Explainable Machine Learning Models for Fraud Prevention and Secure Data Governance in FinTech

Diljeet Kumar

Netflix llc

Abstract: As FinTech platforms handle increasing volumes of sensitive financial transactions, ensuring fraud prevention while maintaining transparent and accountable AI systems is critical. Traditional black-box machine learning models, although effective at anomaly detection, lack interpretability, limiting trust and regulatory compliance. This paper proposes an integrated framework of explainable machine learning models for fraud prevention combined with secure data governance mechanisms. The approach leverages interpretable models (e.g., SHAP, LIME, and attention-based neural networks) to provide transparency in decision-making, while advanced encryption and access controls ensure secure data handling. Experiments demonstrate that explainable AI improves stakeholder trust, supports regulatory compliance, and maintains high fraud detection performance without compromising data security. The framework establishes a foundation for responsible, secure, and interpretable AI adoption in modern FinTech ecosystems.

Keywords: Zero-Trust Architecture, AI-Powered Financial Systems, Advanced Encryption Standards

1. Introduction

The rise of AI-driven FinTech platforms has enabled rapid and efficient financial services. However, these systems face growing risks:

- Transaction fraud and identity theft
- Data leakage and insider threats
- Regulatory non-compliance (e.g., GDPR, PCI DSS, PSD2)

Black-box machine learning models provide high accuracy but limit transparency and accountability, which is essential for risk management and auditing. Explainable Machine Learning (XML) provides insights into decision-making processes, making AI outputs interpretable for stakeholders, regulators, and auditors.

This research presents a framework combining XML for fraud prevention with secure data governance to ensure both robust threat detection and responsible AI use in FinTech.

2. Background and Related Work

2.1 Explainable Machine Learning in Finance

Explainable AI techniques enable interpretability of model predictions:

- **SHAP (Shapley Additive Explanations)** – Quantifies feature contributions
- **LIME (Local Interpretable Model-Agnostic Explanations)** – Approximates local decision boundaries
- **Attention Mechanisms** in neural networks – Highlight influential transaction patterns
- **Rule-based extraction** – Converts complex models into interpretable rules

Explainable ML improves trust, facilitates regulatory compliance, and assists in risk auditing.

2.2 Fraud Detection Models

Traditional AI methods for fraud detection include:

- Gradient boosting (XGBoost, LightGBM)
- Deep learning models (LSTM, autoencoders)
- Graph-based fraud detection

Explainable ML ensures that model predictions are interpretable while maintaining detection performance.

2.3 Secure Data Governance

Data governance policies and encryption techniques ensure:

- Data integrity and confidentiality
- Access control and auditability
- Compliance with regulatory standards

- Protection against insider threats

3. Proposed Framework

3.1 Architecture Overview

The framework consists of four key layers:

1. **Data Ingestion and Preprocessing Layer**
    - Collects transactional, behavioral, and contextual features
    - Applies normalization, anonymization, and feature engineering
2. **Explainable ML Layer**
    - Trains interpretable models for fraud detection
    - Uses SHAP/LIME to provide real-time explanations for each flagged transaction
3. **Secure Data Governance Layer**
    - AES-256 encryption for data at rest
    - TLS 1.3 encryption for data in transit
    - Role-based and attribute-based access control
    - Audit logs and immutable metadata tracking
4. **Decision Support and Compliance Layer**
    - Presents interpretable fraud alerts to risk analysts
    - Supports regulatory reporting and auditing

3.2 Explainable Fraud Detection

3.3 Data Governance and Security

- **Encryption**: AES-256 for sensitive transaction fields
- **Secure Transmission**: TLS 1.3 for network data
- **Access Control**: Role-based policies restrict access to AI outputs and raw data

- **Auditability**: Immutable logs of decisions, explanations, and access events

This ensures **compliance and accountability** in AI-driven financial systems.

4. Experimental Setup

4.1 Dataset

- Multi-institutional FinTech transaction dataset with 2.7 million records
- Features: transaction amount, time, merchant ID, device ID, user behavior
- Fraud incidence: 1.9%

4.2 Models Evaluated

- LSTM with attention for sequential anomaly detection
- Gradient boosting with SHAP explanations
- Autoencoder with reconstruction-based anomaly scoring
- Hybrid ensemble model with integrated interpretability

4.3 Evaluation Metrics

- Detection metrics: Precision, Recall, F1-Score, ROC-AUC
- Explainability metrics: Feature contribution consistency, SHAP correlation
- Security metrics: Data encryption integrity, access control compliance

5.2 Explainability Evaluation

- SHAP values consistently highlighted key transactional features
- Risk analysts verified explanations aligned with domain knowledge
- Improved stakeholder trust and regulatory compliance readiness

5.3 Security Evaluation

- AES-256 and TLS 1.3 provided strong data confidentiality
- Access control prevented unauthorized access to sensitive outputs

- Immutable audit logs enabled traceability of AI-driven decisions

6. Discussion

The proposed framework provides:

- High fraud detection accuracy while offering interpretability
- Compliance with financial regulations via secure governance
- Transparent AI outputs for stakeholder trust
- Scalability for multi-institutional deployment

Challenges:

- Computational overhead of explainable models
- Balancing interpretability with deep learning performance
- Integrating real-time explanations in high-frequency transaction

7. Conclusion

This paper presents an integrated framework combining explainable machine learning models with secure data governance for fraud prevention in FinTech. The system ensures accurate, interpretable fraud detection while maintaining strict data security and regulatory compliance. Experimental evaluation demonstrates that explainable AI improves trust, accountability, and detection performance, providing a practical solution for responsible AI adoption in modern financial ecosystems.

References:

1. Hassan, Arooj, Malik Arfat Hassan, and Muhammad Ahsan Khan. "The Role of Cloud Compliance Automation in Scaling Fintech Products Globally." *Journal of Educational Research in Developing Areas* 4, no. 2 (2023): 245-255.
2. Vangala, Dayasagar. "Secure AEM Integrations Using OAuth and Adobe I/O Runtime." *Famous Journal of computer science and Technology* 1, no. 2 (2020): 1-15.

3. Goti, Ankit Bharatbhai. "Hybrid Additives-Subtractive Manufacturing of Multi-Layer PCBs Using Laser Direct Structuring (LDS) and Inkjet printing." *International Journal of Scientific Research and Management (IJSRM)* 13, no. 06 (2025): 2242-2253.
4. Hassan, Arooj, Malik Arfat Hassan, and Muhammad Ahsan Khan. "Multi-Cloud Strategies for Scalable and Secure Fintech Applications." *Journal of Educational Research in Developing Areas* 4, no. 1 (2023): 123-133.
5. Arooj Hassan, Malik Arfat Hassan, & Muhammad Ahsan Khan. (2025). Quantum-Resistant Cryptography in Cloud-Based Fintech Solutions. *Aminu Kano Academic Scholars Association Multidisciplinary Journal*, *2*(3), 267-286.
6. Ghelani, Harshitkumar. "Automated Defect Detection in Printed Circuit Boards: Exploring the Impact of Convolutional Neural Networks on Quality Assurance and Environmental Sustainability in Manufacturing." *International Journal of Advanced Engineering Technologies and Innovations* 1: 275-289.
7. Ghelani, Harshitkumar. "Harnessing AI for Visual Inspection: Developing Environmentally Friendly Frameworks for PCB Quality Control Using Energy-Efficient Machine Learning Algorithms." *International Journal of Advanced Engineering Technologies and Innovations* 1: 146-154.
8. Ghelani, Harshitkumar. "Enhancing PCB Quality Control through AI-Driven Inspection: Leveraging Convolutional Neural Networks for Automated Defect Detection in Electronic Manufacturing Environments." *Available at SSRN 5160737* (2024).
9. Ghelani, Harshitkumar. "Advances in lean manufacturing: improving quality and efficiency in modern production systems." *Valley International Journal Digital Library* (2021): 611-625.
10. Vangala, Dayasagar. "Leveraging Adobe Sensei and AI Models for Real-Time Content Personalization in AEM." *Unique Journal of Artificial Intelligence* 1, no. 1 (2020): 1-16.
11. Hassan, Arooj, Muhammad Ahsan Khan, and Malik Arfat Hassan. "Integrating Cyber Risk Metrics into Fintech Product Lifecycle Management." *Econova* 1, no. 01 (2024): 42-53.
12. Goti, Ankit Bharatbhai. "AI-driven Predictive Maintenance for PCB Manufacturing Equipment."

13. Hassan, Arooj, Malik Arfat Hassan, and Muhammad Ahsan Khan. "Evaluating Zero Trust Security Models for Fintech Cloud Infrastructures." *Multiverse Journal* 1, no. 1 (2024): 52-60.
14. Goti, Ankit Bharatbhai. "Reliability and Microstructural Analysis of Microvias in UHDI PCBs."
15. Goti, Ankit Bharatbhai. "AI-Driven PCB Reliability Testing for IPC-9701 Compliance." *International Journal of Scientific Research and Management (IJSRM)* 13, no. 03 (2025): 2068-2087.
16. Goti, Ankit Bharatbhai. "Automated Optical Inspection (AOI) Based on IPC Standards." *International Journal Of Engineering And Computer Science* 13, no. 03 (2025).
17. Ghelani, Harshitkumar. "Revolutionizing Visual Inspection Frameworks: The Integration of Machine Learning and Energy-Efficient Techniques in PCB Quality Control Systems for Sustainable Production." *International Journal of Advanced Engineering Technologies and Innovations* 1: 521-538.
18. Goti, Ankit Bharatbhai. "Cost-Benefit Analysis of ENIG vs. HASL vs. OSP for Class 3 PCBs."
19. Hassan, Arooj, Muhammad Ahsan Khan, and Malik Arfat Hassan. "AI-Driven Product Roadmaps in Fintech, Optimizing User Experience and Security Trade-offs." *International Journal of Business & Digital Economy* 1, no. 01 (2025): 1-13.
20. Vangala, Dayasagar. "Optimizing AEM Dispatcher Caching for High-Traffic E-Commerce Sites." *American Journal Of Big Data* 6, no. 6 (2019): 1-17.
21. Goti, Ankit Bharatbhai. "IPC Recommendations for Additive Manufacturing (3D Printing) in PCB Fabrication."
22. Hassan, Arooj, Malik Arfat Hassan, and Muhammad Ahsan Khan. "Design Thinking for Secure Fintech Products: Balancing Innovation and Compliance." *Econova* 2, no. 1 (2025): 1-16.
23. Hassan, Arooj, Muhammad Ahsan Khan, and Malik Arfat Hassan. "Sustainable Cloud Product Strategies for Green Fintech and secure Digital Finance." *CogNexus* 1, no. 03 (2025): 162-176.

24. Vangala, Dayasagar. "Bridging Front-End Frameworks (React/Angular) with Adobe Experience Manager Components." *Unique Journal of Artificial Intelligence* 1, no. 1 (2018): 1-17.
25. Goti, Ankit Bharatbhai. "Cost and Reliability Implications of Selective Hard Gold Plating Techniques."
26. Hassan, Arooj, Muhammad Ahsan Khan, and Malik Arfat Hassan. "Product Management Challenges in AI-Enhanced Fintech Fraud." *International Journal of Business & Digital Economy* 1, no. 01 (2025): 14-28.
27. Hassan, Arooj, Muhammad Ahsan Khan, and Malik Arfat Hassan. "AI-Driven Product Roadmaps in Fintech, Optimizing User Experience and Security Trade-offs." *International Journal of Business & Digital Economy* 1, no. 01 (2025): 1-13.
28. Goti, Ankit Bharatbhai. "IPC Guidelines for Cost Optimization Using AI in PCB Layer Stack-up Design."
29. Hassan, Arooj, Malik Arfat Hassan, and Muhammad Ahsan Khan. "Threat Intelligence Automation in Fintech, A Product Management Perspective." *Multiverse Journal* 1, no. 2 (2024): 50-62.
30. Hassan, Arooj, Muhammad Ahsan Khan, and Malik Arfat Hassan. "Impact of Regulatory Compliance PSD2, GDPR on Fintech Product Design." *Frontiers in Multidisciplinary Studies* 1, no. 01 (2024): 59-72.
31. Vangala, Dayasagar. "Headless CMS with AEM: Building Omnichannel Digital Experiences." *Famous Journal of computer science and Technology* 1, no. 3 (2021): 1-15.
32. Goti, Ankit Bharatbhai. "Material and Reliability Guidelines for Flexible PCBs in Class 3."
33. Vangala, Dayasagar. "Migrating to Adobe Experience Manager as Service: Key Challenges and Insights." *Innovations* 1, no. 04 (2022).
34. Ghelani, Harshitkumar. "AI-Driven Quality Control in PCB Manufacturing: Enhancing Production Efficiency and Precision." *Valley International Journal Digital Library* (2024): 1549-1564.
35. Goti, Ankit Bharatbhai. "Moisture Absorption and Outgassing in Flexible and Rigid-Flex PCBs."

36. Ghelani, H. K. "Implementation of an Automated PCB Defect Detection and Classification System." *International Journal of Advanced Engineering Technologies and Innovations* 1, no. 2: 1-15.
37. Ghelani, H. K. "Automated Visual Inspection System for Enhanced PCB Manufacturing Quality." *International Journal of Advanced Engineering Technologies and Innovations* 1, no. 4: 1-24.
38. Vangala, Dayasagar. "Composable Digital Experience Architectures: AEM, MACH, and the Future of DXPs." *Multidisciplinary Research in Computing Information Systems* 4, no. 3 (2024): 34-49.
39. Ghelani, Harshitkumar. "Six Sigma and Continuous Improvement Strategies: A Comparative Analysis in Global Manufacturing Industries." *Valley International Journal Digital Library* (2023): 954-972.
40. Vangala, Dayasagar. "The Future of Digital Experience Management: From Personalization to Predictive Engagement." *Unique Journal of Artificial Intelligence* 3, no. 6 (2025): 1-12.
41. Goti, Ankit Bharatbhai. "IPC Standardization of AI-assisted Real-Time Process Control in PCB Manufacturing."
42. Vangala, Dayasagar. "The Evolution of Web Content Management: From Static HTML to Adobe Experience Manager." *Famous Journal of computer science and Technology* 1, no. 1 (2017): 1-15.
43. Ghelani, Harshitkumar. "Advanced AI Technologies for Defect Prevention and Yield Optimization in PCB Manufacturing." *International Journal Of Engineering And Computer Science* 13, no. 10 (2024).
44. Vangala, Dayasagar. "Integrating Generative AI with AEM for Dynamic Content Generation." *Famous Journal of computer science and Technology* 2, no. 6 (2024): 1-16.
45. Goti, Ankit Bharatbhai. "3D-Printed Multi-Layer PCBs: Evaluating the Structural Integrity and Electromagnetic Compatibility of Additively Manufactured Circuits." *International Journal Of Engineering And Computer Science* 13, no. 06 (2025).
46. Vangala, Dayasagar. "Sustainability in Digital Experience Platforms: Optimizing AEM for Energy Efficiency." *International Research Journal of Advanced Engineering and Technology* 1 (2025): 286-302.

47. Ghelani, HarshitKumar. "The Evolution of Ransomware: Trends and Countermeasures." (2025).
48. Ghelani, H. "Sustainable manufacturing engineering: enhancing product quality through green process innovations." *Int. J. Eng. Comput. Sci* 11 (2024): 25632-25649.
49. Hassan, Arooj, Malik Arfat Hassan, and Muhammad Ahsan Khan. "Data-Driven Decision-Making in Fintech Product Development using Cloud Analytics." *Multiverse Journal* (2025): 37-50.S