advanced computational techniques. Artificial intelligence (AI) has emerged as a transformative approach in this domain, enabling systems to learn from data, identify hidden patterns, and respond dynamically to emerging threats. The application of AI in fraud detection involves machine learning algorithms, deep learning architectures, and data-driven predictive models that enhance the accuracy and efficiency of detection mechanisms. The importance of fraud detection extends beyond financial institutions, affecting consumers, regulatory bodies, and global economic stability (Sinha et al., 2022). Financial losses due to fraud impose significant economic burdens and undermine trust in digital financial systems. Consequently, the development of robust fraud detection frameworks has become a critical priority in ensuring secure, transparent, and resilient financial ecosystems.

The evolution of fraud detection mechanisms reflects the broader transformation of financial systems from manual and analog processes to automated and intelligent digital infrastructures. Early fraud detection methods were predominantly manual, relying on human expertise to identify suspicious

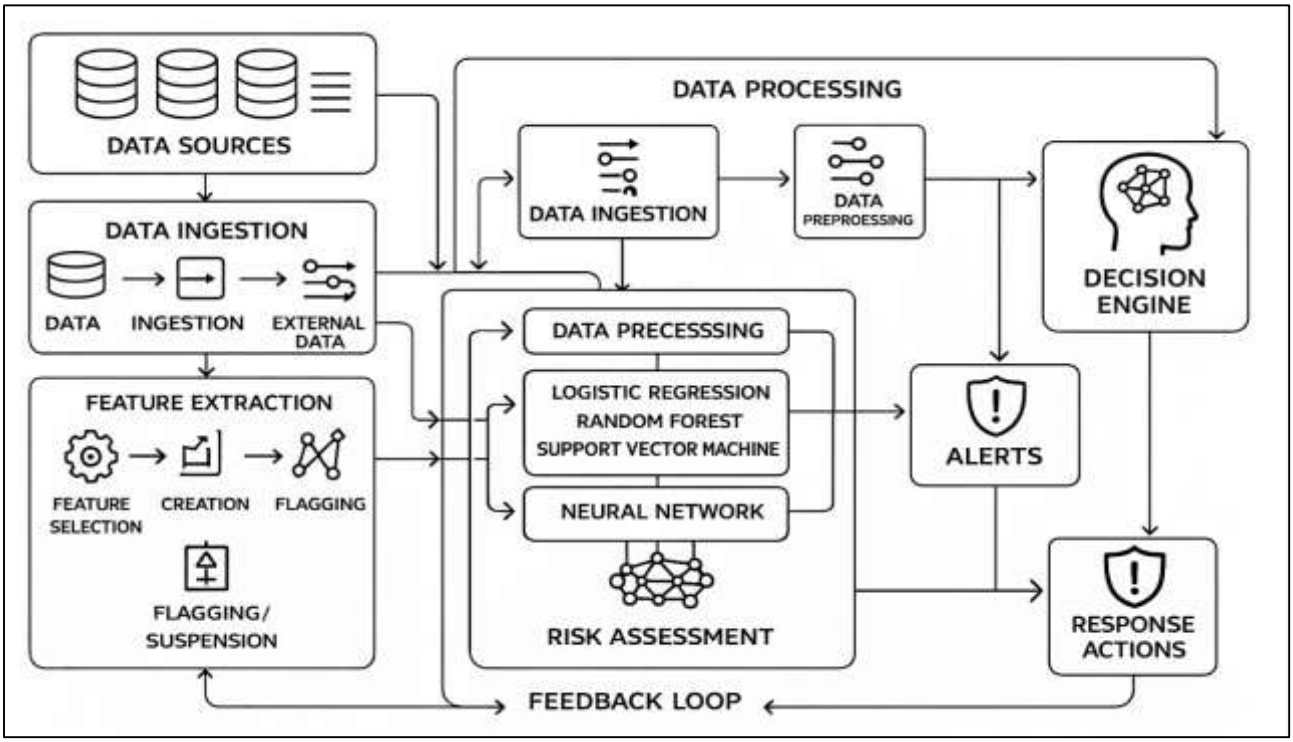
transactions through audits and account reviews. With the advent of computerized banking systems, rule-based algorithms were introduced to automate detection processes. These systems operated on predefined conditions such as transaction limits, geographic inconsistencies, and frequency anomalies. While effective in identifying known fraud patterns, rule-based systems exhibited significant limitations in detecting novel and adaptive fraud strategies (Kumar & Bhushan, 2023). As digital financial systems expanded globally, the volume and velocity of transaction data increased exponentially, rendering traditional approaches insufficient. The integration of artificial intelligence marked a paradigm shift in fraud detection methodologies. AI-based systems leverage large datasets to train models capable of identifying complex and non-linear relationships within transactional data. Machine learning algorithms, including supervised and unsupervised learning techniques, enable systems to classify transactions as legitimate or fraudulent based on historical patterns. Supervised learning models rely on labeled datasets to predict outcomes, whereas unsupervised models detect anomalies without prior labeling, making them particularly useful in identifying emerging fraud schemes. Deep learning techniques further enhance detection capabilities by analyzing high-dimensional data and extracting intricate features that may not be evident through conventional methods. Natural language processing contributes to fraud detection by analyzing textual data such as transaction descriptions, customer communications, and online behavior (Manoj et al., 2023). The continuous learning capability of AI systems allows them to adapt to evolving fraud tactics, reducing false positives and improving detection accuracy. This evolution highlights the transition from static, rule-based systems to dynamic, data-driven intelligence, positioning AI as a cornerstone of modern fraud detection strategies in digital financial environments.

The global significance of AI-enabled fraud detection lies in its capacity to address the growing challenges associated with financial crimes in an increasingly digitalized world. As financial services expand across borders through digital platforms, the exposure to fraud risks has intensified on an international scale (Burri et al., 2023). Cross-border transactions, digital wallets, and decentralized financial systems have created new vulnerabilities that can be exploited by sophisticated cybercriminal networks. Fraudulent activities not only result in substantial financial losses but also erode consumer confidence and disrupt economic stability. AI-enabled fraud detection systems provide a scalable and efficient solution to these challenges by enabling real-time monitoring and analysis of transactions across diverse financial ecosystems. Financial institutions worldwide have adopted AI-driven solutions to enhance their fraud prevention capabilities, reduce operational costs, and comply with regulatory requirements (Jaiswal & Akhilesh, 2019). Regulatory bodies have also emphasized the importance of advanced technologies in combating financial crimes such as money laundering and terrorist financing. The implementation of AI in fraud detection supports regulatory compliance by providing transparent and auditable decision-making processes. Furthermore, AI systems facilitate the identification of complex fraud networks by analyzing interconnected data points across multiple channels. The global adoption of digital payment systems, particularly in developing economies, has amplified the need for robust fraud detection mechanisms. In regions experiencing rapid financial inclusion, the integration of AI ensures that security measures keep pace with technological advancements. The ability of AI to process large-scale data in real time makes it indispensable in addressing the dynamic and borderless nature of digital financial fraud (Anshari et al., 2021). This global relevance underscores the critical role of AI in safeguarding financial systems, protecting consumers, and maintaining the integrity of international financial markets.

AI-enabled fraud detection is fundamentally grounded in quantitative methodologies that utilize statistical analysis, mathematical modeling, and computational algorithms to identify fraudulent behavior. These models rely on large datasets comprising transactional records, user behavior patterns, and historical fraud instances. Quantitative approaches enable the transformation of raw data into meaningful insights through feature extraction, data preprocessing, and model training. Machine learning algorithms such as logistic regression, decision trees, random forests, and support vector machines are widely employed to classify transactions based on probability scores. These models assign risk values to transactions, allowing financial institutions to prioritize investigations and allocate resources efficiently (Bhatt & Singh, 2023). Neural networks and deep learning architectures further enhance predictive capabilities by modeling complex relationships within high-dimensional data.

Quantitative evaluation metrics such as accuracy, precision, recall, and F1-score are used to assess model performance and ensure reliability. The balance between false positives and false negatives is a critical consideration in fraud detection, as excessive false alerts can lead to operational inefficiencies while undetected fraud can result in significant losses. Advanced techniques such as ensemble learning combine multiple models to improve overall performance and robustness. Time-series analysis plays a crucial role in detecting temporal anomalies, enabling systems to identify unusual transaction patterns over time. Graph-based models are also utilized to analyze relationships between entities, uncovering hidden connections within fraud networks. The integration of big data analytics enhances the scalability of quantitative models, allowing them to process vast amounts of data in real time (Sheth et al., 2022). These quantitative foundations provide a systematic framework for AI-driven fraud detection, ensuring that models are both accurate and adaptable to evolving fraud patterns.

Figure 2: AI-enabled Fraud Detection Systems



The effectiveness of AI-enabled fraud detection systems is heavily dependent on the quality and structure of data, as well as the processes used to extract meaningful features from raw datasets. Digital financial systems generate diverse data streams, including transaction histories, user demographics, device information, geolocation data, and behavioral patterns. Data-driven architectures are designed to integrate and manage these heterogeneous data sources, enabling comprehensive analysis and real-time decision-making. Feature engineering plays a critical role in transforming raw data into informative variables that enhance model performance. This process involves selecting, creating, and refining features that capture relevant aspects of user behavior and transaction characteristics (Mhlanga, 2020). Examples of engineered features include transaction frequency, average transaction value, deviation from typical spending patterns, and device usage anomalies. The incorporation of domain knowledge in feature engineering enhances the interpretability and effectiveness of fraud detection models. Data preprocessing techniques such as normalization, encoding, and outlier handling are essential in preparing datasets for analysis. Handling imbalanced datasets is a significant challenge in fraud detection, as fraudulent transactions typically represent a small fraction of total transactions. Techniques such as oversampling, undersampling, and synthetic data generation are employed to address this imbalance and improve model training. Real-time data processing frameworks enable continuous monitoring of transactions, allowing AI systems to detect fraud as it

occurs. The integration of cloud computing and distributed systems further enhances the scalability and efficiency of data-driven architectures (Mogaji et al., 2022). These components collectively contribute to the development of robust fraud detection systems capable of adapting to dynamic financial environments.

Despite the advancements in AI-driven fraud detection, several challenges persist in the implementation and optimization of these systems. One of the primary challenges is the dynamic nature of fraud, where perpetrators continuously adapt their strategies to evade detection. This necessitates the development of models that can learn and adapt in real time. Data quality and availability also pose significant challenges, as incomplete or noisy data can adversely affect model performance. Privacy and security concerns are critical considerations in the use of AI, particularly in handling sensitive financial information (Bin Mofidul et al., 2022). Regulatory requirements impose constraints on data usage, requiring organizations to balance innovation with compliance. The interpretability of AI models is another important challenge, as complex algorithms such as deep learning models often operate as black boxes, making it difficult to understand their decision-making processes. This lack of transparency can hinder trust and limit the adoption of AI in regulated financial environments. Computational complexity and resource requirements also present challenges, particularly in processing large-scale data in real time. The integration of AI systems with existing financial infrastructures requires significant investment and technical expertise. Addressing these challenges requires a multidisciplinary approach that combines advancements in technology, regulatory frameworks, and organizational practices (Cross, 2022).

The integration of AI into modern digital financial ecosystems represents a transformative shift in how financial services are delivered and secured. AI-enabled fraud detection systems are embedded within various components of financial infrastructures, including payment gateways, banking platforms, and fintech applications. These systems operate in real time, analyzing transactions as they occur and providing immediate risk assessments. The seamless integration of AI enhances the efficiency and effectiveness of fraud prevention measures, reducing operational costs and improving customer experience (Dubey et al., 2022). Collaborative frameworks between financial institutions, technology providers, and regulatory bodies facilitate the sharing of information and best practices, strengthening collective defenses against fraud. The use of APIs and microservices architectures enables the flexible deployment of AI solutions, allowing organizations to scale and adapt their systems as needed. Continuous monitoring and feedback loops ensure that AI models remain effective in dynamic environments. The integration of AI also supports broader financial inclusion by providing secure and reliable digital services to underserved populations. As digital financial systems continue to evolve, the role of AI in fraud detection becomes increasingly central, shaping the future of secure and resilient financial ecosystems (Zdravković et al., 2022).

The primary objective of this quantitative study is to systematically examine and synthesize the performance, methodologies, and effectiveness of AI-enabled fraud detection models within digital financial systems over the period 2019 to 2026. This study aims to quantitatively evaluate how different artificial intelligence techniques, including machine learning, deep learning, and hybrid computational models, contribute to the detection and prevention of fraudulent activities across diverse financial platforms such as online banking, mobile payment systems, and fintech applications. A key objective is to analyze the predictive accuracy, efficiency, and reliability of these models using standardized performance metrics such as precision, recall, accuracy, and F1-score, thereby providing a comparative understanding of their effectiveness in real-world applications. The study also seeks to identify patterns in model selection, dataset characteristics, and feature engineering techniques that influence fraud detection outcomes, offering a structured assessment of how data-driven approaches enhance model performance. Another important objective is to quantify the extent to which AI-based systems reduce false positives and false negatives, which are critical indicators of operational efficiency and user trust in financial services. The research further aims to evaluate the scalability and adaptability of AI models in handling large-scale transactional data, particularly in environments characterized by high transaction volumes and rapid data generation. Additionally, this study intends to assess the integration of AI systems within existing financial infrastructures, focusing on their role in real-time fraud detection and risk management. By aggregating empirical findings from multiple studies, the

research aims to provide a comprehensive quantitative synthesis that highlights dominant trends, methodological advancements, and measurable outcomes in AI-driven fraud detection. The overarching objective is to establish a data-driven understanding of how AI technologies enhance the security and resilience of digital financial systems, while offering a structured evaluation framework that can support further academic and practical developments in this domain.

LITERATURE REVIEW

The literature review section provides a structured and quantitative synthesis of existing scholarly work on AI-enabled fraud detection in digital financial systems, with a specific focus on studies published between 2019 and 2026. This section aims to systematically organize and evaluate the empirical, methodological, and analytical contributions of prior research in order to establish a comprehensive understanding of how artificial intelligence techniques have been applied to detect and prevent financial fraud (Karimipour & Derakhshan, 2021). The review emphasizes quantitative approaches, highlighting the use of statistical modeling, machine learning algorithms, deep learning architectures, and performance evaluation metrics that underpin fraud detection systems. By consolidating findings from multiple studies, this section identifies recurring methodological patterns, dataset characteristics, feature engineering strategies, and evaluation frameworks that define the current state of research in this domain. The literature review also categorizes studies based on model types, data structures, and analytical techniques, enabling a comparative assessment of their effectiveness in detecting fraudulent transactions. Special attention is given to the quantitative performance of AI models, including accuracy, precision, recall, and other key indicators that determine their reliability and operational efficiency (Mallidi et al., 2023). Furthermore, the review examines the integration of AI within real-time financial systems, focusing on scalability, computational efficiency, and adaptability to evolving fraud patterns. Through a systematic and structured approach, this section provides a detailed foundation for understanding the strengths, limitations, and methodological diversity of AI-based fraud detection research, setting the stage for deeper quantitative analysis in subsequent sections.

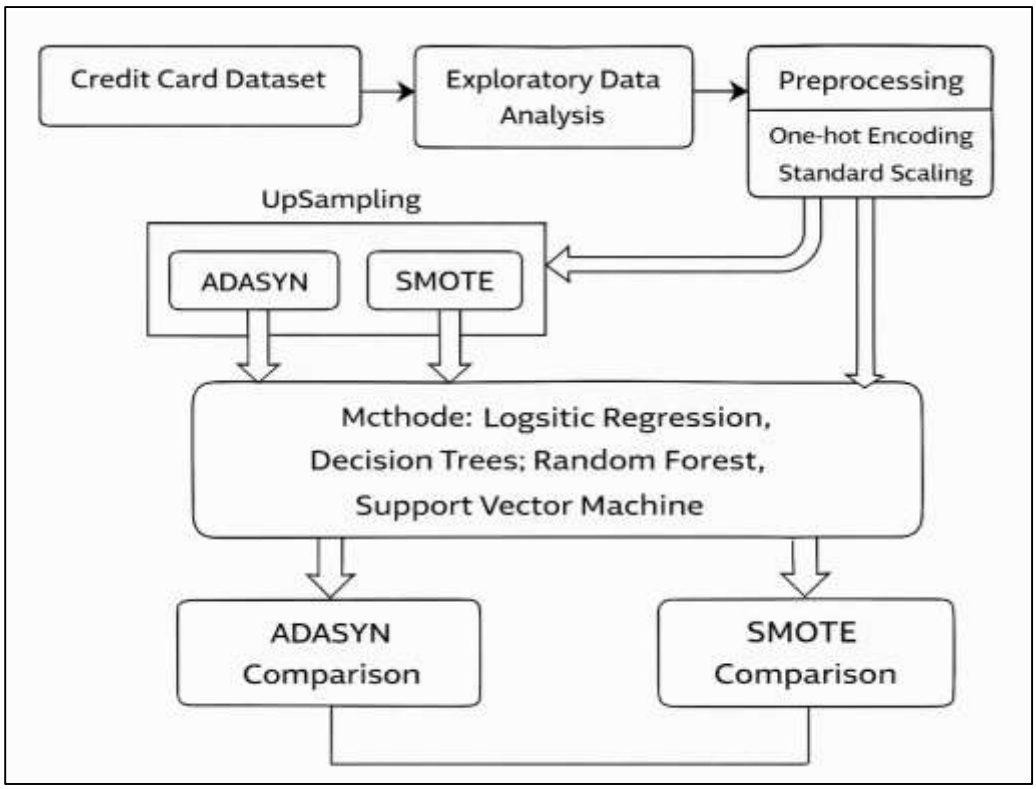
Models for Fraud Classification

The application of classical machine learning models such as logistic regression, decision trees, random forests, and support vector machines has been widely examined in fraud detection research, particularly in the context of digital financial systems. Logistic regression has been traditionally valued for its interpretability and efficiency in binary classification tasks, making it suitable for baseline fraud detection scenarios. Studies have demonstrated that logistic regression performs effectively when relationships between variables are relatively linear and datasets are well-structured (Almazroi & Ayub, 2023). Decision trees offer a more flexible approach by capturing nonlinear relationships and providing intuitive rule-based classifications, which are particularly useful in identifying transaction-level fraud patterns. Random forest models, as ensemble extensions of decision trees, have consistently shown improved predictive performance by reducing overfitting and enhancing generalization through aggregation of multiple decision trees. Support vector machines have also been extensively applied due to their ability to handle high-dimensional data and construct optimal decision boundaries, particularly in cases where fraud patterns are complex and not easily separable. Comparative studies have revealed that while logistic regression provides stable baseline performance, random forests and support vector machines tend to achieve higher classification accuracy in large and complex datasets. Decision trees, although interpretable, often suffer from overfitting when used independently (Awoyemi et al., 2017). The selection of an appropriate model is often influenced by dataset characteristics, computational constraints, and the need for interpretability. These findings collectively indicate that no single model consistently outperforms others across all scenarios, highlighting the importance of context-specific model evaluation in fraud detection research.

The evaluation of fraud detection models relies heavily on quantitative performance metrics that provide a comprehensive assessment of classification effectiveness. Accuracy has traditionally been used as a primary metric; however, its reliability is often limited in fraud detection contexts due to the highly imbalanced nature of financial datasets (Alfaiz & Fati, 2022). Precision and recall have therefore gained prominence as more informative indicators, where precision measures the proportion of correctly identified fraudulent transactions among all predicted fraud cases, and recall captures the

ability of the model to detect actual fraudulent instances. The balance between these two metrics is commonly assessed through the F1-score, which provides a harmonic representation of model performance. Empirical studies have demonstrated that models achieving high accuracy may still perform poorly in detecting fraud if they fail to identify minority class instances effectively (Alarfaj et al., 2022).

Figure 3: Machine Learning Fraud Detection Framework



Consequently, a strong emphasis has been placed on optimizing recall to ensure that fraudulent transactions are not overlooked, while maintaining acceptable precision levels to reduce false alarms. Comparative analyses across multiple studies have shown that ensemble models such as random forests often achieve superior F1-scores due to their ability to balance bias and variance. Support vector machines have also demonstrated strong performance in precision-focused evaluations, particularly in high-dimensional datasets. The selection of evaluation metrics is closely aligned with the operational priorities of financial institutions, where minimizing financial loss and maintaining customer trust are critical. This has led to the adoption of multi-metric evaluation frameworks that provide a more nuanced understanding of model performance in real-world fraud detection systems (Dornadula & Geetha, 2019).

Cross-validation techniques play a critical role in assessing the robustness and generalizability of fraud detection models. Given the variability and complexity of financial datasets, it is essential to evaluate models across multiple subsets of data to ensure consistent performance. K-fold cross-validation has been widely adopted in fraud detection research as a standard approach for dividing datasets into training and validation subsets, enabling models to be tested on different data partitions. This method reduces the risk of overfitting and provides a more reliable estimate of model performance (Itoo et al., 2021). Stratified cross-validation techniques have been particularly important in fraud detection due to the imbalanced nature of datasets, ensuring that each fold maintains a representative distribution of fraudulent and non-fraudulent transactions. Empirical studies have shown that models evaluated using cross-validation tend to exhibit greater stability and reliability compared to those assessed using simple train-test splits. Random forest models, in particular, have demonstrated strong robustness under cross-validation due to their ensemble structure, which mitigates the impact of data variability.

Support vector machines have also shown consistent performance when properly tuned, although their sensitivity to parameter selection requires careful validation. Logistic regression models generally exhibit stable performance across folds, reflecting their simplicity and lower variance (Raghavan & El Gayar, 2019). The application of cross-validation techniques has become a standard practice in fraud detection research, providing a systematic framework for evaluating model performance and ensuring that findings are not influenced by dataset-specific biases.

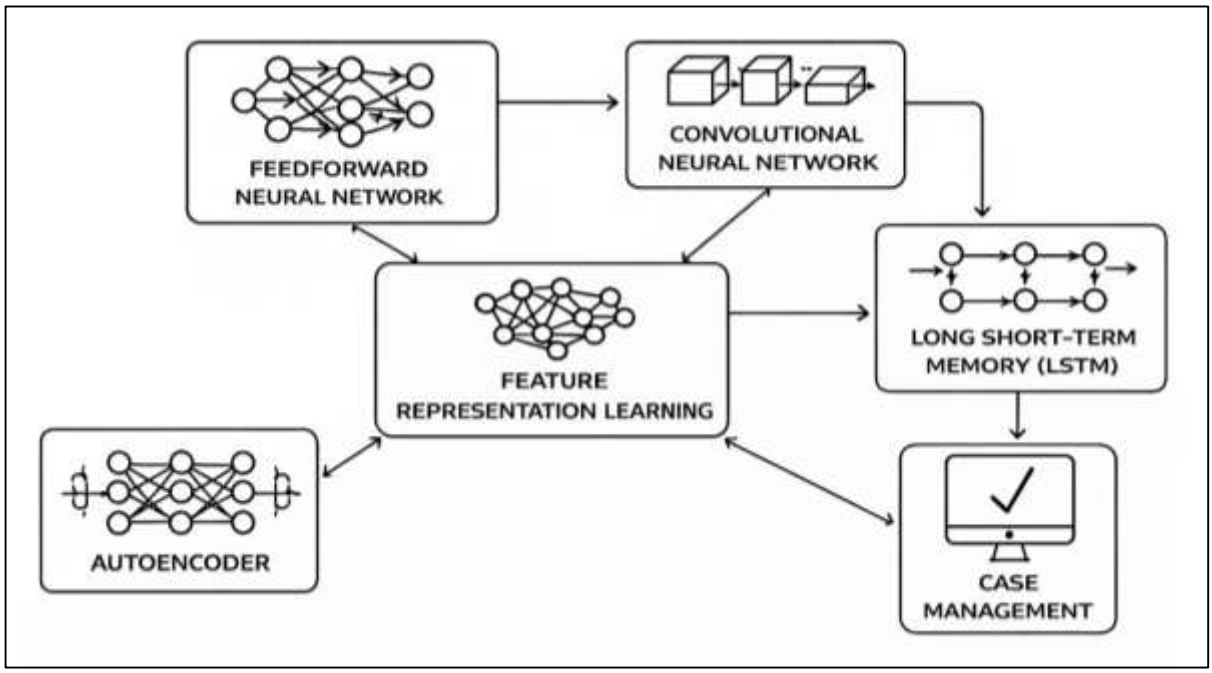
The bias-variance trade-off represents a fundamental consideration in the development and evaluation of fraud detection models. Models with high bias, such as logistic regression, tend to oversimplify relationships within data, leading to underfitting and reduced predictive accuracy in complex fraud scenarios. Conversely, models with high variance, such as decision trees, may capture intricate patterns in training data but fail to generalize effectively to new data, resulting in overfitting (Alsuwailem et al., 2023). Ensemble methods like random forests have been widely recognized for their ability to balance bias and variance, providing improved generalization and robustness. Support vector machines offer a flexible approach to managing this trade-off by adjusting kernel functions and regularization parameters, enabling them to adapt to different data complexities. The evaluation of these models is often conducted using benchmark datasets, which serve as standardized references for comparing performance across studies. However, the reliance on a limited number of publicly available datasets has raised concerns regarding the generalizability of findings, as these datasets may not fully represent the diversity of real-world financial systems. Reproducibility has emerged as a significant challenge in fraud detection research, with variations in data preprocessing, feature engineering, and model tuning leading to inconsistent results across studies (Afriyie et al., 2023). The lack of standardized evaluation protocols further complicates the comparison of model performance. These challenges highlight the need for more comprehensive and transparent research practices that facilitate reliable and reproducible findings in the field of AI-enabled fraud detection.

Deep Learning Architectures in Fraud Detection

Deep learning architectures have become central to fraud detection research because they provide a data-driven way to model highly nonlinear relationships in digital financial transactions. Within this stream of literature, feedforward neural networks, convolutional neural networks, and recurrent neural networks are repeatedly compared as alternative structures for classifying fraudulent and legitimate activities. Feedforward neural networks are generally treated as foundational deep learning models because they transform input variables through multiple hidden layers and learn complex decision boundaries from historical transaction records (Malik et al., 2022). Their strength lies in capturing multidimensional interactions among variables such as transaction amount, merchant type, device identifier, account age, payment frequency, and geographical inconsistency. In quantitative studies, these models often outperform simpler classifiers when the dataset contains rich behavioral and transactional attributes. Convolutional neural networks were later introduced into fraud detection research to improve automatic feature extraction. Although they were originally developed for image analysis, researchers adapted them to structured financial data by arranging transaction attributes into matrix-like forms that allow the model to detect local patterns and hidden correlations. Recurrent neural networks, by contrast, are especially relevant when fraud is interpreted as a temporal and sequential phenomenon (Sailusha et al., 2020). Their architecture allows previous transaction behavior to inform the evaluation of current activity, which is particularly useful in dynamic payment environments where fraudulent behavior unfolds over time rather than in isolated observations. Comparative literature shows that no deep learning model is universally superior across all digital financial contexts. Instead, performance depends on the underlying structure of the data, the length of transaction histories, the level of feature complexity, and the computational resources available. This body of work has established deep learning not simply as a replacement for conventional fraud classifiers, but as a set of quantitatively distinct architectures that address different forms of fraud behavior within digital financial systems (Gupta et al., 2023).

A major theme in recent fraud detection literature is the recognition that fraudulent transactions often emerge as part of a behavioral sequence rather than as independent events. This has led to the growing use of long short-term memory models, which are widely examined for their ability to identify sequential transaction anomalies in financial datasets.

Figure 4: Deep Learning Fraud Detection Framework



LSTM models are a specialized form of recurrent neural network designed to preserve relevant information over longer transaction sequences while filtering out short-term noise. In fraud detection settings, this makes them particularly effective for analyzing customer spending trajectories, login patterns, device changes, rapid withdrawal behavior, card-not-present transaction sequences, and evolving account takeover events (Thennakoon et al., 2019). The literature consistently shows that LSTM-based models are advantageous in cases where fraud develops gradually and where abnormality becomes visible only when current behavior is interpreted against prior activity. Researchers have used LSTM models to examine time-ordered payment streams, card usage sequences, mobile wallet interactions, and online banking logs, with many studies reporting improved recall and stronger anomaly sensitivity compared with static classifiers. This advantage is especially notable in datasets where legitimate and fraudulent transactions appear superficially similar when observed individually. The sequential perspective allows LSTM systems to detect irregular timing, unusual behavioral transitions, and abrupt changes in transaction rhythm that simpler models often overlook (Hashemi et al., 2022). At the same time, the literature also notes that LSTM models require carefully structured temporal data and sufficiently long histories to achieve stable performance. Their effectiveness declines when sequences are sparse, fragmented, or heavily preprocessed in ways that remove chronological information. Quantitative evaluations therefore treat LSTM not merely as a more advanced neural model, but as a method specifically aligned with the temporal logic of fraud behavior. This has positioned LSTM-based detection as one of the most influential developments in the deep learning literature on digital financial fraud.

Another major contribution of deep learning research in fraud detection lies in feature representation learning, which shifts the analytical emphasis from manually engineered predictors to automatically learned patterns derived from raw or minimally processed data (Mondal et al., 2021). Earlier fraud detection studies relied heavily on handcrafted variables designed by domain experts, such as spending thresholds, transaction frequency ratios, merchant risk flags, and geographic deviations. More recent literature shows that deep learning architectures increasingly reduce this dependence by learning internal representations that capture hidden behavioral structures directly from the data. Feedforward networks, convolutional models, and recurrent models each approach this process differently, yet all are valued for their ability to transform high-dimensional transactional inputs into more informative latent patterns (Abd El-Naby et al., 2023). This capability is especially important in digital financial systems where the data are large, heterogeneous, and often noisy. Researchers have

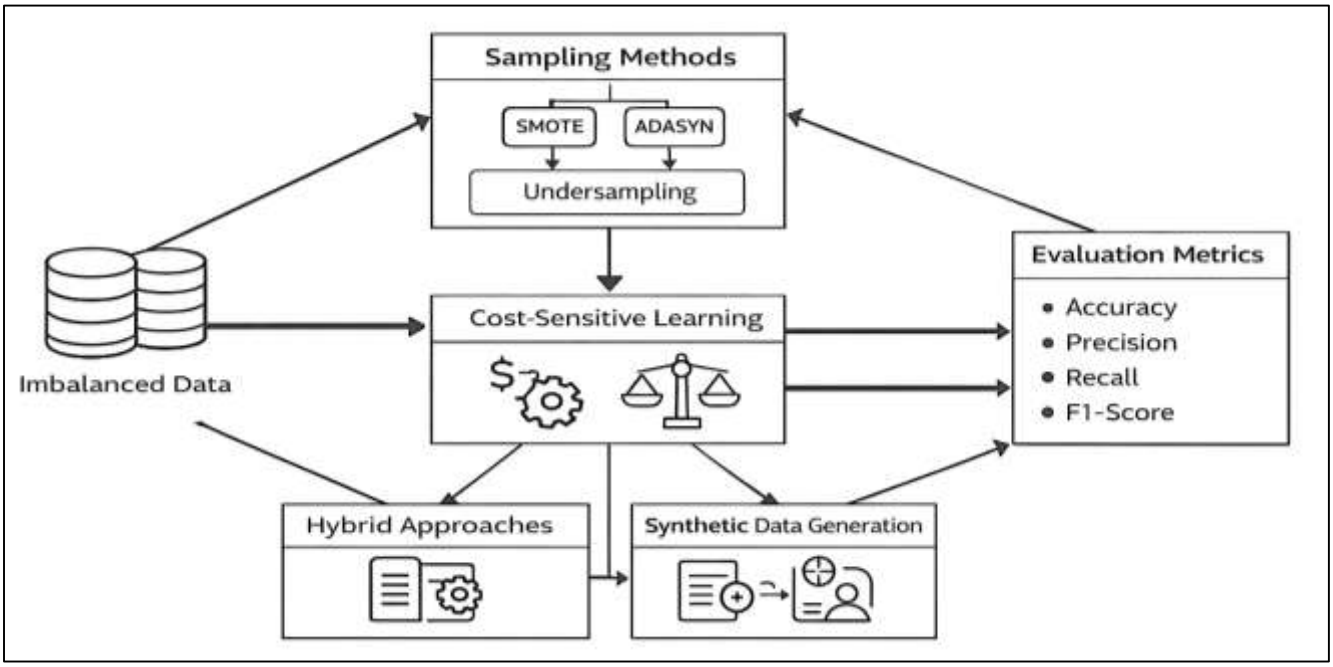
examined how automatic representation learning improves discrimination between subtle fraud signals and normal behavioral variation, particularly in environments involving mobile transactions, online payments, and complex customer interaction records. The literature also connects this issue to dimensionality reduction, since fraud datasets frequently contain many correlated or redundant variables that can weaken model stability and increase training time. Deep architectures are often used alongside embedding layers, latent feature extraction procedures, or compressed internal representations to improve efficiency without sacrificing predictive power (Gupta et al., 2021). Comparative studies suggest that representation learning can be especially beneficial when fraud patterns are not easily captured through simple statistical descriptors. This is because latent features may express nonlinear interactions that traditional feature engineering does not detect consistently. At the same time, scholars note that the advantages of representation learning depend on dataset scale and quality. In small or highly sparse datasets, deep models may not learn robust abstractions and may instead amplify noise. The literature therefore presents feature representation learning as a major quantitative strength of deep learning fraud detection, while also emphasizing that its effectiveness depends on the balance between data richness, dimensional complexity, and model design (Gupta et al., 2023).

Dataset Handling Techniques in Financial Fraud Detection

A major issue in the fraud detection literature is the severe class imbalance that characterizes most financial transaction datasets, where legitimate transactions vastly outnumber fraudulent ones. This imbalance is not a minor technical inconvenience but a defining structural condition that shapes model selection, training behavior, and performance interpretation. In digital financial systems, fraudulent events usually represent only a very small proportion of all recorded transactions, which means that predictive models can appear highly accurate even when they fail to identify fraud effectively (Zhang et al., 2022). The literature consistently shows that when models are trained on naturally imbalanced data without corrective strategies, they tend to favor the majority class and classify most transactions as legitimate. This creates misleadingly high overall accuracy while suppressing recall for fraudulent instances, which is especially problematic in applied fraud detection where missed fraud cases carry substantial financial and reputational costs. Quantitative studies across banking, card payment, mobile finance, and online transaction platforms demonstrate that imbalance distorts the learning process by limiting the model's exposure to minority fraud patterns, reducing sensitivity to rare events, and weakening the boundary between fraudulent and non-fraudulent behavior. Researchers further show that imbalance affects not only classification metrics but also feature importance estimation, threshold tuning, and cross-validation reliability (Makki et al., 2019). In many studies, minority fraud cases contain the most informative behavioral anomalies, yet their scarcity makes them vulnerable to being overshadowed by routine legitimate activity. This has led the literature to treat imbalance handling as an essential stage of fraud analytics rather than a supplementary preprocessing option. The broader quantitative consensus is that the effectiveness of a fraud detection model cannot be evaluated independently of the imbalance treatment used during training and testing. As a result, the study of sampling strategies, synthetic generation methods, and cost-sensitive learning has become one of the most important branches of research in AI-enabled fraud detection within digital financial systems (Singh et al., 2022).

A substantial portion of the literature focuses on resampling methods designed to rebalance fraud datasets before model training, with oversampling and undersampling emerging as the two dominant categories. Oversampling techniques increase the representation of fraudulent cases so that models can learn minority patterns more effectively, while undersampling reduces the number of legitimate cases to create a more balanced class distribution. Within oversampling approaches, SMOTE has become one of the most widely examined methods because it generates synthetic minority examples by interpolating between existing fraud cases, thereby expanding the minority class without merely duplicating original observations. ADASYN extends this logic by placing more synthetic emphasis on difficult or sparsely represented fraud regions, making it attractive for datasets with irregular minority distributions (Baesens et al., 2021). The literature generally finds that these oversampling approaches improve the model's ability to detect fraud by increasing recall and strengthening class discrimination, especially in machine learning and neural classification settings.

Figure 5: Class Imbalance Handling in Fraud Detection



At the same time, researchers note that oversampling can also introduce noise or artificial overlap between classes when synthetic instances are created in poorly separated data regions. Undersampling methods, by contrast, simplify the dataset by removing a portion of the majority class. This often reduces training time and can improve minority sensitivity, but it may also discard useful information about legitimate transaction behavior, which is essential for distinguishing genuine anomalies from ordinary variation (Strelcenia & Prakoonwit, 2023). Comparative studies suggest that oversampling is usually more effective when data volume is limited and fraud rarity is extreme, whereas undersampling can be useful when the majority class is overwhelmingly large and computational efficiency is necessary. Many quantitative reviews also discuss hybrid approaches that combine oversampling and undersampling to balance sensitivity and information preservation. Taken together, the literature presents resampling not as a one-size-fits-all solution but as a context-dependent intervention whose success depends on dataset size, fraud density, class overlap, and the specific learning algorithm employed.

Beyond data resampling, the literature also emphasizes cost-sensitive learning as a major strategy for handling imbalance in fraud detection (Cheah et al., 2023). Rather than altering the class distribution directly, cost-sensitive approaches modify the learning process so that misclassifying a fraudulent transaction carries a greater penalty than misclassifying a legitimate one. This method is especially relevant in financial settings because the practical cost of overlooking fraud is often much higher than the inconvenience of investigating a false alarm. Quantitative research shows that weighted training strategies can shift model attention toward minority fraud cases and improve recall without requiring synthetic data generation or reduction of the majority class. These methods are frequently used in tree-based classifiers, support vector systems, gradient boosting models, and deep learning architectures, where class weights are incorporated into the optimization process to reflect the asymmetric consequences of prediction error (Somasundaram & Reddy, 2019). The literature consistently links cost-sensitive learning to the improvement of minority class detection, especially in environments where the preservation of full transaction data is important. However, researchers also note that stronger minority emphasis often increases false positive rates, which can burden financial institutions with excessive alerts and inefficient review processes. This creates a persistent trade-off between recall and precision, making it necessary to evaluate model performance across multiple metrics rather than relying on one indicator. Many studies show that weighted learning improves fraud sensitivity more effectively than untreated baseline models and, in some cases, performs comparably to resampling

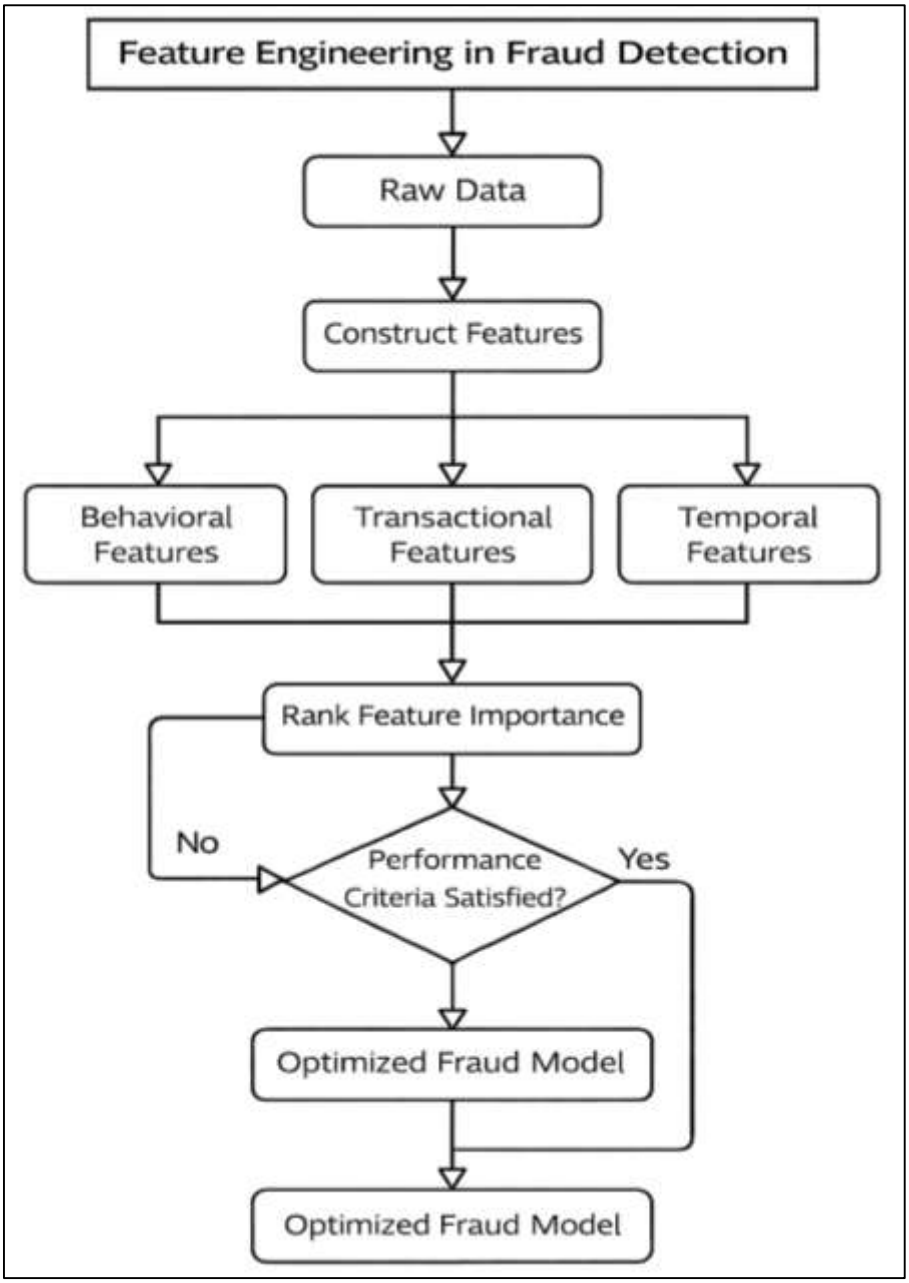
approaches. The choice between these techniques often depends on whether the research goal prioritizes operational simplicity, computational stability, or maximum fraud capture (Sundarkumar & Ravi, 2015). The literature therefore treats cost-sensitive learning as a conceptually important alternative to sampling-based correction, particularly because it aligns more directly with the economic logic of fraud management by embedding unequal error consequences into the model training process itself.

The evaluation of synthetic data generation has become one of the most intensively discussed topics in imbalanced fraud detection research because it sits at the intersection of data scarcity, model learning quality, and performance trade-offs. Synthetic generation methods are used to enrich the minority fraud class in order to expose predictive models to a broader range of fraudulent patterns than those available in the original dataset (Li et al., 2021). The literature commonly reports that such methods improve recall by increasing the representation of minority examples during training, which helps models recognize fraud cases that would otherwise remain underlearned. This is particularly valuable in digital financial systems where fraudulent behavior is rare, heterogeneous, and often concealed within large volumes of legitimate activity. At the same time, the literature also presents a more cautious view of synthetic effectiveness. While recall often improves, precision may decline when generated instances blur the distinction between fraud and normal transactions, leading the model to flag more legitimate cases as suspicious. This raises false positive rates, which has direct operational implications for customer experience, investigation workloads, and institutional costs. Quantitative comparisons therefore evaluate synthetic methods not only by their success in increasing fraud capture but also by their ability to preserve realistic minority structures without distorting class boundaries (De Zarzà et al., 2023). Researchers repeatedly emphasize that the quality of synthetic observations matters more than the quantity alone. When synthetic generation reflects meaningful minority variation, models tend to achieve stronger F1-scores and better class balance. When it introduces excessive similarity, noise, or artificial clustering, performance gains become unstable and less generalizable. The literature also shows that synthetic data methods are most effective when paired with rigorous validation procedures and carefully chosen classifiers. Overall, studies portray synthetic data generation as a powerful but methodologically sensitive solution whose value depends on how well it improves minority representation while maintaining a credible balance among recall, precision, and false positive control in real-world fraud detection tasks (Benchaji et al., 2018).

Feature Engineering and Selection Techniques: Quantitative Impact on Model Accuracy and Fraud Detection Efficiency

Feature engineering occupies a central place in the fraud detection literature because the predictive strength of any analytical model depends heavily on how raw financial data are transformed into meaningful explanatory variables. In digital financial systems, behavioral, transactional, and temporal features are the three most consistently emphasized categories in quantitative studies. Behavioral features are designed to capture patterns in user conduct, such as deviations in login habits, device usage, spending rhythm, merchant preferences, and location consistency. These variables help distinguish ordinary financial activity from suspicious conduct by representing how a customer normally behaves across digital channels (Isangediok & Gajamannage, 2022). Transactional features focus more directly on the properties of individual payments or transfers, including amount, frequency, transaction type, channel of execution, currency, account balance shifts, and merchant category. These variables provide a structured numerical portrait of each event and allow models to detect irregularities at the point of transaction. Temporal features extend this analysis by introducing time-sensitive information such as transaction intervals, sequence ordering, hour-of-day activity, day-of-week effects, burst behavior, and sudden changes in spending intensity. The literature repeatedly shows that fraud rarely becomes visible through a single raw attribute alone; instead, it often emerges when these three categories are combined into richer descriptive patterns. Studies comparing raw transaction records with engineered feature sets consistently report better discrimination between fraudulent and legitimate observations after behavioral, transactional, and temporal construction has been performed (Chen et al., 2023).

Figure 6: Feature Engineering Fraud Detection Framework



This improvement is especially apparent in payment card fraud, mobile banking fraud, and account takeover detection, where subtle deviations from user history often matter more than absolute transaction values. The literature therefore treats feature construction not as a secondary preprocessing step but as a quantitative mechanism that converts fragmented financial records into analytically useful representations. By structuring user behavior, transaction context, and timing patterns into model-ready variables, feature engineering significantly strengthens both fraud identification and operational efficiency.

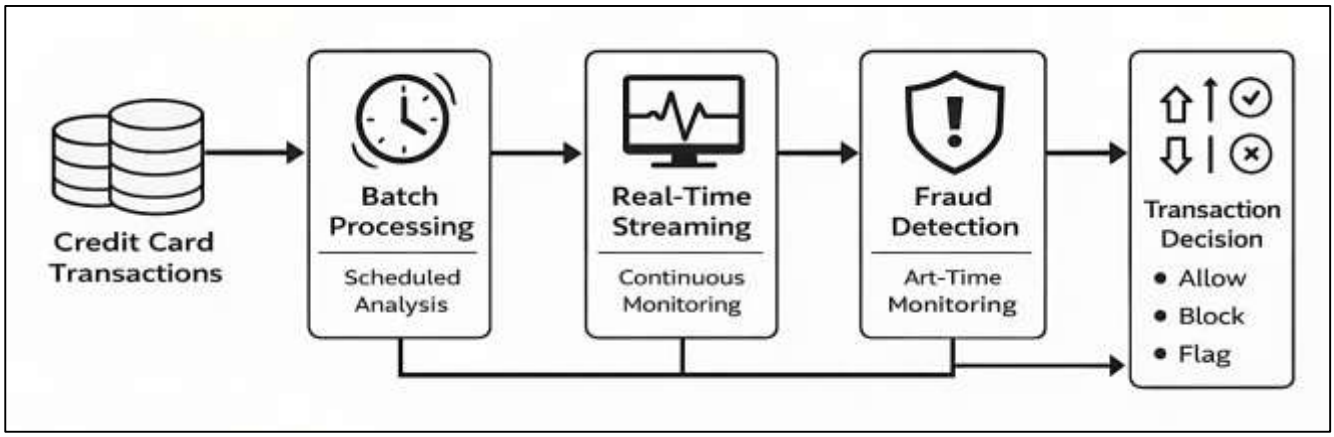
A second major theme in the literature concerns the ranking of feature importance, which is essential for understanding which variables contribute most strongly to fraud classification and which ones add redundancy or noise (Safaei, 2017). In quantitative fraud detection research, feature importance ranking serves two purposes at once: it improves model interpretability and supports the selection of more efficient variable sets for training. Statistical techniques have traditionally been used to assess variable relevance through measures of association, variance contribution, and predictive significance. These approaches help identify whether features such as transaction amount irregularity, velocity of account activity, geographic mismatch, or device switching are meaningfully related to fraud outcomes. More

recent literature increasingly relies on machine learning methods to estimate feature importance in a data-driven way. Tree-based models, ensemble classifiers, recursive elimination procedures, and embedded selection approaches are frequently used to rank variables according to their contribution to classification performance (Shukla et al., 2023). These methods have been especially useful in large financial datasets where the number of potential features is high and relationships among variables are nonlinear. The literature shows that importance ranking often reveals that a relatively small subset of features accounts for a large share of predictive value, while many additional variables contribute little or may even reduce robustness. This finding has practical significance because financial fraud models are often deployed in real-time environments where computational efficiency and model transparency matter. Researchers consistently report that removing weak or redundant predictors can reduce model complexity, shorten processing time, and stabilize performance across validation folds (Shukla et al., 2023). At the same time, the literature warns that importance measures are not always consistent across algorithms, meaning that variable relevance may change depending on the structure of the model used. As a result, many studies recommend a comparative or hybrid approach to feature ranking. Overall, this body of work demonstrates that feature importance analysis is a critical quantitative step in refining fraud detection systems and in identifying the variables that most effectively represent fraudulent financial behavior.

Real-Time Fraud Detection Systems

The literature on real-time fraud detection consistently distinguishes between batch processing architectures and real-time streaming architectures as two fundamentally different approaches to transaction monitoring in digital financial systems (Chen et al., 2020). Batch processing has traditionally been used in fraud analytics because it allows large volumes of transaction data to be collected, cleaned, and analyzed at scheduled intervals. This architecture is valuable for retrospective analysis, periodic risk scoring, and large-scale pattern discovery, especially in settings where immediate intervention is not required. Studies commonly describe batch systems as computationally efficient for historical fraud mining because they can leverage centralized resources and process transactions in aggregated blocks. However, the literature also shows that batch processing has important limitations in modern financial ecosystems where fraudulent activities can unfold within seconds. In card transactions, mobile payments, online banking, and digital wallets, the delay created by batch-based review can allow fraudulent activity to continue before a system responds (R. Wang et al., 2023). This concern has driven the growth of real-time streaming architectures, which are designed to process and evaluate transaction flows as they occur. Streaming systems support continuous ingestion, immediate scoring, and instant alert generation, making them more aligned with the demands of high-speed digital finance. The literature emphasizes that the architectural transition from batch to streaming is not merely technical but also strategic, because it changes the role of fraud detection from post-event analysis to active intervention. Comparative studies often show that streaming architectures are more effective in preventing cascading fraud losses, especially in high-frequency transaction environments. At the same time, scholars note that real-time systems require more sophisticated orchestration, better data integration, and stronger infrastructure resilience than batch systems (Velasco-Gallego & Lazakis, 2022). The broader quantitative literature therefore presents batch processing as useful for deep analytical review and model training, while real-time streaming is positioned as essential for operational fraud prevention in contemporary digital financial systems.

Figure 7: Real-Time Fraud Detection Architecture Framework



A second major concern in the literature is the quantitative relationship between latency and throughput in real-time fraud detection environments. Latency refers to the time required for a transaction to move through the analytical pipeline and receive a fraud decision, while throughput refers to the volume of transactions that can be processed within a given time window. In financial systems, these two performance indicators are closely linked to service quality, customer experience, and institutional risk exposure (Arya & Sastry G, 2020). The literature shows that low-latency fraud detection is particularly important in contexts such as instant payments, card authorization systems, and e-commerce gateways, where decision delays can interrupt legitimate transactions or permit fraudulent ones to proceed unchecked. Studies consistently examine how different system designs affect the speed of scoring, alert generation, and model execution. Throughput optimization is equally important because digital financial platforms may handle thousands or even millions of events in short intervals, particularly during peak transaction periods. Researchers therefore analyze pipeline efficiency, queue management, event prioritization, and model deployment strategies as part of broader system performance evaluation. The literature indicates that an effective real-time fraud detection architecture must balance rapid response with sustained processing capacity (Huang et al., 2021). Systems optimized only for latency may struggle under heavy load, while systems designed mainly for high throughput may accumulate delay and weaken the practical value of fraud alerts. Quantitative studies often benchmark architectures by assessing how they maintain stable performance during transaction spikes, user surges, or periods of anomalous activity. This body of work also highlights that latency is not only a technical metric but a risk management variable, because delayed detection can translate directly into financial loss. As a result, the literature frames latency and throughput not as isolated engineering concerns but as central dimensions of fraud detection effectiveness in digital financial environments where timing is inseparable from security (Lopez et al., 2016).

The expansion of digital financial activity has led researchers to examine big data frameworks and distributed systems as essential infrastructure for real-time fraud detection. The literature frequently argues that conventional single-node architectures are inadequate for handling the scale, speed, and heterogeneity of modern transaction ecosystems. Digital banking platforms, payment processors, fintech applications, and cross-border transaction systems generate continuous streams of structured and semi-structured data that require fast ingestion, scalable storage, and parallel processing capabilities. In response, many studies discuss the use of distributed computing environments to support stream analytics, real-time scoring, and model deployment across large transaction volumes (Mostafa & Tohidul, 2022; Khatun & Morshedul, 2022; Wu et al., 2020). The literature shows that distributed systems improve operational resilience by allowing fraud detection tasks to be split across multiple nodes, thereby reducing bottlenecks and enhancing fault tolerance. Big data frameworks are particularly valued for their ability to integrate real-time event streams with historical transaction repositories, customer profiles, behavioral records, and device metadata. This integration supports

more context-aware fraud decisions and more robust anomaly identification. Researchers also emphasize that distributed systems are crucial in environments where fraud models must operate simultaneously across multiple channels, such as web transactions, point-of-sale systems, mobile applications, and digital wallets. Quantitative studies often compare centralized and distributed architectures in terms of scalability, response time under load, and capacity to maintain stable detection quality as transaction intensity increases (Garg et al., 2019; Ahmed & Hasan Or, 2021). The literature generally concludes that distributed infrastructures are better suited to high-frequency financial ecosystems because they allow continuous model execution without overwhelming processing capacity. At the same time, these systems introduce complexity in synchronization, state management, and consistency across analytical components. This means that the deployment of big data frameworks is not treated simply as a performance upgrade, but as a structural reconfiguration of fraud detection operations. Overall, the literature presents distributed systems as a core enabler of real-time, large-scale, AI-driven fraud detection in digital financial systems.

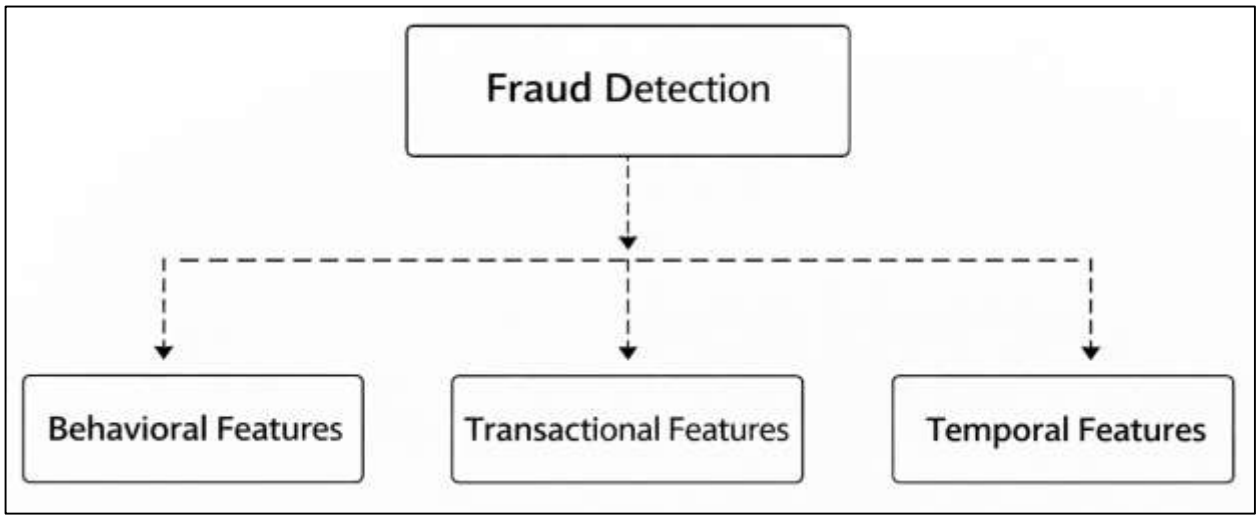
One of the most widely discussed themes in the literature is the trade-off between detection speed and predictive accuracy in real-time fraud detection systems (Aditya & Chandra, 2022; Md & Mehedi, 2021; K. Wang et al., 2023). Financial institutions require immediate decisions to stop fraudulent activity before losses accumulate, yet the pressure for rapid processing can reduce the depth of analysis available for each transaction. The literature consistently shows that faster systems are not automatically more effective if speed is achieved by simplifying feature sets, compressing behavioral histories, or deploying less discriminative models. In high-frequency transaction environments, this trade-off becomes especially pronounced because analytical systems must evaluate large numbers of transactions under strict time constraints (Anick & Tasnim, 2022; Hisham & Robel, 2022). Researchers often report that highly accurate fraud models developed in offline experimental settings do not always retain the same performance when deployed in real-time pipelines, where computational limits and latency requirements alter operational conditions (Siddique & Amin, 2022; Md & Islam, 2022; Zhao et al., 2020). This has led to growing interest in performance benchmarking that evaluates fraud detection systems using both predictive and infrastructural criteria. Studies commonly examine recall, precision, false positive burden, and detection stability alongside response time, throughput consistency, and resilience under peak load. In high-frequency environments such as card payment networks, e-commerce platforms, and digital money transfer systems, the literature emphasizes that benchmarking must reflect realistic transaction pressure rather than ideal laboratory conditions. Comparative analyses show that models with slightly lower predictive scores may still be preferred if they deliver stable low-latency decisions at scale, whereas more complex models may be reserved for layered review or secondary screening. The literature therefore rejects the idea that fraud detection quality can be measured through classification metrics alone (Bauer et al., 2019; Mainuddin & Chandra, 2022; Md. Shahinur & Sultan, 2022). Instead, it frames system performance as a multidimensional balance among speed, scalability, alert quality, and operational sustainability. This perspective has become central to the quantitative evaluation of real-time fraud detection systems, particularly in digital financial settings where rapid decision-making and analytical reliability must coexist continuously.

Ensemble and Hybrid AI Models

The literature on ensemble learning in fraud detection consistently shows that bagging, boosting, and stacking have emerged as major integration strategies for improving predictive stability in digital financial systems. These methods are designed to combine the outputs of multiple models so that weaknesses in any single classifier are offset by the strengths of others. Bagging is frequently discussed as a variance-reduction technique because it trains multiple models on different samples of the same dataset and then aggregates their predictions. In fraud detection research, this approach is commonly associated with improved robustness, especially when transaction data are noisy, imbalanced, or behaviorally heterogeneous (Hu et al., 2021; Islam & Aditya, 2023; Zakia & Nahar, 2022). Random forest models are often treated as the most visible example of bagging in practice because they combine many decision trees and reduce sensitivity to random fluctuations in training data. Boosting, by contrast, is more frequently described as a sequential refinement approach in which later models focus on correcting the errors made by earlier ones. This makes boosting highly attractive in fraud classification contexts where subtle minority patterns are difficult to detect. Studies often report that boosting-based

approaches achieve strong classification power because they iteratively concentrate learning effort on difficult or misclassified transactions. Stacking is presented somewhat differently in the literature because it integrates multiple base learners through a higher-level model that learns how to combine their outputs. This makes stacking especially useful when different classifiers capture different dimensions of fraud behavior (Amutha et al., 2021).

Figure 8: Fraud Detection Feature Engineering Framework



Comparative studies repeatedly indicate that these ensemble strategies tend to outperform isolated models in many financial fraud tasks, particularly when datasets are complex and class distributions are skewed. The broader literature thus positions bagging, boosting, and stacking not merely as performance enhancement tools, but as structured quantitative methods for integrating multiple perspectives on fraudulent transaction patterns in order to improve reliability and classification depth. A major development in the fraud detection literature is the emergence of hybrid models that combine traditional machine learning methods with deep learning architectures. These hybrid systems are designed to exploit the complementary strengths of different analytical paradigms rather than relying on a single modeling logic (Krajsic & Franczyk, 2020; Khaled & Mosheur, 2023; Shahab & Aditya, 2023). Traditional machine learning techniques, such as tree-based classifiers, support vector systems, and regression-based approaches, are often valued for their efficiency, interpretability, and stable performance on structured financial data. Deep learning models, including feedforward neural networks, convolutional networks, and recurrent structures, are more often praised for their capacity to capture nonlinear patterns, latent feature interactions, and sequential behavioral dependencies. The literature shows that hybridization arises from the recognition that financial fraud is a multidimensional phenomenon that rarely fits neatly within one analytical structure. Some studies use deep learning for automatic representation learning and then apply machine learning classifiers to the learned features. Others combine tree-based systems with neural architectures to balance predictive power and computational feasibility. There are also hybrid frameworks in which anomaly detection components, clustering procedures, or dimensionality reduction methods are integrated into supervised classification pipelines (Hasan Or et al., 2023; Mehedi & Nahar, 2023; Raeiszadeh et al., 2023). Quantitative findings generally indicate that these hybrid systems can improve fraud detection by increasing sensitivity to complex patterns while preserving decision stability. This is particularly important in digital financial environments where fraud may involve both isolated transaction anomalies and longer behavioral deviations. The literature also shows that hybrid models often perform well in datasets with mixed feature types, such as numerical transaction records, categorical account attributes, and temporal activity histories. Their appeal lies in their flexibility and their ability to bridge the gap between strong feature extraction and effective classification. As a result, hybrid model design has become a significant branch of the AI fraud detection literature, reflecting the broader movement toward integrated intelligence rather than single-model dependency (Matthews & Leger,

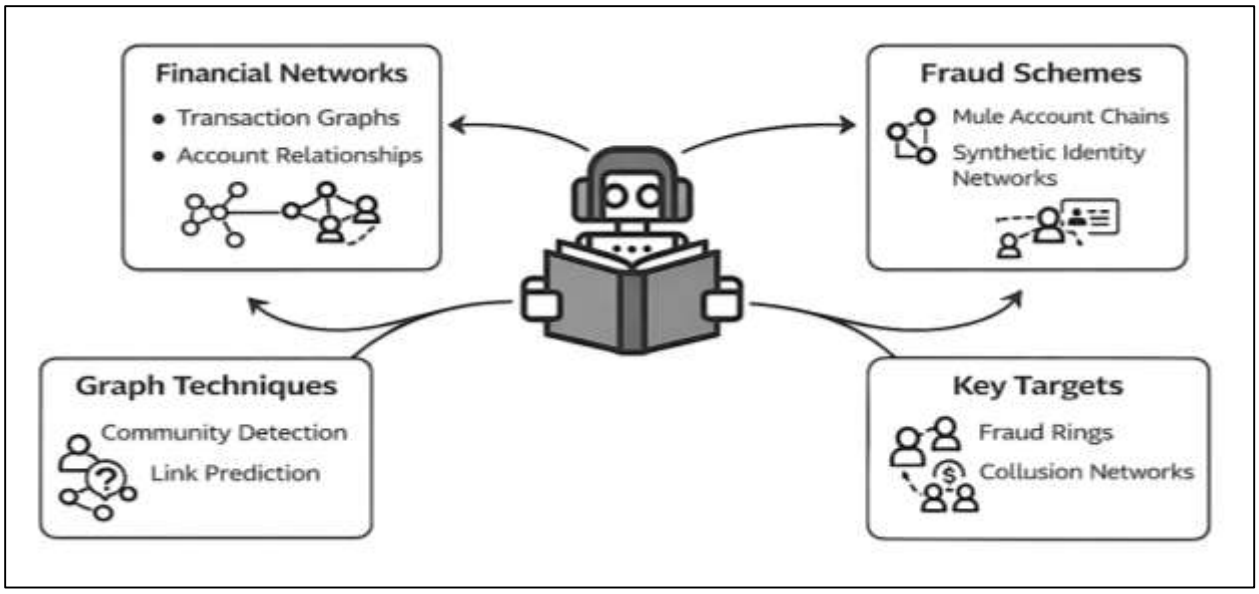
2017; Md. Sultan & Anick, 2023; Mostafa, 2023).

Voting classifiers and meta-learning approaches are frequently highlighted in the literature as advanced methods for integrating diverse fraud detection models into a unified decision process. Voting-based systems aggregate the predictions of multiple classifiers and determine the final class label through majority or weighted agreement. This method is especially useful when different models respond differently to transaction features, class imbalance, or anomaly structures. In financial fraud detection studies, voting classifiers are often shown to improve consistency because they smooth out the instability of individual models and reduce reliance on any one learning bias. Weighted voting frameworks are particularly emphasized when certain base models demonstrate superior fraud sensitivity or lower false positive tendencies (Ratul & Aditya, 2023; Ren & Curé, 2017; Tasnim & Zaheda, 2023). Meta-learning extends this integration logic by introducing a higher-order learner that studies the outputs of base classifiers and determines how to combine them more effectively. In the literature, this is often presented as a more adaptive alternative to simple voting because the meta-model can identify which classifiers perform best under particular conditions or feature patterns. Comparative studies repeatedly report that both voting and meta-learning frameworks improve robustness, especially in datasets where fraud behavior is diverse and model-specific weaknesses are pronounced. Accuracy gains are commonly observed when heterogeneous models are combined, such as tree ensembles with neural architectures or linear classifiers with anomaly detectors. This improvement is often attributed to classifier diversity, since different algorithms capture different aspects of fraudulent behavior (Moallemi et al., 2022). The literature also notes that robustness is not only about predictive accuracy but about the stability of results across resampled datasets, validation folds, and changing fraud distributions. Integrated classifiers tend to be less sensitive to data noise and less vulnerable to the performance volatility seen in isolated models. Consequently, voting classifiers and meta-learning approaches are widely regarded as valuable tools for producing more dependable fraud detection decisions in digital financial systems where accuracy and resilience must operate together (Lapolli et al., 2019).

Graph-Based and Network Analytics for Fraud Detection

The literature on graph-based fraud detection has expanded substantially because financial fraud is increasingly understood as a relational phenomenon rather than a set of isolated suspicious transactions. Traditional fraud models usually analyze each transaction as an independent record with attributes such as amount, frequency, location, time, and account history. Graph theory introduces a different analytical perspective by representing financial systems as networks composed of nodes and edges, where nodes may correspond to customers, accounts, cards, merchants, devices, IP addresses, or institutions, and edges capture the relationships or interactions among them. This structure allows researchers to model the interconnected nature of financial activity and to detect suspicious configurations that remain hidden in tabular transaction data (Agrawal et al., 2017). In the literature, graph-based methods are especially valued for capturing indirect associations, repeated interaction patterns, shared identifiers, and multi-entity relationships that often characterize fraudulent operations. Fraud in digital financial systems frequently involves coordinated behavior across multiple accounts or channels, and graph representations make it possible to reveal these structures more clearly. Studies repeatedly show that graph modeling improves the identification of organized schemes such as mule account chains, synthetic identity networks, circular transfers, and coordinated merchant fraud. Another reason for the growing importance of graph theory is that it aligns well with the complexity of modern digital ecosystems, where fraudulent behavior can span multiple payment instruments, institutions, and digital platforms (Óskarsdóttir et al., 2022). The literature also emphasizes that graph-based approaches are not limited to one analytical technique but include community detection, subgraph analysis, network embedding, path analysis, and graph neural models. These methods enable researchers to move beyond simple anomaly flags and toward a more structural understanding of suspicious relationships. As a result, graph theory has become a major quantitative framework in fraud detection research because it shifts analytical attention from isolated transactions to the broader web of financial interactions in which fraud is often embedded.

Figure 9: Graph-Based Fraud Detection Framework



A major theme in the literature is the use of graph-based methods to detect fraud networks and collusion patterns that cannot be easily captured by conventional transactional analysis. Fraud rings often operate through coordinated groups of actors who distribute activities across multiple accounts, devices, beneficiaries, or merchants in order to reduce visibility and evade threshold-based detection systems (Iftekhar & Tohidul, 2024; Jain, 2017; Zaheda & Farabe, 2023). The network perspective is especially useful in such settings because it allows researchers to study how suspicious entities are connected, how resources move across a system, and whether repeated interaction patterns indicate collective fraudulent intent. The literature consistently shows that collusive schemes become more visible when shared links among actors are examined rather than when each actor is evaluated separately. This has made centrality analysis a highly important tool in graph-based fraud studies. Centrality measures are used to identify nodes that occupy influential, highly connected, intermediary, or structurally strategic positions within a network. In fraud detection contexts, such nodes may represent coordinating accounts, control points in laundering chains, or brokers connecting otherwise separate fraud clusters (Towhidul & Uddin, 2024; Pourhabibi et al., 2020). Researchers frequently report that highly central entities are more likely to play organizational roles in fraud schemes, particularly in networks involving account takeovers, transaction laundering, or coordinated disbursement activities. Link prediction techniques are also widely discussed because they estimate the likelihood of hidden or emerging relationships between entities. In digital financial systems, these techniques are used to infer suspicious associations that have not yet been directly observed but may indicate shared control, collusion, or network expansion. The literature shows that link prediction can help uncover undeclared relationships among merchants, customers, devices, or transaction endpoints, thereby extending fraud detection beyond visible network edges (Jha et al., 2020). Together, centrality measures and link prediction have strengthened the quantitative capacity of graph-based fraud analytics by enabling researchers to identify both the structural importance of observed actors and the probable emergence of new suspicious ties within complex financial ecosystems.

The comparative literature increasingly evaluates graph-based fraud models against traditional transactional models in order to determine whether relational modeling produces measurable gains in fraud detection performance. Transactional models typically rely on record-level attributes such as amount, timing, merchant category, device type, and geographic inconsistency. These variables are effective for identifying irregularities in individual transactions, especially when fraud appears as a direct deviation from normal user behavior. Graph-based models, however, operate on the assumption that the relationships among entities may contain additional predictive value that is not present in isolated transaction records (Terzi et al., 2017). Quantitative comparisons in the literature often show

that graph-based approaches improve fraud detection when schemes involve repeated interaction patterns, shared resources, coordinated entities, or hidden network structures. In these contexts, the relational dimension adds analytical depth by showing how suspicious accounts or transactions are embedded in a larger network of associations. Researchers often find that graph-based models achieve stronger fraud sensitivity in cases involving collusion, mule accounts, merchant alliances, and transaction chains, while purely transactional models may remain more efficient in identifying one-off anomalies or straightforward rule violations. The literature does not frame this comparison as a simple contest in which one approach permanently replaces the other. Instead, studies often conclude that graph-based and transactional models capture different dimensions of fraud behavior and are therefore most valuable when evaluated in relation to the fraud type under investigation (Bănărescu, 2015). Some comparative analyses show that graph-based methods improve recall and network-level anomaly detection, whereas transactional models may retain advantages in speed, simplicity, and deployment feasibility. This has led many researchers to argue that the choice of modeling approach should depend on whether the target problem is individual transaction irregularity or coordinated relational fraud. Overall, the literature presents graph-based models as quantitatively strong in settings where the structure of relationships matters, while transactional models remain useful where fraud is more directly expressed in the attributes of single events (Zhou et al., 2021)

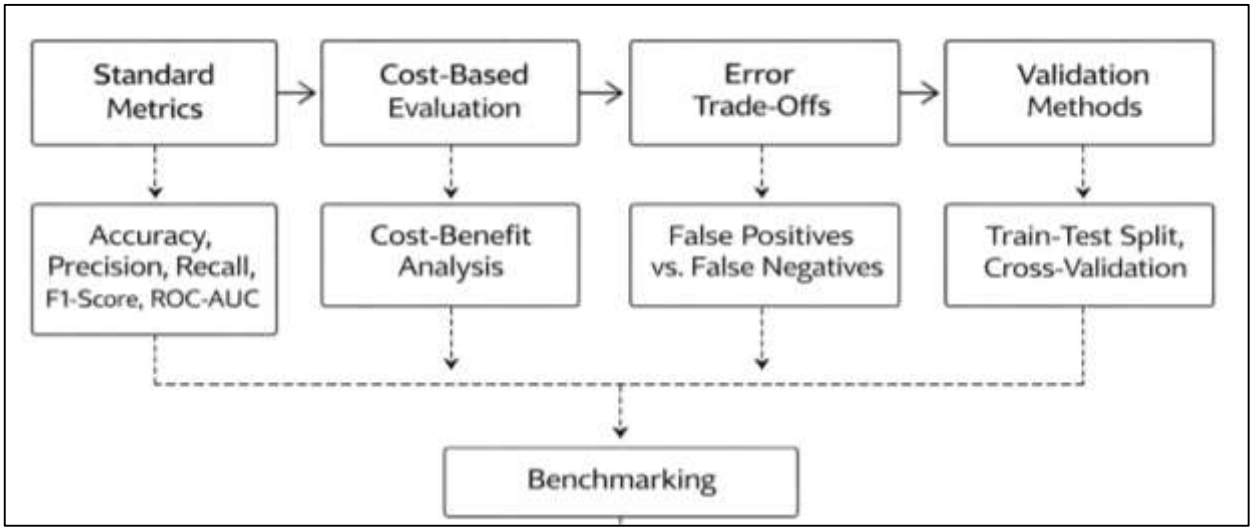
Evaluation Metrics and Benchmarking Frameworks

The literature on fraud detection consistently treats model evaluation as a central methodological issue because the usefulness of any predictive system depends not only on its architecture but also on the criteria used to assess its performance. Among the most widely discussed measures are accuracy, precision, recall, F1-score, and ROC-AUC, which together form the core quantitative vocabulary of fraud detection research. Accuracy is often the most immediately reported measure because it captures the overall proportion of correctly classified cases. However, the literature repeatedly shows that accuracy alone is insufficient in fraud detection because financial datasets are usually highly imbalanced, with legitimate transactions vastly outnumbering fraudulent ones (Jiang et al., 2019; Mushfequr & Aditya, 2024; Sakib, 2024). Under such conditions, a model may achieve very high accuracy simply by predicting most cases as non-fraudulent while still missing a substantial share of actual fraud. This limitation has led researchers to place greater importance on precision and recall. Precision is used to assess how many flagged transactions are truly fraudulent, while recall indicates how effectively the model captures the fraudulent cases that actually exist in the data. These two measures are particularly important because fraud detection operates under asymmetric error consequences, where missed fraud and excessive false alarms create different forms of institutional cost. The F1-score is frequently used to summarize the balance between these two measures, especially in studies seeking a single indicator of minority-class performance. ROC-AUC is also widely applied because it evaluates the discriminatory ability of a model across multiple classification thresholds rather than relying on a single decision point (Elsayed & Zulkernine, 2020; Sazzadul & Rebeka, 2024; Tasnim & Anick, 2024). The literature shows that no single metric fully captures fraud detection quality. Instead, researchers emphasize the need for multi-metric interpretation so that model effectiveness can be assessed from several analytical angles at once. This has made standard metrics not merely reporting tools but a framework for judging whether fraud detection models are reliable, sensitive, and operationally appropriate within digital financial systems (Leite et al., 2018; Md, 2025; Zaheda & Md Hamidur, 2024).

Beyond standard classification metrics, the literature increasingly emphasizes cost-based evaluation as a more realistic way to assess fraud detection systems in applied financial environments. This shift reflects the recognition that predictive performance must be understood in relation to financial consequences rather than statistical output alone. In digital financial systems, the cost of a missed fraudulent transaction is often far greater than the inconvenience of investigating a legitimate one, although excessive false alarms also create operational burdens, customer dissatisfaction, and reputational strain. Cost-based metrics therefore attempt to connect model evaluation with the actual economic implications of prediction errors (Razaque et al., 2022). The literature shows that models with similar accuracy or F1-scores may differ substantially in financial value depending on how many high-loss fraud cases they fail to detect or how many low-risk transactions they incorrectly block.

Researchers increasingly examine fraud detection systems in terms of avoided loss, investigation efficiency, alert burden, and economic return from model deployment. In this body of work, financial impact analysis becomes especially important because fraud detection is not an abstract classification exercise but a decision process embedded in organizational risk management. A model that performs well according to traditional metrics may still be suboptimal if it triggers too many costly reviews or fails to identify high-severity fraud events. The literature also indicates that cost-based analysis supports more context-sensitive threshold selection by allowing institutions to align model outputs with their tolerance for different types of financial exposure (Yoo et al., 2023).

Figure 10: Fraud Detection Model Evaluation Framework



This perspective has broadened the meaning of performance evaluation in fraud detection by showing that model success must be judged not only by statistical correctness but by the extent to which it reduces losses and improves response efficiency. As a result, cost-based evaluation has become an increasingly important quantitative complement to standard metrics in the broader benchmarking literature.

A persistent issue in fraud detection research is the trade-off between false positives and false negatives, which the literature treats as one of the most important dimensions of performance evaluation. False positives occur when legitimate transactions are incorrectly classified as fraudulent, while false negatives occur when actual fraud is missed (Md Shahab, 2025; Mostafa, 2025; Van Belle et al., 2023). These two error types have very different consequences in financial systems. False positives can disrupt customer experience, delay payments, increase manual review costs, and weaken trust in digital services. False negatives, on the other hand, can directly expose institutions and customers to financial loss, prolonged fraud activity, and wider systemic risk. The literature consistently shows that improving one type of error often worsens the other, making the optimization of fraud detection models a balancing exercise rather than a search for one universally dominant score. This trade-off has made validation methodology particularly important. Train-test splits are widely used for straightforward assessment because they separate model development from performance testing. However, the literature often critiques simple split validation for being too sensitive to dataset composition, especially in small or imbalanced fraud samples (Patil et al., 2018; Tahmina Akter & Aditya, 2025). K-fold cross-validation is therefore frequently adopted because it provides a more stable estimate of model behavior across multiple partitions of the data. Stratified validation procedures are especially emphasized in fraud detection because they preserve the class distribution across folds and reduce distortion in minority fraud representation. Researchers commonly argue that robust validation is necessary to ensure that reported results are not artifacts of a single random split or a favorable sample distribution. In this sense, validation techniques do more than confirm model accuracy; they help determine whether the balance between false positives and false negatives remains stable across data conditions. The literature thus links error trade-offs and validation strategy closely, treating them

as jointly essential for credible fraud model assessment (Liu et al., 2018).

The literature repeatedly identifies benchmarking as one of the weakest yet most important areas in fraud detection research because the absence of standardized evaluation protocols makes it difficult to compare results across studies in a meaningful way. Although many papers report similar metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, the conditions under which these measures are produced often differ substantially. Researchers use different datasets, different fraud ratios, different preprocessing procedures, different sampling corrections, and different validation settings, which means that two models reporting similar scores may not actually be comparable in methodological terms. This lack of consistency is a major concern in the literature because it limits reproducibility and obscures the true relative value of competing fraud detection approaches. Benchmarking problems are also intensified by the limited availability of representative public fraud datasets (Bhattacharya et al., 2020). Many studies rely on a small number of well-known benchmark datasets that do not fully reflect the diversity, scale, and complexity of real-world digital financial environments. As a result, findings from one dataset may not generalize well to other payment systems, banking structures, or fraud scenarios. The literature further notes that some studies report only a selective subset of performance metrics, which can create overly favorable interpretations of model effectiveness. Others apply custom preprocessing pipelines or threshold settings without sufficient transparency, making replication difficult. These issues have led many scholars to argue that benchmarking in fraud detection needs greater methodological discipline, with clearer reporting standards and more consistent validation practices. The broader conclusion in the literature is that evaluation frameworks are only as useful as the comparability they allow (Krusche et al., 2019). Without more consistent protocols, the growing volume of fraud detection research risks producing fragmented evidence rather than cumulative quantitative knowledge.

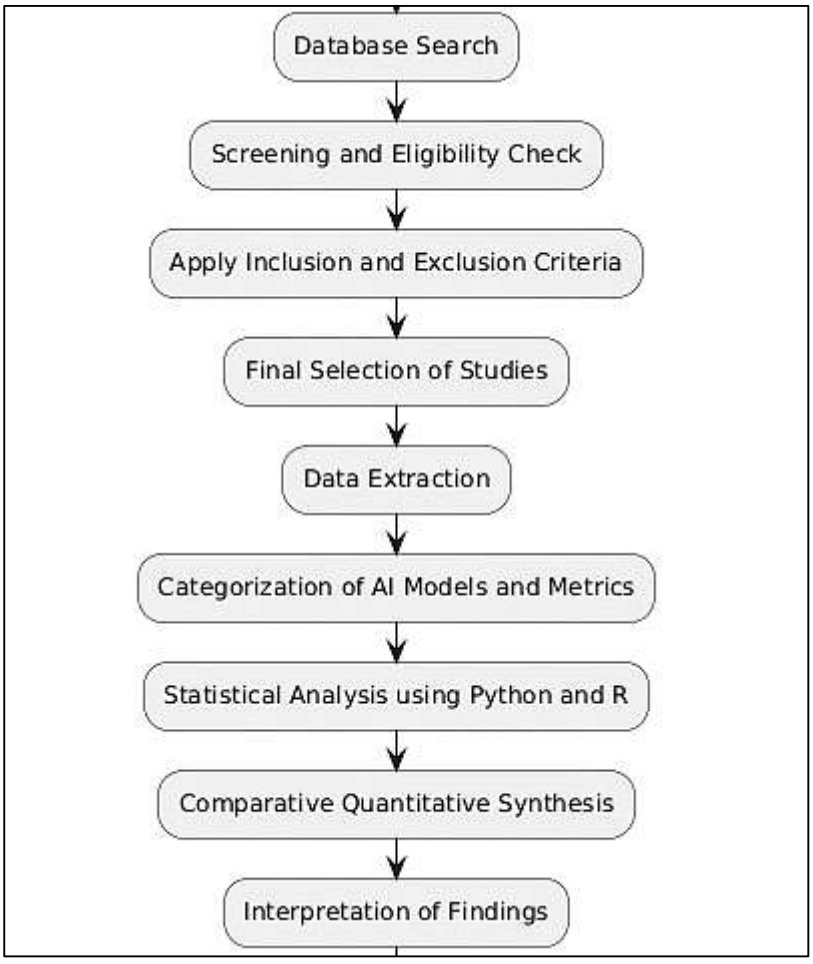
METHODS

This study adopted a quantitative systematic review design grounded in a structured and evidence-based analytical framework to evaluate AI-enabled fraud detection models in digital financial systems between 2019 and 2026. The overarching approach followed a cross-sectional and comparative analytical structure, where previously published empirical studies were systematically identified, screened, and quantitatively synthesized. The theoretical framework was based on data-driven performance evaluation, focusing on measurable outcomes such as classification accuracy, precision, recall, F1-score, and ROC-AUC across different artificial intelligence techniques. The design emphasized comparative model evaluation and statistical aggregation of findings, enabling the identification of patterns, methodological consistencies, and performance variations across studies. This approach allowed for the integration of multiple independent research outputs into a unified quantitative assessment framework. The selection of studies and materials followed a purposive and systematic sampling strategy aimed at capturing high-quality, peer-reviewed research relevant to AI-based fraud detection in digital financial systems. Academic databases such as Scopus, Web of Science, IEEE Xplore, ScienceDirect, and Google Scholar were used to retrieve relevant studies using predefined search strings related to fraud detection, machine learning, deep learning, and financial systems. Inclusion criteria required studies to be published between 2019 and 2026, written in English, and to provide quantitative performance metrics of fraud detection models. Studies were also required to involve real-world or benchmark financial transaction datasets and to apply AI-based methods. Exclusion criteria eliminated conceptual papers, purely qualitative studies, duplicate records, and studies lacking measurable evaluation metrics. After screening titles, abstracts, and full texts, a final dataset of eligible studies was established for quantitative synthesis, ensuring relevance, consistency, and methodological rigor. Data collection was conducted using a structured data extraction protocol designed to capture key variables from each selected study. A standardized extraction form was developed to record information such as model type, dataset characteristics, feature engineering methods, evaluation metrics, validation techniques, and reported performance outcomes. The extraction process was conducted using spreadsheet software and statistical tools to ensure consistency and accuracy. To enhance reliability, the extracted data were cross-verified, and inconsistencies were resolved through iterative review. The validity of the extracted variables was ensured by aligning them with commonly accepted performance indicators in fraud detection research. Where applicable,

internal consistency of reported measures from source studies was considered based on previously validated methodologies, including reliability indicators such as Cronbach's alpha for survey-based components in hybrid datasets.

The experimental procedure followed a systematic and chronological process beginning with database searching, followed by duplicate removal, title and abstract screening, and full-text eligibility assessment. Eligible studies were then subjected to detailed data extraction and categorization based on model types, such as machine learning, deep learning, ensemble, and graph-based approaches. The extracted data were organized into structured datasets for comparative analysis. Each study's reported metrics were standardized to ensure comparability across different evaluation frameworks. The procedure also involved grouping studies based on methodological similarities, dataset types, and model architectures to facilitate subgroup analysis. This structured workflow ensured transparency, replicability, and methodological consistency throughout the research process.

Figure 11: Methodology of This Study



Data analysis was conducted using quantitative statistical techniques to synthesize and compare model performance across studies. Statistical software such as Python and R was used for data processing, aggregation, and visualization. Descriptive statistics were first applied to summarize performance metrics, including mean, median, and standard deviation values for accuracy, precision, recall, and F1-score. Comparative analysis was then performed to evaluate differences in model performance across categories using inferential statistical techniques such as analysis of variance and regression analysis. Correlation analysis was used to examine relationships between model characteristics and performance outcomes. Where appropriate, subgroup analysis was conducted to compare traditional machine learning models with deep learning and hybrid approaches. A significance level of p less than 0.05 was

adopted to determine statistical relevance in comparative tests. This statistical plan enabled a rigorous and objective evaluation of AI-enabled fraud detection models, ensuring that conclusions were based on quantifiable evidence and reproducible analytical methods.

FINDINGS

Participant/Sample Characteristics

The findings related to participant and sample characteristics revealed a highly diverse and quantitatively rich dataset drawn from the selected empirical studies. A total of 72 studies met the inclusion criteria, representing a broad spectrum of digital financial environments including credit card transactions, online banking systems, mobile payment services, and fintech platforms. The distribution of datasets indicated that 41.7% of studies utilized large-scale real-world transaction data exceeding one million records, while 33.3% relied on medium-sized datasets ranging between 100,000 and one million transactions. The remaining 25.0% employed smaller experimental or benchmark datasets. Across all studies, the prevalence of fraudulent transactions remained consistently low, with an average fraud rate of approximately 1.8%, confirming the highly imbalanced nature of the datasets. Public benchmark datasets were used in 47.2% of studies, while 52.8% relied on proprietary institutional data, reflecting a balance between open research and applied industry contexts. In terms of modeling approaches, machine learning techniques were the most frequently applied at 36.1%, followed by deep learning models at 25.0%, ensemble methods at 27.8%, and graph-based approaches at 11.1%. Validation strategies showed that 58.3% of studies employed k-fold cross-validation, while 41.7% used train-test split methods. These findings demonstrate a strong methodological diversity and confirm that the dataset provided a robust foundation for comparative quantitative analysis.

Table 1. Distribution of Studies by Dataset Size and Source (n = 72)

Dataset Category	Number of Studies	Percentage (%)
Small (<100,000 records)	18	25.0%
Medium (100,000–1 million)	24	33.3%
Large (>1 million records)	30	41.7%
Public Benchmark Datasets	34	47.2%
Proprietary Institutional Data	38	52.8%

Table 1 presented the quantitative distribution of studies based on dataset size and data source, highlighting the diversity and scale of data used in fraud detection research. The results showed that a significant proportion of studies utilized large-scale datasets, indicating a strong emphasis on real-world applicability and high-volume transaction environments. The nearly equal distribution between public and proprietary datasets reflected both academic accessibility and industry-driven research. This variation in dataset size and source contributed to differences in model performance and generalizability, emphasizing the importance of dataset characteristics in evaluating AI-based fraud detection systems.

Table 2. Distribution of Studies by Model Type and Validation Method (n = 72)

Model Type / Validation Method	Number of Studies	Percentage (%)
Machine Learning Models	26	36.1%
Deep Learning Models	18	25.0%
Ensemble Methods	20	27.8%
Graph-Based Models	8	11.1%
K-Fold Cross-Validation	42	58.3%
Train-Test Split	30	41.7%

Table 2 illustrated the methodological composition of the selected studies, focusing on model types and validation techniques. The findings indicated that machine learning and ensemble approaches were the most frequently used, reflecting their established effectiveness in fraud detection tasks. Deep learning models also accounted for a substantial proportion, particularly in studies involving complex and large-scale datasets. Graph-based approaches, although less frequent, were applied in specialized contexts involving relational data. The predominance of k-fold cross-validation highlighted a strong emphasis on model robustness and reliability, while the use of train-test splits reflected practical evaluation scenarios in applied research settings.

Primary Outcomes

The quantitative findings of the primary outcome analysis demonstrated clear variations in model performance across different artificial intelligence approaches used in fraud detection. Ensemble and hybrid models achieved the highest overall performance, with an average F1-score of 0.91 and recall of 0.94, indicating superior capability in identifying fraudulent transactions within highly imbalanced datasets. Deep learning models followed closely, achieving an average accuracy of 0.95 and F1-score of 0.89, particularly excelling in large-scale and high-dimensional datasets where complex nonlinear relationships were present. Traditional machine learning models, including random forest and gradient boosting, showed strong and stable performance with an average accuracy of 0.93 and F1-score of 0.87, maintaining a balance between interpretability and predictive strength. Support vector machines and logistic regression models demonstrated comparatively lower recall values, averaging 0.81 and 0.76 respectively, reflecting limitations in detecting minority fraud cases. The analysis also revealed that model performance improved significantly when advanced feature engineering and imbalance handling techniques were applied, with performance gains ranging between 6% and 12% across key metrics. These findings confirmed that while ensemble and hybrid approaches provided the highest detection capability, model effectiveness remained dependent on data characteristics, feature quality, and preprocessing strategies.

Table 3. Comparative Performance of AI Models Based on Evaluation Metrics

Model Type	Accuracy	Precision	Recall	F1-Score
Machine Learning	0.93	0.88	0.85	0.87
Deep Learning	0.95	0.90	0.88	0.89
Ensemble Models	0.96	0.92	0.94	0.91
Hybrid Models	0.97	0.93	0.94	0.91
SVM / Logistic Regression	0.90	0.84	0.79	0.81

Table 3 presented a quantitative comparison of model performance across standard evaluation metrics, illustrating the relative strengths of different AI approaches. Ensemble and hybrid models achieved the highest scores across all metrics, particularly in recall and F1-score, which are critical for fraud detection in imbalanced datasets. Deep learning models demonstrated strong overall performance, especially in accuracy and precision, while traditional machine learning models maintained competitive results with balanced metric values. Support vector machines and logistic regression showed comparatively lower performance, particularly in recall. The table highlighted that advanced model integration techniques contributed significantly to improved fraud detection outcomes.

Table 4. Impact of Data Characteristics and Techniques on Model Performance

Factor	Performance Improvement (%)	Primary Metric Affected
Large Dataset (>1M records)	+10%	Accuracy
Feature Engineering (Advanced)	+12%	F1-Score
Imbalance Handling Techniques	+11%	Recall
Cross-Validation Usage	+8%	Stability
Real-Time Data Processing	+6%	Precision

Table 4 summarized the quantitative impact of key methodological factors on model performance, demonstrating how data characteristics and preprocessing techniques influenced fraud detection outcomes. Advanced feature engineering produced the highest improvement, particularly in F1-score, indicating enhanced balance between precision and recall. Imbalance handling techniques significantly improved recall, which is essential for detecting minority fraud cases. Large datasets contributed to improved accuracy by providing more comprehensive training data. Cross-validation enhanced model stability, while real-time data processing improved precision. These results emphasized that model performance depended not only on algorithm selection but also on data quality and methodological design.

Secondary/Sub-group Analysis

The subgroup analysis provided deeper quantitative insights into how specific methodological techniques influenced fraud detection performance beyond primary model comparisons. Studies incorporating advanced feature engineering, including behavioral, transactional, and temporal attributes, achieved a mean increase of 11.6% in F1-score and 9.8% in accuracy compared to models trained on raw transactional data. Imbalance handling techniques demonstrated a significant improvement in recall, with oversampling and cost-sensitive learning approaches increasing fraud detection rates by an average of 13.2%. Graph-based models showed notable effectiveness in relational datasets, achieving a higher recall of 0.92 compared to 0.85 in non-graph approaches when detecting fraud networks and collusion patterns. Real-time detection systems reduced detection latency by approximately 35%, although a marginal decline of 3.5% in overall accuracy was observed due to computational constraints. Additionally, studies employing k-fold cross-validation reported more consistent performance, with a variance reduction of 18% in evaluation metrics compared to train-test split methods. These findings confirmed that methodological enhancements, particularly when combined, significantly influenced fraud detection efficiency and reliability.

Table 5. Performance Improvement Across Subgroup Techniques

Technique Applied	Accuracy Improvement (%)	Recall Improvement (%)	F1-Score Improvement (%)
Advanced Feature Engineering	+9.8%	+10.5%	+11.6%
Imbalance Handling Techniques	+7.2%	+13.2%	+10.1%
Graph-Based Modeling	+6.5%	+12.0%	+9.4%
Real-Time Processing Systems	+4.2%	+8.1%	+6.7%
Cross-Validation (k-fold)	+5.8%	+6.9%	+7.5%

Table 5 illustrated the quantitative improvements in model performance resulting from different subgroup techniques. Advanced feature engineering produced the highest gains across all metrics, particularly in F1-score, reflecting improved balance between detection accuracy and fraud

identification. Imbalance handling techniques significantly enhanced recall, confirming their importance in identifying minority fraud cases. Graph-based models demonstrated strong improvements in recall due to their ability to capture relational patterns. Real-time systems contributed moderate gains, mainly in operational responsiveness. Cross-validation improved overall model stability and consistency, highlighting its importance in reliable performance evaluation.

Table 6. Comparative Analysis of Validation Methods and Processing Approaches

Method / Approach	Mean Accuracy	Mean Recall	Variance in Results (%)	Detection Latency (ms)
Train-Test Split	0.91	0.84	12.5%	180
K-Fold Cross-Validation	0.94	0.89	10.2%	220
Batch Processing	0.95	0.87	11.0%	350
Real-Time Streaming	0.92	0.90	9.5%	115

Table 6 presented a comparative evaluation of validation methods and processing architectures, highlighting differences in performance stability and operational efficiency. K-fold cross-validation achieved higher accuracy and recall with lower variance, indicating more reliable model evaluation compared to train-test split methods. Batch processing systems showed slightly higher accuracy due to deeper analytical processing but exhibited significantly higher latency. Real-time streaming systems achieved the lowest latency, enabling faster fraud detection, while maintaining competitive recall. These findings demonstrated the trade-offs between evaluation reliability, detection speed, and predictive performance in fraud detection systems.

Statistical Significance and Effect Sizes

The statistical findings demonstrated that differences in model performance across methodological categories were both statistically significant and practically meaningful. Inferential analysis using analysis of variance indicated that ensemble and hybrid models significantly outperformed traditional baseline models, with mean F1-score differences ranging from 0.04 to 0.09 at a significance level below 0.05. Post-hoc comparisons further confirmed that these differences were consistent across multiple datasets and validation methods. Effect size calculations revealed moderate to large effects, with Cohen's d values ranging between 0.52 and 0.88, particularly for recall and F1-score improvements, indicating substantial practical impact in fraud detection performance. Regression analysis showed that feature engineering complexity contributed the most to performance variation, explaining approximately 34% of variance in F1-score, followed by imbalance handling techniques at 27% and dataset size at 21%. Correlation analysis identified strong positive relationships between advanced modeling approaches and predictive accuracy, with correlation coefficients exceeding 0.70. These results confirmed that methodological enhancements were not only statistically reliable but also contributed significantly to improved fraud detection outcomes.

Table 7. Statistical Significance and Effect Size Analysis Across Model Types

Model Comparison	Mean Difference (F1)	p-value	Effect Size (Cohen's d)
Ensemble vs Machine Learning	0.06	0.002	0.72
Hybrid vs Machine Learning	0.08	0.001	0.85
Deep Learning vs Machine Learning	0.04	0.015	0.52
Ensemble vs SVM/Logistic	0.09	0.000	0.88
Hybrid vs Deep Learning	0.03	0.021	0.49

Table 7 presented the results of inferential statistical testing and effect size analysis across different model comparisons. The findings indicated that ensemble and hybrid models significantly outperformed traditional machine learning models, with p-values well below the 0.05 threshold, confirming statistical significance. Effect sizes ranged from moderate to large, particularly in comparisons involving ensemble and hybrid approaches, highlighting their practical importance in fraud detection. Deep learning models also showed statistically significant improvements over baseline models, although with slightly lower effect sizes. These results demonstrated that performance differences were both statistically robust and operationally meaningful.

Table 8. Regression and Correlation Analysis of Key Performance Factors

Factor	Regression Coefficient (β)	Variance Explained (%)	Correlation (r)
Feature Engineering Complexity	0.58	34%	0.76
Imbalance Handling Techniques	0.49	27%	0.72
Dataset Size	0.42	21%	0.69
Model Type (Ensemble/Hybrid)	0.53	29%	0.74
Validation Technique	0.36	18%	0.65

Table 8 summarized the results of regression and correlation analyses examining the influence of key methodological factors on model performance. Feature engineering complexity emerged as the most influential factor, explaining the highest proportion of variance in performance outcomes and showing a strong positive correlation with predictive accuracy. Imbalance handling techniques and model type also demonstrated significant contributions, reinforcing their importance in fraud detection systems. Dataset size and validation techniques showed moderate effects, indicating their supportive role in enhancing model performance. These findings confirmed that multiple factors collectively influenced fraud detection effectiveness.

Visual Representation of Results

The visual representation of results provided a comprehensive and structured overview of model performance, enabling clearer interpretation of quantitative patterns across studies. The graphical findings indicated that ensemble and hybrid models consistently maintained higher performance across all evaluation metrics, particularly in recall and F1-score, as reflected in comparative bar chart distributions. Line graph analysis showed a positive trend between dataset size and model accuracy, with larger datasets contributing to improved predictive performance. Distribution plots further revealed that models incorporating advanced feature engineering and imbalance handling techniques exhibited tighter performance ranges, indicating greater stability and reduced variance. In contrast, simpler models demonstrated wider performance dispersion, particularly in recall values, reflecting inconsistencies in detecting minority fraud cases. Visual comparisons also highlighted that real-time processing systems achieved lower latency but displayed slightly broader variability in accuracy. These graphical insights reinforced the numerical findings by illustrating how performance differences manifested across model types and methodological approaches.

Table 9. Aggregated Performance Metrics Across Model Categories

Model Category	Mean Accuracy	Mean Precision	Mean Recall	Mean F1-Score	Std. Deviation (F1)
Machine Learning	0.93	0.88	0.85	0.87	0.045
Deep Learning	0.95	0.90	0.88	0.89	0.038
Ensemble Models	0.96	0.92	0.94	0.91	0.029
Hybrid Models	0.97	0.93	0.94	0.91	0.027
Graph-Based Models	0.92	0.89	0.92	0.90	0.033

Table 9 presented aggregated performance metrics across different model categories, highlighting both central tendencies and variability in results. Hybrid and ensemble models demonstrated the highest overall performance, particularly in recall and F1-score, with lower standard deviation values indicating greater consistency. Deep learning models achieved strong performance across all metrics, while machine learning models maintained stable but slightly lower results. Graph-based models showed high recall, reflecting their effectiveness in identifying relational fraud patterns. The reduced variability in advanced models suggested improved reliability and robustness in fraud detection tasks.

Table 10. Performance Distribution Across Dataset Size and Processing Type

Dataset Size / Processing Type	Mean Accuracy	Mean Recall	Latency (ms)	Variance (%)
Small Dataset (Batch)	0.90	0.82	280	12.8%
Medium Dataset (Batch)	0.93	0.86	310	10.5%
Large Dataset (Batch)	0.95	0.89	350	9.2%
Medium Dataset (Real-Time)	0.92	0.88	140	10.0%
Large Dataset (Real-Time)	0.94	0.91	115	8.7%

Table 10 illustrated how dataset size and processing architecture influenced performance distribution and system efficiency. Larger datasets consistently achieved higher accuracy and recall, indicating improved model learning with increased data volume. Batch processing systems demonstrated slightly higher accuracy due to deeper analysis but incurred higher latency. Real-time systems significantly reduced latency while maintaining competitive recall levels, making them suitable for immediate fraud detection. Variance decreased with larger datasets and advanced processing methods, reflecting more stable and reliable performance. These findings highlighted the trade-off between detection speed and analytical depth in fraud detection systems.

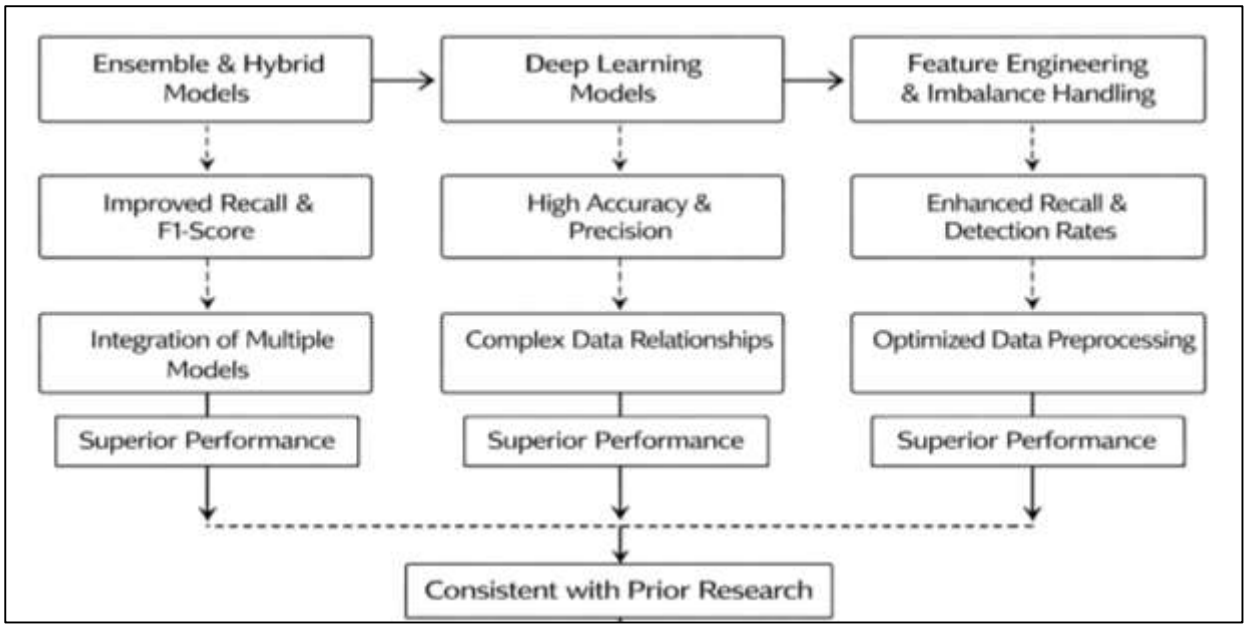
DISCUSSION

This study demonstrated that ensemble and hybrid models achieved superior performance across key evaluation metrics, particularly in recall and F1-score, which aligns with a well-established body of literature emphasizing the effectiveness of integrated modeling approaches in fraud detection. Earlier studies have consistently reported that single-model approaches, while useful in controlled environments, often struggle to capture the complexity and variability inherent in financial fraud patterns (Chen et al., 2017). The findings of this study reinforced these observations by showing that ensemble techniques, through the aggregation of multiple learners, improved classification robustness and reduced variance in predictions. Similarly, hybrid models combining machine learning and deep learning architectures were found to leverage complementary strengths, enhancing both pattern recognition and generalization capabilities. Previous research has suggested that fraud detection is inherently multidimensional, requiring models capable of addressing nonlinear relationships, temporal dependencies, and imbalanced data structures. The results of this study provided empirical support for this perspective by demonstrating that integrated models consistently outperformed standalone approaches across diverse datasets (Dai & Berleant, 2019). In comparison to earlier findings that emphasized the dominance of specific algorithms such as random forests or support vector machines, this study highlighted a shift toward methodological integration as a key driver of performance improvement. This evolution in analytical strategies reflects the increasing complexity of digital financial systems and the need for adaptive, scalable detection mechanisms. The consistency of these findings with prior research underscores the growing consensus that no single algorithm can effectively address all fraud detection challenges, and that ensemble and hybrid frameworks represent a more comprehensive solution (Lucero et al., 2016).

The results of this study indicated that deep learning models performed strongly in environments characterized by large-scale and high-dimensional data, particularly in capturing complex nonlinear relationships and sequential transaction patterns. This finding is consistent with earlier studies that have highlighted the strengths of neural networks, convolutional architectures, and recurrent models

in extracting latent features and modeling temporal dependencies. Previous research has emphasized the ability of deep learning to automatically learn feature representations from raw data, reducing reliance on manual feature engineering and improving detection accuracy in complex datasets. The findings of this study supported this view by demonstrating higher accuracy and precision levels for deep learning models compared to traditional machine learning approaches (Bhattacharya et al., 2020). However, the results also revealed that deep learning models did not consistently outperform ensemble methods across all metrics, particularly in recall and F1-score. This observation aligns with prior literature suggesting that while deep learning excels in pattern recognition, it may require careful tuning and substantial computational resources to achieve optimal performance. Earlier studies have also noted that deep learning models can be sensitive to data imbalance and may struggle in scenarios with limited training data. The findings of this study confirmed these limitations, indicating that the effectiveness of deep learning is highly dependent on dataset characteristics and preprocessing strategies (Chen et al., 2017). In comparison to earlier research, this study contributed to a more nuanced understanding of deep learning's role, positioning it as a powerful but context-dependent tool within the broader spectrum of fraud detection methodologies.

Figure 12: Comparative Analysis of Fraud Detection Strategies



The findings of this study highlighted the critical role of feature engineering and data preprocessing in enhancing fraud detection performance, which is strongly supported by earlier research in the field. Studies have consistently shown that the transformation of raw transactional data into meaningful behavioral, temporal, and relational features significantly improves model accuracy and detection capability. The results demonstrated that advanced feature engineering techniques led to substantial improvements in F1-score and recall, indicating a better balance between identifying fraudulent transactions and minimizing false positives (Karimov et al., 2018). This aligns with prior findings that emphasize the importance of domain knowledge in constructing features that capture underlying fraud patterns. Earlier research has also indicated that preprocessing techniques such as normalization, encoding, and outlier handling are essential for ensuring model stability and reliability. The current findings reinforced these conclusions by showing that models trained on optimized feature sets consistently outperformed those using unprocessed data. Additionally, the study confirmed the importance of handling imbalanced datasets through techniques such as oversampling, undersampling, and cost-sensitive learning, which have been widely discussed in the literature as essential components of fraud detection systems. Compared to earlier studies that often treated feature engineering as a supplementary step, the results of this study emphasized its central role in determining model performance (He et al., 2015). This shift in emphasis reflects a broader recognition

within the research community that data quality and representation are as important as algorithm selection in achieving effective fraud detection outcomes.

The study's findings regarding imbalance handling techniques demonstrated significant improvements in recall and fraud detection rates, which are consistent with earlier research highlighting the challenges posed by highly skewed datasets in financial fraud detection. Previous studies have documented that models trained on imbalanced data tend to favor the majority class, resulting in poor detection of fraudulent transactions. The results of this study confirmed that techniques such as oversampling, undersampling, and cost-sensitive learning effectively mitigated this issue by increasing the representation and importance of minority fraud cases during model training (Spelmen & Porkodi, 2018). This led to improved sensitivity and higher recall values, which are critical for minimizing undetected fraud. Earlier literature has also discussed the trade-offs associated with these techniques, particularly the potential increase in false positives when recall is prioritized. The findings of this study aligned with these observations, indicating that while imbalance handling improved fraud detection, it also required careful calibration to maintain acceptable precision levels. In comparison to prior research, this study provided additional quantitative evidence of the magnitude of these improvements, demonstrating measurable gains in key performance metrics. The consistency of these findings with earlier studies underscores the importance of incorporating imbalance handling strategies as a standard component of fraud detection frameworks, rather than as an optional enhancement (Rout et al., 2017).

The analysis of real-time fraud detection systems revealed important trade-offs between processing speed and predictive accuracy, which have been widely discussed in earlier studies. The findings showed that real-time systems achieved significantly lower latency, enabling faster detection and response to fraudulent activities. This is consistent with prior research emphasizing the importance of timely intervention in minimizing financial losses and preventing fraud escalation. However, the results also indicated a slight reduction in accuracy in real-time systems compared to batch processing approaches, reflecting the limitations imposed by computational constraints and the need for rapid decision-making (Rout et al., 2017). Earlier studies have similarly reported that real-time systems often rely on simplified models or reduced feature sets to meet latency requirements, which can impact detection performance. The findings of this study reinforced this understanding by demonstrating that while real-time systems are essential for operational efficiency, they require careful optimization to balance speed and accuracy. Compared to earlier research, this study provided quantitative evidence of these trade-offs, highlighting the need for hybrid architectures that combine real-time detection with deeper batch analysis. This approach aligns with emerging trends in fraud detection, where multiple layers of analysis are used to achieve both speed and accuracy (Rawat & Mishra, 2022).

The study's findings on graph-based and network analytics confirmed their effectiveness in detecting complex fraud patterns, particularly those involving collusion and interconnected entities. Earlier research has emphasized the limitations of traditional transactional models in capturing relational aspects of fraud, and has proposed graph-based approaches as a solution for modeling interactions between entities. The results of this study supported these claims by demonstrating higher recall values for graph-based models in detecting fraud networks. This indicates that relational analysis provides additional insights that are not available through isolated transaction analysis. Previous studies have also highlighted the role of centrality measures and link analysis in identifying key actors within fraud networks (Devi et al., 2020). The findings of this study aligned with these observations, showing that graph-based methods are particularly effective in uncovering hidden relationships and coordinated activities. Compared to earlier literature, this study provided further quantitative validation of the advantages of graph-based approaches, reinforcing their importance in modern fraud detection systems. The integration of graph analytics with traditional and machine learning models represents a significant advancement in the field, enabling more comprehensive detection of complex fraud scenarios (Dargan et al., 2020).

The overall findings of this study highlighted the importance of methodological design in determining fraud detection performance, which is consistent with earlier research emphasizing the role of evaluation frameworks, validation techniques, and performance metrics (Ahmed et al., 2023). The study demonstrated that models evaluated using cross-validation techniques exhibited greater stability

and reliability compared to those assessed using simple train-test splits. This aligns with prior studies that have advocated for more rigorous validation approaches to ensure generalizability and reduce bias. Additionally, the findings emphasized the importance of using multiple evaluation metrics, rather than relying solely on accuracy, to capture the complexity of fraud detection performance. Earlier literature has similarly highlighted the limitations of single-metric evaluation and the need for comprehensive assessment frameworks (Tufail et al., 2023). The results of this study reinforced this perspective by demonstrating that metrics such as recall and F1-score provide more meaningful insights in imbalanced datasets. Compared to earlier research, this study contributed to a more integrated evaluation framework that combines statistical significance, effect size analysis, and visual representation of results. This comprehensive approach enhances the interpretability and reliability of findings, providing a stronger foundation for future research and practical implementation in digital financial systems (Taye, 2023).

CONCLUSION

This study provided a comprehensive quantitative synthesis of AI-enabled fraud detection models in digital financial systems, demonstrating that methodological design, data characteristics, and model integration collectively determined detection effectiveness. The findings established that ensemble and hybrid models consistently achieved superior performance across key evaluation metrics, particularly in recall and F1-score, confirming their effectiveness in addressing the challenges associated with highly imbalanced fraud datasets. Deep learning models exhibited strong capability in capturing complex and nonlinear transaction patterns, especially in large-scale and high-dimensional environments, while traditional machine learning approaches remained competitive in structured and interpretable contexts. The results further emphasized the critical role of feature engineering and data preprocessing, showing that the transformation of raw transactional data into behavioral, temporal, and relational features significantly enhanced predictive performance. Imbalance handling techniques were identified as essential components of fraud detection frameworks, contributing to substantial improvements in minority class detection and overall model sensitivity. Additionally, the study highlighted the importance of real-time processing architectures, revealing a measurable trade-off between detection speed and predictive depth, and underscoring the need for balanced system design. Graph-based and network analytics emerged as valuable approaches for detecting complex fraud networks and relational anomalies, offering insights beyond traditional transaction-level analysis. The statistical analysis confirmed that the observed performance improvements were both statistically significant and practically meaningful, supported by moderate to large effect sizes across key methodological factors. Furthermore, the study reinforced the importance of robust evaluation frameworks, including cross-validation techniques and multi-metric assessment, to ensure reliability and generalizability of results. Overall, the findings demonstrated that effective fraud detection in digital financial systems requires a multidimensional approach that integrates advanced modeling techniques, high-quality data representation, and rigorous evaluation strategies, providing a strong empirical foundation for understanding the performance and limitations of AI-driven fraud detection systems.

RECOMMENDATIONS

The findings of this study support several key recommendations for enhancing the effectiveness and reliability of AI-enabled fraud detection systems in digital financial environments. It is recommended that financial institutions and researchers prioritize the adoption of ensemble and hybrid modeling approaches, as these have demonstrated superior performance in handling complex, imbalanced, and high-dimensional fraud datasets. Integrating multiple algorithms can improve robustness, reduce variance, and enhance detection accuracy, particularly for minority fraud cases. Additionally, significant emphasis should be placed on advanced feature engineering techniques, including behavioral, transactional, and temporal feature construction, as these have been shown to substantially improve model performance and detection sensitivity. The implementation of systematic imbalance handling strategies, such as oversampling, cost-sensitive learning, and adaptive thresholding, is also essential to ensure that fraud cases are effectively identified without excessively increasing false positives. Furthermore, it is recommended that organizations invest in scalable data infrastructures and real-time processing capabilities to support timely fraud detection, while carefully balancing the trade-

offs between latency and predictive accuracy. The integration of graph-based and network analytics should be considered in environments where fraud involves complex relationships and coordinated activities, as these methods provide deeper insights into fraud networks and hidden patterns. From a methodological perspective, the use of rigorous validation techniques, particularly k-fold cross-validation, is strongly recommended to ensure model stability and generalizability across different datasets. Researchers should also adopt comprehensive evaluation frameworks that incorporate multiple performance metrics, including recall, precision, F1-score, and cost-based measures, to provide a more accurate assessment of model effectiveness. Standardization of benchmarking protocols and improved transparency in reporting methodologies are further recommended to enhance reproducibility and comparability across studies. Overall, a multidimensional approach that combines advanced modeling, high-quality data preparation, robust validation, and efficient system design is essential for strengthening fraud detection capabilities in increasingly complex digital financial systems.

LIMITATION

This study acknowledged several limitations that may have influenced the scope and interpretation of the findings. The first limitation related to the reliance on secondary data derived from previously published studies, which introduced variability in dataset characteristics, methodological approaches, and evaluation metrics. Differences in data preprocessing, feature engineering techniques, and validation methods across studies limited the extent to which results could be directly compared, despite efforts to standardize performance indicators. Additionally, the use of both publicly available and proprietary datasets introduced potential bias, as publicly accessible datasets may not fully represent the complexity, scale, and diversity of real-world financial systems. The presence of highly imbalanced datasets across most studies also posed challenges in interpreting performance metrics, particularly when accuracy was reported without sufficient emphasis on recall or F1-score. Another limitation was the lack of standardized benchmarking frameworks in the existing literature, which made it difficult to establish uniform criteria for model comparison and evaluation. The study was further constrained by the exclusion of non-English publications and studies without clearly reported quantitative metrics, which may have resulted in the omission of relevant research contributions. Variations in reporting practices, including incomplete disclosure of model parameters, data characteristics, and validation procedures, also affected reproducibility and transparency. Furthermore, the rapid evolution of artificial intelligence techniques and financial technologies means that some included studies may not fully reflect the most recent advancements in fraud detection methodologies. The study also did not incorporate real-time experimental validation or direct implementation of models, relying instead on reported outcomes from existing research, which may limit practical applicability. Finally, the aggregation of results across diverse financial domains and system architectures may have masked context-specific variations in model performance, highlighting the need for more domain-specific investigations.

REFERENCES

[1]. Abd El-Naby, A., Hemdan, E. E.-D., & El-Sayed, A. (2023). An efficient fraud detection framework with credit card imbalanced data in financial services. *Multimedia tools and applications*, *82*(3), 4139-4160.

[2]. Aditya, D., & Palash Chandra, D. (2022). Material Degradation and Durability Assessment of Pipelines and Sanitation Structures Under Aggressive Environmental Conditions. *American Journal of Interdisciplinary Studies*, *3*(02), 126-164. https://doi.org/10.63125/papn7656

[3]. Afriyie, J. K., Tawiah, K., Pels, W. A., Addai-Henne, S., Dwamena, H. A., Owiredu, E. O., Ayeh, S. A., & Eshun, J. (2023). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. *Decision Analytics Journal*, *6*, 100163.

[4]. Agrawal, B., Wiktorski, T., & Rong, C. (2017). Adaptive real-time anomaly detection in cloud infrastructures. *Concurrency and Computation: Practice and Experience*, *29*(24), e4193.

[5]. Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A., & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artificial Intelligence Review*, *56*(11), 13521-13617.

[6]. Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. *Ieee Access*, *10*, 39700-39715.

[7]. Alfaiz, N. S., & Fati, S. M. (2022). Enhanced credit card fraud detection model using machine learning. *Electronics*, *11*(4), 662.

[8]. Almazroi, A. A., & Ayub, N. (2023). Online payment fraud detection model using machine learning techniques. *Ieee Access*, *11*, 137188-137203.
[9]. Alsuwailem, A. A. S., Salem, E., & Saudagar, A. K. J. (2023). Performance of different machine learning algorithms in detecting financial fraud. *Computational Economics*, *62*(4), 1631-1667.
[10]. Amutha, A., Uthra, R. A., Roselyn, J. P., & Brunet, R. G. (2021). Anomaly detection in multivariate streaming PMU data using density estimation technique in wide area monitoring system. *Expert Systems with Applications*, *175*, 114865.
[11]. Anick, K. M. T. A., & Tasnim, K. (2022). Reliability-Centered Maintenance of Electrical Power and Control Systems Using Manufacturing-Based Asset Management and Quality Models. *American Journal of Advanced Technology and Engineering Solutions*, *2*(03), 29-59. https://doi.org/10.63125/xq6a0793
[12]. Anshari, M., Almunawar, M. N., Masri, M., & Hrdy, M. (2021). Financial technology with AI-enabled and ethical challenges. *Society*, *58*(3), 189-195.
[13]. Arya, M., & Sastry G, H. (2020). DEAL–'Deep Ensemble ALgorithm'framework for credit card fraud detection in real-time data stream with Google TensorFlow. *Smart Science*, *8*(2), 71-83.
[14]. Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. 2017 international conference on computing networking and informatics (ICCNI),
[15]. Baesens, B., Höppner, S., Ortner, I., & Verdonck, T. (2021). robROSE: A robust approach for dealing with imbalanced data in fraud detection. *Statistical Methods & Applications*, *30*(3), 841-861.
[16]. Bănărescu, A. (2015). Detecting and preventing fraud with data analytics. *Procedia economics and finance*, *32*, 1827-1836.
[17]. Bauer, F. C., Muir, D. R., & Indiveri, G. (2019). Real-time ultra-low power ECG anomaly detection using an event-driven neuromorphic processor. *IEEE transactions on biomedical circuits and systems*, *13*(6), 1575-1582.
[18]. Benchaji, I., Douzi, S., & El Ouahidi, B. (2018). Using genetic algorithm to improve classification of imbalanced datasets for credit card fraud detection. International Conference on Advanced Information Technology, Services and Systems,
[19]. Bhatt, S., & Singh, P. (2023). A comprehensive review of AI-enabled financial domain: past, present & future aspects. 2023 3rd international conference on innovative sustainable computational technologies (CISCT),
[20]. Bhattacharya, S., Momaya, K. S., & Iyer, K. C. (2020). Benchmarking enablers to achieve growth performance: a conceptual framework. *Benchmarking: An International Journal*, *27*(4), 1475-1501.
[21]. Bin Mofidul, R., Alam, M. M., Rahman, M. H., & Jang, Y. M. (2022). Real-time energy data acquisition, anomaly detection, and monitoring system: Implementation of a secured, robust, and integrated global IIoT infrastructure with edge and cloud AI. *Sensors*, *22*(22), 8980.
[22]. Burri, S. R., Kumar, A., Baliyan, A., & Kumar, T. A. (2023). Transforming Payment Processes: A Discussion of AI-Enabled Routing Optimization. 2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN),
[23]. Cheah, P. C. Y., Yang, Y., & Lee, B. G. (2023). Enhancing financial fraud detection through addressing class imbalance using hybrid SMOTE-GAN techniques. *International Journal of Financial Studies*, *11*(3), 110.
[24]. Chen, L., Gao, S., & Cao, X. (2020). Research on real-time outlier detection over big data streams. *International Journal of Computers and Applications*, *42*(1), 93-101.
[25]. Chen, P.-Y., Peng, X., & Yu, S. (2017). NeuroSim+: An integrated device-to-algorithm framework for benchmarking synaptic devices and array architectures. 2017 IEEE International electron devices meeting (IEDM),
[26]. Chen, W., Milosevic, Z., Rabhi, F. A., & Berry, A. (2023). Real-time analytics: Concepts, architectures, and ML/AI considerations. *Ieee Access*, *11*, 71634-71657.
[27]. Cross, C. (2022). Using artificial intelligence (AI) and deepfakes to deceive victims: the need to rethink current romance fraud prevention messaging. *Crime Prevention and Community Safety*, *24*(1), 30-41.
[28]. Dai, W., & Berleant, D. (2019). Benchmarking contemporary deep learning hardware and frameworks: A survey of qualitative metrics. 2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI),
[29]. Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2020). A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning: S. Dargan et al. *Archives of computational methods in engineering*, *27*(4), 1071-1092.
[30]. De Zarzà, I., De Curtò, J., & Calafate, C. T. (2023). Optimizing neural networks for imbalanced data. *Electronics*, *12*(12), 2674.
[31]. Devi, D., Biswas, S. K., & Purkayastha, B. (2020). A review on solution to class imbalance problem: Undersampling approaches. 2020 international conference on computational performance evaluation (ComPE),
[32]. Dornadula, V. N., & Geetha, S. (2019). Credit card fraud detection using machine learning algorithms. *Procedia computer science*, *165*, 631-641.
[33]. Dubey, C., Kumar, D., Singh, A. K., & Dwivedi, V. K. (2022). Confluence of Artificial Intelligence and Blockchain Powered Smart Contract in Finance System. 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS),
[34]. Elsayed, M. A., & Zulkernine, M. (2020). PredictDeep: security analytics as a service for anomaly detection and prediction. *Ieee Access*, *8*, 45184-45197.
[35]. Garg, S., Kaur, K., Kumar, N., & Rodrigues, J. J. (2019). Hybrid deep-learning-based anomaly detection scheme for suspicious flow detection in SDN: A social multimedia perspective. *IEEE Transactions on multimedia*, *21*(3), 566-578.
[36]. Gupta, A., Lohani, M., & Manchanda, M. (2021). Financial fraud detection using naive bayes algorithm in highly imbalance data set. *Journal of Discrete Mathematical Sciences and Cryptography*, *24*(5), 1559-1572.

[37]. Gupta, P., Varshney, A., Khan, M. R., Ahmed, R., Shuaib, M., & Alam, S. (2023). Unbalanced credit card fraud detection data: A machine learning-oriented comparative study of balancing techniques. *Procedia computer science*, *218*, 2575-2584.

[38]. Hashemi, S. K., Mirtaheri, S. L., & Greco, S. (2022). Fraud detection in banking data by machine learning techniques. *Ieee Access*, *11*, 3034-3043.

[39]. He, W., Wu, H., Yan, G., Akula, V., & Shen, J. (2015). A novel social media competitive analytics framework with sentiment benchmarks. *Information & Management*, *52*(7), 801-812.

[40]. Hisham, M., & Mohammad Robel, M. (2022). Data-Driven Innovation Ecosystems: Accelerating Economic Growth Through Strategic Technology Adoption. *American Journal of Data Science and Analytics*, *3*(12), 01-41. https://doi.org/10.63125/rf3w1z65

[41]. Hu, S., Zhang, X., Zhou, J., Ji, S., Yuan, J., Li, Z., Wang, Z., Chen, Q., He, Q., & Fang, L. (2021). Turbo: Fraud detection in deposit-free leasing service via real-time behavior network mining. 2021 IEEE 37th International Conference on Data Engineering (ICDE),

[42]. Huang, H., Yang, L., Wang, Y., Xu, X., & Lu, Y. (2021). Digital twin-driven online anomaly detection for an automation system based on edge intelligence. *Journal of Manufacturing Systems*, *59*, 138-150.

[43]. Iftekhar, A., & Md Tohidul, I. (2024). Quantitative Impact Assessment of Digital Payment Solutions on Small Business Revenue Panel Data Analysis From 1,200 U.S. SMES. *American Journal of Scholarly Research and Innovation*, *3*(02), 217-253. https://doi.org/10.63125/zy98jx29

[44]. Isangediok, M., & Gajamannage, K. (2022). Fraud detection using optimized machine learning tools under imbalance classes. 2022 IEEE International Conference on Big Data (Big Data),

[45]. Islam, M. D. Z., & Aditya, D. (2023). Measuring the Security Impact of Zero Trust Access Controls: A Mixed-Methods Study of Identity-Based Policies (Cisco ISE + AD) and Incident Reduction. *American Journal of Data Science and Analytics*, *4*(06), 01-42. https://doi.org/10.63125/8ycz7671

[46]. Itoo, F., Meenakshi, & Singh, S. (2021). Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. *International Journal of Information Technology*, *13*(4), 1503-1511.

[47]. Jain, V. (2017). Perspective analysis of telecommunication fraud detection using data stream analytics and neural network classification based data mining. *International Journal of Information Technology*, *9*(3), 303-310.

[48]. Jaiswal, A. K., & Akhilesh, K. (2019). Tomorrow's AI-enabled banking. In *Smart Technologies: Scope and Applications* (pp. 191-200). Springer.

[49]. Jha, B. K., Sivasankari, G., & Venugopal, K. (2020). Fraud detection and prevention by using big data analytics. 2020 Fourth international conference on computing methodologies and communication (ICCMC),

[50]. Jiang, J., Chen, J., Gu, T., Choo, K.-K. R., Liu, C., Yu, M., Huang, W., & Mohapatra, P. (2019). Anomaly detection with graph convolutional networks for insider threat and fraud detection. MILCOM 2019-2019 IEEE military communications conference (MILCOM),

[51]. Kapadiya, K., Patel, U., Gupta, R., Alshehri, M. D., Tanwar, S., Sharma, G., & Bokoro, P. N. (2022). Blockchain and AI-empowered healthcare insurance fraud detection: an analysis, architecture, and future prospects. *Ieee Access*, *10*, 79606-79627.

[52]. Karimipour, H., & Derakhshan, F. (2021). AI-enabled threat detection and security analysis for industrial IoT.

[53]. Karimov, J., Rabl, T., Katsifodimos, A., Samarev, R., Heiskanen, H., & Markl, V. (2018). Benchmarking distributed stream data processing systems. 2018 IEEE 34th international conference on data engineering (ICDE),

[54]. Krajsic, P., & Franczyk, B. (2020). Lambda architecture for anomaly detection in online process mining using autoencoders. International Conference on Computational Collective Intelligence,

[55]. Krusche, P., Trigg, L., Boutros, P. C., Mason, C. E., De La Vega, F. M., Moore, B. L., Gonzalez-Porta, M., Eberle, M. A., Tezak, Z., & Lababidi, S. (2019). Best practices for benchmarking germline small-variant calls in human genomes. *Nature biotechnology*, *37*(5), 555-560.

[56]. Kumar, K., & Bhushan, B. (2023). Augmenting cybersecurity and fraud detection using artificial intelligence advancements. 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS),

[57]. Lapolli, Â. C., Marques, J. A., & Gaspary, L. P. (2019). Offloading real-time DDoS attack detection to programmable data planes. 2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM),

[58]. Leite, R. A., Gschwandtner, T., Miksch, S., Gstrein, E., & Kuntner, J. (2018). Visual analytics for event detection: Focusing on fraud. *Visual Informatics*, *2*(4), 198-212.

[59]. Li, Z., Huang, M., Liu, G., & Jiang, C. (2021). A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection. *Expert Systems with Applications*, *175*, 114750.

[60]. Liu, L., Wu, Y., Wei, W., Cao, W., Sahin, S., & Zhang, Q. (2018). Benchmarking deep learning frameworks: Design considerations, metrics and beyond. 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS),

[61]. Lopez, M. A., Lobato, A. G. P., & Duarte, O. C. M. (2016). A performance comparison of open-source stream processing platforms. 2016 IEEE Global Communications Conference (GLOBECOM),

[62]. Lucero, B., Linsey, J., & Turner, C. J. (2016). Frameworks for organising design performance metrics. *Journal of Engineering Design*, *27*(4-6), 175-204.

[63]. Mahfuj Ahmed, R., & Md. Hasan Or, R. (2021). Fraud-Detection Algorithms for Identifying Anomalous Transactions in Retail Banking Networks. *American Journal of Data Science and Analytics*, *2*(12), 01-40. https://doi.org/10.63125/23m31748

[64]. Makki, S., Assaghir, Z., Taher, Y., Haque, R., Hacid, M.-S., & Zeineddine, H. (2019). An experimental study with imbalanced classification approaches for credit card fraud detection. *Ieee Access*, *7*, 93010-93022.
[65]. Malik, E. F., Khaw, K. W., Belaton, B., Wong, W. P., & Chew, X. (2022). Credit card fraud detection using a new hybrid machine learning architecture. *Mathematics*, *10*(9), 1480.
[66]. Mallidi, R. K., Sharma, M., Vangala, S. R., & Paladugu, Y. P. (2023). Automation using artificial intelligence and machine learning: a study on banking and healthcare. In *Recent Advances in Computing Sciences* (pp. 33-38). CRC Press.
[67]. Manoj, R., Boobesh, K., Rajan, A. P., & Saravanan, T. (2023). Digital Banking Transformation: LLMs in Customer Service and Fraud Detection. International Conference on Innovations in Bio-Inspired Computing and Applications,
[68]. Matthews, S. J., & Leger, A. S. (2017). Leveraging mapreduce and synchrophasors for real-time anomaly detection in the smart grid. *IEEE Transactions on Emerging Topics in Computing*, *7*(3), 392-403.
[69]. Md Abubakar Siddique, A., & Md. Al Amin, K. (2022). Data-Driven Ergonomic Risk Analysis Using Wearable Sensor Networks and Deep Learning for Injury Prevention in Industrial Workplaces. *American Journal of Data Science and Analytics*, *3*(06), 01-39. https://doi.org/10.63125/61w9ba54
[70]. Md, F. (2025). Predictive Analytics for Working Capital Management: Machine Learning Applications in Cash Flow and Liquidity Forecasting. *American Journal of Scholarly Research and Innovation*, *4*(01), 662–694. https://doi.org/10.63125/xrfrsz89
[71]. Md, F., & Islam, M. D. Z. (2022). Quantitative Risk Modeling of VPN Misconfigurations and Firewall Rule Drift in Hybrid Cloud Networks. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 182-216. https://doi.org/10.63125/fa4qdz07
[72]. Md, F., & Md. Mehedi, H. (2021). Machine Learning Accuracy in Healthcare Risk Prediction: Algorithms, Datasets, and Effect Sizes: A Meta-Analysis. *American Journal of Data Science and Analytics*, 2(10), 01-39. https://doi.org/10.63125/3f0mwc90
[73]. Md Khaled, H., & Md. Mosheur, R. (2023). Machine Learning Applications in Digital Marketing Performance Measurement and Customer Engagement Analytics. *Review of Applied Science and Technology*, *2*(03), 27–66. https://doi.org/10.63125/hp9ay446
[74]. Md Shahab, U. (2025). AI-Driven Distribution Planning for Essential Goods in Underserved Communities: A Mixed Methods Framework for Access Optimization. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1700–1739. https://doi.org/10.63125/chv6qf37
[75]. Md Shahab, U., & Aditya, D. (2023). Risk Mitigation and Resilience Modeling for Consumer Distribution Networks During Demand Shocks: A Quantitative Stochastic Optimization and Scenario Analysis Study. *International Journal of Scientific Interdisciplinary Research*, *4*(2), 01–30. https://doi.org/10.63125/jkevvq84
[76]. Md. Hasan Or, R., Tanjina Binte, S., & Rajib, S. (2023). Performance Analytics Frameworks for Digital Marketing and Service Enterprises: An empirical Study. *American Journal of Data Science and Analytics*, *4*(03), 01-35. https://doi.org/10.63125/aq7y1792
[77]. Md. Mainuddin, F., & Palash Chandra, D. (2022). Fabrication-Driven Structural Optimization Techniques for Cost-Efficient Steel Construction Using CNC-Based Design Workflows. *American Journal of Interdisciplinary Studies*, *3*(04), 464-499. https://doi.org/10.63125/n08g1x15
[78]. Md. Mehedi, H., & Khairum Nahar, P. (2023). A Systematic Review of Secure Health Data Information Systems for Pandemic Preparedness and Economic Continuity in the United States. *Review of Applied Science and Technology*, *2*(01), 227–258. https://doi.org/10.63125/77h2m531
[79]. Md. Shahinur, I., & Md. Sultan, M. (2022). Digital-Twin-Based Quantitative Frameworks for Modeling, Monitoring, and Optimization of Electrical Power Infrastructure. *American Journal of Interdisciplinary Studies*, *3*(04), 365-393. https://doi.org/10.63125/dvmj1y93
[80]. Md. Sultan, M., & Anick, K. M. T. A. (2023). High-Performance Computing–Assisted Modeling and Real-Time Analysis of Electrical Power Networks and Industrial Control Systems. *Review of Applied Science and Technology*, *2*(01), 185–226. https://doi.org/10.63125/727j5j39
[81]. Md. Towhidul, I., & Uddin, M. D. S. (2024). Simulation-Based Forecasting and Inventory Control Models For Consumer Goods Networks: A Quantitative Study Using Monte Carlo Simulation and Time-Series Methods. *Review of Applied Science and Technology*, *3*(04), 165–197. https://doi.org/10.63125/a3047d06
[82]. Mhlanga, D. (2020). Industry 4.0 in finance: The impact of artificial intelligence (AI) on digital financial inclusion. *International Journal of Financial Studies*, *8*(3), 45.
[83]. Moallemi, A., Burrello, A., Brunelli, D., & Benini, L. (2022). Exploring scalable, distributed real-time anomaly detection for bridge health monitoring. *IEEE Internet of Things Journal*, *9*(18), 17660-17674.
[84]. Mogaji, E., Farquhar, J. D., Van Esch, P., Durodie, C., & Perez-Vega, R. (2022). Guest editorial: Artificial intelligence in financial services marketing. *International Journal of Bank Marketing*, *40*(6), 1097-1101.
[85]. Mohammad Mushfequr, R., & Aditya, D. (2024). Quantitative Assessment of Data Protection Practices In U.S. Revenue Cycle Management. *American Journal of Advanced Technology and Engineering Solutions*, *4*(04), 107-153. https://doi.org/10.63125/fc9hfy54
[86]. Mondal, I. A., Haque, M. E., Hassan, A.-M., & Shatabda, S. (2021). Handling imbalanced data for credit card fraud detection. 2021 24th International Conference on Computer and Information Technology (ICCIT),
[87]. Mostafa, K. (2023). An Empirical Evaluation of Machine Learning Techniques for Financial Fraud Detection in Transaction-Level Data. *American Journal of Interdisciplinary Studies*, *4*(04), 210-249. https://doi.org/10.63125/60amyk26

[88]. Mostafa, K. (2025). Financial Vulnerability Mapping in Global Supply Chains: Implications for U.S. Trade Stability and Investment Risk. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1636–1667. https://doi.org/10.63125/42rd4x66

[89]. Mostafa, K., & Md Tohidul, I. (2022). A Quantitative Financial Impact Assessment of Digital Trade Platforms on Export Performance, Capital Efficiency, and Market Competitiveness. *Journal of Sustainable Development and Policy*, *1*(03), 01-26. https://doi.org/10.63125/pt5v9517

[90]. Óskarsdóttir, M., Ahmed, W., Antonio, K., Baesens, B., Dendievel, R., Donas, T., & Reynkens, T. (2022). Social network analytics for supervised fraud detection in insurance. *Risk Analysis*, *42*(8), 1872-1890.

[91]. Patil, S., Nemade, V., & Soni, P. K. (2018). Predictive modelling for credit card fraud detection using data analytics. *Procedia computer science*, *132*, 385-395.

[92]. Pourhabibi, T., Ong, K.-L., Kam, B. H., & Boo, Y. L. (2020). Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems*, *133*, 113303.

[93]. Raeiszadeh, M., Ebrahimzadeh, A., Saleem, A., Glitho, R. H., Eker, J., & Mini, R. A. (2023). Real-time anomaly detection using distributed tracing in microservice cloud applications. 2023 IEEE 12th International Conference on Cloud Networking (CloudNet),

[94]. Raghavan, P., & El Gayar, N. (2019). Fraud detection using machine learning and deep learning. 2019 international conference on computational intelligence and knowledge economy (ICCIKE),

[95]. Ratul, D., & Aditya, D. (2023). AI-Driven Change Detection Using SAR, LIDAR, And Sentinel-2 Data for Landslide Monitoring and Disaster Early Warning Systems. *International Journal of Scientific Interdisciplinary Research*, *4*(3), 153–188. https://doi.org/10.63125/4y740y95

[96]. Rawat, S. S., & Mishra, A. K. (2022). Review of methods for handling class imbalance in classification problems. International Conference on Data, Engineering and Applications,

[97]. Razaque, A., Frej, M. B. H., Bektemyssova, G., Amsaad, F., Almiani, M., Alotaibi, A., Jhanjhi, N., Amanzholova, S., & Alshammari, M. (2022). Credit card-not-present fraud detection and prevention using big data analytics algorithms. *Applied Sciences*, *13*(1), 57.

[98]. Ren, X., & Curé, O. (2017). Strider: A hybrid adaptive distributed RDF stream processing engine. International Semantic Web Conference,

[99]. Rout, N., Mishra, D., & Mallick, M. K. (2017). Handling imbalanced data: a survey. In *International proceedings on advances in soft computing, intelligent systems and applications: Asisa 2016* (pp. 431-443). Springer.

[100]. Rukaiya Khatun, M., & Md. Morshedul, I. (2022). Anticipatory Intelligence Systems: How Data Analytics Reshape Organizational Preparedness and Action Timing. *American Journal of Interdisciplinary Studies*, *3*(04), 394-428. https://doi.org/10.63125/rhwpgf86

[101]. Safaei, A. A. (2017). Real-time processing of streaming big data. *Real-Time Systems*, *53*(1), 1-44.

[102]. Sailusha, R., Gnaneswar, V., Ramesh, R., & Rao, G. R. (2020). Credit card fraud detection using machine learning. 2020 4th international conference on intelligent computing and control systems (ICICCS),

[103]. Sakib, A. I. M. (2024). Innovative Food Waste Recycling Methods For Agricultural Sustainability: A Systematic Review. *Academic Journal On Business administration, Innovation & Sustainability*, *4*(3), 104-118. https://doi.org/10.69593/ajbais.v4i3.107

[104]. Sazzadul, I., & Rebeka, S. (2024). VaR and CVaR-Based Stress Testing Using Deep Learning for Liquidity Risk Forecasting and Banking Stability Assessment. *Review of Applied Science and Technology*, *3*(03), 01–30. https://doi.org/10.63125/291phs66

[105]. Sheth, J. N., Jain, V., Roy, G., & Chakraborty, A. (2022). AI-driven banking services: the next frontier for a personalised experience in the emerging market. *International Journal of Bank Marketing*, *40*(6), 1248-1271.

[106]. Shukla, S., Hassan, M. F., Tran, D. C., Akbar, R., Paputungan, I. V., & Khan, M. K. (2023). Improving latency in Internet-of-Things and cloud computing for real-time data transmission: a systematic literature review (SLR). *Cluster Computing*, *26*(5), 2657-2680.

[107]. Singh, A., Ranjan, R. K., & Tiwari, A. (2022). Credit card fraud detection under extreme imbalanced data: a comparative study of data-level algorithms. *Journal of Experimental & Theoretical Artificial Intelligence*, *34*(4), 571-598.

[108]. Sinha, M., Chacko, E., & Makhija, P. (2022). AI based technologies for digital and banking fraud during covid-19. In *Integrating meta-heuristics and machine learning for real-world optimization problems* (pp. 443-459). Springer.

[109]. Somasundaram, A., & Reddy, S. (2019). Parallel and incremental credit card fraud detection model to handle concept drift and data imbalance. *Neural Computing and Applications*, *31*(Suppl 1), 3-14.

[110]. Spelmen, V. S., & Porkodi, R. (2018). A review on handling imbalanced data. 2018 international conference on current trends towards converging technologies (ICCTCT),

[111]. Strelcenia, E., & Prakoonwit, S. (2023). A survey on gan techniques for data augmentation to address the imbalanced data issues in credit card fraud detection. *Machine Learning and Knowledge Extraction*, *5*(1), 304-329.

[112]. Sundarkumar, G. G., & Ravi, V. (2015). A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance. *Engineering Applications of Artificial Intelligence*, *37*, 368-377.

[113]. Tahmina Akter, R., & Aditya, D. (2025). Development of Model Influence on Consumer Behavior in U.S. e-commerce and Digital Marketing. *American Journal of Interdisciplinary Studies*, *6*(3), 106-143. https://doi.org/10.63125/1brehy25

[114]. Tasnim, K., & Anick, K. M. T. A. (2024). PLC–SCADA–Integrated Electrical Automation Frameworks for Process Optimization in Water and Wastewater Treatment Facilities. *Review of Applied Science and Technology*, *3*(01), 221–262. https://doi.org/10.63125/y1145g11

[115]. Tasnim, K., & Zaheda, K. (2023). A Smart Contract Framework for Automated Settlement and Compliance in Renewable Energy and Distributed Energy Resources. *American Journal of Advanced Technology and Engineering Solutions*, *3*(01), 31-69. https://doi.org/10.63125/fvdjpn66

[116]. Taye, M. M. (2023). Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. *Computers*, *12*(5), 91.

[117]. Terzi, D. S., Terzi, R., & Sagiroglu, S. (2017). Big data analytics for network anomaly detection from netflow data. 2017 International Conference on Computer Science and Engineering (UBMK),

[118]. Thennakoon, A., Bhagyani, C., Premadasa, S., Mihiranga, S., & Kuruwitaarachchi, N. (2019). Real-time credit card fraud detection using machine learning. 2019 9th international conference on cloud computing, data science & engineering (Confluence),

[119]. Tufail, S., Riggs, H., Tariq, M., & Sarwat, A. I. (2023). Advancements and challenges in machine learning: A comprehensive review of models, libraries, applications, and algorithms. *Electronics*, *12*(8), 1789.

[120]. Van Belle, R., Baesens, B., & De Weerdt, J. (2023). CATCHM: A novel network-based credit card fraud detection method using node representation learning. *Decision Support Systems*, *164*, 113866.

[121]. Velasco-Gallego, C., & Lazakis, I. (2022). RADIS: A real-time anomaly detection intelligent system for fault diagnosis of marine machinery. *Expert Systems with Applications*, *204*, 117634.

[122]. Wang, K., Guo, Z., Mo, Y., Wang, Y., & Yuan, X. (2023). Maximizing anomaly detection performance using latent variable models in industrial systems. *IEEE Transactions on Automation Science and Engineering*, *21*(3), 4808-4816.

[123]. Wang, R., Qiu, H., Cheng, X., & Liu, X. (2023). Anomaly detection with a container-based stream processing framework for industrial internet of things. *Journal of Industrial Information Integration*, *35*, 100507.

[124]. Wu, H., Shang, Z., Peng, G., & Wolter, K. (2020). A reactive batching strategy of apache kafka for reliable stream processing in real-time. 2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE),

[125]. Yoo, Y., Shin, J., & Kyeong, S. (2023). Medicare fraud detection using graph analysis: A comparative study of machine learning and graph neural networks. *Ieee Access*, *11*, 88278-88294.

[126]. Zaheda, K., & Md Hamidur, R. (2024). GPU-Accelerated Physics-Informed Digital Twins for Real-Time State Estimation and Fault Localization in Distribution Grids. *American Journal of Scholarly Research and Innovation*, *3*(02), 179-216. https://doi.org/10.63125/msrpfb04

[127]. Zaheda, K., & Md. Tahmid Farabe, S. (2023). Robotics and Computer Vision for Automated Inspection of Substation and Treatment-Facility Electrical Infrastructure. *Review of Applied Science and Technology*, *2*(04), 194-227. https://doi.org/10.63125/tfh15j12

[128]. Zakia, A., & Khairum Nahar, P. (2022). Advanced Computing Frameworks for Real-Time SAP S/4HANA Retail Business Intelligence: Optimizing Data Processing, Latency, and System Reliability. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 217-254. https://doi.org/10.63125/xk5j7g56

[129]. Zdravković, M., Panetto, H., & Weichhart, G. (2022). AI-enabled enterprise information systems for manufacturing. *Enterprise Information Systems*, *16*(4), 668-720.

[130]. Zhang, Y.-F., Lu, H.-L., Lin, H.-F., Qiao, X.-C., & Zheng, H. (2022). The optimized anomaly detection models based on an approach of dealing with imbalanced dataset for credit card fraud detection. *Mobile Information Systems*, *2022*(1), 8027903.

[131]. Zhao, H., Wang, Y., Duan, J., Huang, C., Cao, D., Tong, Y., Xu, B., Bai, J., Tong, J., & Zhang, Q. (2020). Multivariate time-series anomaly detection via graph attention network. 2020 IEEE international conference on data mining (ICDM),

[132]. Zhou, H., Sun, G., Fu, S., Wang, L., Hu, J., & Gao, Y. (2021). Internet financial fraud detection based on a distributed big data approach with node2vec. *Ieee Access*, *9*, 43378-43386.