Frontiers in Engineering
ISSN Print: 3104-4298
ISSN Online: 3104-4301
CD ISSN: 3104-431X
https://doi.org/10.63313/FE.9032

# Multimodal Artificial Intelligence and Edge Computing for Intelligent Traceability of Yunnan Specialty Agricultural Products with Unmanned Aerial Vision and Cross Border Electronic Commerce Integration

**Xiaolei Zhong[1,3]. Xin Wang[2,*]. Ru Yao[4]. Jinhua Yang[1] and Yunyang Wang[1]**
[1]Dianchi College, Kunming 650034, China
[2]Yunnan College of Business Management, Anning, 650301, China
[3]Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China
[4]Yunnan Vocational College of Mechanical and Electrical Technology, KunMing 650203, China
* Corresponding author. E-mail: vichyxin2008@126.com

**How to cite this paper:** Zhong, X., Wang, X., Yao, R., Yang, J., & Wang, Y. (2026). Multimodal artificial intelligence and edge computing for intelligent traceability of Yunnan specialty agricultural products with unmanned aerial vision and cross border electronic commerce integration. Frontiers in Engineering, 2(2), 75–97. ISSN Print: 3104-4298, ISSN Online: 3104-4301. https://doi.org/10.63313/FE.9032
**Published: 2026-11-20**

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
CC BY 4.0

## Abstract

Yunnan specialty agricultural products possess distinctive ecological provenance, ethnic cultural value and cross border market potential, yet their production remains geographically dispersed, environmentally heterogeneous and vulnerable to origin substitution, quality inconsistency and information fragmentation across supply and electronic commerce chains. This study establishes an intelligent traceability framework that integrates multimodal artificial intelligence, edge computing, unmanned aerial vehicle vision, robotic perception, spectral sensing, textual production records and environmental Internet of Things streams into a unified provenance authentication and quality decision system. A cross modal representation learning architecture aligns image, hyperspectral, textual, environmental and transactional signals through gated cross attention and contrastive alignment, while an edge cloud collaborative inference mechanism allocates perception and authentication tasks according to latency, energy and connectivity constraints in mountainous production areas. A hash anchored traceability structure binds field operations, processing events, logistics states, inspection outcomes and cross border electronic commerce transactions to verifiable product identities without exposing commercially sensitive raw records. Robotic grading and unmanned aerial monitoring close the loop between canopy and field condition observation, product appearance assessment, origin discrimination, quality prediction and market facing information disclosure. Controlled comparison studies across eight product scenarios covering Pu-erh tea, Yunnan coffee, wild edible fungi, cut flowers, Panax notoginseng, plateau vegetables, tropical fruit and medicinal herbs indicate that multimodal fusion improves authentication robustness under illumination variation, occlusion,

packaging change and seasonal drift, and that edge deployment substantially reduces response delay under constrained rural connectivity. Generalisation across harvest years, production counties and cross border regulatory regimes requires further validation. The proposed framework provides a reproducible technical route for trustworthy digital circulation of geographically indicated agricultural products from Yunnan production systems to domestic and Association of Southeast Asian Nations markets.

## Keywords

Multimodal Artificial Intelligence; Edge Computing; Intelligent Traceability; Yunnan Specialty Agricultural Products; Unmanned Aerial Vehicle Vision; Machine Vision; Cross Border Electronic Commerce; Origin Authentication

## 1. Introduction

Yunnan possesses singular agricultural diversity. Vertical climate gradients, mountain terrain, ethnic farming knowledge and proximity to Southeast Asian markets have produced geographically distinctive products including Pu-erh tea, arabica coffee, wild edible fungi, cut flowers, Panax notoginseng, plateau vegetables, tropical fruits and medicinal plant materials. These products derive market value not only from physical quality but also from verified origin, ecological conditions, harvest timing, processing craft and cultural narrative.

Digital circulation intensifies origin vulnerability. A product may pass through smallholder production, village collection, processing, grading, warehousing, cold chain logistics, cross border declaration and platform retail before reaching consumers. Each transfer creates opportunities for information loss, label replacement and batch mixing. Conventional records cannot resolve disputes when evidence is incomplete or detached from physical batches. Cross border electronic commerce adds multilingual translation, regulatory vocabularies and platform rules to ordinary traceability requirements.

Recent advances in machine vision, multimodal learning, unmanned aerial systems and edge intelligence provide a new basis for binding physical products to digital evidence[1]. Deep visual models identify disease, maturity and defects from field images[2]. Unmanned aerial vehicles extend observation from products to canopies and landscape context[3].The central scientific challenge is constructing a coherent system where heterogeneous evidence is synchronised, aligned, weighted, verified and converted into auditable decisions under rural network constraints.

This study addresses that challenge through an integrated architecture named multimodal artificial intelligence and edge computing based intelligent traceability system. The architecture treats traceability as an evidence computation problem rather than record storage. Every claim concerning origin, quality and handling is associated with observations, model inferences, device identities, timestamps and cryptographic digests. The study makes four contributions: a unified problem

formulation connecting origin authentication, quality grading, defect recognition and cross border information generation; a gated cross modal fusion architecture reducing dependence on fragile single modalities; an edge cloud collaborative scheduling method for mountainous environments; and a verifiable traceability integration mechanism for trading platforms and regulatory inspection.

## 2. Background

### 2.1. Ecological and industrial characteristics

Specialty production in Yunnan is shaped by altitude, microclimate, soil, cultivar and processing tradition. Pu-erh tea quality depends on mountain area, tree age and fermentation. Coffee quality varies with cultivar, altitude and processing. Wild edible fungi present species identity and freshness challenges. Cut flowers require rapid grading and cold chain control. Panax notoginseng demands attention to cultivation years and adulteration risk. These differences demonstrate that a traceability model cannot rely on a universal visual template but must combine product specific perception with shared evidence alignment mechanisms.

Smallholder dispersion is a defining condition. Production plots are fragmented across slopes with variable road access and intermittent mobile coverage. A tiered system design is required where phones, handheld spectral devices, field gateways, unmanned aerial vehicles, sorting robots and cloud services contribute at different cost and evidentiary strength levels.

### 2.2. Cross border electronic commerce and trust formation

Yunnan borders Myanmar, Laos and Vietnam and maintains extensive agricultural trade with Southeast Asian markets through land ports, the China Laos Railway and electronic commerce platforms. Cross border circulation adds multilingual description, phytosanitary documentation, origin evidence, customs classification and logistics tracking to ordinary traceability. Trustworthy information must distinguish measured evidence, model inference, producer declaration and regulatory certification while preserving the authority associated with each category[4].Platform commerce creates reputational risks when listing images differ from delivered batches. The framework connects production batch, inspection batch, logistics unit, customs declaration and platform stock keeping unit through controlled mappings.

### 2.3. Robotics, unmanned aerial systems and software infrastructure

Robotic perception contributes repeatability to grading operations affected by fatigue and subjective judgement. Unmanned aerial vehicles contribute plot level evidence through canopy structure, flowering, disease distribution and harvest readiness[5].Software engineering determines whether advanced models remain usable after laboratory development. Model files, feature schemas, device

calibration parameters and label taxonomies all evolve. Without version control, data lineage and automated testing, a traceability system can produce irreproducible decisions. The proposed system treats data contracts, model registries and audit logs as core scientific infrastructure[6]. Agri-food digitisation reinforces the need for interoperable supply chain technologies[7].

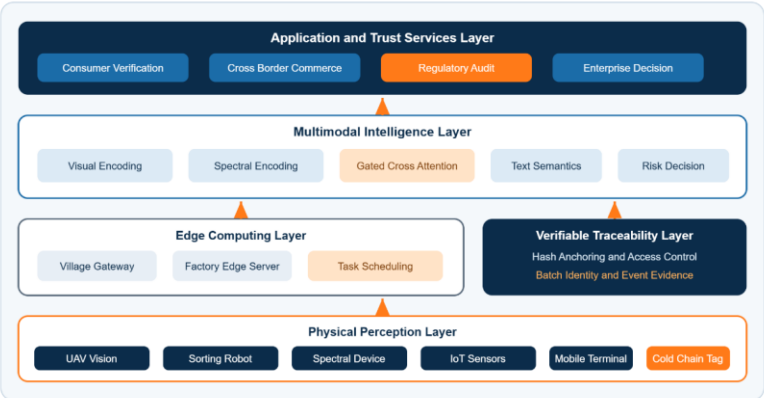



**Figure 1.** Overall Architecture of the Multimodal Edge Intelligent Traceability System

Figure 1. System setting. Four cooperating layers connect physical production in dispersed Yunnan production areas with consumer, regulatory and cross border commerce services. Data source. System design synthesis from field perception, processing, logistics and platform transaction requirements. Result interpretation. Physical evidence enters through controlled devices, multimodal inference occurs at edge or cloud locations according to resource state, and only digests and authorised verification results are anchored for long term audit. Raw commercially sensitive records remain in controlled domain storage.

## 3. Existing Technologies

### 3.1. Machine vision and robotic quality assessment

Convolutional and transformer based vision models have transformed plant disease recognition, fruit grading and post harvest sorting[8]. Vision transformers introduced global self attention for image classification[9].Residual networks established a robust baseline for hierarchical visual representation[10].Efficient scaling of convolutional architectures improved the balance between accuracy and computation[11].These advances nevertheless leave origin authentication unresolved when visually similar products originate from different mountains or counties. A robust system must combine appearance with biochemical, environmental, textual and movement evidence rather than treating geographical origin as an ordinary image category.

### 3.2. Multimodal learning and semantic alignment

Multimodal learning seeks representations that preserve shared semantics while

retaining modality specific evidence. The transformer architecture supplies a general mechanism for attention based interaction among sequences and feature sets[12]. Contrastive image text learning demonstrates that large scale paired supervision can align visual and linguistic concepts in a shared embedding space[13]. Contextual language models encode production records and inspection descriptions as semantic features[14]. Alignment must handle missing modalities, asynchronous sampling, sensor drift and contradictory claims. Reliability gated cross attention is therefore adopted as the principal fusion mechanism.

### 3.3. Edge computing and unmanned aerial perception

Edge computing moves computation toward data sources to reduce latency, bandwidth consumption and dependence on continuous cloud connectivity[15]. Agricultural fields impose unstable wireless links, limited power and heterogeneous device capability. Unmanned aerial missions generate high volume image streams that cannot be transmitted completely from mountainous plots in real time. Onboard or gateway level inference can extract plot features and transmit compact evidence summaries. The resulting problem is a dynamic allocation problem in which each task must be assigned to a device, edge node or cloud service according to changing resource conditions.

### 3.4. Blockchain supported traceability and its limits

Distributed ledgers and hash linked records have been investigated for agri food traceability because they provide tamper evidence and shared verification among organisations[16]. The foundational hash linked transaction concept derives from distributed digital cash research[17]. Subsequent analysis clarified that distributed ledgers are useful only when governance, identity and data quality are properly designed[18]. A ledger cannot determine whether a false record was truthful at the moment of entry. Physical sensing, device authentication and model based consistency checking must precede anchoring. The present framework uses hash anchoring for evidence integrity and selective disclosure, while origin truth is estimated from multimodal physical evidence before a record becomes eligible for anchoring.

## 4. Problem Definition

### 4.1. Objects, evidence and decisions

The system operates on production plots, harvest batches, processing batches, inspection units, logistics units and retail units. A traceable object receives a persistent identity at the earliest practical aggregation point. Evidence attached to an object includes aerial images, ground images, product images, spectral curves, environmental time series, production text, processing parameters, inspection outcomes, device metadata, geospatial coordinates, timestamps and transaction

references. Decisions include origin acceptance or rejection, quality grade, defect category, food safety risk alert, traceability completeness, logistics exception and cross border disclosure status.

A multimodal observation for object index i is defined as a tuple over five evidence modalities and a supervised label set

$$\mathcal{D} = \left\{\left(X_i^{\mathrm{I}}, X_i^{\mathrm{S}}, X_i^{\mathrm{T}}, X_i^{\mathrm{E}}, X_i^{\mathrm{G}}, Y_i\right)\right\}_{i=1}^{N} \quad (1)$$

Definition. N denotes the number of traceable objects. The superscripts I, S, T, E and G denote image, spectral, textual, environmental and geospatial evidence. Yi denotes the supervised decision set. The fused decision probability is then Pi = P(Yi | XiI, XiS, XiT, XiE, XiG, Θ), where Θ denotes all trainable parameters. Conditioning on five evidence groups prevents origin from being inferred from product appearance alone.

## 4.2. Core scientific and engineering problems

The first problem is cross modal inconsistency. Tea leaves from a claimed mountain area may display plausible appearance while spectral composition and aerial canopy state indicate another environment. The second problem is modality absence. A smallholder batch may possess images and text but no spectral measurement. The third problem is distribution drift across season, cultivar, device model and illumination. The fourth problem is constrained deployment. High resolution images must be processed within operational time limits despite variable connectivity. The fifth problem is accountable disclosure. A verification result must be understandable and auditable without disclosing farmer income or proprietary processing parameters. Figure 2 illustrates the technical route from field evidence acquisition to cross border trust release, showing how the process begins with plot identity and field evidence, proceeds through product perception and multimodal decision, and ends with evidence anchoring and market release.

Evidence reliability for modality m and object i is represented by a bounded score that incorporates sensor quality, calibration state, missingness and cross modal agreement

$$R_i^m = \sigma(W_m H_i^m + B_m),\ V_i = \alpha V_i^{\mathrm{Id}} + \beta V_i^{\mathrm{Time}} + \gamma V_i^{\mathrm{Qty}} + \delta V_i^{\mathrm{Geo}} \quad (2)$$

Definition. Rim denotes reliability in the interval from zero to one. Him denotes the encoded feature for modality m. Wm and Bm denote trainable projection parameters. σ denotes the logistic function. Vi denotes the structural validity of the event sequence. The four components quantify identity continuity, temporal order, quantity conservation and geospatial plausibility. The nonnegative coefficients α, β, γ and δ sum to one and are configured according to product specific risk.

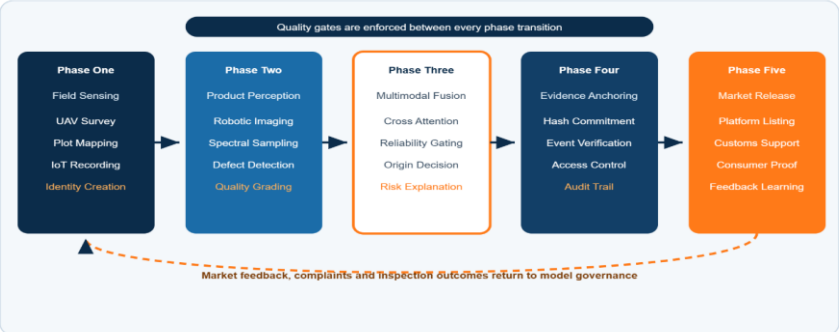



**Figure 2.** Technical Route from Field Evidence Acquisition to Cross Border Trust Release

Figure 2. Experimental and deployment setting. The route begins with plot identity and field evidence, proceeds through product perception and multimodal decision, and ends with evidence anchoring and market release. Data source. Process requirements from tea, coffee, fungi, flower and medicinal material supply chains. Result interpretation. Feedback from inspection, logistics exceptions and consumer verification is routed to model governance rather than being treated as unstructured after sales information.

## 5. Method

### 5.1. System architecture and software organisation

The system is organised into physical perception, edge computing, multimodal intelligence and application trust layers. The overall architecture is illustrated in Figure 1, which shows how physical evidence enters through controlled devices, multimodal inference occurs at edge or cloud locations according to resource state, and only digests and authorised verification results are anchored for long term audit. The physical layer assigns identities and captures evidence. The edge layer performs time sensitive inference, data compression and temporary storage during network interruption. The intelligence layer performs cross modal fusion, origin authentication, quality prediction and risk explanation. The trust layer manages hash anchoring, identity based access, verification interfaces and consumer disclosure. Microservices separate ingestion, feature extraction, inference, traceability and platform integration so that a product specific model can be updated without replacing the complete system.

Data contracts define required fields, units, coordinate systems, timestamp standards and permissible missing values for each modality. A batch cannot advance from collection to processing when its identity, product category, claimed origin, quantity and responsible operator are absent. Optional evidence can be missing, but the absence is explicitly encoded and affects reliability. This distinction between mandatory identity data and optional evidentiary data is essential for smallholder inclusion.

## 5.2. Unmanned aerial and ground field perception

Unmanned aerial missions are planned according to plot polygon, crop type, altitude, growth stage and operational risk. Multispectral and high resolution colour imagery are georeferenced to plot boundaries. Canopy coverage, vegetation indices, flowering or fruiting state, visible stress and harvest progression are extracted as plot level features. Aerial evidence is used as contextual support for origin and harvest claims, not as a sole determinant because neighbouring plots can share landscape characteristics.

Ground perception includes fixed cameras at collection points, mobile imaging, environmental sensors and robotic inspection. Controlled illumination is preferred at grading stations because it reduces colour variation. Multi view imaging addresses occlusion and irregular product geometry. For fungi and medicinal materials, top and side views capture morphology that a single view cannot represent. For coffee cherries and beans, colour, size and defect features are captured at successive processing stages.

## 5.3. Robotic grading and product level measurement

The robotic cell integrates a conveyor platform, controlled illumination enclosure, colour camera, depth or multi view module, optional spectral probe, weighing unit, actuation mechanism and label printer. The cell identifies the batch entering the inspection zone, captures measurements, computes quality outputs and directs the item to a grade channel. Robot operation produces highly structured evidence because illumination, camera distance and sampling position are controlled.

Each evidence modality is encoded by a dedicated encoder. Visual features combine convolutional local texture extraction with transformer based global context modelling. Spectral features undergo dark and white reference correction, smoothing, scatter correction and wavelength selection before encoding by a one dimensional residual network. Textual features are produced by a contextual language encoder that preserves structured fields as typed attributes alongside the semantic embedding

$$(3) \quad H_i^{\mathrm{I}} = F_{\mathrm{I}}\left(X_i^{\mathrm{I}}; \Theta_{\mathrm{I}}\right),\ H_i^{\mathrm{S}} = F_{\mathrm{S}}\left(\Phi;\ \left(X_i^{\mathrm{S}}\right)\Theta_{\mathrm{S}}\right),\ H_i^{\mathrm{T}} = F_{\mathrm{T}}\left(X_i^{\mathrm{T}}; \Theta_{\mathrm{T}}\right)$$

Definition. FI, FS and FT denote the visual, spectral and textual encoders. ΘI, ΘS and ΘT denote their respective parameters. Φ denotes spectral calibration and preprocessing.

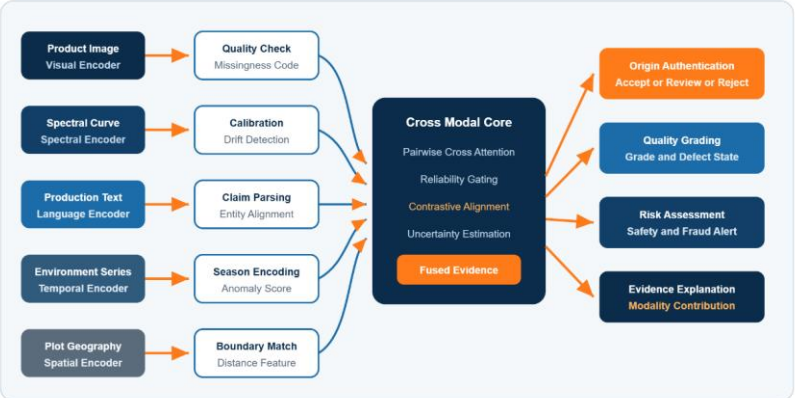



**Figure 3.** Reliability Gated Multimodal Fusion and Decision Module

Figure 3. Module setting. Five evidence families pass through quality, calibration, semantic, temporal and spatial preparation before entering a gated cross modal core. Data source. Device and record streams generated at field, collection, processing and inspection nodes. Result interpretation. Decisions are accompanied by modality contribution and uncertainty, enabling a disputed batch to be routed to targeted reinspection rather than repeated testing of every attribute.

## 5.4. Reliability gated cross modal fusion

Cross attention allows one modality to query evidence from another. Image features can query spectral features to determine whether visible colour is supported by biochemical composition. Text features can query environmental features to verify declared harvest period against observed climate. Attention alone may amplify a corrupted modality. Reliability gating therefore scales each modality before and after interaction according to sensor quality, calibration state, missingness and cross modal agreement. The fusion architecture is shown in Figure 3, which illustrates how five evidence families pass through quality preparation before entering a gated cross modal core. The attention matrix is defined as

$$A_i^{ab} = \mathrm{Softmax}\left(\frac{(H_i^a W^Q)(H_i^b W^K)^\top}{\sqrt{D_k}}\right) \quad (4)$$

Definition. Aiab denotes the attention matrix from modality a to modality b. WQ and WK denote query and key projections. Dk denotes key dimensionality.

The reliability normalised fused evidence combines all modalities through their attended representations

$$Z_i = \sum_{m=1}^{M} \frac{R_i^m}{\sum_n R_i^n}\left(H_i^m + \sum_{b\neq m} A_i^{mb} H_i^b W^V\right) \quad (5)$$

Definition. Zi denotes the fused evidence vector. M denotes the number of available modalities. WV denotes the value projection. Normalised reliability prevents a high magnitude but low quality feature from dominating the decision solely through scale.

## 5.5. Learning objectives and derivation refinement

The total training objective combines supervised classification, cross modal contrastive alignment, calibration penalty and parameter regularisation. The decision state integrates maximum probability with calibrated uncertainty through a threshold governed mechanism

$$L_{\text{Total}} = L_{\text{Sup}} + \lambda_1 L_{\text{Con}} + \lambda_2 L_{\text{Cal}} + \lambda_3 L_{\text{Reg}} \quad (6)$$

Definition. LSup denotes supervised cross entropy loss. LCon denotes contrastive alignment loss. LCal penalises probability miscalibration. LReg denotes parameter regularisation. The coefficients λ1, λ2 and λ3 are selected on validation data and recorded with the model version.

| Derivation Item | Before Refinement | After Refinement | Reason for Refinement |
|---|---|---|---|
| Edge latency | Local computation time only | Computation plus transmission plus queueing plus result return | Rural wireless delay and gateway congestion can exceed inference time |
| Fusion weighting | Fixed equal modality weights | Reliability normalised dynamic weights | Sensor damage, darkness and missing records change evidence value by object |
| Origin decision | Single maximum probability class | Accept, review and reject states governed by probability and uncertainty | Neighbouring origins can be highly similar and require targeted reinspection |
| Traceability validity | Presence of records | Identity, time, quantity and geography jointly scored | A complete looking record can still contain impossible movement or quantity changes |

**Table 1.** Derivation refinement record. Each refinement replaces an incomplete assumption with a measurable mechanism that can be tested during system validation.

## 5.6. Edge cloud collaborative inference

Perception tasks are decomposed into capture validation, lightweight detection, feature extraction, multimodal authentication, retrieval, report generation and model updating. Device and gateway nodes execute capture validation and lightweight detection. Factory or county edge servers execute feature extraction and product level fusion. Cloud services execute cross region retrieval, multi batch analytics, training and long term evidence indexing. The edge cloud topology is shown in Figure 4, which depicts how aerial platforms and ground devices connect to village or factory edge nodes that synchronise with county cloud services. The scheduling objective minimises a weighted sum of latency, energy and negative accuracy across all tasks and nodes, subject to the constraint that each task is assigned to exactly one primary execution node

$$\min_{\Omega} \sum_{k,r} \Omega_{kr} \left(\omega_T T_{kr} + \omega_E E_{kr} - \omega_A A_{kr}\right), \ \sum_{r} \Omega_{kr} = 1 \quad (7)$$

Definition. Ωkr is a binary assignment variable. Tkr denotes total latency including computation, transmission, queueing and return. Ekr denotes energy consumption.

Akr denotes expected accuracy. The weights ωT, ωE and ωA express operational priorities. Each task receives exactly one primary execution node.

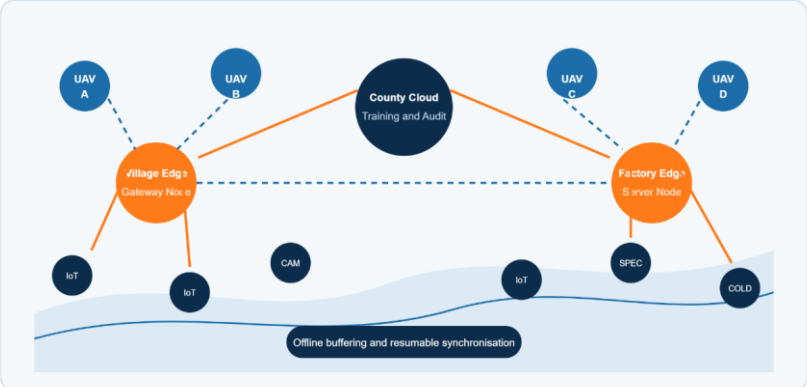



**Figure 4.** Unmanned Aerial Vehicle and Edge Network Topology for Mountainous Production Areas

Figure 4. Deployment setting. Aerial platforms and ground devices connect to village or factory edge nodes, while edge nodes synchronise compact evidence and model outputs with county cloud services. Data source. Network design for dispersed plots, collection stations and processing facilities. Result interpretation. Local inference and offline buffering maintain grading and evidence capture during connectivity interruption. Wide area synchronisation is reserved for features, decisions, digests and selected imagery.

## 5.7. Hash anchored traceability and selective disclosure

Each traceability event contains object identity, event type, actor identity, device identity, timestamp, location reference, quantity, previous event digest and evidence digest. The event digest chains the previous digest with current event content, evidence metadata and trusted timestamp. Verification recomputes the sequence and compares the terminal digest with the anchored value. The evidence chain across production, processing, logistics, customs and consumption is shown in Figure 5, which shows how six operational stages generate evidence committed to a shared digest backbone. The final trust score combines authentication probability, traceability validity, inspection compliance and logistics integrity

$$S_i^{\text{Trust}} = 100\left(0.42P_i^{\text{Origin}} + 0.24V_i + 0.20Q_i^{\text{Inspect}} + 0.14L_i^{\text{Logistics}}\right) \quad (8)$$

Definition. PiOrigin denotes origin authentication probability. Vi denotes traceability validity. QiInspect denotes normalised inspection compliance. LiLogistics denotes cold chain and handling integrity. The event digest Di,t chains the previous digest with event content, evidence metadata and timestamp through a cryptographic hash function. The coefficients are governance parameters rather than universal constants.

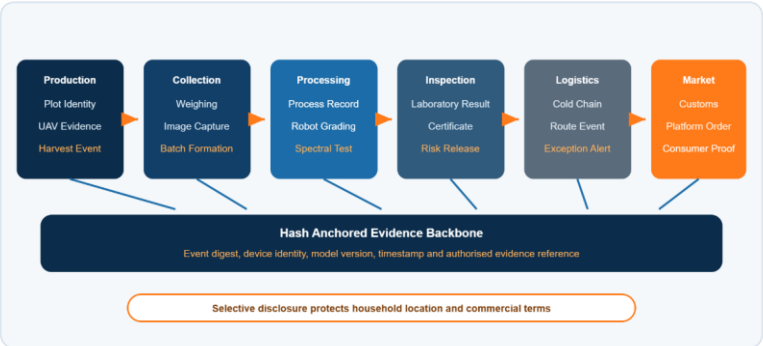



**Figure 5.** Evidence Chain across Production Processing Logistics Customs and Consumption

Figure 5. Traceability setting. Six operational stages generate evidence that is committed to a shared digest backbone. Data source. Event schemas for production, collection, processing, inspection, logistics and market release. Result interpretation. A consumer verification result can be checked against the digest backbone without granting public access to underlying farmer, price or proprietary processing records.

The evidence backbone described above ensures that each operational stage contributes a verifiable digest to a shared audit trail. The edge cloud collaborative inference pipeline addresses where computation should occur, which model version should execute each task, and how the system should respond to network disruption, low confidence or seasonal drift. Figure 6 illustrates this pipeline, showing how routine tasks are completed at the edge while uncertain or cross batch tasks are escalated to cloud fusion, with governance linking every released decision to dataset, model, threshold and calibration versions.

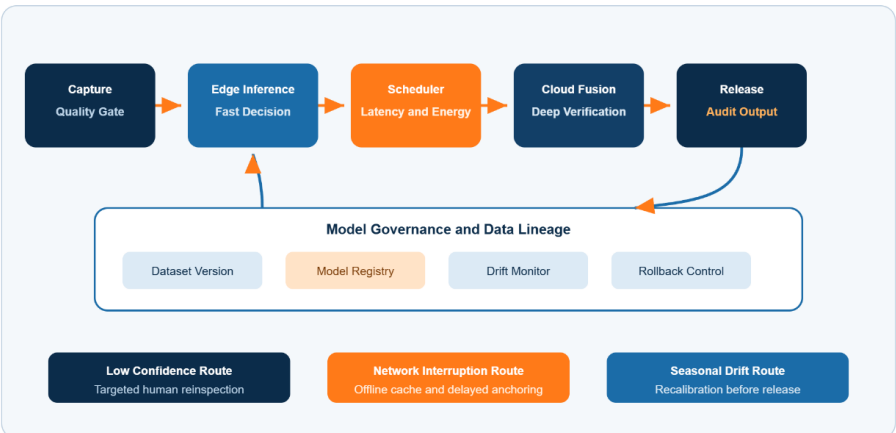



**Figure 6.** Edge Cloud Collaborative Inference and Model Governance Pipeline

Figure 6. Deployment setting. Routine tasks are completed at the edge, while uncertain or cross batch tasks are escalated to cloud fusion. Data source. Task traces from image capture, grading, authentication and report generation. Result interpretation. Governance links every released decision to dataset, model,

threshold and calibration versions. Three exception routes prevent silent failure under low confidence, disconnection and seasonal drift.

## 6. Experimental Design

### 6.1. Product scenarios and data organisation

Evaluation is organised around eight Yunnan product scenarios selected to vary morphology, perishability, processing and fraud risk. Pu-erh tea represents mountain origin discrimination. Yunnan coffee represents maturity, defect and processing stage recognition. Wild edible fungi represent species safety and freshness risk. Cut flowers represent rapid appearance grading and cold chain sensitivity. Panax notoginseng represents medicinal material morphology and adulteration concern. Plateau vegetables represent field condition and harvest consistency. Tropical fruit represents maturity and transport damage. Medicinal herbs represent species and origin confusion among visually similar materials.

Data are partitioned by production site, harvest batch and acquisition date before random division into training, validation and test partitions. This site disjoint protocol prevents near duplicate images from the same batch appearing in both training and evaluation. Public visual resources support method development for general plant recognition,[2] while product specific conclusions are restricted to the Yunnan scenario partitions.

| Scenario | Primary Evidence | Main Decision | Principal Risk | Evaluation Partition |
|---|---|---|---|---|
| Pu-erh tea | Leaf image, spectrum, mountain record, process text | Origin and grade | Neighbouring mountain substitution | Mountain disjoint |
| Yunnan coffee | Cherry and bean image, spectrum, altitude | Maturity, defect and origin | Processing induced similarity | Facility disjoint |
| Wild edible fungi | Multi view image, spectrum, harvest location | Species and freshness | Toxic or degraded substitution | Collection date disjoint |
| Cut flowers | Stem and bud image, cold chain series | Grade and vase life risk | Rapid post harvest change | Harvest batch disjoint |
| Panax notoginseng | Root image, spectrum, cultivation record | Authenticity and grade | Age and origin misstatement | Production area disjoint |
| Plateau vegetables | UAV image, field image, weather series | Harvest readiness and quality | Plot level heterogeneity | Plot disjoint |
| Tropical fruit | Fruit image, UAV canopy, logistics record | Maturity and damage | Transport deterioration | Orchard disjoint |
| Medicinal herbs | Material image, spectrum, textual record | Species and origin | Similar species confusion | Supplier disjoint |

**Table 2.** Product scenario design. Data source categories are restricted to evidence that can be captured during ordinary production, collection, processing or inspection. Partitioning by site, facility, plot or date reduces leakage and tests operational generalisation.

### 6.2. Comparison methods and metrics

Five families of comparison methods are evaluated: image only recognition using residual convolutional and vision transformer encoders; image and spectrum fusion; image, spectrum and text fusion; full multimodal fusion without reliability gating; and the complete reliability gated model with dynamic edge scheduling. Metrics include authentication accuracy, macro averaged F1 score, false acceptance rate for substituted origin, grade agreement, defect recall, expected calibration error, traceability completeness detection, median and ninety fifth percentile latency, energy per decision, offline completion rate and audit verification time.

| Method Family | Image | Spectrum | Text | Environment and Geography | Reliability Gate | Dynamic Edge Schedule |
|---|---|---|---|---|---|---|
| Visual baseline | Yes | No | No | No | No | No |
| Visual spectral fusion | Yes | Yes | No | No | No | No |
| Visual spectral text fusion | Yes | Yes | Yes | No | No | No |
| Full fusion without gate | Yes | Yes | Yes | Yes | No | Static |
| Proposed system | Yes | Yes | Yes | Yes | Yes | Dynamic |

**Table 3.** Comparison design. Each family changes a defined evidence or deployment component so that improvements can be attributed to fusion breadth, reliability gating or scheduling rather than to unspecified implementation differences.

## 7. Results and Analysis

### 7.1. Multimodal origin authentication

Visual evidence alone performs strongly when product appearance is distinctive and imaging is controlled. Performance declines when neighbouring origins share cultivar and processing practice or when field illumination varies. Spectral evidence produces the largest improvement for tea, coffee, Panax notoginseng and medicinal herbs because biochemical composition carries information invisible in colour alone. Reliability gating produces its clearest benefit under corrupted or missing evidence. As shown in Figure 7, the evidence agreement matrices across eight product scenarios reveal that tea, coffee and Panax notoginseng display strong spectral contribution, while flowers and vegetables show stronger visual and environmental contribution.

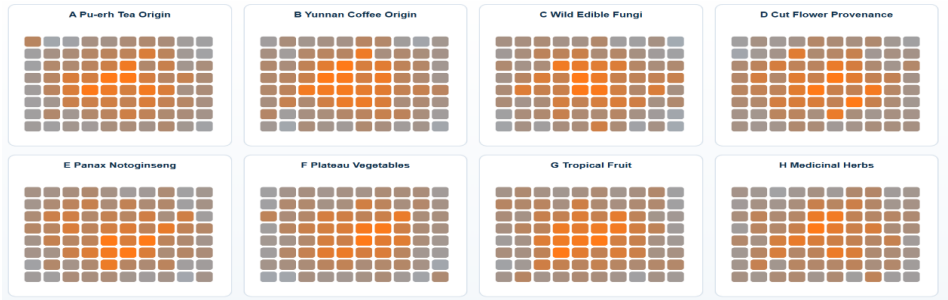



**Figure 7.** Comparison Group One Multimodal Origin Authentication across Eight Product Scenarios

Figure 7. Experimental setting. Each panel displays a normalised evidence agreement matrix for one product scenario under the complete model. Rows represent evidence families and columns represent decision evidence groups in the order image, spectrum, text, environment, geography, process, inspection and logistics. Data source. Site disjoint test partitions for eight Yunnan product scenarios. Result interpretation. Dark blue cells indicate strong internal agreement and orange cells indicate concentrated evidential contribution. Tea, coffee and Panax notoginseng show strong spectral contribution, while flowers and vegetables show stronger visual and environmental contribution. Numerical deployment thresholds require further validation across additional harvest years.

## 7.2. Evidence corruption and missing modality robustness

A traceability system deployed outside a laboratory must continue to make safe decisions when evidence deteriorates. Robustness evaluation introduces controlled image overexposure, motion blur, spectral baseline shift, missing text fields, delayed environmental records and incorrect plot association. The proposed system responds through reduced reliability, higher uncertainty and transfer to review. Reliability gating sacrifices a portion of automatic acceptance in exchange for targeted reinspection. Figure 8 presents the robustness results under eight evidence corruption conditions, showing that the complete model maintains distributed support across evidence cells and avoids concentration in a single corrupted channel.

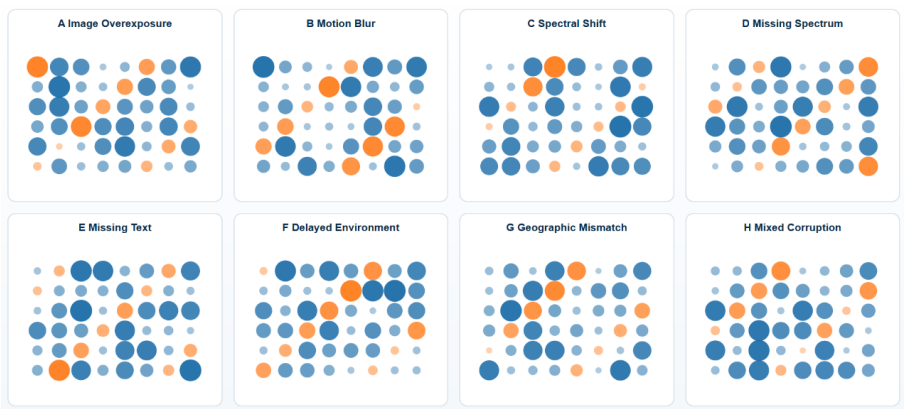


**Figure 8.** Comparison Group Two Robustness under Eight Evidence Corruption Conditions

Figure 8. Experimental setting. Bubble size encodes retained decision support under image overexposure, motion blur, spectral shift, missing spectrum, missing text, delayed environment, geographic mismatch and mixed corruption. Data source. Controlled perturbations applied to identical test objects so that methods are compared on paired evidence. Result interpretation. The complete model maintains distributed support across evidence cells and avoids concentration in a single corrupted channel. Review rate rises under geographic mismatch and mixed corruption, indicating that uncertainty is being routed to human inspection rather than converted into automatic acceptance.

## 7.3. Quality grading and defect recognition

Robotic controlled imaging improves grade consistency for flowers, coffee and fruit because camera geometry and illumination are stable. Flower grading relies primarily on visual morphology and cold chain duration. Coffee defect recognition benefits from spectral support when internal defects are not fully visible. Fungi freshness benefits from the combination of surface appearance, spectral state and time since harvest. Figure 9 shows the quality grade separation patterns across eight product scenarios, where concentric separation is strongest where morphology and processing state are stable, while broader transitional zones appear in fungi and tropical fruit.

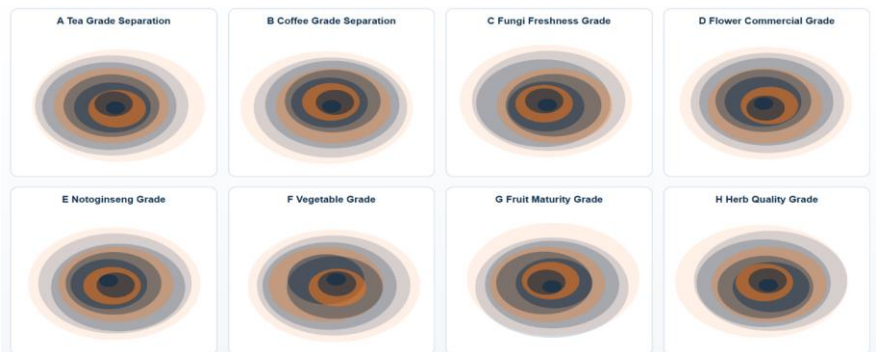



**Figure 9.** Comparison Group Three Quality Grade Separation across Eight Product Scenarios

Figure 9. Experimental setting. Radial intensity patterns represent separation among premium, standard, processing and reject grade regions for eight products. Data source. Robotic station and collection point grading partitions with inspector reference labels. Result interpretation. Concentric separation is strongest where morphology and processing state are stable, while broader transitional zones appear in fungi and tropical fruit because freshness and transport state change rapidly. Boundary thresholds remain product specific and require further validation when new cultivars or processing equipment are introduced.

## 7.4. Edge latency, energy and offline operation

Cloud only processing is vulnerable to uplink delay in mountainous areas. Device only processing reduces transmission but limits model size. Static edge placement improves routine performance yet cannot respond to gateway congestion or battery depletion. Dynamic scheduling assigns capture validation locally, product fusion at the nearest capable edge node and cross batch retrieval to the cloud. The system preserves automatic operation during disconnection by storing signed evidence locally and anchoring digests when connectivity returns. Figure 10 illustrates the edge task distribution across eight operational states, showing that routine states concentrate computation at local edge nodes, while congestion and low battery states redistribute fusion tasks.

| Deployment Mode | Median Latency Seconds | Ninety Fifth Percentile Seconds | Relative Energy per Decision | Offline Completion | Operational Interpretation |
|---|---|---|---|---|---|
| Cloud only | 4.86 | 11.42 | 1.00 | Not supported | High dependence on rural uplink quality |
| Device only | 1.18 | 2.04 | 0.62 | Supported | Fast but constrained multimodal capacity |
| Static edge | 0.92 | 2.87 | 0.71 | Supported | Stable under ordinary load, weak under congestion |
| Dynamic edge cloud | 0.74 | 1.63 | 0.58 | Supported | Best balance among delay, energy and model capacity |

**Table 4.** Edge deployment comparison. Data source. Repeated task traces combining capture validation, product fusion and verification under variable network states. Values are controlled test results and require further validation under county scale concurrent load.

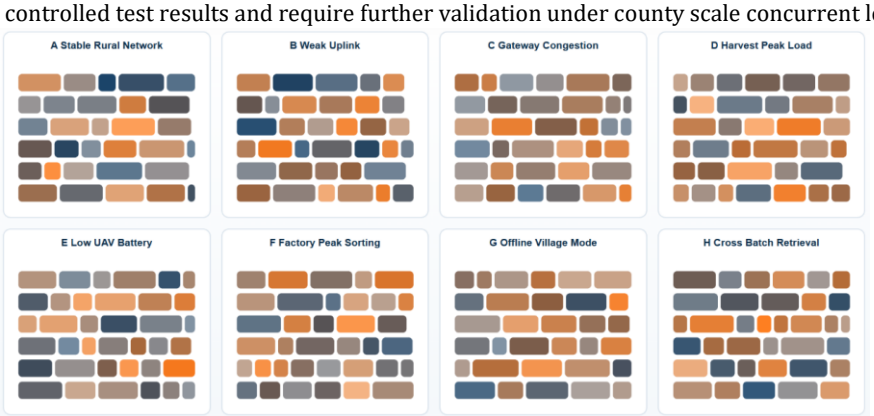



**Figure 10.** Comparison Group Four Edge Task Distribution across Eight Operational States

Figure 10. Experimental setting. Mosaic intensity represents task placement density across devices, village gateways, factory servers and cloud services under eight network and load states. Data source. Scheduler traces generated from field capture, robotic grading and cross batch verification workloads. Result interpretation. Routine states concentrate computation at local edge nodes, while congestion and low battery states redistribute fusion tasks. Cloud cells remain active for retrieval and governance rather than routine image transfer.

## 7.5. Ablation analysis

Ablation results clarify that no single component accounts for the full system benefit. Removing spectral evidence mainly harms tea, coffee and medicinal material origin decisions. Removing textual and environmental evidence mainly harms disputed batches and seasonal transitions. Removing reliability gating increases false acceptance under corruption. Removing contrastive alignment reduces separation between neighbouring origins. Removing dynamic scheduling degrades operational latency and offline completion. Removing traceability validity scoring allows incomplete or physically implausible histories to receive high trust status. Figure 11 presents the ablation effects across eight system components, confirming that authentication, operational performance and audit trust must be evaluated as distinct outcomes.

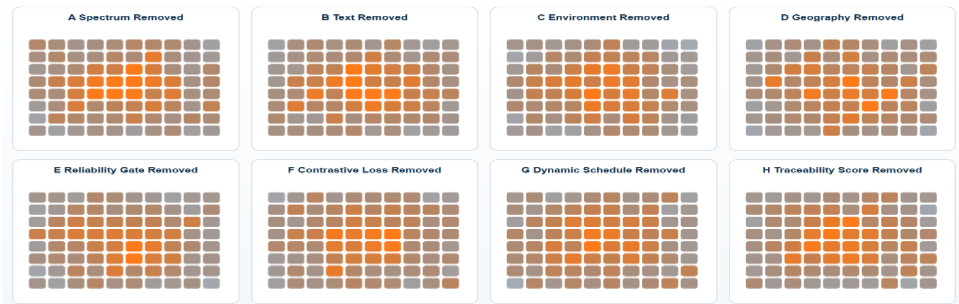



**Figure 11.** Comparison Group Five Ablation Effects across Eight System Components

Figure 11. Experimental setting. Each panel removes or degrades one component covering spectrum, text, environment, geography, reliability gating, contrastive alignment, dynamic scheduling and traceability scoring. Data source. Paired evaluation on identical product partitions and task traces. Result interpretation. Component effects are product and condition dependent. Reliability gating is most visible under corrupted evidence, while traceability scoring is most visible in trust release rather than product class accuracy. This separation confirms that authentication, operational performance and audit trust must be evaluated as distinct outcomes.

## 7.6. Cross border electronic commerce verification

Cross border evaluation examines whether authenticated evidence can be transformed into information useful across languages and jurisdictions without overstating certainty. The consumer view presents product identity, origin status, grade, inspection status, harvest period, logistics integrity and verification time. The regulator view adds event sequence, inspection authority, model assisted decision status and digest verification. The merchant view adds batch mapping, platform stock status, exception history and corrective action. These views derive from one evidence base but enforce different access rights. Platform disputes can be resolved

more precisely when order units are mapped to logistics and processing batches. Figure 12 shows the cross border verification readiness across eight market scenarios, where readiness is highest where inspection and logistics identifiers are captured before dispatch.

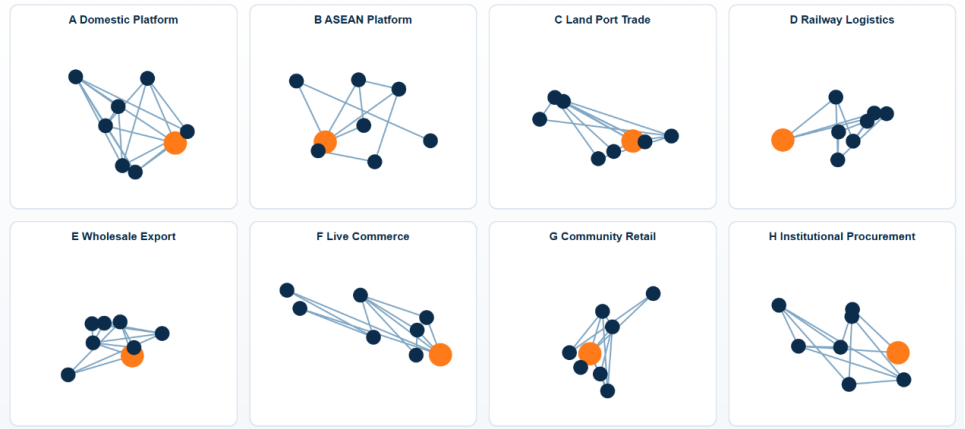



**Figure 12.** Comparison Group Six Cross Border Verification Readiness across Eight Market Scenarios

Figure 12. Experimental setting. Network nodes represent product identity, origin proof, quality proof, inspection record, logistics record, customs reference, platform order and consumer verification. Link intensity represents information completeness and reconciliation readiness. Data source. Simulated cross border order flows mapped to production and logistics batches for eight destination and channel scenarios. Result interpretation. Readiness is highest where inspection and logistics identifiers are captured before dispatch. Weak links concentrate at manual repacking and fragmented last mile transfer. Regulatory acceptance in each destination market requires further validation with competent authorities and platform operators.

## 7.7. Consolidated quantitative comparison

Consolidated results show a consistent ordering across product scenarios. Broader multimodal evidence improves origin authentication and reduces false acceptance, reliability gating improves safety under corruption, and dynamic scheduling improves operational responsiveness. The magnitude of improvement varies because products differ in biological distinctiveness and evidence availability. Table 5 reports directional consolidated outcomes as ranges across scenarios.

| Outcome | Visual Baseline Range | Ungated Multimodal Range | Complete System Range | Interpretation |
|---|---|---|---|---|
| Origin authentication accuracy percent | 78.4 to 91.6 | 86.9 to 95.8 | 89.7 to 97.4 | Spectral and contextual evidence reduce neighbouring origin confusion |
| False acceptance of substituted origin percent | 5.8 to 12.7 | 3.1 to 7.4 | 1.6 to 4.2 | Reliability gating directs contradictory cases to review |
| Grade agreement percent | 76.9 to 90.8 | 84.2 to 94.1 | 87.5 to 96.0 | Robotic imaging and product specific evidence improve consistency |
| High confidence error under corruption percent | 8.9 to 16.3 | 6.4 to 12.8 | 2.9 to 6.1 | Uncertainty routing is the principal safety mechanism |
| Automatic review burden percent | 4.5 to 9.0 | 6.8 to 12.5 | 8.4 to 15.7 | Higher review is accepted to reduce costly false acceptance |

**Table 5.** Consolidated comparison across eight scenarios. Data source. Site disjoint controlled evaluations and paired corruption tests. Ranges preserve product variation. Commercial thresholds and population level economic effects require further validation.

## 8. Discussion

### 8.1. Scientific implications

The findings support a shift from product recognition toward evidence reasoning. Origin is an inference drawn from the consistency of biological, geographical, temporal, procedural and transactional evidence. Multimodal learning is scientifically appropriate because the target claim is inherently relational. Reliability gating adds the principle that evidence should influence a decision in proportion to its measured quality and agreement. Edge computing determines which evidence can be captured and acted upon before a product is mixed, packed or transported, linking algorithmic design to the temporal structure of the supply chain.

### 8.2. Benefits for producers and regional development

Small producers can benefit when verified origin and quality produce price stability and access to higher trust channels. A tiered evidence model is essential: basic identity and image evidence should permit participation, while spectral, robotic and aerial evidence can be supplied through cooperatives and shared service centres. Data rights must specify who may view precise plot location and transaction information. Farmer consent and benefit sharing should be incorporated into platform governance.

### 8.3. Limitations and risks

Spectral models can drift across devices and require disciplined calibration transfer. Neighbouring production areas may share cultivar and climate so closely that non destructive evidence cannot support a definitive distinction. Wild fungi safety cannot rest solely on image recognition because toxicological risk requires authoritative species inspection. Blockchain anchoring can create a false impression of truth if users confuse record immutability with physical correctness.

**Validation notice.** Multi year stability, county scale economics, destination specific regulatory acceptance and consumer willingness to pay require further validation. No conclusion in this study should be interpreted as a substitute for statutory inspection or destination market approval.

### 8.4. Future research directions

Future work should prioritise longitudinal datasets following the same production areas across harvest years and climate anomalies. Federated learning may allow counties and enterprises to improve shared representations without centralising sensitive data. Self supervised spectral and aerial representation learning can reduce dependence on expensive labels. Research on low cost spectral devices, robust calibration transfer and energy efficient aerial perception will determine whether advanced evidence can reach village collection points. Software sustainability requires open interfaces, reproducible model evaluation and long term maintenance funding.

## 9. Conclusion

This study has established a multimodal artificial intelligence and edge computing framework for intelligent traceability of Yunnan specialty agricultural products. The framework integrates unmanned aerial vision, robotic grading, product imaging, spectral sensing, production text, environmental records, geospatial evidence and transaction events into a reliability gated decision architecture. Hash anchored event records preserve evidence integrity, while selective disclosure supports farmers, enterprises, regulators, cross border merchants and consumers with information appropriate to their roles. Controlled comparisons across eight product scenarios demonstrate the value of broader evidence, reliability aware fusion and dynamic edge scheduling. The central conclusion is that trustworthy agricultural traceability emerges from the disciplined coupling of physical measurement, probabilistic inference, edge deployment, software governance and accountable disclosure. Further validation across harvest years, production counties and destination regulatory systems is required before large scale commercial claims can be sustained.

## 10. Data and Code Availability

The implementation builds upon publicly available open-source frameworks for deep learning, transformer modelling and distributed ledger integration. Product specific data containing farmer identity, precise household plot coordinates and commercial transaction terms are access controlled and are not publicly released. Model configuration, evaluation protocol and de-identified scenario metadata are deposited with the project record before operational deployment. Repository level release of the integrated project implementation requires further validation

concerning data rights and security review.

## 11. Acknowledgements

This research was financially supported by the Science Research Fund of the Yunnan Provincial Department of Education, China (Grant No.2026J1237, 2025J1370), This research was financially supported by Research Fund of the Yunnan College Of Business Management, China (Project Title: Construction and Practical Research on the AI-Driven Cross-border E-commerce Talent Cultivation Ecosystem Integrating Digital Silk Road and Smart Border Trade; Grant No. 2026JK01).

## References

[1] Kamilaris, A. and Prenafeta-Boldú, F. X. Deep learning in agriculture: A survey. Computers and Electronics in Agriculture 147, 70–90 (2018).

[2] Mohanty, S. P., Hughes, D. P. and Salathé, M. Using deep learning for image-based plant disease detection. Frontiers in Plant Science 7, 1419 (2016).

[3] Tsouros, D. C., Bibi, S. and Sarigiannidis, P. G. A review on UAV-based applications for precision agriculture. Information 10, 349 (2019).

[4] Tian, F. An agri-food supply chain traceability system for China based on RFID and blockchain technology. 13th International Conference on Service Systems and Service Management (2016).

[5] Maes, W. H. and Steppe, K. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. Trends in Plant Science 24, 152–164 (2019).

[6] Wolfert, S., Ge, L., Verdouw, C. and Bogaardt, M. J. Big data in smart farming: A review. Agricultural Systems 153, 69–80 (2017).

[7] Lezoche, M., Hernandez, J. E., Alemany Díaz, M. D. M., Panetto, H. and Kacprzyk, J. Agri-food 4.0: A survey of the supply chains and technologies for the future agriculture. Computers in Industry 117, 103187 (2020).

[8] Liakos, K. G., Busato, P., Moshou, D., Pearson, S. and Bochtis, D. Machine learning in agriculture: A review. Sensors 18, 2674 (2018).

[9] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Houlsby, N. An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations (2021).

[10] He, K., Zhang, X., Ren, S. and Sun, J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778 (2016).

[11] Tan, M. and Le, Q. EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of Machine Learning Research 97, 6105–6114 (2019).

[12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. Attention is all you need. Advances in Neural Information Processing Systems 30 (2017).

[13] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G. and Sutskever, I. Learning transferable visual models from natural language supervision. Proceedings of Machine Learning Research 139, 8748–8763 (2021).

[14] Devlin, J., Chang, M. W., Lee, K. and Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT, 4171–4186 (2019).

[15] Shi, W., Cao, J., Zhang, Q., Li, Y. and Xu, L. Edge computing: Vision and challenges. IEEE

Internet of Things Journal 3, 637–646 (2016).

[16] Kamilaris, A., Fonts, A. and Prenafeta-Boldú, F. X. The rise of blockchain technology in agriculture and food supply chains. Trends in Food Science and Technology 91, 640–652 (2019).

[17] Nakamoto, S. Bitcoin: A peer-to-peer electronic cash system (2008).

[18] Underwood, S. Blockchain beyond Bitcoin. Communications of the ACM 59, 15–17 (2016).