Rule-Based Hybrid AI–Ml Models for Fraud Detection: A Systematic Review and Thematic Analysis

Hema Verma
Research Scholar
Department of Computer Science & Applications
Maharshi Markandeshwar (Deemed to be University) ,
Mullana

Dr. Meenakshi Sharma
Associate Professor
Department of MMICT&BM
Maharshi Markandeshwar (Deemed to be University) ,
Mullana

Abstract

The rapid expansion of digital transactions across banking, insurance, e-commerce, and government platforms has intensified both the scale and sophistication of fraudulent activities. Traditional rule-based fraud detection systems, while transparent and compliant with regulatory requirements, lack adaptability to emerging and evolving fraud patterns. In contrast, standalone machine learning (ML) models provide strong predictive capabilities but often suffer from limited interpretability, regulatory resistance, and vulnerability to data drift. To overcome these limitations, rule-based hybrid Artificial Intelligence–Machine Learning (AI–ML) models have emerged as a robust approach for final fraud detection by integrating expert knowledge with data-driven intelligence. This study systematically examines and analyses existing rule-based hybrid AI–ML models, focusing on their conceptual foundations, architectural designs, integration mechanisms, and application domains. Adopting a systematic literature review methodology, the study categorizes hybrid models into sequential, parallel, rule-augmented, and explainable architectures. The analysis demonstrates that hybrid AI–ML models consistently outperform standalone approaches by enhancing detection accuracy, reducing false positives, and ensuring explain ability and regulatory compliance. The study concludes that rule-based hybrid AI–ML frameworks are essential for developing scalable, transparent, and trustworthy fraud detection systems in high-stakes digital environments.

Keywords: Fraud Detection; Hybrid AI–ML Models; Rule-Based Systems; Machine Learning; Explainable Artificial Intelligence; Financial Crime

Introduction

The rapid digitalization of financial services andcommercial transactions has fundamentally transformed the way organizations operate acrossbanking,insurance,e-commerce, telecommunications, and government sectors. While digital platforms have improved operationalefficiencyandcustomerconvenience, they have simultaneously created complex and dynamic environments that are increasingly vulnerable to fraudulent activities. Financial fraud, including credit card fraud, identity theft, insurance fraud, and transaction laundering, continues to impose significant economic losses and reputational damage on organizations worldwide. As fraudsters adopt sophisticated and adaptive strategies, the need for accurate, scalable, and trustworthy fraud detection systems has become a critical priority.

Traditionally, fraud detection has relied heavily on rule-based systems developed from expert knowledge, predefined thresholds, and regulatory guidelines. These systems offer high levels of transparency, interpretability, and auditability, making them particularly suitable for compliance-driven domains.

DOI: https://doi.org/10.5281/zenodo.18350970

However, static rule-based approaches struggle to cope with evolving fraud patterns, high-dimensional data, and the increasing volume and velocity of digital transactions. Frequent manual rule updates, limited adaptability, and high false-positive rates significantly constrain their effectiveness in modern digital ecosystems.

In response to these limitations, machine learning (ML) techniques have gained widespread adoption in fraud detection applications. ML models, including decision trees, support vector machines, ensemble methods, and deep learning architectures, are capable of identifying complex non-linear patterns and subtle anomalies within large datasets. Despite their strong predictive performance, standalone ML models often function as black boxes, raising concerns related to explainability, fairness, regulatory compliance, and accountability. Moreover, ML-based systems are susceptible to data drift, adversarial manipulation, and operational risks in high-stakes financial environments.

To address the complementary limitations of rule-based and ML-based approaches, rule-based hybrid Artificial Intelligence–Machine Learning (AI–ML) models have emerged as a promising solution for final fraud detection. These hybrid frameworks integrate domain-specific rules with data-driven learning algorithms to enhance detection accuracy while preserving transparency and interpretability. By leveraging expert knowledge alongside adaptive learning mechanisms, hybrid AI–ML models enable organizations to balance performance, compliance, and trustworthiness in real-world fraud detection systems.

Recent academic and industry research has proposed a variety of hybrid architectures, including sequential models where rules act as pre-filters or post-validators, parallel models combining rule and ML outputs, rule-augmented learning systems, and explainable hybrid frameworks designed to meet regulatory requirements. However, existing studies remain fragmented, with limited systematic synthesis of model structures, integration strategies, and application outcomes. A comprehensive analysis of these hybrid models is essential to understand their strengths, limitations, and practical relevance across different fraud detection contexts.

Against this background, the present study aims to systematically review and analyse existing rule-based hybrid AI–ML models for final fraud detection. By examining their conceptual foundations, architectural designs, integration mechanisms, and application domains, this study seeks to provide a structured understanding of hybrid fraud detection approaches. The findings contribute to both theory and practice by offering insights into model selection, system design, and future research directions for developing robust, explainable, and scalable fraud detection systems.

Review of Literature

The growing complexity of financial transactions and the rapid evolution of fraudulent techniques have driven extensive academic and industry research on fraud detection systems. Early studies primarily relied on rule-based approaches, while recent advances have emphasized machine learning (ML) and hybrid Artificial Intelligence–Machine Learning (AI–ML) models. This section critically reviews existing literature on fraud detection with a specific focus on rule-based systems, ML-based approaches, and hybrid AI–ML frameworks.

Rule-Based Fraud Detection Systems

Rule-based fraud detection systems represent one of the earliest and most widely adopted approaches in financial crime prevention. These systems operate on predefined rules derived from expert knowledge, regulatory requirements, and historical fraud patterns. Their primary strength lies in transparency, interpretability, and ease of regulatory compliance, making them suitable for high-stakesfinancial environments. Sundararamaiah et al. (2024) highlighted the continued relevance of rule-based models, particularly in scenarios where explainability and auditability are critical. However, the authors noted that static rules often fail to capture emerging fraud patterns, leading to increased false positives and reduced detection efficiency.

Despite these limitations, rule-based systems remain an essential component of modern fraud detection architectures. Wahid and Hassini (2024) demonstrated that embedding expert-defined rules within advanced detection frameworks enhances system stability and

DOI: https://doi.org/10.5281/zenodo.18350970

operational reliability, especially in invoicing and enterprise fraud contexts. Nevertheless, the inability of purely rule-based systems to adapt dynamically to evolving fraud tactics has necessitated the integration of learning-based methods.

Machine Learning-Based Fraud Detection Approaches

Machine learning techniques have gained significant prominence in fraud detection due to their ability to process large volumes of transactional data and identify complex, non-linear relationships. Supervised learning algorithms such as decision trees, support vector machines, random forests, and neural networks have been widely applied to detect fraudulent patterns. Maheshwari et al. (2023) proposed a deep neural network-based approachenhancedwithattentionmechanisms,demonstratingimproved detection accuracy for credit card fraud. Similarly, Btoush et al. (2025) employed a hybrid ensemble of ML and deep learning models, achieving superior performance in cyber fraud detection.

Systematic reviews further support the effectiveness of ML-based approaches. Yanto et al. (2024) and Yussiff et al. (2024) identified ensemble and deep learning models as consistently outperforming traditional statistical methods. However, these studies also emphasized critical challenges, including model opacity, data imbalance, concept drift, and difficulties in regulatory acceptance. Yuhertiana and Amin (2023) underscored that the lack of explainability in ML models limits their adoption in compliance-driven financial institutions.

Hybrid AI–ML Models for Fraud Detection

To address the complementary weaknesses of rule-based and ML-based systems, hybrid AI–ML models have emerged as a promising solution. These models integrate expert-defined rules with adaptive learning algorithms to enhance both accuracy and interpretability. Wahid and Hassini (2024) proposed an augmented hybrid framework that combines rule validation with ML-based anomaly detection, demonstrating improved robustness and reduced false-positive rates. Similarly, Reddy et al. (2023) highlighted that hybrid ML techniques outperform standalone models in detecting complex fraud patterns.

Hybrid architectures vary in their integration strategies. Sequential hybrid models use rule-based filters before or after ML classification, while parallel models combine outputs from both systems. Albone (2024) explored the integration of hybrid ML models with graph databases, emphasizing the importance of relational data in detecting organized fraud networks. Baisholan et al. (2025) introduced an interpretable hybrid framework designed to address data imbalance while maintaining transparency, reinforcing the role of explainable AI in fraud detection.

Explainability, Compliance, and Regulatory Considerations

Explainability has become a central concern in fraud detection research, particularly in regulated financial environments. Hybrid AI–ML models inherently support explainability by embedding rule-based reasoning within predictive systems. Baisholan et al. (2025) demonstrated that interpretable hybrid models not only improve user trust but also facilitate regulatory audits. The systematic review by HumanitiesandSocialSciences Communications (2024) further emphasized that explainable and transparent systems are essential for ethical and compliant AI deployment in financial services.

Regulatory frameworks increasingly demand accountability, fairness, and transparency in automated decision-making systems. Hybrid models, by combining symbolic rules with data-driven intelligence, offer a practical pathway to meeting these requirements while maintaining high detection performance.

Research Gap

Although existing studies provide valuable insights into rule-based, ML-based, and hybrid fraud detection models, the literature remains fragmented with respect to systematic classification and comparative analysis of hybrid architectures. Most studies focus on performance optimization or specific application domains, with limited emphasis on integration strategies, explainability, and final decision-making mechanisms. Furthermore, comprehensive reviews that synthesize architectural designs, application contexts, and regulatory implications of rule-based hybrid AI–ML models remain scarce. This gap underscores the need for a structured and

DOI: https://doi.org/10.5281/zenodo.18350970

systematic analysis of existing hybrid models for final fraud detection, which the present study aims to address.

Statement of the Problem

The rapid growth of digital transactions across financial and commercial platforms has led to a corresponding increase in the frequency, complexity, and sophistication of fraudulent activities. Financial institutions and digital service providers are under increasing pressure to deploy fraud detection systems that are not only accurate but also transparent, explainable, and compliant with evolving regulatory frameworks. Traditional rule-based fraud detection systems, although reliable and interpretable, lack adaptability to emerging fraud patterns and often result in high false-positive rates. Conversely, standalone machine learning (ML) models offer improved detection accuracy but suffer from limited interpretability, regulatory resistance, and vulnerability to data drift and operational risks. While recent research has proposed rule-based hybrid Artificial Intelligence–Machine Learning (AI–ML) models to address these challenges, existing studies remain fragmented and context-specific. There is a lack of systematic analysis that consolidates and evaluates the diverse hybrid architectures, integration strategies, and application outcomes reported in the literature. Moreover, limited attention has been given to understanding how hybrid models function in final fraud detection stages, where decision explainability, accountability, and compliance are critical.

The absence of a comprehensive and structured review of existing rule-based hybrid AI–ML models creates a knowledge gap for researchers and practitioners seeking to design robust, scalable, and trustworthy fraud detection systems. This gap necessitates a systematic examination and analysis of existing hybrid AI–ML models to identify their strengths, limitations, and practical relevance in high-stakes fraud detection environments. The present study seeks to address this problem by providing a structured synthesis of existing literature on rule-based hybrid AI–ML models for final fraud detection.

Objectives of the Study

The primary objective of this study is to systematically analyse existing rule-based hybrid Artificial Intelligence–Machine Learning (AI–ML) models for final fraud detection. The specific objectives are to:

- Review and synthesize existing literature on rule-based, machine learning, and hybrid AI–ML fraud detection approaches.
- Classify hybrid AI–ML models based on their architectural and integration strategies.
- Evaluate the effectiveness of hybrid models in terms of accuracy, false-positive reduction, explainability, and regulatory compliance.
- Identify key challenges and research gaps in the design and application of rule-based hybrid AI–ML fraud detection systems.

Hypotheses

Based on the objectives and scope of the study, the following hypotheses are proposed:

- H1: Rule-based hybrid AI–ML models demonstrate higher fraud detection accuracy compared to standalone rule-based and standalone machine learning models.
- H2: The integration of rule-based systems with machine learning techniques significantly reduces false-positive rates in fraud detection.
- H3: Rule-based hybrid AI–ML models provide greater explainability and transparency than standalone machine learning models, thereby improving regulatory compliance.
- H4: Hybrid AI–ML architectures incorporating explainable components are more effective for final fraud detection in high-stakes financial environments than non-explainable models.

Conceptual Framework

The conceptual framework of the present study is designed to illustrate the structural relationships between fraud detection approaches, hybrid AI–ML architectures, and final fraud detection outcomes. The framework integrates rule-based systems and machine learning models as complementary components within a unified fraud detection architecture, emphasizing their combined influence on detection performance, explainability, and regulatory compliance.

DOI: https://doi.org/10.5281/zenodo.18350970

At the foundational level, the framework comprises two primary input components: rule-based systems and machine learning models. Rule-based systems encapsulate domain expertise, regulatory rules, and predefined thresholds that ensure transparency, auditability, and compliance. Machine learning models, on the other hand, contribute adaptive learning capabilities, pattern recognition, and predictive intelligence derived from transactional data. Individually, each approach exhibits inherent limitations; however, their integration forms the basis of hybrid AI–ML fraud detection models.

The framework categorizes rule-based hybrid AI–ML architectures into four distinct integration strategies: sequential, parallel, rule-augmented, and explainable hybrid models. In sequential architectures, rule-based filters operate either before or after ML classification to validate or refine detection outcomes. Parallel architectures combine the outputs of rule-based and ML systems through weighted scoring or ensemble mechanisms. Rule-augmented models embed expert rules directly into the learning process, while explainable hybrid models incorporate interpretability mechanisms to support transparency and accountability.

These hybrid architectures act as the core processing layer, influencing key fraud detection performance outcomes, namely detection accuracy, false-positive reduction, system robustness, explainability, and regulatory compliance. The effectiveness of final fraud detection decisions is determined by how effectively hybrid models balance predictive performance with transparency and compliance requirements.

Hypothesis Mapping within the Conceptual Framework

The proposed hypotheses are directly aligned with the relationships depicted in the conceptual framework. H1 posits that the integration of rule-based systems and ML models within hybrid architectures leads to higher fraud detection accuracy compared to standalone approaches. H2 is mapped to the framework's outcome layer, suggesting that hybrid integration reduces false-positive rates through rule validation and adaptive learning. H3 aligns with the explainability dimension, proposing that hybrid models enhance transparencyandregulatory compliance relative to black-box ML models. Finally, H4 corresponds to the explainable hybrid architecture, asserting that explainability-oriented hybrid models are more effective for final fraud detection in high-stakes financial environments.

Overall, the conceptual framework provides a structured representation of how rule-based hybrid AI–ML models contribute to improved final fraud detection outcomes. It serves as a theoretical foundation for systematically analysing existing hybrid models and supports the evaluation of their effectiveness, limitations, and practical relevance in fraud detection systems.

Research Methodology

This study adopts a systematic literature review (SLR) methodology to examine and analyse existing rule-based hybrid Artificial Intelligence–Machine Learning (AI–ML) models for final fraud detection. A systematic approach ensures transparency, reproducibility, and methodological rigor in synthesizing prior research, making it suitable for conceptual and review-based studies published in Scopus-indexed journals.

Research Design

The study follows a qualitative and analytical research design based on secondary data sources. The systematic review process was conducted in accordance with established review guidelines, including the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework. The review process comprised four key stages: identification, screening, eligibility assessment, and inclusion of relevant studies.

Data Sources

Relevant literature was collected from reputed academic databases and digital libraries, including Scopus, Web of Science, IEEE Xplore, SpringerLink, ScienceDirect, and MDPI. These databases were selected due to their extensive coverage of peer-reviewed journals in artificial intelligence, machine learning, and financial fraud detection.

Search Strategy

A structured keyword-based search strategy was employed to identify relevant studies. The

DOI: https://doi.org/10.5281/zenodo.18350970

search strings included combinations of the following keywords: fraud detection, rule-based systems, machine learning, hybrid AI–ML models, explainable AI, and financial crime. Boolean operators (AND, OR) were used to refine the search results. The search was limited to articles published in English between 2018 and 2025 to ensure relevance to recent technological developments.

Inclusion and Exclusion Criteria

To ensure the quality and relevance of the reviewed studies, explicit inclusion and exclusion criteria were applied. Studies were included if they: (i) focused on fraud detection using rule-based, machine learning, or hybrid AI–ML approaches; (ii) were published in peer-reviewed journals or conference proceedings; and (iii) provided methodological or architectural insights into fraud detection models. Studies were excluded if they: (i) lacked a clear methodological description; (ii) were non-peer-reviewed sources such as blogs or white papers; or (iii) focused on non-financial fraud domains without relevance to transactional fraud detection.

Study Selection Process

The initial database search yielded a broad set of articles. Duplicate records were removed, and titles and abstracts were screened for relevance. Full-text screening was then conducted to assess methodological quality and alignment with the study objectives. Only studies meeting all inclusion criteria were retained for final analysis.

Data Extraction and Analysis

Data were systematically extracted from the selected studies using a predefined review protocol. Extracted information included publication details, fraud detection domain, model type, hybrid integration strategy, performance metrics, explainability features, and reported limitations. A thematic synthesis approach was employed to classify and analyse hybrid AI–ML models into sequential, parallel, rule-augmented, and explainable architectures.

Validity and Reliability

To enhance the validity and reliability of the review, multiple databases were consulted, and clearly defined selection criteria were applied. Consistency in data extraction and classification was maintained through iterative cross-checking of extracted information. The systematic approach adopted in this study minimizes selection bias and enhances the credibility of the findings.

Data Analysis And Results

The data analysis was conducted using a thematic synthesis approach to systematically evaluate and integrate findings from the selected studies on rule-based hybrid Artificial Intelligence–Machine Learning (AI–ML) models for fraud detection. Thematic synthesis enabled the identification of recurring patterns, architectural strategies, performance outcomes, and implementation challenges across diverse application contexts. Based on iterative coding and comparison of the reviewed studies, four major analytical themes emerged.

Theme 1: Architectural Integration of Rule-Based and Machine Learning Models

A dominant theme across the reviewed literature is the architectural integration of rule-based systems with machine learning models. Studies consistently classified hybrid AI–ML frameworks into sequential, parallel, rule-augmented, and explainable hybrid architectures.Sequentialarchitectures commonly employ rule-based filters as pre-processing or post-processing layers to validate ML predictions, thereby reducing false positives and improving decision reliability. Parallel architectures combine rule-based scores and ML outputs through ensemble or weighted decision mechanisms to enhance detection robustness. Rule-augmented models embed expert rules directly into the learning process, while explainable hybrid models incorporate interpretability techniques to support transparency.

The analysis indicates that no single architecture is universally optimal; rather, the effectiveness of a hybrid model depends on the fraud domain, data characteristics, and regulatory requirements. However, studies consistently report improved performance when rule-based logic is systematically integrated with adaptive learning models.

Theme 2: Performance Enhancement in Fraud Detection

DOI: https://doi.org/10.5281/zenodo.18350970

Performance improvement emerged as a central outcome of hybrid AI–ML integration. The majority of reviewed studies reported higher detection accuracy and lower false-positive rates for hybrid models compared to standalone rule-based or ML-based approaches. Rule-based validation mechanisms were found to eliminate spurious ML predictions, while ML components enhanced the system's ability to detect complex and evolving fraud patterns.

Thematic evidence supports Hypotheses H1 and H2, demonstrating that hybrid models outperform traditional approaches in final fraud detection. Studies focusing on credit card fraud, invoicing fraud, and e-commerce fraud consistently highlighted the superior predictive capability and operational efficiency of hybrid AI–ML frameworks.

Theme 3: Explainability and Regulatory Compliance

Explainability and regulatory compliance constitute a critical theme, particularly in high-stakes financial environments. The reviewed literature emphasizes that hybrid AI–ML models inherently offer better interpretability than black-box ML models due to the presence of explicit rule-based reasoning. Explainable hybrid architectures were shown to facilitate auditability, accountability, and regulatory approval.

Several studies reported that financial institutions prefer hybrid models for final decision-making stages, as they provide traceable explanations for fraud alerts. This theme provides strong support for Hypotheses H3 and H4, highlighting the importance of explainable hybrid frameworks in meeting ethical and regulatory requirements.

Theme 4: Challenges and Limitations of Hybrid AI–ML Models

Despite their advantages, the reviewed studies also identified several challenges associated with hybrid AI–ML fraud detection systems. Key limitations include the complexity of model integration, increased system maintenance costs, and the need for continuous rule updates to address evolving fraud tactics. Additionally, data imbalance, concept drift, and scalability issues remain persistent challenges.

The thematic synthesis reveals that while hybrid models improve detection outcomes, their effectiveness depends on careful system design, regular monitoring, and alignment with organizational and regulatory contexts. These findings underscore the need for standardized evaluation frameworks and adaptive hybrid architectures.

Overall, the thematic synthesis demonstrates that rule-based hybrid AI–ML models offer a balanced approach to final fraud detection by combining predictive accuracy with explainability and compliance. The results confirm that hybrid architectures are more effective than standalone approaches in detecting complex fraud patterns while maintaining transparency. The findings also highlight existing research gaps related to model standardization, scalability, and automated rule learning, which present opportunities for future research.

Discussion

This section discusses the findings of the systematic review in relation to the proposed hypotheses. By synthesizing evidence from existing studies, the discussion evaluates the effectiveness, interpretability, and practical relevance of rule-based hybrid Artificial Intelligence–Machine Learning (AI–ML) models for final fraud detection.

H1: Rule-based hybrid AI–ML models demonstrate higher fraud detection accuracy compared to standalone rule-based and standalone machine learning models.

The thematic synthesis provides strong support for H1, as the majority of reviewed studies reported superior detection accuracy for hybrid AI–ML models. Hybrid frameworks leverage the complementary strengths of rule-based systems and ML algorithms, enabling them to capture both explicit fraud patterns and complex, data-driven anomalies. Sequential and parallel hybrid architectures were particularly effective in improving overall predictive performance by integrating expert-defined rules with adaptive learning mechanisms. These findings align with prior research indicating that hybridization enhances robustness and reduces the limitations associated with isolated detection approaches.

DOI: https://doi.org/10.5281/zenodo.18350970

H2: The integration of rule-based systems with machine learning techniques significantly reduces false-positive rates in fraud detection.

Evidence from the reviewed literature supports H2,ashybrid models consistently demonstrated lower false-positive rates than standalone ML models. Rule-based validation layers were found to play a critical role in filtering erroneous predictions generated by ML algorithms, particularly in imbalanced datasets common to fraud detection. By incorporating domain expertise and regulatory constraints, hybrid models improved decision reliability and operational efficiency. These results underscore the practical importance of hybrid AI–ML systems in minimizing unnecessary fraud alerts and associated investigation costs.

H3: Rule-based hybrid AI–ML models provide greater explainability and transparency than standalone machine learning models, thereby improving regulatory compliance.

The findings strongly support H3, highlighting explainability as a key advantage of hybrid AI–ML models. The presence of rule-based components enables transparent reasoning and traceable decision paths, addressing concerns related to black-box ML models. Several studies emphasized that explainable hybrid systems facilitate regulatory audits, enhance user trust, and support ethical decision-making. This aligns with regulatory expectations in financial sectors, where transparency and accountability are essential for automated decision systems.

H4: Hybrid AI–ML architectures incorporating explainable components are more effective for final fraud detection in high-stakes financial environments than non-explainable models.

The results of the thematic analysis provide substantial support for H4. Explainable hybrid architectures were frequently preferred for final fraud detection stages, where decisions have significant financial and legal implications. These models balance predictive performance with interpretability, enabling organizations to justify decisions and comply with regulatory standards. The discussion reveals that explainability-oriented hybrid models are particularly effective in high-stakes environments such as banking and insurance, where trust and compliance are paramount.

Collectively, the hypothesis-wise discussion confirms that rule-based hybrid AI–ML models offer a comprehensive and effective solution for final fraud detection. By integrating expert knowledge with adaptive learning, hybrid models address the limitations of traditional approaches while enhancing accuracy, transparency, and compliance. However, the discussion also highlights ongoing challenges related to system complexity, maintenance, and scalability. Addressing these challenges requires continued research into adaptive hybrid architectures, automated rule learning, and standardized evaluation frameworks.

Conclusion

This study systematically examined and analysed existing rule-based hybrid Artificial Intelligence–Machine Learning (AI–ML) models for final fraud detection through a comprehensive review of the literature. The findings demonstrate that hybrid AI–ML frameworks effectively address the limitations of standalone rule-based and machine learning approaches by combining predictive accuracy with transparency, explainability, and regulatory compliance. By integrating expert-defined rules with adaptive learning mechanisms, hybrid models enhance fraud detection performance while maintaining trust and accountability in high-stakes financial environments.

The thematic synthesis and hypothesis-wise discussion confirm that rule-based hybrid AI–ML models consistently outperform traditional approaches in terms of detection accuracy and false-positive reduction. Moreover, the incorporation of explainable components enables hybrid systems to meet regulatory and ethical requirements, making them particularly suitable for final decision-making stages in fraud detection workflows. The analysis further highlights that sequential, parallel, rule-augmented, and explainable hybrid architectures offer flexible design choices depending on application context and organizational constraints.

Despite their advantages, the study identifies ongoing challenges related to system complexity, scalability, rule maintenance, and adaptability to evolving fraud patterns. These

DOI: https://doi.org/10.5281/zenodo.18350970

challenges underscore the need for standardizedevaluationframeworks, automated rule learning mechanisms, and continuous model monitoring. Overall, the study concludes that rule-based hybrid AI–ML models represent a robust and sustainable approach for final fraud detection, offering valuable insights for researchers, practitioners, and policymakers seeking to develop transparent, scalable, and trustworthy fraud detection systems.

Managerial Implications

The findings of this study offer several important managerial implications for organizations involved in fraud risk management, particularly in banking, insurance, e-commerce, and digital payment platforms. First, managers should recognize that rule-based hybrid AI–ML models provide a balanced approach to fraud detection by combining predictive accuracy with transparency. Adopting hybrid architectures can help organizations reduce false positives, improve operational efficiency, and minimize investigation costs while maintaining compliance with regulatory requirements.

Second, the study highlights the importance of explainability in fraud detection systems. Managers responsible for risk, compliance, and technology functions should prioritize hybrid models that incorporate explainable components to ensure accountability and facilitate regulatory audits. Transparent decision-making mechanisms enhance stakeholder trust and support ethical AI adoption in high-stakes environments.

Third, the classification of hybrid architectures presented in this study can guide managerial decision-making in system design and implementation. Sequential and parallel hybrid models may be suitable for organizations seeking incremental upgrades to existing rule-based systems, while rule-augmented and explainable hybrid frameworks may be more appropriate for advanced, data-driven fraud detection strategies.

Finally, managers should invest in continuous monitoring, rule optimization, and skill development to address challenges related to system complexity and evolving fraud patterns. Cross-functional collaboration between domain experts, data scientists, and compliance teams is essential for maximizing the effectiveness of hybrid AI–ML fraud detection systems and ensuring their long-term sustainability.

Policy Implications

The findings of this study have significant policy implications for regulators, financial authorities, and policymakers overseeing the deployment of artificial intelligence in fraud detection systems. The demonstrated effectiveness of rule-based hybrid AI–ML models suggests that regulatory frameworks should encourage the adoption of hybrid approachesthat balance predictive performance with transparency and accountability. Policymakers can promote responsible AI adoption by recognizing hybrid models as compliant alternatives to black-box machine learning systems in high-stakes financial applications.

Regulatory guidelines should emphasize the importance of explainability and auditability in automated fraud detection systems. By supporting the integration of rule-based logic within AI-driven frameworks, policymakers can ensure that automated decisions remain traceable, fair, and aligned with ethical standards. This is particularly relevant in the context of data protection laws, consumer rights regulations, and financial compliance requirements.

Furthermore, policymakers should facilitate standardization in the evaluation and reporting of AI-based fraud detection systems. Establishing common benchmarks and disclosure requirements for hybrid AI–ML models can enhance transparency, comparability, and trust across financial institutions. Support for research and innovation in explainable and adaptive hybrid architectures can also strengthen national and international efforts to combat financial crime in increasingly digital economies.

Future Research Directions

Future studies should empirically validate rule-based hybrid AI–ML models using real-world transactional datasets across multiple fraud domains. Research on automated rule-learning, adaptive hybrid architectures, and concept drift management is needed to enhance scalability and responsiveness. Further work should also focus on advanced

DOI: https://doi.org/10.5281/zenodo.18350970

explainability techniques and privacy-preserving hybrid frameworks to support ethical, transparent, and compliant fraud detection systems.

References

Wahid, D. F., & Hassini, E. (2024). An augmented AI-based hybrid fraud detection framework for invoicing platforms. *Applied Intelligence*, *54*(2), 1297–1310. https://doi.org/10.1007/s10489-023-05223-x

Btoush, E., Zhou, X., Gururajan, R., Chan, K. C., & Alsodi, O. (2025). Achieving excellence in cyber fraud detection: A hybrid ML + DL ensemble approach for credit cards. *Applied Sciences*, *15*(3), 1081. https://doi.org/10.3390/app15031081

Yanto, Y., Lisah, L., & Tandra, R. (2024). The best machine learning model for fraud detection in the banking sector: A systematic literature review. *ECo-Buss*, *7*(2), 1361–1384. https://doi.org/10.32877/eb.v7i2.1474

Maheshwari, V. C., Osman, N. A., & Aziz, N. (2023). A hybrid approach adopted for credit card fraud detection based on deep neural networks and attention mechanism. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *32*(1), 315–331. https://doi.org/10.37934/araset.32.1.315331

Albone, A. (2024). Optimization of fraud detection model with hybrid machine learning and graph database. *Engineering, Mathematics and Computer Science Journal*, *6*(1), 13–17. https://doi.org/10.21512/emacsjournal.v6i1.10744

Reddy, B. R., Rajesh, N., Anand, K. V., & Srikanth, G. (2023). Fraud transaction detection approach using machine learning hybrid techniques. *International Journal of Scientific Research in Science and Technology*, *10*(2), 90–96. https://doi.org/10.32628/IJSRST52310213

Yussiff, A.-S., Prikutse, L. F., Asuah, G., Yussiff, A.-L., Dortey Tetteh, E., & Ibrahim, N. (2024). The best machine learning model for fraud detection on e-platforms: A systematic literature review. *Computer Science and Information Technologies*, *5*(2), 195–204. https://doi.org/10.11591/csit.v5i2.pp195-204

Yuhertiana, I., & Amin, A. H. (2023). Artificial intelligence-driven approaches for financial fraud detection: A systematic literature review. *KnE Social Sciences*, *9*(20), Article 16551. https://doi.org/10.18502/kss.v9i20.16551

Baisholan, N., Dietz, J. E., Gnatyuk, S., Turdalyuly, M., Matson, E. T., & Baisholanova, K. (2025). FraudX AI: An interpretable machine learning framework for credit card fraud detection on imbalanced datasets. *Computers*, *14*(4), 120. https://doi.org/10.3390/computers14040120

Financial fraud detection based on machine learning: A systematic literature review. (2024). *Humanities and Social Sciences Communications*. https://doi.org/10.1057/s41599-024-03606-0

Sundararamaiah, M., Nagarajan, S. K. S., Mudunuru, K. R., & Remala, R. (2024). Unifying AI and rule-based models for financial fraud detection. *International Journal of Computer Trends and Technology*, *72*(12), 61–68. https://doi.org/10.14445/22312803/IJCTT-V72I12P107

DOI: https://doi.org/10.5281/zenodo.18350970