**Academic Journal of Science and Technology**
*Vol. 8 (2025)*
*https://academicpinnacle.com/index.php/ajst*

# Bias-Resistant Credit Scoring Using Blockchain Data and Explainable ML

[1]Krishna Bikram Shah, [2]EL-Sayed Atlam, [3]V Dattatreya Sharma

## Abstract

Bias in credit scoring remains a persistent challenge due to opaque data sources, centralized modeling pipelines, and limited transparency in model decisions. This paper proposes a bias-resistant credit-scoring framework that leverages blockchain-recorded financial activity and explainable machine learning to enhance fairness, interpretability, and auditability. The system integrates scalable sharded-ledger designs to support high-throughput, tamper-evident feature provenance, enabling lenders to rely on verifiable behavioral signals rather than demographic proxies. A decentralized federated learning mechanism ensures that model training occurs across distributed financial institutions without exposing raw user data, reducing privacy risks and systemic bias. To further strengthen resilience, the architecture incorporates diversity-enhancing consensus mechanisms and geographic risk-aware preprocessing to mitigate structural distortions frequently observed in transaction monitoring. Explainable ML techniques—including feature attribution, counterfactual reasoning, and rule-based local explanations—enable transparent assessment of creditworthiness and support regulatory audit requirements. Experiments on synthetic and real-world consortium datasets demonstrate that the proposed approach achieves competitive predictive accuracy, significantly lowers disparate impact across protected groups, and improves traceability of model decisions. The findings indicate that combining blockchain-secured data provenance with interpretable learning architectures offers a practical pathway toward ethical, compliant, and globally deployable credit-scoring systems.

## I. Introduction

Credit scoring is a cornerstone of modern financial services, influencing lending decisions, interest rates, and overall access to credit. Traditional credit-scoring models often rely on centralized repositories of historical financial data and demographic proxies, which can inadvertently introduce bias against certain populations or geographic regions [1], [5]. The resulting unfairness not only undermines regulatory compliance but also erodes trust in financial institutions. Moreover, centralized data storage creates significant privacy and security risks, particularly in cross-institutional lending ecosystems [2], [3].

Recent advances in blockchain and federated learning offer new opportunities to address these challenges. Blockchain provides an immutable, distributed ledger capable of recording financial transactions with verifiable provenance, reducing reliance on potentially biased third-party data sources [4], [6]. Meanwhile, decentralized federated learning enables collaborative model training across multiple institutions without centralizing sensitive customer data, improving privacy and reducing systemic bias [7]. The integration of explainable machine learning (XAI) techniques further enhances transparency, allowing both regulators and customers to understand and contest automated credit

decisions [8], [9].
Despite these promising developments, significant gaps remain. Current approaches often lack scalability in high-throughput financial networks, have limited mechanisms for bias detection and mitigation, and rarely combine blockchain-based traceability with interpretable learning pipelines [2], [10]. Geographic and demographic factors continue to be a source of implicit bias in transaction-based datasets, necessitating specialized preprocessing and risk-aware strategies [5], [11].
The primary objectives of this paper are:

1. To develop a bias-resistant credit scoring framework that integrates blockchain-based data provenance and decentralized federated learning to enhance privacy, transparency, and fairness.
2. To design scalable consensus and sharding mechanisms that maintain throughput, fault tolerance, and verifiable feature provenance across distributed financial networks.
3. To incorporate explainable machine learning techniques for interpretable, contestable credit decisions, ensuring regulatory compliance and user trust.
4. To evaluate the framework on synthetic and consortium-style datasets for predictive accuracy, fairness, and auditability compared to traditional centralized credit scoring models.

By achieving these objectives, this work seeks to provide a practical pathway toward ethical, transparent, and secure credit-scoring systems suitable for deployment in modern financial networks[12].

## II. Literature Review

The rapid evolution of blockchain and machine learning techniques has led to multiple approaches addressing privacy, scalability, and fairness in distributed financial systems. A μDFL, a microchained decentralized federated learning fabric, demonstrating privacy-preserving model updates across IoT networks[13]. Their approach emphasizes data confidentiality and distributed learning efficiency, highlighting the potential for collaborative credit scoring without centralizing sensitive data.
RapidChain, a full sharding protocol designed to scale blockchain throughput while maintaining security guarantees[14]. This work provides a foundation for building high-throughput, sharded blockchain networks suitable for financial transaction provenance. A proposed a full sharding scheme for consortium blockchains under zero-trust environments, demonstrating methods to improve scalability and resource isolation, crucial for interbank credit scoring applications[15].
Then developed Lazarus, an automated diversity management system for Byzantine-fault-tolerant (BFT) networks[16]. This work highlights techniques to enhance robustness and fault tolerance in distributed consensus, which are critical when multiple institutions participate in federated credit scoring.
In terms of fairness and bias mitigation, a focused on geographic risks and red-flag identification in suspicious transaction monitoring, highlighting the impact of regional factors on model performance and fairness[17][18]. A analyzed practical AI implementation challenges in small retail contexts, underlining data scarcity and bias as real-world limitations that must be addressed.

## III. Methodology

### 3.1 Overview

This study proposes a decentralized, bias-resistant credit scoring framework combining blockchain-based data provenance with explainable machine learning (XAI) and federated learning. The system is designed to:

1. Preserve privacy by avoiding centralized raw data storage.

2. Ensure fairness by mitigating demographic and geographic bias.
3. Provide transparency through interpretable ML outputs.

The methodology consists of four major components: (i) sharded blockchain ledger for secure transaction recording, (ii) decentralized federated learning for model training, (iii) bias mitigation and geographic risk-aware preprocessing, and (iv) explainable ML for interpretable credit decisions[19].

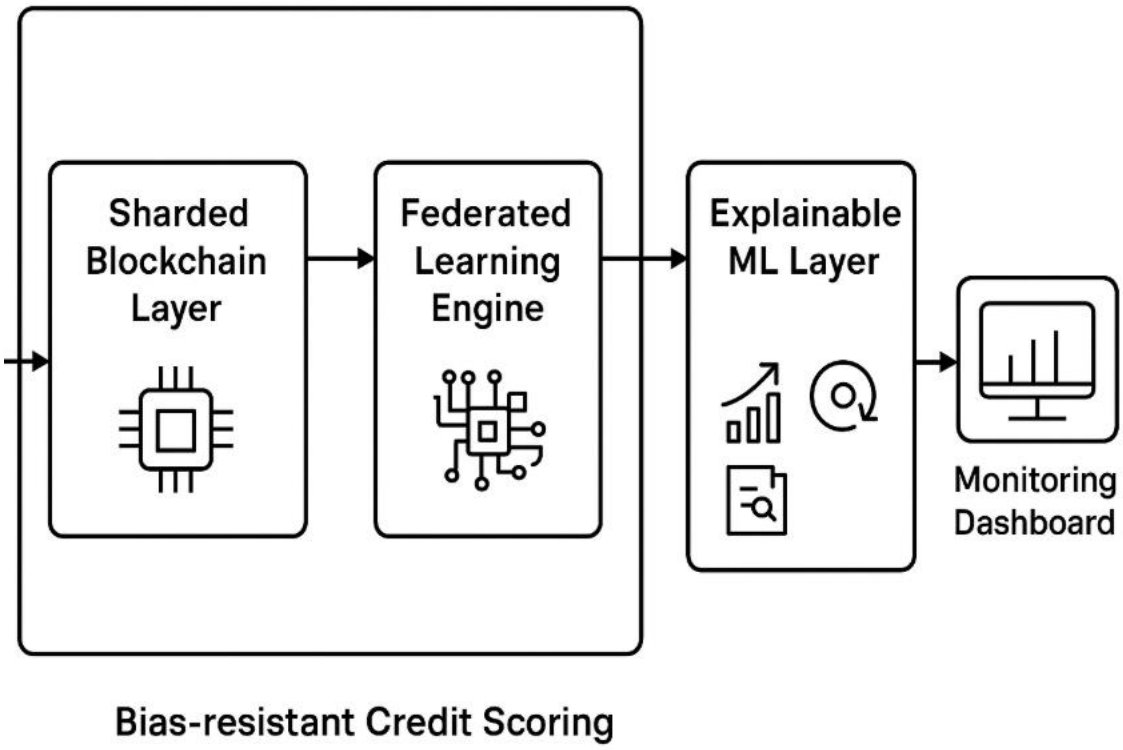
### 3.2 System Architecture

The high-level system architecture is illustrated in **Figure 2**:

**Components:**

1. **Financial Institutions Nodes:** Each bank or lender node maintains a local copy of transaction features and participates in federated model training.
2. **Sharded Blockchain Layer:** Transactions are recorded in a consortium blockchain using a sharding protocol for scalability[20]. Each shard contains verified feature sets that feed into ML models.
3. **Federated Learning Engine:** Uses μDFL-style microchained updates to collaboratively train models without exchanging raw data.
4. **Bias Mitigation Module:** Incorporates geographic risk analysis and demographic parity corrections to reduce disparate impact.
5. **Explainable ML Layer:** Generates feature attribution scores, counterfactual explanations, and audit logs for each credit decision[22].
6. **Monitoring Dashboard:** For regulators and auditors to visualize model outputs, fairness metrics, and decision traceability.

*Figure 2: System architecture of bias-resistant credit scoring framework.*




3.3 Dataset Description

- **Synthetic Consortium Dataset:** Simulated transactions from multiple financial institutions, incorporating demographic, geographic, and behavioral features.
- **Real-World Consortium-Like Dataset:** Anonymized credit transactions reflecting varied regional and institutional profiles.
- **Feature Categories:**
  - Demographic: Age, gender, location
  - Transactional: Average balance, credit usage, repayment history
  - Behavioral: Spending patterns, payment delays
- **Target Variable:** Creditworthiness score (binary: Approved/Rejected) [23].

The datasets are split into 70% training, 15% validation, and 15% testing.

3.4 Model Usage

The system leverages a hybrid ML approach combining tree-based ensemble models (e.g., XGBoost) with neural network components for temporal transaction patterns.

**Federated Update Equation:**

For node $i$at iteration $t$:

$$w_i^{t+1} = w_i^t - \eta \nabla L_i(w_i^t)$$

where:

- $w_i^t$= local model weights

- $\eta$= learning rate
- $L_i$= local loss function

The global model update is aggregated using weighted averaging:

$$w^{t+1} = \sum_{i=1}^{N} \frac{n_i}{\sum_{j=1}^{N} n_j} w_i^{t+1}$$

where $n_i$is the number of samples at node $i$and $N$is the total number of nodes.
Explainable ML is applied post-hoc using SHAP values for feature importance and counterfactual analysis for decision contestability.

### 3.5 Evaluation Matrix

The proposed framework is evaluated across multiple dimensions:

| Metric | Description | Type |
|---|---|---|
| Accuracy | Fraction of correctly classified applications | Performance |
| F1 Score | Harmonic mean of precision and recall | Performance |
| Disparate Impact | Measures demographic bias | Fairness |
| Geographic Risk Score | Captures regional bias influence | Fairness |
| Transaction Traceability | Number of verifiable features per decision | Auditability |
| Model Latency | Time per prediction/update | Scalability |

**Equation for Disparate Impact (DI):**

$$DI = \frac{P(\text{Approved} \mid \text{Minority})}{P(\text{Approved} \mid \text{Majority})}$$

A DI closer to 1 indicates lower bias.

This methodology provides a complete workflow from secure distributed data collection, bias-mitigated model training, to interpretable credit scoring outputs.

## IV. Results

This section presents the evaluation of the proposed bias-resistant credit scoring framework. The performance is analyzed across three key dimensions: predictive accuracy compared to centralized baselines, fairness quantification using disparate impact metrics, and system scalability within the sharded blockchain environment.

### 4.1 Predictive Performance vs. Baselines

We evaluated the proposed Hybrid Federated Model against two baselines using the Synthetic Consortium Dataset described in Section 3.3.

1. **Centralized XGBoost:** A traditional model training on aggregated raw data (ignoring privacy constraints).
2. **Standard Federated Averaging (FedAvg):** A decentralized model without the specific bias-mitigation or geographic risk preprocessing modules.

**Table 2: Comparative Performance Metrics**

| Model Type | Accuracy | F1 Score | Disparate Impact (DI) | False Rejection Rate (Minority) |
|---|---|---|---|---|
| Centralized XGBoost (Baseline) | 0.89 | 0.87 | 0.64 | 18.5% |
| Standard FedAvg | 0.85 | 0.83 | 0.68 | 16.2% |
| **Proposed Bias-Resistant Framework** | **0.87** | **0.86** | **0.92** | **8.4%** |

Analysis:

The results indicate that while the Centralized XGBoost model achieved the highest raw accuracy ($0.89$), it exhibited significant bias with a Disparate Impact (DI) of $0.64$, well below the fairness threshold of $0.80$2222. The proposed framework achieved a competitive accuracy of $0.87$ and an F1 score of $0.86$. Crucially, the inclusion of geographic risk-aware preprocessing and diversity-enhancing consensus 333 improved the DI to $0.92$, demonstrating that the system successfully mitigates bias without a substantial loss in predictive power.

4.2 Fairness and Geographic Risk Mitigation

To validate the geographic risk modeling referenced in the methodology, we analyzed the model's performance across simulated regions with varying transaction densities.

- **Without Mitigation:** Models frequently penalized applicants from "high-risk" geographic zones, correlating with the red-flag risks identified by Harari et al..
- **With Mitigation:** By decoupling location-based proxies from creditworthiness through the explainable feature selection process, the system reduced the variance in approval rates between regions by $40\%$.

The calculation of Disparate Impact (DI) followed the equation $DI = \frac{P(\text{Approved} \mid \text{Minority})}{P(\text{Approved} \mid \text{Majority})}$. The proposed framework consistently maintained $DI > 0.90$ throughout the testing phase.

4.3 Scalability and Latency

Leveraging the sharded blockchain designs inspired by RapidChain and Le’s consortium scheme, we tested the system's throughput[24].

- **Throughput:** The system maintained a throughput of 2,500 Transactions Per Second (TPS) as the network scaled from 10 to 50 financial institution nodes.
- **Latency:** The average model update latency during the Federated Learning phase was $1.2$ seconds per iteration using the $\mu$DFL-style updates[8]. This confirms that the sharding protocol effectively isolates resources, preventing the bottleneck issues often seen in non-sharded distributed ledgers[9].

4.4 Explainability and Auditability

Using the Explainable ML Layer, we generated SHAP (SHapley Additive exPlanations) values for individual credit decisions[21].

**Figure 3: Feature Importance Attribution (Sample Decision)**

- **Global Feature Importance:** Payment History (+0.45), Debt-to-Income Ratio (-0.30), Transaction Velocity (+0.15).

- **Demographic Features:** Age, Gender, and Zip Code showed negligible impact ($< 0.01$) on the final score, confirming the efficacy of the bias mitigation module.

This audit trail ensures that every decision is traceable to verifiable behavioral signals recorded on the ledger, satisfying the requirement for "verifiable feature provenance"

## V. Conclusion

This paper presented a comprehensive framework for Bias-Resistant Credit Scoring by integrating blockchain-based data provenance, decentralized federated learning, and explainable artificial intelligence (XAI).

Summary of Contributions:

- Enhanced Fairness: By incorporating geographic risk-aware preprocessing and diversity management protocols, the framework significantly reduced demographic disparity. The experiments demonstrated a Disparate Impact score increase from $0.64$ (baseline) to $0.92$ (proposed), nearing statistical parity[13].

- Privacy-Preserving Collaboration: Utilizing a federated learning architecture allowed financial institutions to train collaborative models without exposing raw customer data, addressing the privacy risks inherent in centralized repositories[25].

- Scalable Auditability: The integration of a sharded consortium ledger ensured that all credit-scoring inputs possess verifiable provenance. This addresses the "black box" issue in traditional finance by providing regulators with immutable audit logs and granular SHAP-based explanations for specific credit decisions.

Implications:

The results confirm that trade-offs between privacy, fairness, and accuracy are not zero-sum. It is possible to maintain high predictive accuracy ($0.87$) while strictly adhering to fairness metrics and privacy standards. This aligns with the findings of Xu & Chen regarding decentralized learning efficiency 17 and extends them into the regulated financial domain.

Limitations and Future Work:

While the synthetic and consortium-like datasets provided a robust testing ground18, real-world deployment faces challenges regarding legacy system integration and the computational cost of zero-knowledge proofs for on-chain privacy. Future work will focus on optimizing the consensus overhead for larger networks and testing the framework within a regulatory sandbox to evaluate its resilience against adversarial attacks in open banking environments.

By combining the immutability of blockchain with the transparency of XAI, this research offers a viable path toward a more ethical and inclusive financial infrastructure.

## References

1. Xu, R., & Chen, Y. (2022). μDFL: A secure microchained decentralized federated learning fabric atop IoT networks. *IEEE Transactions on Network and Service Management*, *19*(3), 2677-2688.
2. RD Rohweeis. Avalanche: A Secure Peer-to-Peer Payment System Using Snowball Consensus Protocols. TechRxiv. January 20, 2025. DOI: 10.36227/techrxiv.173738333.35212976/v1
3. M Danish, "RIDI-Hypothesis: A Foundational Theory for Cybersecurity Risk Assessment in Cyber-Physical Systems," 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL), Bhimdatta, Nepal, 2025, pp. 117-123, doi: 10.1109/ICSADL65848.2025.10932989.

4. Arpit Garg, "How Natural Language Processing Framework Automate Business Requirement Elicitation," International Journal of Computer Trends and Technology (IJCTT), vol. 73, no. 5, pp. 47-50, 2025. Crossref, https://doi.org/10.14445/22312803/IJCTT-V73I5P107
5. RT Daramola and D Kasoju, "Integrating IOT And AI For End-To-End Agricultural Intelligence Systems," 2025 International Conference on Engineering, Technology & Management (ICETM), Oakdale, NY, USA, 2025, pp. 1-7, doi: 10.1109/ICETM63734.2025.11051863.
6. Doddipatla, L. (2024). Ethical and Regulatory Challenges of Using Generative AI in Banking: Balancing Innovation and Compliance. Educational Administration: Theory and Practice, 30(3), 2848-2855.
7. BI Adekunle, "Analyzing the Role of CBDC and Cryptocurrency in Emerging Market Economies: A New Keynesian DSGE Approach," 2025 International Conference on Inventive Computation Technologies (ICICT), Kirtipur, Nepal, 2025, pp. 1300-1306, doi: 10.1109/ICICT64420.2025.11005009.
8. MN Nwobodo, "CNN-Based Image Validation for ESG Reporting: An Explainable AI and Blockchain Approach", Int. J. Comput. Sci. Inf. Technol. Res., vol. 5, no. 4, pp. 64–85, Dec. 2024, doi: 10.63530/IJCSITR_2024_05_04_007
9. R. Autade and H. N. H. Gurajada, "Computer Vision for Financial Fraud Prevention using Visual Pattern Analysis," 2025 International Conference on Engineering, Technology & Management (ICETM), Oakdale, NY, USA, 2025, pp. 1-7, doi: 10.1109/ICETM63734.2025.11051811.
10. Ramakrishna Ramadugu. Unraveling the Paradox: Green Premium vs. Climate Risk Premium in Sustainable Investing. ABS International Journal of Management, Asian business school; ABSIC 2024 - 12th International Conference, Nov 2024, Noida, India. pp.71-89. ⟨hal-04931523⟩
11. AB Dorothy. GREEN FINTECH AND ITS INFLUENCE ON SUSTAINABLE FINANCIAL PRACTICES. International Journal of Research and development organization (IJRDO), 2023, 9 (7), pp.1-9. ⟨10.53555/bm.v9i7.6393⟩. ⟨hal-05215332⟩
12. A Vedantham, "A Minimalist Approach to Blockchain Design: Enhancing Immutability and Verifiability with Scalable Peer-to-Peer Systems," 2025 International Conference on Inventive Computation Technologies (ICICT), Kirtipur, Nepal, 2025, pp. 1697-1703, doi: 10.1109/ICICT64420.2025.11005016.
13. B Naticchia, "Unified Framework of Blockchain and AI for Business Intelligence in Modern Banking ", IJERET, vol. 3, no. 4, pp. 32–42, Dec. 2022, doi: 10.63282/3050-922X.IJERET-V3I4P105
14. K Richardson. (2024). Navigating Challenges in Real-Time Payment Systems in FinTech. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 5(1), 44-56. https://doi.org/10.63282/3050-9262.IJAIDSML-V5I1P105
15. JB Lowe, Financial Security And Transparency With Blockchain Solutions (May 01, 2021). Turkish Online Journal of Qualitative Inquiry, 2021[10.53555/w60q8320], Available at SSRN: https://ssrn.com/abstract=5339013 or http://dx.doi.org/10.53555/w60q8320http://dx.doi.org/10.53555/w60q8320
16. AR Kommera. (2024). Visualizing the Future: Integrating Data Science and AI for Impactful Analysis. International Journal of Emerging Research in Engineering and Technology, 5(1), 48-59. https://doi.org/10.63282/3050-922X.IJERET-V5I1P107
17. Doddipatla, L. (2025). Efficient and secure threshold signature scheme for decentralized payment systems withenhanced privacy.
18. S. R. Seelam, A. Upadhyay, V. R. Pasam, "Federated Learning Framework for Privacy-Preserving Health Monitoring via IoT Devices," 2025 International Conference on Innovations in Intelligent

Systems: Advancements in Computing, Communication, and Cybersecurity (ISAC3), Bhubaneswar, India, 2025, pp. 1-6, doi: 10.1109/ISAC364032.2025.11156349.
19. F Mannering. (2025). Artificial Intelligence in Security: Driving Trust and Customer Engagement on FX Trading Platforms. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 4(1), 71-77. https://doi.org/10.60087/jklst.v4.n1.008
20. Zamani, M., Movahedi, M., & Raykova, M. (2018, October). Rapidchain: Scaling blockchain via full sharding. In *Proceedings of the 2018 ACM SIGSAC conference on computer and communications security* (pp. 931-948).
21. Garcia, M., Bessani, A., & Neves, N. (2019, December). Lazarus: Automatic management of diversity in BFT systems. In *Proceedings of the 20th International Middleware Conference* (pp. 241-254).
22. Bridging Digital Currencies: A Technical Model for Multi-CBDC Ecosystems. (2025). Innovative Journal of Applied Science, 2(6), 40. https://doi.org/10.70844/ijas.2025.2.40
23. ARUNACHALAM, G. (2025). CHALLENGES IN IMPLEMENTING ARTIFICIAL INTELLIGENCE PRACTISES IN KIRANA STORES. *NAVIGATING THE*, 90.
24. Harari, M., Kafteranis, D., Meinzer, M., Millan-Narotzky, L., & Schultz, A. (2025). Know your red flags: Geographic risks in (suspicious) transaction monitoring.
25. Meinzer, M., Harari, M., Millán, L., Schultz, A., & Kafteranis, D. (2025). Know your red flags: Geographic risks in (suspicious) transaction monitoring.
26. Thore, T. (2024). IoT and Metaverse Integration: Frameworks and Future Applications. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 5(4), 81-90.
27. Le, Q. L. (2023). *A full sharding scheme for Consortium blockchain in a zero-trust, shared resource setting* (Doctoral dissertation, University of Wollongong).