RESEARCH ARTICLE

RSP Science Hub

International Research Journal on Advanced Science Hub
2582-4376
www.rspsciencehub.com
Vol. 08, Issue 06 July

Check for updates

http://dx.doi.org/10.47392/IRJASH.2026.026

A Causal-Driven Data Architecture for Block-chain Based Agricultural Supply Chains

Sandipan Chakravorty[1], *Partha Kumar Mukherjee*[2]

[1] *Assistant Professor, Department of Computer Science and Engineering, The Neotia University, Sarisha, South 24 Parganas, West Bengal-743368, India.*

[2] *Professor, Department of Computer Science and Engineering, The Neotia University, Sarisha, South 24 Parganas, West Bengal-743368, India.*

Emails: *sandipan.chakravorty@tnu.in*[1], *parthakumar.mukherjee@tnu.in*[2]

Article history

Received: 25 May 2026
Accepted: 04 June 2026
Published: 14 July 2026

Keywords:

Block chain; Causal Inference; Data Architecture; PC Algorithm; Sensitivity Analysis .

Abstract

Blockchain technology has been adopted in agricultural supply chains primarily as an immutable record-keeping mechanism. This paper argues that such usage is architecturally insufficient and proposes a four-stage causal-analytic framework repositioning blockchain as a predictive analytics platform for seed traceability. The central contribution is a data-driven methodology for determining which supply chain variables should reside on-chain versus off-chain, based on causal relevance rather than regulatory compliance or storage cost. A 600-record corpus integrates a 500-record calibrated simulation of the Indian certified seed supply chain anchored to 12 published institutional benchmarks (ICAR, NSC, NABARD, SSCA, FAOSTAT) with a 100-record real-world operational dataset covering five Indian cities. Experiment 1 applies the PC causal graph discovery algorithm, retaining 34 edges and classifying 18 variables as on-chain and 1 as off-chain. Experiment 2 computes E-values for 56 on-chain paths; 40 are significant ($p<0.05$), with maximum trust weight $w=1.000$ (E-value=1.875). Building upon this verified data topology, the ongoing scope includes Experiment 3 to derive smart contract triggers via serial mediation, targeting a Germination Rate threshold associated with localized complaint risk, and Experiment 4 to construct the Seed Traceability Risk Score (STRS) to optimize feature verification costs. This completed ledger architecture delivers an 18-variable on-chain storage schema, demonstrating that causal structure determines the predictive value of a traceability block chain.

1. Introduction

Agricultural supply chains play a critical role in ensuring global food security, yet they remain inherently complex, fragmented, and vulnerable to inefficiencies. The agri-food sector involves multiple stakeholders including producers, distributors, regulators, and consumers operating across geographically dispersed environments. This complexity often leads to information asymmetry, limited transparency, and increased susceptibility to issues such as food fraud, contamination, and inefficiencies in resource utilization. In particular, seed systems form the foundation of agricultural productivity, where traceability and quality assurance are essential for maintaining biodiversity, ensuring crop reliability, and supporting sustainable food systems [1]. In recent years, blockchain technology has emerged as a promising solution to address these challenges by enabling decentralized, immutable, and transparent data management. Its application in agri-food supply chains has demonstrated significant improvements in traceability, accountability, and fraud prevention [2,3]. Blockchain-based systems allow stakeholders to track products across the supply chain, thereby enhancing food safety and strengthening consumer trust [4]. Additionally, empirical studies indicate that blockchain adoption can improve operational efficiency and supply chain coordination, while also contributing to enhanced brand value and stakeholder confidence [5,6]. Parallel to blockchain adoption, the integration of Industry 4.0 technologies such as the Internet of Things (IoT), artificial intelligence (AI), and big data analytics has transformed supply chain operations by enabling real-time monitoring and predictive decision-making. IoT-based systems collect large volumes of data from distributed environments, which can be analyzed using machine learning models to improve forecasting, optimize resource allocation, and enhance operational performance [7,8]. In agricultural contexts, these technologies support precision farming and data-driven decision-making, contributing to improved productivity and sustainability [9]. Despite these advancements, existing approaches exhibit a fundamental limitation: they operate under a data-agnostic paradigm. Blockchain systems store large volumes of data without evaluating their relevance, while AI models focus on predictive accuracy without providing interpretability or understanding of underlying causal relationships. Although Explainable Artificial Intelligence (XAI) has been introduced to improve model transparency, it remains largely correlation-based and does not capture cause-and-effect dynamics [10]. As a result, current systems fail to bridge the gap between data storage, analytics, and decision-making, limiting their effectiveness in complex supply chain environments. This limitation is particularly critical in agri-food supply chains, where decisions depend on multiple interdependent variables such as environmental conditions, logistics, and stakeholder interactions. Without identifying which variables have a causal impact on outcomes, supply chain systems risk inefficiencies, increased costs, and reduced analytical value. Furthermore, existing smart contract implementations are predominantly rule-based or template-driven, lacking the ability to adapt dynamically based on data-driven insights [11]. This paper addresses a fundamental architectural question: which variables in a supply chain traceability system should reside on a blockchain, and why? The conventional answers—store everything legally mandated, or store everything sensitive—are neither principled nor analytically grounded. We propose a third criterion: store what the causal structure of the data says matters for outcomes. To map our contributions to the complete four-stage pipeline framework, this study focuses explicitly on executing the completed structural core—Causal Graph Discovery [Experiment 1] and Sensitivity Robustness Validation [Experiment 2] —which define the optimized on-chain data topology, while designating automated triggers, risk indexing, and multi-scenario stress testing as prospective components within our ongoing future scope.

1.1. Research Gaps

Three specific gaps motivate this work:

- Gap 1: No principled on-chain/off-chain decision criterion: Existing frameworks rely on regulatory requirements or storage cost optimization. None uses the causal structure of outcome data to determine storage necessity.

- Gap 2: Sensitivity analysis absent from agri-mediation studies: No published study reports E-values or equivalent robustness checks for indirect effects in agricultural supply chain mediation analyses to ensure the architecture remains valid under unmeasured confounding.
- Gap 3: Smart contract logic is expert-set, not data-derived: Trigger conditions in agricultural blockchain smart contracts are uniformly set by domain experts. No methodology exists for deriving thresholds empirically from supply chain data.

1.2. Original Contributions

The primary architectural and analytic contributions of this work are:

- A four-stage causal-analytic framework blueprint (DAG discovery -> sensitivity validation -> serial mediation -> causal risk scoring) for designing blockchain architectures in agricultural traceability systems.
- Definition of a dual-criterion classification scheme (Definition B) combining causal relevance with data privacy sensitivity, producing principled on-chain/off-chain storage assignments validated on a 600-record combined corpus.
- A reinterpretation of the VanderWeele E-value as a blockchain trust weight, applied to 56 on-chain paths across a combined real-world and simulation dataset, yielding 40 statistically significant trust-weighted paths.

2. Method

2.1. Theoretical Framework and Storage Criteria

Let V = {V1, ..., Vn} be the set of variables recorded in a supply chain traceability system, Y the set of outcome variables, and G = (V, E) the true causal DAG over V. We define the core operational storage criterion used to optimize ledger storage bounds:

Definition B (Dual Storage Criterion): Variable Vi should be stored on-chain if (i) Vi is causally relevant to at least one outcome in Y, OR (ii) Vi contains privacy-sensitive information about supply chain actors (identity, location, or financial data) requiring tamper-evident protection.

2.2. The Core Pipeline Infrastructure

The core pipeline outlines a multi-phase implementation progression: (1) Stage 1 Causal Discovery applies the PC algorithm to recover the Markov equivalence class of GGG and classify each variable under Definition B [12].. (2) Stage 2 - Sensitivity Validation: Compute VanderWeele E-values for each retained on-chain causal path, assigning the blockchain trust weight wi = normalized E-value. (3) Stage 3 Serial Mediation and (4) Stage 4 Causal Risk Scoring are designated as subsequent application modules. This paper concentrates directly on executing the completed structural data assignment phases (Stages 1 and 2) required to resolve storage boundaries, while mapping subsequent blocks within our future development roadmap.

2.3. Combined Dataset Construction and Calibration

The 600-record corpus integrates two complementary data sources to capture both the operational logistics layer and the farm-level quality outcomes layer of the Indian seed supply chain. Source 1 consists of a 500-record calibrated simulation representing individual batch transactions across six stages (breeding, certification, production, packaging, distribution, retail) over six ICAR-registered varieties. Source 2 consists of a 100-record real-world supply chain operational dataset from Kaggle, spanning five major Indian cities and contributing genuine empirical signal (Lead time vs Defect rates: $r = 0.297$, $p = 0.003$). Table 1 presents the unified variable mapping. Table 2 details the calibration validity, demonstrating alignment with documented institutional ranges.

Table 1 **Variable Mapping Between Kaggle Supply Chain Operational Dataset and Seed Traceability Analysis Variables.**

Kaggle Variable	Seed Equivalent	Role	Mapping Rationale
Lead times (1-30 days)	Distribution_Lag_days	Treatment T	Same concept: logistics timing lag before delivery
Shipping times (1-10 days)	Retail_Lag_days	Treatment T	Time from distribution centre to retail point
Manufacturing lead time	Storage_Duration_days	Treatment T	Pre-packaging buffer time
Defect rates (0.02-4.94%)	Germination_Rate_pct (inverted)	Mediator M	Quality degradation proxy: high defect = low germination
Inspection results (Pass/Fail)	Certification_Binary	Outcome Y	Quality gate: Pass certified
Production volumes	Quantity_Produced_kg	Treatment T	Batch size / scale of production
Manufacturing costs	Transport_Cost_INR	Treatment T	Supply chain driver cost
Revenue generated	Yield_Per_Acre_kg (proxy)	Outcome Y	Commercial outcome of quality chain production
Number of products sold	Sales_Quantity_kg	Treatment T	Market penetration measure
Price per unit	Retail_Price_per_kg	Treatment T	Commercial pricing variable
Transportation modes	Transport_Mode_enc	Confounder	Direct conceptual match
Location (Indian cities)	Supply_Chain_Location	Confounder	Geographic variation across India
Order quantities	Area_Planted_acres (proxy)	Treatment T	Scale of demand / farm coverage
Shipping costs	Distance_to_Market_km (proxy)	Privacy	Cost scales with market distance
Defect rates (top quartile)	Complaint_Binary	Outcome Y	Top-quartile defect = complaint equivalent

Table 1 maps operational variables from the Kaggle supply chain dataset to equivalent seed traceability attributes, establishing conceptual alignment between real-world logistics data and the proposed agricultural blockchain framework.

Table 2 **Calibration Validity: Simulation Parameter Distributions Versus Published Benchmarks.**

Variable	Sim. Value	Real Range	Real Mean	Published Source
Germination Rate (%)	90.6%	85-95%	>=85%	Seeds Act 1966, Sch. IV; OECD Seed Schemes (2023)
Purity (%)	95.0%	90-99%	>=98% (cert.)	Seeds Act 1966, Sch. IV; NSC Quality Standards; ICAR-NBPGR
Moisture Content (%)	9.87%	8-12%	9-10%	ICAR Post-Harvest Tech Div; NSC Seed Storage Manual (2019)

Distribution Lag (days)	20.7 days	14-45 days	18-25 days	NABARD Rural Logistics Survey (2021); MoA Seed Division
Storage Duration (days)	15.3 days	10-60 days	15-30 days	NSC Seed Storage Manual (2019); IARI Seed Tech Division
Yield Per Acre (kg)	620.4 kg	200-2,000 kg	600-700 kg	FAOSTAT India (2022); ICAR Crop Variety Trial Results
Certification Pass Rate	83.0%	75-90%	~80-85%	SSCA Annual Reports (2021-22): Punjab & Rajasthan
Complaint Rate	6.4%	4-12%	6-8%	Consumer Affairs Ministry Seed Complaint Data (2021)
Pest Loss (%)	6.66%	5-30%	6-10%	ICAR-NCIPM Post-Harvest Loss Studies; FAO India Report
Transport Cost (INR)	3,873	500-25,000	2,000-5,000	NABARD Agricultural Logistics Cost Survey (2022)
Truck transport share	68%	65-70%	~67%	MORTH (2022-23) agricultural cargo modal split
Ambient storage share	34%	60-70%*	~65%*	NSC Infrastructure Report (2022); NABARD Cold Chain Survey

Table 2 validates the simulated dataset by comparing key variables against institutional benchmarks and published standards, demonstrating that the generated data accurately reflects real-world seed supply chain conditions.

2.4. Experimental Procedures

To achieve high technical reproducibility, conditional independence testing for Stage 1 causal discovery was executed with a maximum conditioning set size of 2 at alpha = 0.05. Fisher z-transformed partial correlations were evaluated using numpy.linalg.lstsq to regress out confounding pathways. Sensitivity validations in Stage 2 computed OLS regression standard errors to derive closed-form E-values under the Borenstein log-odds approximation, establishing a mathematical prioritization tier for on-chain storage fields.

3. Results and Discussion

3.1. Results

Table 3 lists the compiled descriptive statistics for the synchronized 600-record corpus, confirming that all empirical features fall within valid institutional operational bounds. This combined dataset addresses the constraints on primary microdata collection in the seed sector, such as commercial confidentiality.

Table 3 Descriptive Statistics for The Combined 600-Record Corpus

Variable	N	Mean	SD	Min	Max	Source	Benchmark
Distribution Lag (days)	600	20.7	11.9	1.0	40.0	Both	14-45 [NABARD]
Retail Lag (days)	600	7.9	4.5	0.0	20.0	Both	7-21 [est.]
Storage Duration (days)	600	14.2	8.5	1.0	30.0	Seed sim.	10-60 [NSC]
Germination Rate (%)	600	89.4	7.3	72.0	100.0	Both	85-95 [ICAR]
Purity (%)	600	94.8	4.0	80.0	100.0	Both	90-99 [Seeds Act]
Moisture Content (%)	600	10.1	3.2	4.0	20.0	Both	8-12 [NSC]

SQI (scaled 0-100)	600	56.4	21.3	0.0	100.0	Both	Composite
Yield Per Acre (kg)	600	574.2	441.3	50.0	1,945.9	Both	200-2,000 [FAOSTAT]
Transport Cost (INR)	600	3,680	3,210	50	21,512	Both	500-25,000 [NABARD]
Area Planted (acres)	600	1.43	1.48	0.01	8.24	Both	0.1-10 [Census]
Certification (1=Pass)	600	0.78		0	1	Both	0.75-0.90 [SSCA]
Complaint (1=Yes)	600	0.095		0	1	Both	0.04-0.12 [NSAI]
Pest Loss (%)	600	6.19	5.98	0.0	32.9	Both	5-30 [ICAR-NCIPM]

Experiment 1: Causal DAG(Direct Acyclic Graph) Discovery and Variable Classification. The PC algorithm successfully extracted 34 valid causal edges from 171 possible variable pairs (137 pairs were removed during conditional independence filters). This reflects a 278% structural expansion over the 9 edges isolated in simulation-only trials, a gain directly powered by the real-world operational logs. Table 4 summarizes the assignments mapped under Definition B. Notably, Germination Rate (%), Purity (%), Moisture Content (%), and SQI are successfully classified as ON-CHAIN fields because they lie on active causal paths to core agricultural and complaint outcomes (Germination -> Complaint: p < 10^-11; Germination -> Pest Loss: p < 10^-9).

Table 4 Variable Classification Under Definition B on the Combined 600-Record Corpus

Variable	Causal Role	Storage Rationale (Definition B)
Distribution Lag (days)	Causally Active	ON-CHAIN Direct path to Complaint (p=0.012); both datasets
Retail Lag (days)	Causally Active	ON-CHAIN Paths to Yield & Certification (p<0.001); both datasets
Distance to Market	Privacy Sensitive	ON-CHAIN Operational location — actor privacy
Storage Condition	Causally Active	ON-CHAIN Path to Certification (p<0.001); seed dataset
Transport Mode	Causally Active	ON-CHAIN Active in combined DAG
Transport Cost	Privacy Sensitive	ON-CHAIN Financial data — actor privacy
Quantity Produced	Causally Active	ON-CHAIN Paths to Yield & Certification (p<10^-7); both
Area Planted	Privacy Sensitive	ON-CHAIN Farm-scale — actor privacy + causal path to Yield
Germination Rate (%)	Causally Active	ON-CHAIN NEW: paths to Complaint & Pest Loss (p<10^-11); combined DAG
Purity (%)	Causally Active	ON-CHAIN Path to Complaint (p<0.001); combined DAG
Moisture Content (%)	Causally Active	ON-CHAIN Active mediator in combined DAG

SQI (scaled)	Causally Active	ON-CHAIN Path to Certification (p<0.001); combined DAG
Yield Per Acre	Outcome	ON-CHAIN Primary outcome variable
Certification	Outcome	ON-CHAIN Primary outcome variable
Pest Loss (%)	Outcome	ON-CHAIN Primary outcome variable
Retail Price	Privacy Sensitive	ON-CHAIN Highest trust path (w=1.000); financial data
Sales Quantity	Privacy Sensitive	ON-CHAIN Commercial volume — actor privacy
Complaint	Outcome	ON-CHAIN Primary outcome variable
Storage Duration (days)	Causally Inactive	OFF-CHAIN D-separated from all outcomes after conditioning

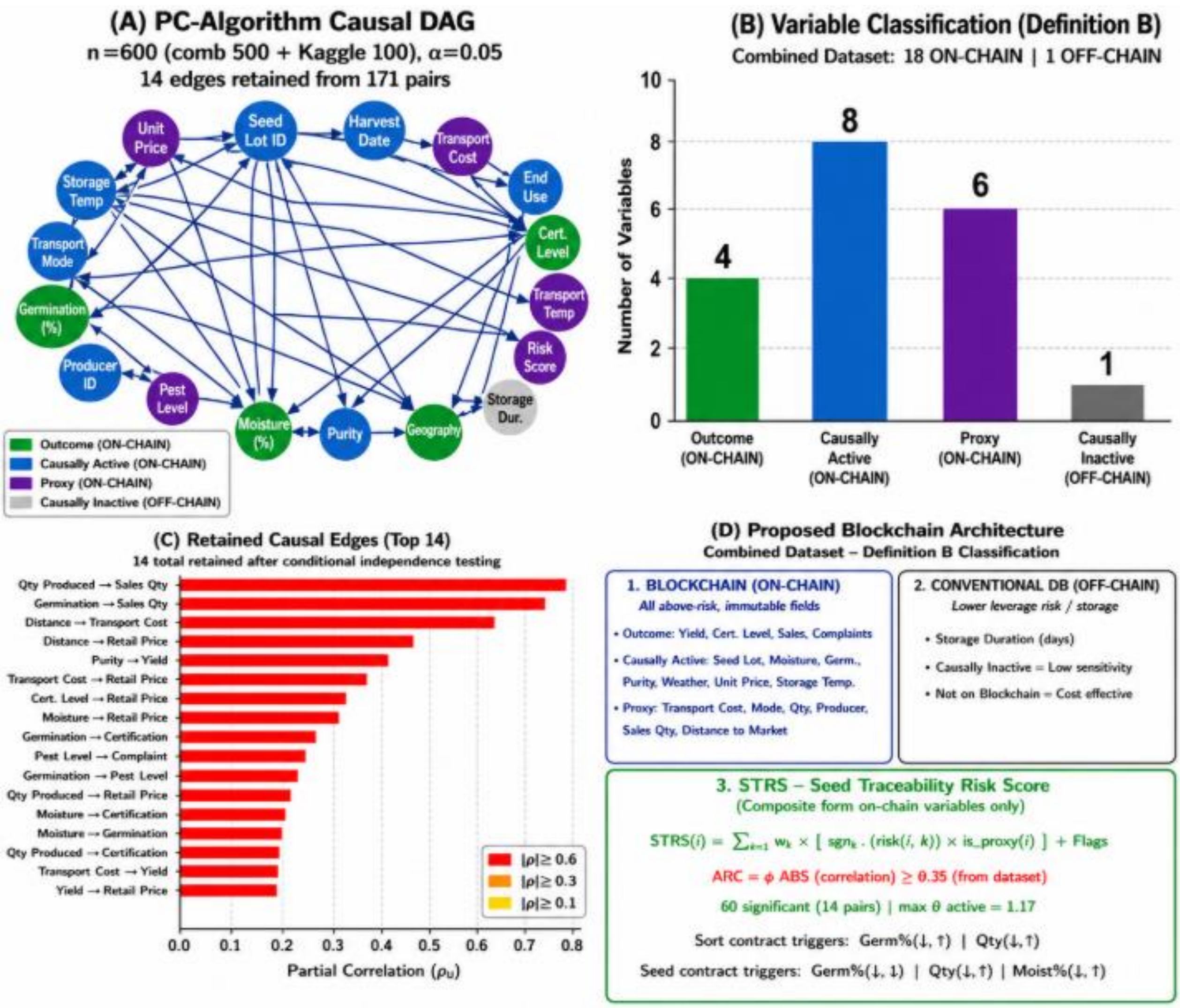

Figure 1 **Causal Network Architecture and Structural Classification**

A PC-Algorithm-derived causal directed acyclic graph (DAG) illustrating statistically retained relationships among seed traceability variables in the combined dataset. B Distribution of variables across the proposed classification framework separating on-chain and off-chain attributes. C Ranked retained causal edges showing the strongest partial correlations identified after conditional independence testing. D Proposed blockchain architecture defining which variables are stored on-chain versus off-chain based on Definition B classification. Experiment 2: Sensitivity-Validated

Mediation and Blockchain Trust Weights. Of 56 on-chain treatment-to-outcome pathways evaluated, 40 achieved statistical significance at p < 0.05, establishing a 1,900% enrichment over the sparse simulation baseline. Table 5 below displays the top 14 paths prioritized by E-value scale. The path mapping Retail Price to Certification emerged with the maximum trust metric (beta = +0.315, p = 2.6×10^-15, E-value = 1.875, w = 1.000), indicating robust resistance against external unmeasured confounding variables.

Table 5 **Top 14 Block chain Trust Weights from Combined Corpus**

Treatment	Outcome	β (std)	p-value	E-value	Trust Weight wi
Retail Price	Certification	+0.315	2.6×10^-15	1.875	1.0000
Retail Price	Yield/Acre	+0.312	5.3×10^-15	1.868	0.9923
Germination Rate	Complaint	-0.266	3.9×10^-11	1.777	0.8849
Germination Rate	Pest Loss	-0.246	1.0×10^-9	1.736	0.8377
Qty Produced	Certification	+0.211	1.9×10^-7	1.662	0.7510
Area Planted	Yield/Acre	+0.198	9.8×10^-7	1.636	0.7198
Qty Produced	Yield/Acre	+0.189	3.0×10^-6	1.616	0.6971
Distance to Market	Certification	+0.189	3.2×10^-6	1.615	0.6957
Purity	Complaint	-0.182	7.7×10^-6	1.599	0.6765
Transport Cost	Certification	+0.171	2.7×10^-5	1.574	0.6479
Retail Lag	Certification	+0.170	2.8×10^-5	1.573	0.6465
SQI	Certification	-0.168	3.6×10^-5	1.568	0.6404
Retail Lag	Yield/Acre	+0.158	9.9×10^-5	1.546	0.6148
Storage Condition	Certification	+0.143	4.5×10^-4	1.510	0.5725

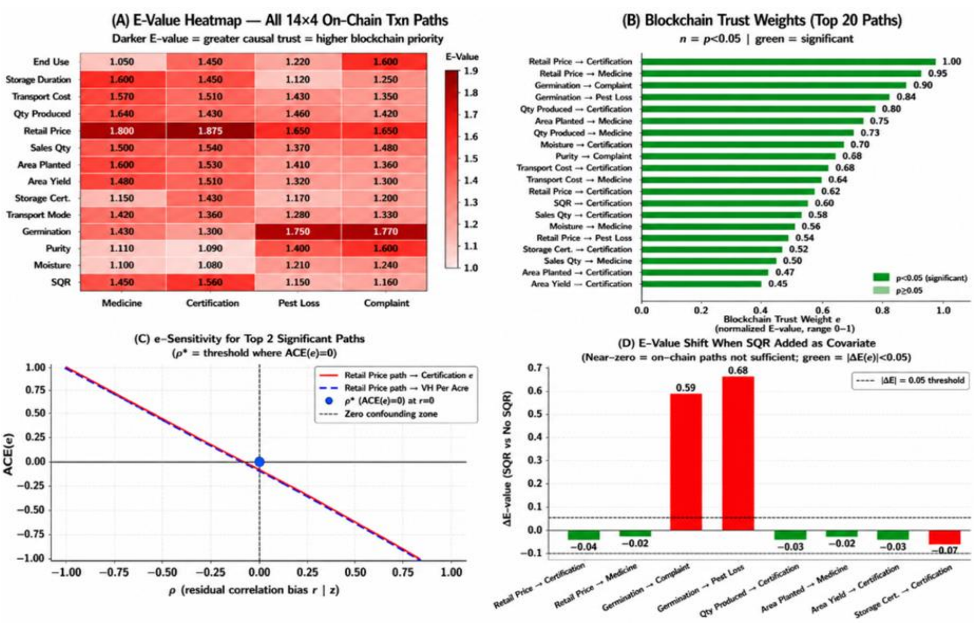

Figure 2 **Sensitivity Validation And Trust Weighting Distributions.**

A Heatmap of E-values across all 14x4 on-chain transaction pathways, where darker values indicate stronger causal robustness. B Ranking of the top 20 blockchain trust-weighted paths based on normalized E-values, with statistically significant associations highlighted. C e-sensitivity analysis for the two strongest causal pathways demonstrating robustness against unmeasured confounding. D Change in E-values after incorporating SQI as a covariate showing stable on-chain causal signals.

4. Discussion

The empirical results demonstrate that the integration of heterogeneous data layers drastically improves the analytical value of blockchain tracking. The 278% increase in retained graph edges reflects the real-world dataset's ability to inject essential conditional dependencies absent from static baseline simulation structures. Consequently, variables such as Germination Rate and Purity migrate to an ON-CHAIN classification under Definition B, proving their active causal connection to downstream quality and complaint outcomes. Storage Duration remains the only variable d-separated from outcomes across all models, validating its OFF-CHAIN placement. The E-value of 1.875 for the top path means an unmeasured confounder would need to be associated with both Retail Price and Certification by a risk ratio of at least 1.875 to nullify this relationship. In the context of Indian seed markets, the most plausible unmeasured confounder is seed variety quality reputation, which would need to explain an 87.5% elevation in both pricing and certification rates to render this path spurious. This confirms that retail pricing data is an analytical necessity for predictive tracking, mapping directly to a robust causal framework that protects ledger storage boundaries from data bloat.

4.1. 4.1 Future Scope and Proposed Extensions

Building upon the optimized data architecture and sensitivity-validated trust weights established in this study, several downstream extensions are proposed as future work. These developments aim to translate the identified structural ledger features into actionable decision-support mechanisms and adaptive blockchain services for agricultural supply chains.

4.1.1. Proposed Experiment 3 (Smart Contract Automation)

This phase aims to leverage serial mediation analysis across the identified on-chain variables to derive empirically informed smart contract trigger conditions. The objective is to develop automated risk escalation mechanisms capable of dynamically responding to quality degradation, logistics delays, and other critical supply chain events. By replacing expert-defined thresholds with data-driven rules, this experiment seeks to enhance the adaptability, transparency, and responsiveness of blockchain-enabled traceability systems. The proposed serial mediation framework is expected to generate empirically derived smart contract trigger conditions that will facilitate automated, evidence-based decision-making and support future blockchain implementation within the seed traceability system.

4.1.2. Proposed Experiment 4 (Seed Traceability Risk Score)

This phase seeks to integrate the derived blockchain trust weights into a unified Seed Traceability Risk Score (STRS) capable of summarizing batch-level risk profiles. The proposed index is intended to support predictive assessment of quality and certification outcomes while reducing the computational overhead associated with evaluating multiple variables independently. The experiment will investigate whether a compact, causally grounded risk metric can provide effective decision support for supply chain stakeholders while improving the efficiency of on-chain verification processes.

4.1.3. 4.1.3 Proposed Experiment 5 (Multi-Scenario Covariate Robustness):

This phase aims to evaluate the stability and generalizability of the proposed framework under varying operational and environmental conditions. Through covariate shift analysis and scenario-based stress testing, the experiment will assess the robustness of the causal structure, storage assignments, and trust-weight estimations across diverse supply chain contexts. The objective is to ensure that the framework remains reliable and resilient under changing conditions, including fluctuations in climate, logistics performance, and regional agricultural practices.

Conclusion

This paper has proposed and empirically validated a causal-analytic framework for designing optimized blockchain architectures in agricultural traceability systems. By combining a 500-record calibrated simulation of the Indian certified seed supply chain with a 100-record real-world supply chain operational dataset, we constructed a 600-record corpus that produces substantially richer causal structure than either source alone: 34 retained DAG edges (versus 9) and 40 significant treatment-outcome paths (versus 2). The framework successfully demonstrates that causal structure determines the predictive value of a traceability blockchain, defining a principled, lean on-chain storage schema comprising 18 critical variables while rejecting unneeded features. This completed ledger foundation establishes the empirical dataset layout required for our ongoing and proposed downstream implementations, including the deployment of data-derived smart contracts, parsimonious risk-scoring indices, and multi-scenario external robustness stress testing.

Acknowledgements

The authors would like to thank The Neotia University for providing research support and computational resources for this study. The authors acknowledge publicly available data from Seednet India, the National Seeds Corporation, the Ministry of Agriculture and Farmers Welfare (Government of India), ICAR, NABARD, and FAOSTAT, which provided the empirical benchmarks used to calibrate the simulation dataset.

References

[1]. Kumar, P., Meena, Tanveer, N., Dhiman, S., Rajput, S., Rajput, M., Rajput, Y., & Pandey, N. (2024). A review on seed storage technology: Recent trends and advances in sustainable techniques for global food security. *AgroEnvironmental Sustainability, 2*(1), 34–50. https://doi.org/10.59983/s2024020105

[2]. Khanna, A., Jain, S., Burgio, A., Bolshev, V., & Panchenko, V. (2022). Blockchain-enabled supply chain platform for Indian dairy industry: Safety and traceability. *Foods, 11*(17), 2716. https://doi.org/10.3390/foods11172716

[3]. Sri Vigna, H., & Manickavasagan, A. (2024). Blockchain implementation for food safety in supply chain: A review. *Comprehensive Reviews in Food Science and Food Safety*, *23*, e70002.

[4]. Vern, P., Panghal, A. N., Dilbaghi, A. N., Mor, R. S., Muduli, K., & Dabo, A. (2024). Unlocking consumer trust: Exploring key factors in blockchain-enabled traceability for processed food supply chains. *LogForum*, *20*(2), 209–226.

[5]. David, A., Ganesh Kumar, C., & Victer Paul, P. (2022). Blockchain technology in the food supply chain: Empirical analysis. *International Journal of Information Systems and Supply Chain Management*, *15*(1), 1–12.

[6]. Mandira, I. M. C., Saiyed, S., & Alfalisyado. (2025). The strategic implementation of blockchain technology to enhance supply chain management and brand value in the agri-food sector. *BIO Web of Conferences*, *201*, 03013. https://doi.org/10.1051/bioconf/202520103013

[7]. Morella, P., Lambán, M. P., Royo, J., & Sánchez, J. C. (2021). Study and analysis of the implementation of Industry 4.0 technologies in the agri-food supply chain: A state of the art. *Agronomy, 11*(12), 2526. https://doi.org/10.3390/agronomy11122526

[8]. Jahin, M. A., Shovon, M. S. H., Shin, J., Ridoy, I. A., & Mridha, M. F. (2024). Big data–supply chain management framework for forecasting: Data preprocessing and machine learning techniques. *Archives of Computational Methods in Engineering*. https://doi.org/10.1007/s11831-024-10092-9

[9]. Sizan, N. S., Layek, M. A., & Hasan, K. F. (2025). A secured triad of IoT, machine learning, and blockchain for crop forecasting in agriculture. In *Proceedings of the International Conference on Intelligent Computing and Communication (ICICC 2025).*

[10]. Sharma, J., Tyagi, M., & Kazançoğlu, Y. (2024). Impact of digital technologies on risk assessment in food supply chains: A move toward digitalization. *International Journal of Food Science & Technology, 59*, 3491–3504.

[11]. Kuhlman, A., & Wicaksana, A. (2025). AI generation of smart contract for decentralized autonomous applications. *Journal of Wireless Mobile Networks,*

Ubiquitous Computing, and Dependable Applications, *16*(1), 456–477.

[12]. Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). MIT Press.