STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–25.
Published online in International Academic Press (www.IAPress.org)

RepViT-M1.1 Based Classification of Arabica Coffee Varieties with Blockchain-Enabled NFT Certification and On-Chain SHA-256 Anti-Duplication on Polygon Network

Zainur Rasyid Ridlo [1*], Dafik [2,5], Silvi Putri Ayu Ningsih [1], Rina Sugiarti Dwi Gita[3], D Salangai Nayagi[4,5]

[1]*Department of Science Education, Faculty of Education, University of Jember, Jember, Indonesia*
[2]*Department of Mathematics, University of Jember, Jember, Indonesia*
[3]*Department of Learning Technology, Universitas PGRI Argopuro Jember, Jember, Indonesia*
[4]*Galgotias University, Greater Noida, Uttar Pradesh, India*
[5]*PUI-PT CGANT, University of Jember, Jember, Indonesia*

Abstract This study aims to develop a system that integrates CNN-based classification with blockchain technology to authenticate Arabica coffee varieties. The system employs RepViT-M1.1, a convolutional network redesigned from a vision transformer perspective to classify six categories: five Arabica varieties (Peaberry, Natural Ijen, Anaerobic Carbonic, Orange Bourbon, and Blue Mountain) and a specific "Non-Coffee" class designed to reject out-of-distribution (OOD) inputs. The classifier was trained on 12,000 images collected in Bondowoso, East Java, using a 70/15/15 data split. Accepted predictions are certified as ERC-721 standard NFTs on the Polygon Amoy testnet via a permissionless, bilingual (Indonesian and English) decentralized application (dApp). A smart contract manages an on-chain SHA-256 hash registry that rejects image hashes already recorded at the consensus layer; the process of linking the hash to the photograph occurs client-side and is noted as a system limitation. The classifier achieved a test accuracy of 99.28%, and the implemented rejection rules successfully rejected 98.00% of non-coffee images while accepting 96.93% of genuine coffee images. All 30 duplicate resubmissions were rejected by the on-chain registry, and the certification process took an average of 14.81 ± 2.04 seconds based on 30 trials covering hashing, inference, IPFS upload, and on-chain confirmation excluding the time required to transfer images to the inference endpoint. In benchmark comparisons against six baseline networks, RepViT-M1.1 delivers competitive accuracy alongside integrated OOD detection-based rejection capabilities and acceptable latency. These results demonstrate that classification incorporating OOD detection and an on-chain anti-duplication mechanism can be combined into an accessible certification workflow; as a proof of concept, the system provides a framework that can be further developed for the large-scale implementation of coffee certification.

Keywords Arabica coffee classification; RepViT; Blockchain; NFT certification; OOD detection; Agricultural traceability

DOI: 10.19139/soic-2310-5070-4372

1. Introduction

Indonesia is the fourth largest coffee producer in the world and one of the largest contributors to coffee production in the world. One of the most famous types is Arabica coffee from the Ijen Plateau in Bondowoso, East Java, which is highly appreciated internationally for its characteristic sensory profile: flowery scents, balanced acidity and a rich taste [1]. For many smallholders coffee is not only a source of money, it is an inheritance [2]. But it's difficult to get premium beans via the supply chain, and one of the most difficult processes is to verify the reported variety

*Correspondence to: Zainur Rasyid Ridlo (Email: zainur.fkip@unej.ac.id). Department of Science Education, Faculty of Education, University of Jember, Jember, Indonesia.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 202x International Academic Press

of a lot. The appearance of Arabica cultivars is often so similar that variety mislabeling and the substitution of cheaper beans, whether intentional or not, usually go undetected by traditional scrutiny [3]. There are real costs: loss of market transparency, lost consumer trust, loss of export competitiveness, and reduced returns for rural communities. The visual assessment of green beans by expert graders is still the main way to verify the declared variety of a coffee lot, a manual procedure that has encouraged the development of computer-vision systems for green bean inspection and grading [4, 5]. Once a supply chain goes through several hands it is hard to keep things consistent, as such decisions are dependent on the individual grader. Deep learning has created new opportunities for image analysis in agriculture. Plant leaves can now be classified and crop disease detected with excellent accuracy using convolutional neural networks (CNNs) and vision transformers (ViTs) [6–8]. However, coffee is an exception. Most of the computer-vision research has focused on defect identification, roast level and overall bean quality, with very little attention to the discrimination of visually identical Arabica varieties [9]. Classification of coffee from bean images has only been recently investigated [10]. A second, equally consequential, drawback is that most models assume that all inputs are from one of the pre-defined training classes, and hence fail on the out-of-distribution (OOD) samples that occur in actuality [11, 12]. In contrast, blockchain is widely employed to make agricultural supply chains traceable, but few studies combine such certification with explicit OOD detection to ensure that only trustworthy forecasts are committed to an immutable, shared record.

Overall, these gaps indicate a lack of integration of the different components. High-resolution categorisation, out-of-distribution (OOD) detectors and blockchain traceability do exist, but they have all been investigated separately and no system has incorporated them into a unified process for coffee verification, to the best of our knowledge. Without the use of OOD rejection, highresolution classification is guarantyd to misclassify non-coffee inputs. OOD-aware models cannot stop double or fraudulent certification because their predictions are not coupled to immutable and tamper-resistant records. Application-layer uniqueness-enforcing blockchain ledgers can be evaded via direct contract calls. This convergence is what inspires this research, and its innovation is not in any single component but its integration into an accessible, OOD-aware and anti-duplication certification system for Indonesian Arabica coffee.

To overcome this gap, we offer an end-to-end system combining artificial intelligence with blockchain to certify Arabica coffee bean varietals, which makes five main contributions. First, our framework uses RepViT-M1.1 for fine-grained classification of Indonesian Arabica varieties, an architecture that, to the best of our knowledge, has not been used for coffee before. Secondly, this approach extends the six-class model with an explicit ‘Non-Coffee’ class, expressly rejecting non-coffee inputs (out-of-distribution, OOD). Thirdly, the framework proposes an ERC-721 smart contract with a SHA-256 hash registry and a permissionless open-mint design that guarantees the uniqueness of hashes at the blockchain consensus layer, stopping the re-registration of same image hashes. Fourth, this research provides an installation-free, bilingual (Indonesian and English) decentralised application (dApp) that is cross-device and emphasises the cultural uniqueness of Ijen Arabica coffee. Fifth, we analyse the framework with an OOD ablation study, a comparison against six regularly used baselines, anti-duplication and perceptual hash tests, and end-to-end latency measurements, including field trials with farmers. In this research, no new CNN architecture or blockchain protocol is proposed, instead, an integrated application system is constructed as a proof of concept.

2. Related Work

2.1. Deep Learning for Coffee Bean Classification

Deep learning has made rapid progress in agricultural image classification [13, 14]. Convolutional models now grade bean quality and detect defects, one study compared five deep networks for coffee-bean species recognition and found InceptionV3 the most accurate at 93% [15], while another developed a model specifically for classifying seventeen defect categories in Thai Arabica green beans [16]. Transfer learning has likewise proven effective for fine-grained agricultural imagery [17]. A structured search of Scopus, IEEE Xplore, and Web of Science (2015–2026), pairing lightweight vision transformers and coffee or Arabica classification with out-of-distribution detection and blockchain certification, shows these capabilities have largely been studied in isolation rather than

combined in one authentication pipeline. Most of these systems cannot explicitly reject out-of-distribution (OOD) inputs [18, 19], we address it by training a dedicated Non-Coffee class on 2,000 heterogeneous non-coffee images [20]. Related studies in coffee and precision agriculture point in the same direction, ranging from microclimate forecasting in coffee agroforestry [21] to soil-moisture prediction for vertical farming [22] and machine-learning-based smart-agroforestry tools [23]. Deep learning has also been mapped as a pedagogical framework in science education, where it intersects computational thinking [24].

2.2. Blockchain for Agricultural Traceability

Blockchain has diffused rapidly across agricultural supply chains [25, 26]. The deployed systems, however, share a recurring weakness, where duplicate prevention exists it is enforced at the application layer through database or API checks and can therefore be bypassed by calling the smart contract directly [27–29]. Survey evidence indicates that only a small fraction of agricultural blockchain deployments achieve full multi-stakeholder integration [30–32], few embed an on-chain hash registry that enforces record uniqueness at the consensus layer. This work addresses that gap by storing SHA-256 fingerprints directly in smart-contract state [33–35], it converts duplicate rejection from an application-level convention into a consensus-layer guarantee [36–38].

2.3. NFT Applications in Agriculture

Certifying agricultural produce with NFTs has attracted growing scholarly interest [39]. Recent work extends ERC-721 NFTs to agricultural and food traceability, including a non-fungible-token and blockchain scheme for cotton-lint traceability [40] and NFT-based certification frameworks for sustainable agriculture [41, 42]. A 2026 review goes further still, positioning NFT-based food certification as a cornerstone of emerging Web 3.0 food-safety architecture [43]. Across these works, however, NFT certification, image-based classification, and out-of-distribution rejection are typically addressed separately rather than within one system. Consistent with the search reported in Section 2.1, we did not find a single system for authenticating coffee varieties that combines three capabilities: CNN-based fine-grained classification, an explicit six-class OOD-aware architecture, and on-chain hash verification.

3. Methods

3.1. System Architecture

The proposed system integrates deep-learning image classification with blockchain infrastructure to form a tamper-resistant, end-to-end certification pipeline for Indonesian Arabica coffee. Conventional variety verification relies on isolated expert judgement or standalone models; this architecture does not. It operates as a seven-layer chain, from image capture to immutable on-chain inscription, and remains usable by smallholders. The design was prompted by three coupled weaknesses of existing certification schemes. The first is the absence of any out-of-distribution (OOD) rejection, which allows closed-set classifiers to misclassify non-coffee objects with high confidence. The second is exposure to duplicate certification, whereby the same image can be registered multiple times. The third is a significant participation barrier imposed by complex wallet configuration and installation-dependent applications. Each layer addresses one or more of these weaknesses through deep-learning inference, cryptographic hashing, or decentralised consensus.

Figure 1 illustrates a seven-layer architecture: (1) image upload via dApp, (2) classification process using a CNN with ‘Non-Coffee’ out-of-class (OOD) rejection, (3) photo storage on IPFS, (4) creation of NFT metadata, (5) SHA-256 anti-duplication verification against the on-chain hash registry, (6) minting of an ERC-721 NFT, and (7) confirmation on the Polygon Amoy blockchain.

Layer 1: Image acquisition via dApp. Certification begins in the dApp, a browser-based client built with Next.js 14 and React 18 and deployed on the Vercel Edge Network. Because it requires no installation, farmers reach the system through any standard web browser [44, 45]. When an image is selected from the gallery or captured with the camera, the Web Cryptography API (SubtleCrypto.digest) computes a SHA-256 fingerprint of the raw

file buffer, a unique 256-bit identifier that persists through the remainder of the pipeline. This hash is submitted to the contract's cekHashFoto() function via a read-only JSON-RPC call, confirming that the image has not been certified before and sparing the cost of running inference on a duplicate. Layer 2: OOD-aware CNN classification

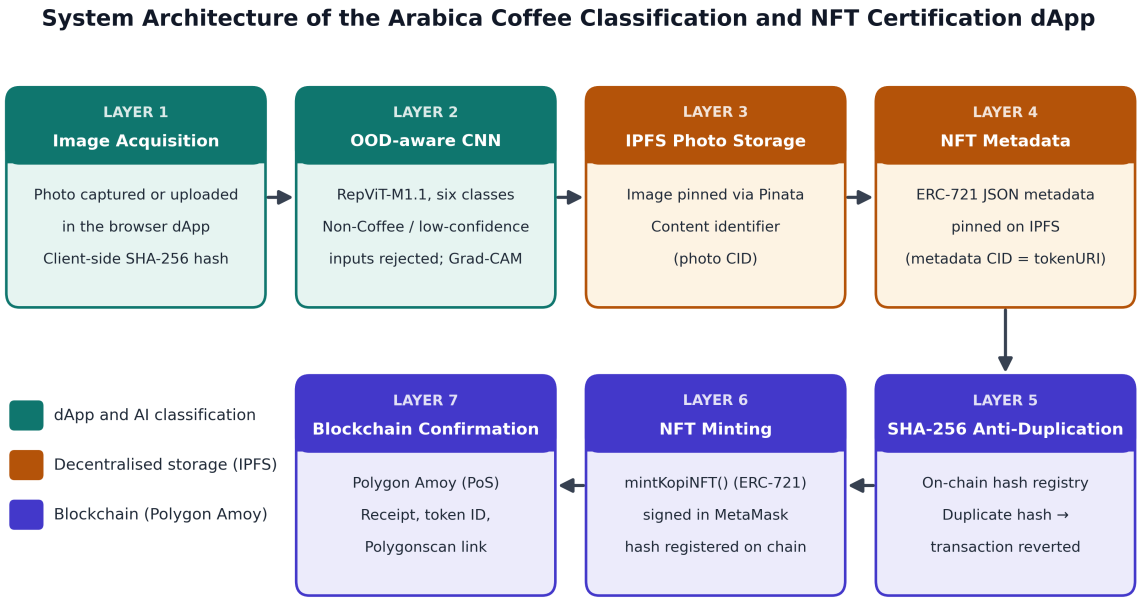


Figure 1. Architecture of the Arabica Coffee Classification and NFT Certification System.

(RepViT-M1.1, six-class). The image is sent to a CNN inference endpoint deployed on Hugging Face Spaces and accessed using the Gradio API. RepViT-M1.1 is a mobile-efficient convolutional network compacted through structural re-parameterization [46,47]. The endpoint resizes the input to 224×224 pixels and returns a six element softmax vector over five varieties of Arabica, Peaberry, Natural Ijen, Anaerobic Carbonic, Orange Bourbon and Blue Mountain, and one Non-Coffee class. The argmax of this vector is the prediction. If the detected class is not a coffee bean, based on the calibrated threshold value $\tau = 0.7957$ (Eq. 10), the processing flow will stop and the system will indicate that it is not a coffee bean. This filtering prevents fraudulent or accidental non-coffee entries from reaching the blockchain stage, a failure scenario often demonstrated by closed-set classification by providing a high confidence level for inputs outside the distribution [48,49]. In addition to predictions, the endpoint computes Grad-CAM saliency maps that highlight the regions of the image that contribute most to the predicted class, whilst also attaching a visual explanation to each image that has undergone the certification process. This enables farmers and auditors to verify that the model is focusing on the coffee beans, rather than the background. [50].

Layer 3: Decentralised photo storage using IPFS (Pinata). After the classification, the original image is uploaded to the InterPlanetary File System (IPFS) via Pinata. IPFS is content addressed, with a Content Identifier (CID) for each file being a cryptographic hash of its bytes [51, 52].This provides built-in tamper detection: any subsequent change will result in a different CID, corrupting the reference stored in the NFT's metadata and revealing the change during verification. Therefore, the content-based addressing scheme maintains the integrity of the saved images, but the long-term availability depends on the content remaining pinned, instead of being automatically distributed by IPFS [53].

Layer 4: NFT metadata construction and IPFS upload. A Next.js API route (/api/upload-ipfs) assembles a JSON metadata document in the ERC-721 format that records the complete certificate: photo CID, predicted variety, the variety identification label, prediction confidence, farmer name, farm location, SHA-256 fingerprint, and Unix timestamp. Uploading this document to IPFS via Pinata yields a second, metadata CID that serves as the NFT tokenURI. Storing the photo (Layer 3) and metadata (Layer 4) as separate IPFS objects follows the ERC-721

specification and enables independent verification: auditors confirm image integrity through the photo CID and certificate integrity through the metadata CID, without accessing the application layer [54].

Layer 5: On-chain SHA-256 anti-duplication. Before minting, the contract's cekHashFoto() view function queries the _hashFotoRegistry mapping, a Solidity mapping(string ⇒ uint256) that indexes the token identifier of every certified image by its SHA-256 hash [55,56]. A non-zero result signals prior certification, and the conflicting token ID is returned to the interface. The definitive rejection, however, occurs on-chain: mintKopiNFT() itself reverts with a FotoDuplikat() error before any state change is committed. This two-stage design, with application-side pre-screening in Layer 1 followed by on-chain enforcement in Layer 5, guarantees at the consensus layer that no two NFTs can register the same SHA-256 string, regardless of the calling interface, a guarantee that application-layer checks alone cannot provide.

Layer 6: NFT minting via ERC-721 smart contract. The mintKopiNFT() function of the KopiArabikaNFT contract accepts nine arguments: farmer wallet address, photo CID, metadata tokenURI, variety string, variety identification label string, farmer name, farm location, confidence as uint256, and the SHA-256 hash string [57,58]. It atomically executes three operations: it registers the SHA-256 hash in _hashFotoRegistry to block future duplicates, mints a new ERC-721 token to the farmer's wallet, and stores every certificate field in the contract for on-chain retrieval. The transaction is signed locally with an ECDSA key in the farmer's MetaMask wallet, so the private key never leaves the device. The gas price is set to at least 30 Gwei to ensure timely inclusion (30 Gwei in all evaluation trials).

Layer 7: Layer 7: Blockchain confirmation and immutable finalization (Polygon Amoy). Signed transactions are published to Polygon Amoy (Chain ID: 80002) via the Alchemy JSON-RPC endpoint. Under Amoy's Proof-of-Stake consensus mechanism, validators confirm and include the transaction in a block; the application waits for a transaction receipt confirming block inclusion and state commitment before displaying the result. Subsequently, the application extracts the Token ID from the ERC-721 Transfer event within the receipt and displays it alongside the transaction hash, the IPFS photo URL, and a Polygonscan link. The resulting record is publicly auditable and immutable by any party, including the deployer [59–61], this record can only be revoked via a dispute function restricted to the owner and persists independently of the web application.

This system guarantees three digital facts recorded at the time of NFT minting. (1) The Arabica variety predicted by the model and the confidence level for a specific photograph, (2) the integrity of the metadata linked via the photograph's SHA-256 hash, (3) the uniqueness of that hash within the blockchain.

3.2. Dataset Collection and Preprocessing

This study utilised five classes of Arabica coffee beans based on five commercial grades, namely Arabica Orange Bourbon, Arabica Blue Mountain, Arabica Peaberry, Arabica Natural Ijen and Arabica Anaerobic Carbonic, all sourced from the Ijen Highlands (Bondowoso, East Java). The training/validation/test data were allocated in a 70%/15%/15% ratio. The data utilised in this study are detailed in Table 1. Variety labelling is determined on the basis of origin, not subjective visual assessment; each photographed batch is sourced directly from coffee producers whose beans are known to be of the Arabica variety, and each image reflects the variety stated on its original bat

Before splitting, the image collection was manually screened to guarantee that only one set contained identical or visibly near-duplicate photographs of the same seed . Thus, in the 70/15/15 split, no image in the training or validation folds was a duplicate or near duplicate of any image in the test set. This filtering prevents the presence of almost duplicate photos that can artificially boost test accuracy [62]. All images were acquired under controlled environmental conditions to ensure consistency across samples. Using a calibrated multi-function environment meter, illuminance was held between 2942 and 3157 lux, ambient temperature between 27.6 °C and 31.7 °C, and relative humidity between 75% and 80%. Standardising these conditions stabilises colour rendition and minimises photometric variability between images.

The non-coffee class consisted of 2,000 photos from two complementary sets. The first group was made of common items from CIFAR-100 (a subset of ImageNet) and synthetically generated backdrops to increase the out-of-distribution coverage of ordinary inputs. The second group includes hard negatives that look like maize kernels, dried berries, and other small grains and particles that look like coffee beans. The reported out-of-distribution (OOD) performance does include such visually perplexing objects, not only those that are easily distinguished [63].

Table 1. Dataset characteristics for 6-class Arabica coffee classification.

Parameter	Details
Total images	12,000 (6 classes including Non-Coffee)
Images per class (coffee)	2,000 per Arabica variety
Number of classes	6 (5 Arabica varieties + 1 Non-Coffee)
Image dimensions	224 × 224 px (after resizing)
Dataset split	70% train / 15% validation / 15% test
Training samples	8,400 (1,400 per class)
Validation samples	1,800 (300 per class)
Test samples	1,800 (300 per class)
Non-Coffee sources	CIFAR-100 + Imagenette + synthetic backgrounds + hard negatives (maize, dried berries, coffee-like seeds)
Capture conditions	Controlled lighting (2942–3157 lux), 27.6–31.7 °C, 75–80% RH, white/neutral background
Geographic origin	Bondowoso, East Java, Indonesia
Altitude range	900–1,500 m above sea level

3.3. CNN Model: RepViT-M1.1 with OOD Architecture

RepViT-M1.1, introduced at CVPR 2024, is an efficient architecture that utilises a lightweight convolutional backbone design, incorporating design choices from vision transformers and employing structural re-parameterisation, thereby achieving high ImageNet accuracy with a compact number of parameters [64–68]. For this study, the classification head was reconfigured to produce six outputs: the five Arabica varieties and an explicit 'Non-Coffee' class.

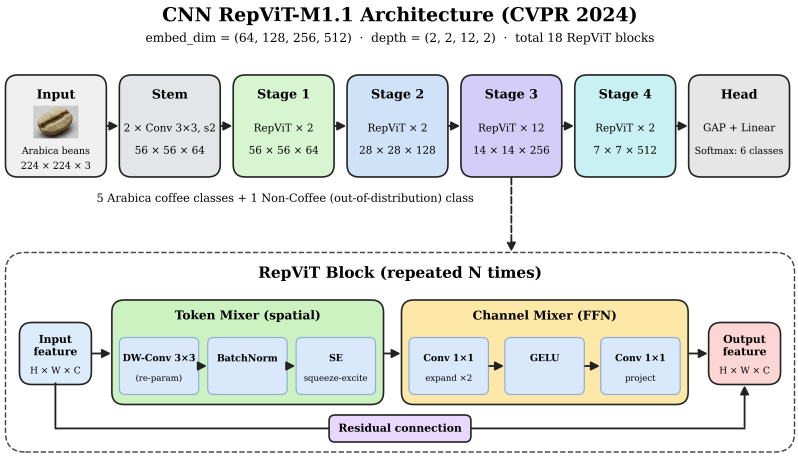


Figure 2. Architecture of the RepViT-M1.1 backbone reconfigured for 6-class classification

Each RepViT block (Figure 2, lower panel) consists of two components. A token mixer captures spatial structure via 3×3 depthwise convolutions, batch normalisation, and a Squeeze-and-Excitation (SE) module. A channel mixer, which is a feed-forward network, applies 1×1 expansion convolutions, GELU activation, and 1×1 projections, each with residual connections. Its key feature is structural re-parameterisation. During training, multiple parallel convolutional branches enrich the representation; at the inference stage, these branches are merged into a single 3×3 convolution, thereby maintaining speed without sacrificing accuracy.

3.3.1. Depthwise Separable Convolution

RepViT-M1.1 decomposes ordinary convolution into depthwise and pointwise convolutions, achieving computational economy without sacrificing representational capacity [46]. For an input $X \in \mathbb{R}^{H\times W\times C}$, depthwise convolution applies a different filter for each channel:

$$Y_c = X * k_{DW,c}, \quad c = 1, 2, \ldots, C \tag{1}$$

where Y_c is the depthwise output for channel c; $k_{DW,c} \in \mathbb{R}^{k\times k\times 1}$ is the per-channel filter; $*$ denotes convolution; and C is the total number of channels. For RepViT-M1.1, $k = 3$ (a 3×3 depthwise kernel).

The pointwise convolution mixes channel information using 1×1 convolution:

$$Z = W_{pw} * Y, \quad W_{pw} \in \mathbb{R}^{1\times1\times C\times C'} \tag{2}$$

where W_{pw} is the 1×1 cross-channel mixing kernel and C' is the number of output channels. Computational complexity is reduced by:

$$\frac{C(\text{DW-Sep})}{C(\text{Standard})} = \frac{1}{k^2} + \frac{1}{C'} \approx \frac{1}{9} \quad (k = 3) \tag{3}$$

This yields ≈8–9× fewer multiply-accumulate operations versus standard convolution for mobile CNN architecture [69].

3.3.2. Structural Re-Parameterization

A key feature of RepViT, adopted from RepVGG, is structural re-parameterization: a multi-branch topology during training is algebraically collapsed into a single-branch network at inference [70]. During training, each convolutional block operates three parallel branches:

$$F(x) = \mathrm{BN}(W_{k\times k} * x) + \mathrm{BN}(W_{1\times1} * x) + \mathrm{BN}(x) \tag{4}$$

where $\mathrm{BN}(\cdot)$ is Batch Normalization; the three terms represent (i) a $k \times k$ convolution branch, (ii) a 1×1 convolution branch, and (iii) an identity shortcut (applied when input/output dimensions match).

At inference deployment, the three branches are algebraically merged into a single equivalent convolution, eliminating branch overhead:

$$F^{\inf}(x) = W_{rep} * x + b_{rep}, \quad W_{rep} = W'_{k\times k} + W'_{1\times1} + I' \tag{5}$$

where W_{rep} and b_{rep} are the merged kernel and bias; W' denotes the BN-fused weights; and I' is the identity kernel zero-padded to $k \times k$. BN parameters are absorbed into the convolutional weights via:

$$W' = \frac{\gamma}{\sqrt{\sigma^2 + \varepsilon_{\mathrm{bn}}}} \cdot W, \quad b' = \beta - \frac{\gamma\mu}{\sqrt{\sigma^2 + \varepsilon_{\mathrm{bn}}}} \tag{6}$$

μ and σ^2 are the running (population) mean and variance; γ, β are the learnable BN scale/shift parameters; and $\varepsilon_{\mathrm{bn}} = 1 \times 10^{-5}$ is a numerical-stability constant.

3.3.3. MetaFormer Token Mixer and Channel Mixer

RepViT adopts the MetaFormer paradigm, separating spatial (token) mixing from channel processing into two sequential sub-blocks per layer:

$$X' = X + \mathrm{TokenMixer}(X); \quad X'' = X' + \mathrm{FFN}(X') \tag{7}$$

where TokenMixer is a 3×3 depthwise convolution with batch normalisation and a squeeze-and-excitation module, and X' and X'' are intermediate feature maps; normalisation is performed by batch normalisation inside

each sub-block, which is folded into the convolution weights at inference (Eq. 6). The Feed-Forward Network (channel mixer) with expansion ratio r is defined as:

$$\mathrm{FFN}(x) = W_2 \cdot \mathrm{GELU}(W_1 x + b_1) + b_2 \tag{8}$$

where $W_1 \in \mathbb{R}^{rd \times d}$ expands the input to the hidden dimension rd, $W_2 \in \mathbb{R}^{d \times rd}$ projects it back, and $\mathrm{GELU}(x) = x\,\Phi(x)$ is the Gaussian Error Linear Unit. For RepViT-M1.1, $r = 2$.

A two-phase transfer learning protocol governed model training [71]. In Phase 1 (15 epochs, head-only), all backbone weights were frozen and only the classification head was optimized using AdamW (learning rate = 1×10^{-3}, weight decay = 0.01) with a CosineAnnealingLR schedule [72, 73]. In Phase 2 (up to 30 epochs, full fine-tuning), the backbone was unfrozen and joint optimization proceeded at a reduced learning rate of 5×10^{-5}, governed by early stopping with patience = 7. A batch size of 32 was maintained throughout. Label smoothing (ε = 0.1) was integrated into the CrossEntropyLoss formulation [74]. The augmentation pipeline comprised random cropping, horizontal and vertical flipping, rotation up to $\pm20°$, ColorJitter, and RandomErasing (p = 0.1) [75, 76].

3.3.4. Classification Head and OOD Detection Formulation

The 6-class classification head applies a softmax function to the final linear layer logits $z \in \mathbb{R}^6$, producing a normalized probability distribution over five Arabica varieties and the Non-Coffee class:

$$p_i = \frac{\exp(z_i)}{\sum_{j=1}^{K} \exp(z_j)}, \quad K = 6 \tag{9}$$

where z_i is the i-th logit, $p_i \in (0, 1)$ with $\sum_i p_i = 1$, and $K = 6$ (five Arabica classes and one Non-Coffee rejection class).

$$\mathrm{Reject}(x) = \mathrm{TRUE} \;\text{ if }\; [\hat{y} = \text{Non-Coffee}] \;\text{ OR }\; [\max_k p_k < \tau], \quad \tau = 0.7957 \tag{10}$$

where τ is calibrated on the validation set rather than chosen by hand. Let $\mathcal{C}_{\mathrm{val}} = \{\max_k p_k(x) : x \in \mathcal{V}_{\mathrm{coffee}}\}$ be the maximum softmax confidences of the 1,500 genuine coffee images in the validation split. The threshold is the empirical quantile

$$\tau = Q_{1-\mathrm{TPR}^*}(\mathcal{C}_{\mathrm{val}}), \qquad \mathrm{TPR}^* = 0.976,$$

that is, the 2.4th percentile of $\mathcal{C}_{\mathrm{val}}$, which by construction accepts at least 97.6% of genuine coffee images; for the six-class model this gives $\tau = 0.7957$. The 97.6% target limits unnecessary rejections of genuine submissions to fewer than one in forty, which a farmer can resolve by resubmitting a photograph. The same target is used for every ablation configuration, placing all variants at a common operating point, following the fixed-TPR convention of OOD evaluation [77]. Because τ is estimated on the validation split only, the held-out test split provides an independent check against overfitting: the test Coffee TPR of 96.93% lies within the 95% sampling interval of the target ([96.8%, 98.4%] for $n = 1{,}500$). The Shannon entropy of the prediction distribution serves as a complementary uncertainty measure, reported by the dApp alongside each prediction (Figure 10):

$$H(p) = -\sum_{k=1}^{6} p_k \log p_k, \quad H_{\max} = \ln(6) = 1.791 \text{ nats} \tag{11}$$

$H(p) \to 0$ for high-confidence (peaked) predictions and $H(p) \to 1.791$ for maximally uncertain (uniform) predictions. Ablation Config C applied an entropy threshold $\xi_5 = 0.8185$ (on the five-class model) in addition to the five-class confidence threshold $\tau_5 = 0.7593$, both calibrated on the same validation split; the finding that Config C adds no OOD recall over Config B confirms the mathematical coupling between softmax confidence and entropy.

3.3.5. Training Loss Function

The model is trained with categorical cross-entropy loss over $N = 8{,}400$ training samples and $K = 6$ classes:

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{6} y_{i,k}\log p_{i,k} \tag{12}$$

where $y_{i,k} \in \{0,1\}$ is the one-hot ground-truth label and $p_{i,k}$ is the predicted probability for class k on sample i. Label smoothing regularization (Eq. 13) reduces overconfident predictions by redistributing probability mass from the target class to non-target classes [74]:

$$\tilde{y}_{i,k} = (1-\varepsilon)\cdot y_{i,k} + \frac{\varepsilon}{K}, \quad \varepsilon = 0.1,\ K = 6 \tag{13}$$

Hard one-hot labels $y_{i,k} \in \{0,1\}$ are replaced by smoothed labels $\tilde{y}_{i,k}$; the target class receives $1 - 0.1 + 0.1/6 \approx 0.917$ instead of 1.0, and each non-target class receives $0.1/6 \approx 0.017$ instead of 0. This discourages overconfident predictions. The two-phase transfer learning strategy formalizes the training procedure:

$$\text{Phase 1: } \theta^*_{cls} = \arg\min_{\theta_{cls}} \mathcal{L}(\theta^{\text{frozen}}_{bb}, \theta_{cls}); \quad \text{Phase 2: } \theta^* = \arg\min_{\theta} \mathcal{L}(\theta_{bb}, \theta_{cls}) \tag{14}$$

where θ_{bb} and θ_{cls} are the backbone and head parameters; hyperparameters are as given above. Figure 3 presents the training curves for both phases.

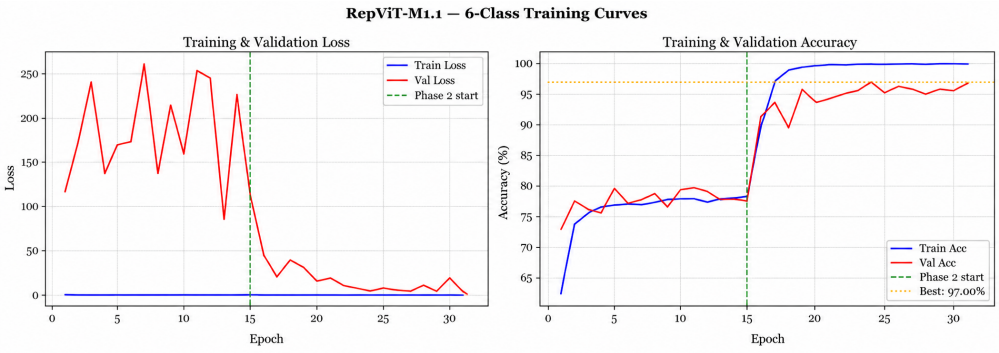


Figure 3. RepViT-M1.1 6-class training curves, validation loss, and validation accuracy

3.4. Smart Contract Design

The smart contract is written in Solidity 0.8.34 with 200 optimisation runs and published under the MIT License [33, 78]. It is deployed on Polygon Amoy (Chain ID: 80002) at address 0x5392C2F10d8Dea3e498726BcB8c806E8DA78834b, with source code publicly verified on Polygonscan. The designation "V3" denotes the third version of this contract; previous prototypes (V1 and V2) validated the core minting workflow, but had the drawback of lacking an on-chain SHA-256 hash registry and the permissionless, open minting model that characterises the current design. As the implemented contract is immutable, any improvements are released as a new version of the contract rather than by editing the existing one. Compared to previous versions, V3 introduces three architectural enhancements: (1) a permissionless open minting model that allows any farmer with a wallet to mint certificates via the mintKopiNFT() function without prior approval from the contract owner, thereby improving accessibility for small-scale farmers; (2) a dedicated farmerWallet field that creates an on-chain audit trail linking each certified lot to the farmer who submitted it; (3) the burnNFT() function, which provides an emergency dispute resolution mechanism. Table 2 summarises all smart contract functions.

Table 2. Smart contract functions in KopiArabikaNFT

Function	Type	Description
mintKopiNFT()	Write	Mint NFT with 9 parameters including petaniWallet; reverts on duplicate with FotoDuplikat() error
burnNFT()	Write	Emergency burn mechanism for dispute resolution (owner only)
cekHashFoto()	Read	Query on-chain SHA-256 registry; returns (sudahAda: bool, tokenIdLama: uint256)
verifikasiKeaslian()	Read	Return the stored photo CID, SHA-256 hash, and farmer wallet for any tokenId
getDataKopi()	Read	Retrieve complete coffee data including hashFoto
totalNFT()	Read	Return total number of minted NFTs

3.5. Three-Layer Anti-Duplication Mechanism

The anti-duplication architecture works on three levels: Layer 1 (SHA-256 cryptographic hashing): This is performed on the client side by using the Web Cryptography API and produces a 256-bit fingerprint that uniquely identifies each picture file. Layer 2 (Perceptual hashing) average hash (aHash) 16×16 pixel canvas reduction; Hamming distance ≤5 bits from 256 total bits flags high-similarity images. Layer 3 (On-chain registry): the SHA-256 hash is stored immutably in the smart contract's hashFotoRegistry mapping upon successful minting, rejecting re-registration of an identical hash at the blockchain consensus layer regardless of the calling application.

3.5.1. SHA-256 Cryptographic Hash Function

The SHA-256 hash function takes an arbitrary-length binary image buffer M and produces a fixed 256-bit digest. The digest is the unique identifier utilised throughout the anti-duplication pipeline [79]:

$$h = \text{SHA-256}(M) : \{0,1\}^* \to \{0,1\}^{256}, \quad |h| = 64 \text{ hexadecimal characters} \tag{15}$$

Here M is the raw image file buffer (of any size) and h the 256-bit fingerprint. Finding a collision, $\text{SHA-256}(M_1) = \text{SHA-256}(M_2)$ with $M_1 \neq M_2$, requires on the order of 2^{128} operations (the birthday bound), which is computationally infeasible. Computed client-side via Web Cryptography API (SubtleCrypto.digest) at Layer 1 of the system pipeline.

The SHA-256 compression function iterates 64 rounds over 512-bit message blocks. Each round t updates the working variables (a to h) via:

$$a_t = T_1 + T_2, \quad T_1 = h_t + \Sigma_1(e_t) + \text{Ch}(e_t, f_t, g_t) + K_t + W_t \tag{16}$$

Here $\Sigma_1(e) = \text{ROTR}^6(e) \oplus \text{ROTR}^{11}(e) \oplus \text{ROTR}^{25}(e)$ is a bitwise rotation function, $\text{Ch}(e,f,g) = (e \wedge f) \oplus (\neg e \wedge g)$ is the bitwise choice function, K_t are the 64 pre-computed round constants, and W_t is the message-schedule word derived from the padded input block.

The IPFS Content Identifier (CID) is derived deterministically from the image content, providing content-addressed storage with intrinsic tamper detection:

$$\text{CID}(I) = \text{Base58}(\text{Multihash}(\text{SHA-256}(I))) \tag{17}$$

Any single-bit modification to image I produces a completely different CID, making tampering immediately detectable during NFT verification. Photo CID and metadata CID are independently stored, enabling auditors to verify either component without access to the application layer.

3.5.2. On-Chain Duplicate Prevention Mechanism

The smart contract implements a permissionless on-chain duplicate prevention mechanism enforced at the blockchain consensus layer. The mint-eligibility predicate MintAllowed, evaluated inside mintKopiNFT() against

the _hashFotoRegistry mapping, formally specifies the condition under which a new certification NFT may be issued:

$$\text{MintAllowed}(h) = \begin{cases} \text{TRUE}, & \text{if Registry}[h] = 0 \text{ (first mint)} \\ \text{FALSE}, & \text{if Registry}[h] \neq 0 \text{ (duplicate)} \end{cases} \tag{18}$$

The Registry denotes the _hashFotoRegistry mapping(string ⇒ uint256), a Solidity state variable in the smart contract; $\text{Registry}[h] = 0$ indicates no prior certification exists; $\text{Registry}[h] \neq 0$ returns the token ID of the conflicting NFT. Token identifiers are assigned sequentially starting from 1, so the value 0 used by the Registry to indicate an unregistered hash will never match a valid token ID. Implementing this check at the consensus layer, rather than the application layer, makes it impossible to re-register a hash that has already been recorded, as the hash is calculated on the client side and passed as an argument; this guarantee applies to the SHA-256 string submitted, not to the photograph itself, which is what ensures the anti-duplication process operates robustly.

3.6. Decentralised Application (dApp)

The system is a dApp, realised as a responsive web application using Next.js 14 and React 18 and deployed on the Vercel Edge Network, without installation on the user's device. Ethers.js v6 allows for on-chain interactions, with transaction signing handled by the user's MetaMask wallet. Pinata's IPFS pinning service gives you decentralised storage, and CNN inference is served from a Gradio API endpoint hosted on Hugging Face Spaces. This architecture is intentionally built as a hybrid system: certificate records and content addressing are decentralised, and hosting, pinning and inference are offered by centralised providers. This compromise is taken to keep the operating costs at zero and not to require infrastructure for small-scaled customers. Latency is measured on the client side using browser performance.now() API.

4. Results and Discussion

4.1. Classification, OOD Ablation, and Calibration of RepViT-M1.1

On the 1,800-image test set, the six-class RepViT-M1.1 attains an overall test accuracy of 99.28%, a weighted F1-score of 99.27%, and macro-averaged precision, recall, and F1 of 99.29%, 99.28%, and 99.27%, respectively (Table 3). All five Arabica varieties are recalled perfectly (100%), with per-variety precision between 98.36% and 99.67% and F1 between 99.17% and 99.83%, indicating that inter-variety discrimination is essentially saturated on this dataset: no genuine coffee image is assigned to another variety. The residual error is confined to the open-set boundary. The Non-Coffee class attains 100% precision but 95.67% recall (F1 = 97.79%): 13 of its 300 images are misclassified as coffee, whereas nothing is ever wrongly labelled Non-Coffee. This asymmetry also explains the slightly sub-unity precision of the coffee classes, since those thirteen leaked Non-Coffee images constitute their only false positives. Wilson 95% confidence intervals on recall [80] are reported in Table 3. These results indicate that the classifier's difficulty lies not in distinguishing visually similar Arabica varieties which it does without error but in rejecting out-of-distribution inputs at the coffee/non-coffee boundary, directly motivating the explicit OOD mechanism. The validation split shows the same error pattern. Its best accuracy of 97.00% (Figure 3) is lower than the test accuracy because all 54 validation errors are Non-Coffee images assigned to coffee classes (Non-Coffee recall 82.00%, Wilson 95% CI [77.3, 85.9]), whereas all 1,500 validation coffee images are classified correctly, as on the test split. The gap between validation and test accuracy therefore reflects differences between the Non-Coffee subsets, which are drawn from heterogeneous sources, rather than confusion between varieties; it also indicates that Non-Coffee recall measured on 300 images should not be generalised beyond the negative sources used here (Section 3.2).

A central design question is how to reject non-coffee images before certification. To quantify the contribution of each OOD component, we ablate four progressive configurations of the classifier on the full six-class test set. Configuration A is the standard five-class softmax classifier, no OOD filtering. It represents the deployment baseline, the circumstance when by accident non-coffee items could be supplied. Configurations B and C add

Table 3. RepViT-M1.1 model performance ($n = 1800$, 300/class)

Class	Precision (%)	Recall (%)	F1 (%)	Support	Recall 95% CI
Arabica Anaerob (Carbonic)	99.01	100.00	99.50	300	[98.7, 100.0]
Arabica Blue Mountain	98.36	100.00	99.17	300	[98.7, 100.0]
Arabica Natural Ijen	99.67	100.00	99.83	300	[98.7, 100.0]
Arabica Orange Bourbon	99.67	100.00	99.83	300	[98.7, 100.0]
Arabica Peaberry	99.01	100.00	99.50	300	[98.7, 100.0]
Non-Coffee	100.00	95.67	97.79	300	[92.7, 97.5]
Macro avg	99.29	99.28	99.27	1800	[98.8, 99.6]
Weighted avg	99.29	99.28	99.27	1800	[98.8, 99.6]

post-hoc rejection to that model, confidence thresholding alone in B, and then confidence plus Shannon-entropy thresholding in C. The proposed design is Configuration D. Here a 6th class Non-Coffee is incorporated in training, so the model learns the coffee vs non-coffee boundary directly, rather than relying on post-hoc thresholds alone; the calibrated confidence threshold is then applied on top. Coffee True Positive Rate (Coffee TPR) is the percentage of true Arabica coffee photos that are correctly categorised and approved by the system. A low value means that the system is rejecting too many legitimate farmer submissions. OOD Recall calculates the percentage of non-coffee photos that the system correctly identifies and rejects [20]. A low OOD recall rate is a security issue and may result in fraudulent submissions or items being incorrectly certified as NFTs [49]. The full ablation results are shown in Table 4. Figure 4 shows the trade-off between Coffee TPR and OOD Recall over the four setups.

Table 4. Ablation of OOD configurations for RepViT-M1.1.

Configuration	Coffee Acc	Coffee TPR	OOD recall	OOD rejected	Δ OOD
A: 5-class, no OOD	99.93%	100.00%	0.00%	0/300	0.00%
B: 5-class + confidence	99.93%	96.60%	53.67%	161/300	+53.67%
C: 5-class + conf. + entropy	99.93%	96.40%	53.67%	161/300	+53.67%
D: 6-class + confidence (proposed)	100.00%	96.93%	98.00%	294/300	+98.00%

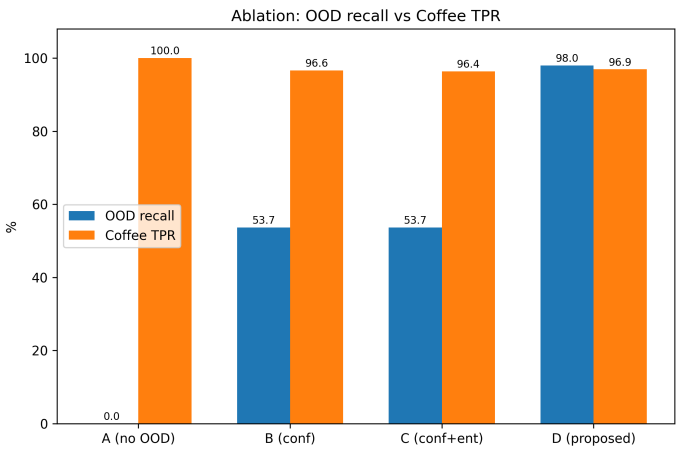


Figure 4. Ablation study visualization across four configurations.

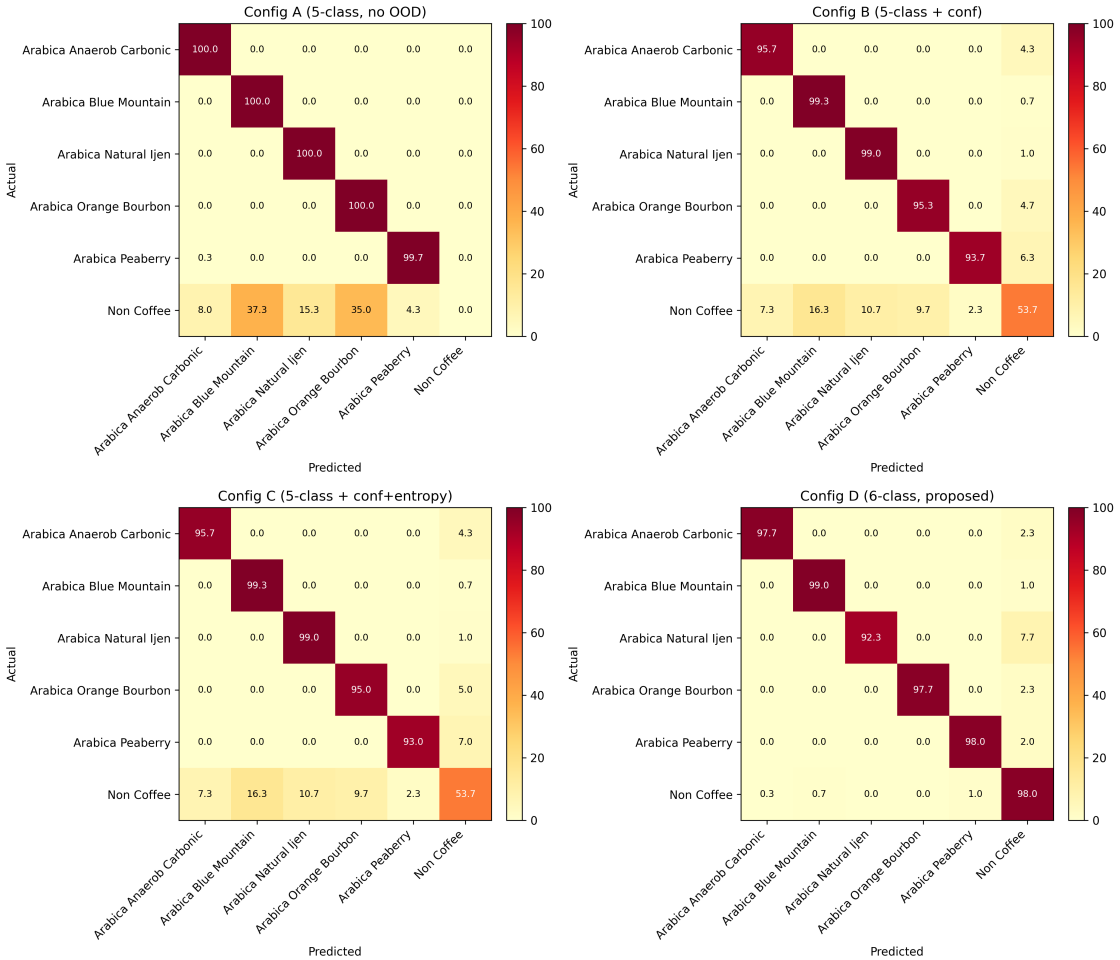


Figure 5. Confusion matrices of the four ablation configurations

Configuration B rejected only 161 of 300 Non-Coffee images (53.67%). These misses expose the ceiling of confidence-based OOD detection, because the score derives from the same softmax that is overconfident on OOD inputs, threshold tuning alone cannot surpass it [48].

Configuration C added Shannon-entropy thresholding ($\xi_5 = 0.8185$) to the confidence rule. For the five-class model, entropy $H = -\sum_i p_i \log p_i$ attains its maximum of $\ln(5) = 1.609$ under a uniform prediction. The two criteria appear complementary. They are not. Config C matched Config B, with an identical 53.67% OOD recall and a slightly lower Coffee TPR (96.40% vs. 96.60%). The reason is the deterministic link between softmax confidence and entropy, a high peak probability bounds the entropy, and vice versa, so the thresholds are redundant; when the remaining mass spreads evenly over the K$-$1 non-maximum classes, entropy is fixed by the confidence value. Varying threshold criteria, without any architectural change, thus cannot raise the OOD ceiling imposed by closed-set training [49, 81, 82].

Configuration D, the proposed six-class model with an explicit Non-Coffee class and the calibrated threshold $\tau = 0.7957$, reached 98.00% OOD recall, rejecting 294 of 300 Non-Coffee images, 44.33 percentage points more than the threshold-only configurations, while accepting 96.93% of genuine coffee images, a higher Coffee TPR than Configurations B and C at the same calibration target. No coffee image was assigned to a wrong variety (Coffee Acc 100.00%). Because the explicit Non-Coffee class already absorbs most out-of-distribution inputs, the threshold only removes residual ambiguous cases. Rejected coffee images are concentrated in Natural Ijen (7.7% rejected),

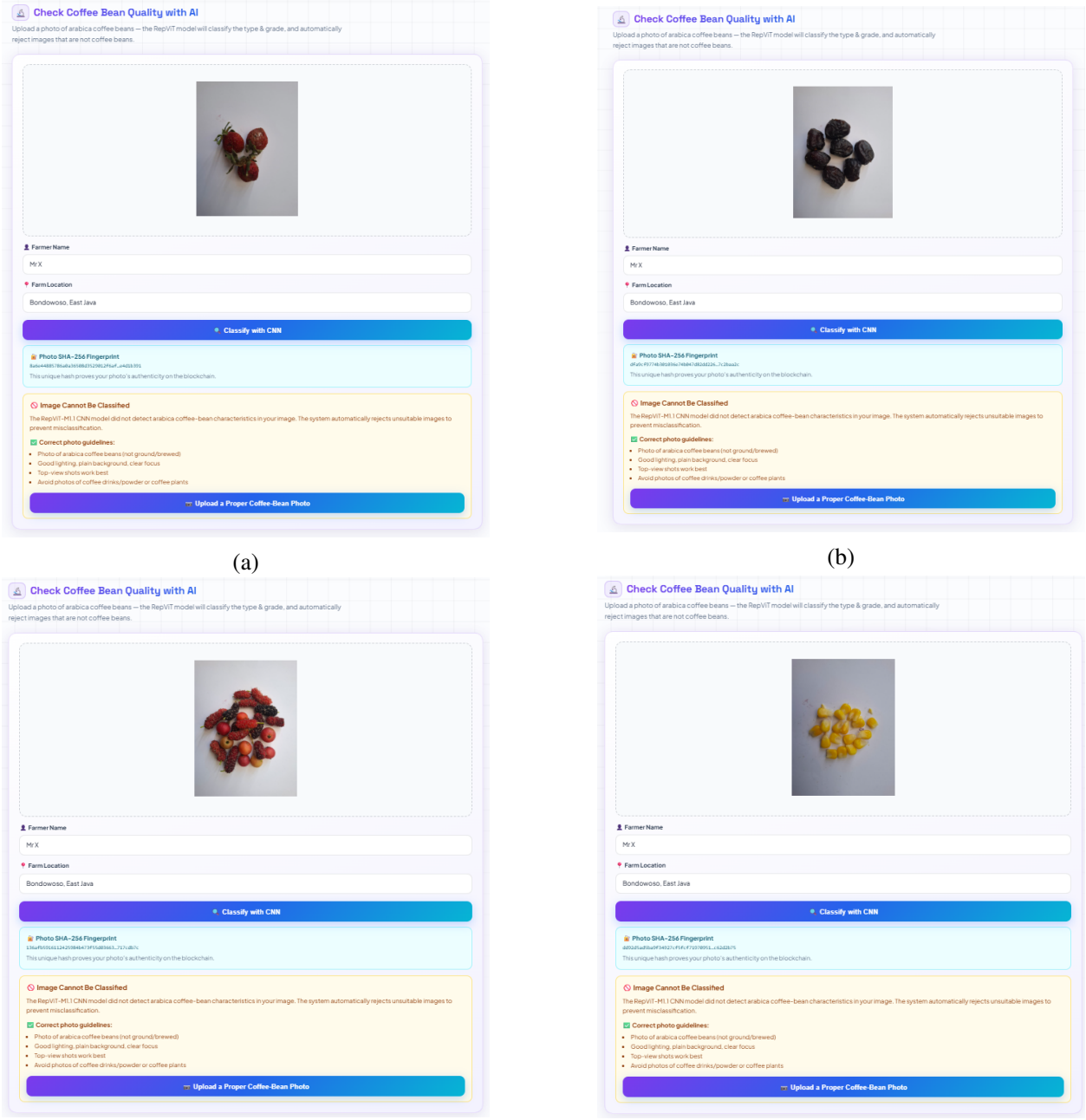


Figure 6. Hard negatives rejected by the deployed system

the class with the lowest mean confidence (87.7%, Figure 7), whereas the other four varieties lose at most 2.3%. The six remaining Non-Coffee misses (2.00%) mark the boundary of supervised OOD detection, which reliably rejects only inputs resembling the negative examples seen in training [83]. Figure 5 presents the per-configuration confusion matrices. Table 5 contrasts raw argmax with the deployed rule on the same trained model. The rule halves the OOD miss rate (4.33% to 2.00%) at the cost of 46 rejected coffee images; counting a coffee image as correct only when accepted with the right variety and a Non-Coffee image only when rejected, the deployed system

is correct on 97.11% of test images. The asymmetry is deliberate: a false acceptance commits a non-coffee image to an immutable record, whereas a false rejection only costs a resubmission.

Table 5. Deployed OOD performance of RepViT-M1.1

Metric	**Raw Argmax**	**Deployed Rule** ($\tau = 0.7957$)
Test samples	1,800	1,800
Overall accuracy	99.28% (1,787/1,800)	**97.11% (1,748/1,800)**
Coffee TPR (accepted)	100.00% (1,500/1,500)	**96.93% (1,454/1,500)**
Non-Coffee recall	95.67% (287/300)	**98.00% (294/300)**
Non-Coffee misses	13/300	**6/300**
OOD miss rate	4.33%	**2.00%**

On the deployed system, coffee-mimicking hard negatives are rejected as "Image Cannot Be Classified" rather than certified (Figure 6). A second finding emerged from the ablation. Configuration D's six-class model tended to be more confident on correctly classified coffee images than the five-class models of Configs A, B, and C, consistent with the redistribution of probability mass described below. That adding a rejection class should raise confidence in positive predictions seems counter-intuitive. It follows, however, from the redistribution of probability mass [83]: the Non-Coffee class absorbs the mass that the five-class model was forced to spread across coffee varieties for ambiguous inputs. Clearly identifiable coffee images therefore receive more concentrated softmax probability, reducing residual uncertainty and yielding more decisive predictions for certification. Figure 7 presents the per-class confidence distributions.

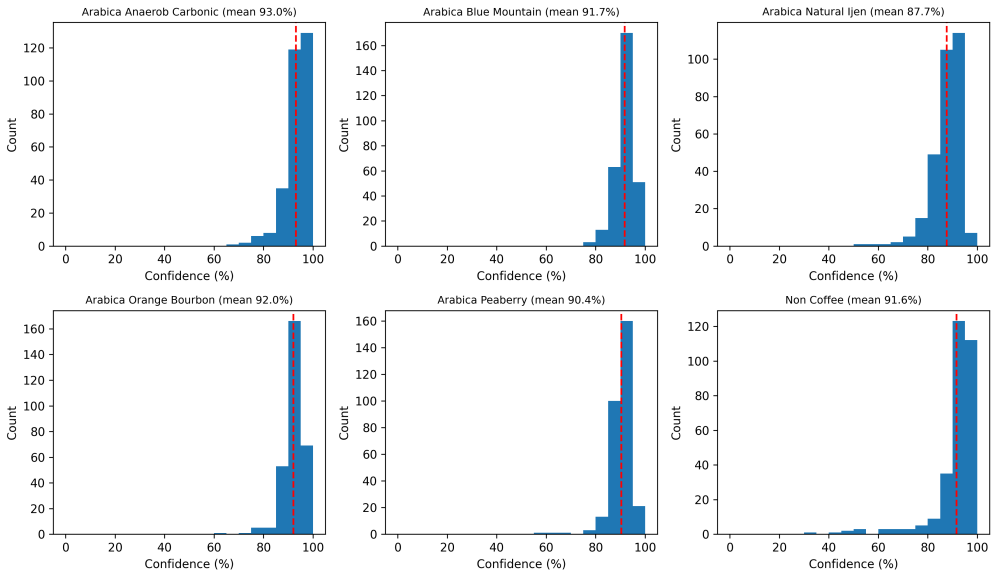


Figure 7. Per-class prediction confidence distributions

Beyond the threshold-specific operating points above, threshold-independent metrics characterise the model's separability and calibration. The classifier attains an AUROC of 0.9815 for the coffee-versus-non-coffee decision and a false-positive rate of 0.00% at 95% true-positive rate [20, 49], confirming strong discrimination between coffee and out-of-distribution inputs. The Expected Calibration Error (15-bin) [84] is 8.27%, indicating moderately overconfident predictions despite label smoothing [84, 85]. Table 6 summarises these metrics.

Table 6. Threshold-independent OOD and calibration metrics

Metric	Value
AUROC (Coffee vs. Non-Coffee)	0.9815
FPR @ 95% TPR	0.00%
Expected Calibration Error (15-bin)	8.27%

4.2. Comparison with Baseline Architectures

To better contextualise RepViT-M1.1 in the current landscape we benchmarked it against six architectures spanning convolutional, transformer, and hybrid designs: a classic CNN (ResNet-50 [86]), a mobile CNN (MobileNetV3-Large [87]), an efficient-scaling CNN (EfficientNet-B0 [88]), a modern convolutional design (ConvNeXt-Tiny [89]), a pure vision transformer (ViT-Small/16 [65]) and a CNN-ViT hybrid (MobileViT-S [67]). All seven models were trained and evaluated under the same conditions, including data split, training protocol, augmentation, optimiser, and early stopping, such that architecture was the only variation [90–92]. Table 7 reports test accuracy, weighted F1, parameters, CPU latency (the relevant setting, since the deployed endpoint runs on a shared CPU), and raw Non-Coffee recall.

Table 7. Comparative performance of CNN/ViT architectures

Model	**Category**	**Params**	**Test Acc**	**F1**	**CPU Lat. (ms)**	**OOD Class**
EfficientNet-B0	Efficient CNN	4.0M	100.00%	100.00%	25.61 ± 0.35	100.00%
ViT-Small/16	Vision Trans.	21.7M	100.00%	100.00%	38.57 ± 1.38	100.00%
MobileNetV3-Large	Mobile CNN	4.2M	100.00%	100.00%	17.47 ± 0.78	100.00%
MobileViT-S	Hybrid CNN+ViT	4.9M	100.00%	100.00%	36.31 ± 1.09	100.00%
ResNet-50	Classic CNN	23.5M	99.89%	99.89%	47.13 ± 0.83	100.00%
ConvNeXt-Tiny	Modern CNN	27.8M	100.00%	100.00%	46.99 ± 0.51	100.00%
RepViT-M1.1 (PROPOSED)	ViT-inspired mobile CNN	7.8M	99.28%	99.27%	41.73 ± 1.52	95.67%

On this controlled dataset all seven architectures saturate EfficientNet-B0, ViT-Small/16, MobileNetV3-Large, MobileViT-S, and ConvNeXt-Tiny reach 100.00% test accuracy and F1, ResNet-50 reaches 99.89%, and RepViT-M1.1 reaches 99.28%. This comparison does not highlight RepViT-M1.1 as the most accurate or most efficient backbone; lighter CNNs such as MobileNetV3-Large match or surpass it in terms of accuracy, raw out-of-distribution (OOD) recall, number of parameters, and CPU latency. RepViT-M1.1 is nevertheless retained as the backbone in use because its accuracy remains competitive and, when combined with the calibrated rejection rules, the model meets the certification requirements.

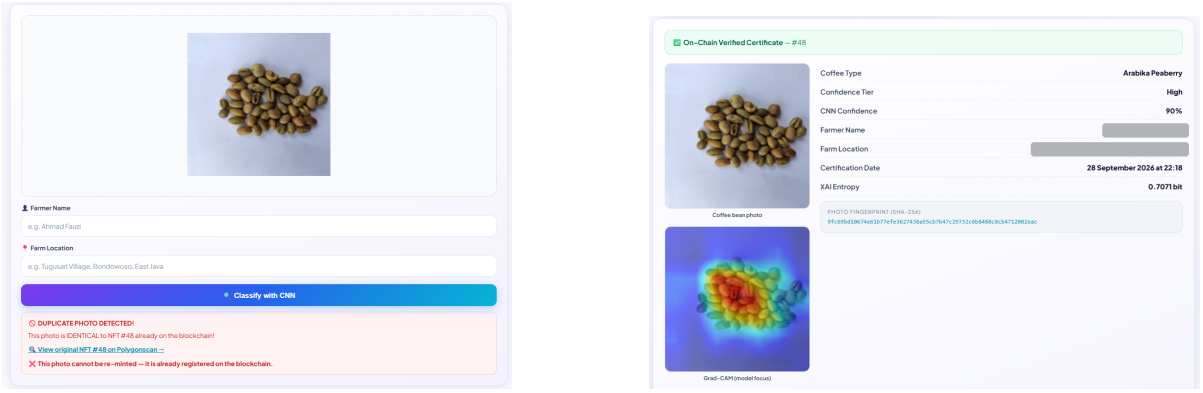
4.3. Blockchain System Performance and Viability

4.3.1. Anti-Duplication Mechanism

The smart contract provides an on-chain SHA-256 anti-duplication registry where each photograph's 256 bit fingerprint is permanently stored in the hashFotoRegistry . This ensures that each token id is uniquely tied to the matching hash. There are two enforcement techniques. Clients can call the public function cekHashFoto() before submitting a transaction to query the registry. The mintKopiNFT() function does an internal check and can reverse with a FotoDuplikat error if a duplicate is found. This check happens at the consensus layer of the blockchain thus it can't be bypassed and no two tokens can have the same SHA-256 string.

The effectiveness of the SHA-256 fingerprint layer was demonstrated through direct testing on the deployed system. An image file already registered as certificate NFT #48 was resubmitted through the interface. The system rejected the submission before minting, and the rejection was specific: it explicitly identified the prior certificate

holding the identical fingerprint, together with a link to its transaction record on the blockchain explorer (Fig. 8a). Following that link displays certificate #48 as verified on-chain, carrying the same SHA-256 digest value (Fig. 8b). The rejection is therefore not merely an interface message but a consequence of a permanent record that any third party can trace and independently re-verify. Across 30 resubmission trials of previously certified images, every attempt was rejected before minting (30/30), and no duplicate token was issued.



(a) (b)

Figure 8. Demonstration of the SHA-256 anti-duplication layer

Post-minting governance is provided by `burnNFT()`. Restricted to the contract owner, it permanently destroys a specified token, removing it from supply and flagging its SHA-256 digest as disputed in the registry without affecting other records. This explicit dispute path distinguishes the contract from naive ERC-721 implementations that cannot amend records after minting, and it addresses the irreversibility concern frequently cited as a barrier to institutional adoption of blockchain-based agricultural certification [93, 94]. Administrative authority in the contract is deliberately minimal: `burnNFT()` is the sole owner-restricted function, while minting and all read operations remain permissionless. The concentration it introduces is nonetheless real, since revocation permanently destroys the token and alters its registry entry. Mitigation does not require redeployment: the inherited `transferOwnership()` function allows the owner role to be reassigned to a multi-signature wallet or governance contract (DAO) at any time, converting unilateral revocation into a threshold decision without altering contract logic [95]. Conversely, `renounceOwnership()` would render the dispute mechanism permanently inoperable, and the owner key is presently held in a single externally owned account, so its loss or compromise carries the same consequence.

Beyond the SHA-256 fingerprint for exact duplicates, the system incorporates a client-side perceptual-hashing layer based on average hash (aHash), with 16×16 pixel reduction yielding a 256-bit descriptor and a Hamming-distance threshold of ≤ 5 bits. The layer targets submissions that cryptographic hashing cannot catch by construction: an image that has been re-encoded, had its metadata stripped, or been marginally altered produces a different SHA-256 digest while remaining visually identical. The perceptual-hash layer is advisory and local in scope. It compares each submission against a single previously processed image retained in browser storage rather than against the on-chain registry, and a match below the distance threshold raises a warning without blocking classification or minting. It therefore complements, but does not extend, the cryptographic uniqueness guarantee enforced on chain. Unlike the agricultural NFT systems reviewed in Section 2.3, the proposed system pairs perceptual and cryptographic hashing.

To assess this layer directly, each of 50 coffee images was subjected to eleven content-preserving or geometric transformations, and the average-hash (aHash, 16×16, 256-bit) Hamming distance between the original and each variant was measured (Table 8). Every transformation altered the file's SHA-256 digest, so the exact-match registry alone would treat each variant as a new image. The perceptual layer nonetheless recognised all re-encoded, re-saved, resized, and darkened copies: JPEG re-compression at qualities 90–50, metadata stripping/re-saving, resize round-trips, and darkening (-20%) were detected in 100% of cases at a median Hamming distance of 0, and

brightening (+20%) in 72%. Geometric edits (5–10% cropping and ±5–10° rotation) exceeded the conservative 5-bit threshold and were not flagged [96], reflecting a conservative threshold that favours a near-zero false-alarm rate over invariance to spatial edits. Across all 2,450 distinct-image pairs the false-positive rate was 0.08% (median Hamming distance 85), confirming that genuinely different beans, including different photographs of the same variety, are not mistaken for duplicates. Because these layers are deterministic, their contribution is shown by such direct tests rather than by an accuracy ablation: the registry rejected all 30 exact resubmissions, SHA-256 alone flagged none of the 550 transformed variants, and adding the perceptual layer flagged 336 of them (61%).

Table 8. Perceptual-hash (aHash, 16×16, 256-bit) evaluation over 50 coffee images.

Transformation	Median Hamming	Detected ($\leq$5/256)	SHA-256 changed
JPEG re-compression (q90)	0	100.0%	100%
JPEG re-compression (q70)	0	100.0%	100%
JPEG re-compression (q50)	0	100.0%	100%
Re-save / metadata strip	0	100.0%	100%
Resize round-trip	0	100.0%	100%
Brightness −20%	0	100.0%	100%
Brightness +20%	4	72.0%	100%
Crop 5%	18	0.0%	100%
Crop 10%	35	0.0%	100%
Rotation ±5°	25–26	0.0%	100%
Rotation +10°	39	0.0%	100%
Distinct images (false positive)	85	0.08%	n/a

4.3.2. Smart Contract Architecture and Verification

KopiArabikaNFT is deployed on the Polygon Amoy testnet, and this choice shapes how the results reported here should be read. The contract source is published and verified on the public block explorer, so its logic can be inspected independently rather than taken on trust. Amoy replicates Polygon PoS mainnet consensus rules, EVM compatibility, and execution semantics, so the functional behaviour observed in this evaluation is expected to carry over to production,although latency under mainnet load may differ. We accordingly present this deployment as a proof of concept, the latency evaluation and the anti-duplication demonstration provide preliminary assurance, while migration to the Polygon mainnet would require a fresh deployment at a new address and, ideally, an independent third-party security audit [97, 99, 101].

A dedicated `petaniWallet` field records the wallet address that submitted each lot, creating an immutable on-chain link between certificate and issuer and establishing a cryptographically verifiable audit trail for supply-chain accountability, independent of any off-chain database. Although the contract does not restrict who may issue, it therefore renders every issuance attributable, and post hoc dispute resolution depends on this property. As deployed contracts are immutable, updates are carried out by deploying a new version of the contract as in the transition to V3 with the registry status migrated as required. The current deployment does not use upgradeable proxies, so each contract version occupies a fixed, independently verified address. Such upgrades are localised to the blockchain layer, as the CNN inference service and the IPFS storage layer do not depend on the contract's internal logic, migration requires only redeployment alongside updates to the interface contract address and ABI, meaning that the classification model, decentralised storage, and the broader Web3 architecture remain unchanged. Production deployments can utilise an upgradeable proxy pattern to apply bug fixes without the need for redeployment, though this does require the existence of an upgrade authority.

4.3.3. Economic Viability

On the Polygon Amoy testnet, each minting transaction incurs a gas fee equivalent to less than USD 0.01, based on a consistently applied gas price of 30 Gwei. As gas on the testnet has no monetary value, this figure is merely an indicative estimate, not a verified cost on the Polygon mainnet. Mainnet fees are paid in POL and increase as network congestion rises, so the actual cost per certificate will vary; a realistic analysis of mainnet costs under varying gas conditions will form part of future research. Table 9 summarises the deployed contract's configuration and measured performance.

Table 9. Blockchain system performance metrics

Parameter	**Value**
Contract address (V3)	0x5392C2F10d8Dea3e498726BcB8c806E8DA78834b
Blockchain network	Polygon Amoy Testnet (Chain ID: 80002)
Token standard	ERC-721 (OpenZeppelin)
Compiler version	Solidity 0.8.34, Optimization 200 runs
Contract verification	Verified on Polygonscan (MIT License)
Mint model	Open mint (permissionless, any wallet)
Gas price override	30 Gwei (Polygon Amoy)
Avg. mint cost	$<$ \$0.01 USD per NFT (testnet estimate; Polygon mainnet cost not yet measured)
IPFS storage	Pinata (image + JSON metadata per NFT)
NFTs minted in latency trials	30
Duplicate attempts blocked	30/30 identical-hash resubmissions rejected by the on-chain registry (100%)
dApp deployment	Vercel Edge Network (kopi-arabika-web3.vercel.app)

4.3.4. End-to-End System Latency

Practical use also requires an acceptable per-sample latency. The two main components were measured separately and behave differently. CNN inference depends on model and hardware: on an NVIDIA L4 GPU (Google Colab) it took 20.83 $\pm$ 16.04 ms over 25 post-warm-up runs (range 12.25–81.54 ms, the spread reflecting latency spikes on shared cloud infrastructure), and 41.73 ms on CPU (Table 7), the setting of the deployed Hugging Face Spaces endpoint (CPU basic tier). Minting, measured over 30 independent trials from a laptop browser, depends on IPFS availability, validator queue depth, and network congestion, and is therefore non-deterministic.

The certification latency is obtained by adding the measured components: client-side SHA-256 hashing (T_{hash}), server-side inference (T_{CNN}), and minting (T_{mint}, IPFS upload plus receipt of the confirmed transaction):

$$T_{total} = T_{hash} + T_{CNN} + T_{mint} = 14.81 \pm 2.04 \text{ s} \quad (n = 30 \text{ trials}) \tag{19}$$

Table 10. End-to-end system latency (n=30 minting trials; CNN inference: 25 post-warm-up runs).

Component	**Mean**	**Std**	**Min**	**Max**
CNN Inference (NVIDIA L4 GPU, Colab)	20.83 ms	16.04 ms	12.25 ms	81.54 ms
IPFS Upload + Blockchain Tx (T_{mint})	14.79 s	2.04 s	10.14 s	18.80 s
Total End-to-End (T_{total})	14.81 s	2.04 s	10.16 s	18.82 s

Minting dominates the total (median 14.86 s, 95th percentile 17.53 s; Figure 9), whereas hashing and inference add only about 0.02 s; replacing GPU with CPU inference leaves T_{total} practically unchanged. All 30 minting trials were run from a laptop browser with MetaMask at a gas price of 30 Gwei, each with a distinct photograph; image transfer to the inference endpoint is excluded. To capture this transfer in practice, the dApp was tested in the field with 20 smallholder farmers, each classifying coffee samples on their own smartphone. The mean wall-clock time

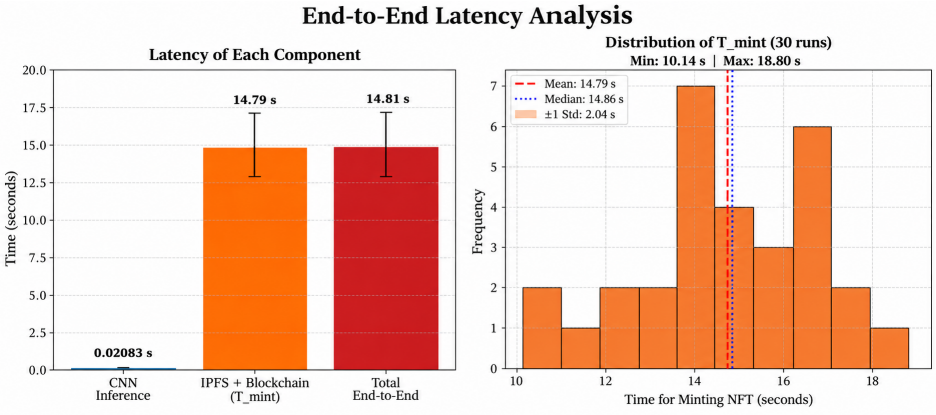


Figure 9. End-to-end latency analysis

from image capture to on-screen classification result was 135.5 s (SD = 9.5 s; median 134.5 s; range 121–153 s; 95% CI [131.3, 139.6]). This time is dominated by image upload, endpoint cold start, and mobile-network latency rather than by model inference.

4.3.5. Deployment, Explainability, and Storage

The system is deployed as a browser-based dApp on the Vercel Edge Network, whose global content delivery keeps load times acceptable across varying connectivity conditions, and requires nothing on the user's device beyond a standard web browser and the MetaMask extension. The CNN inference endpoint is publicly accessible on Hugging Face Spaces, allowing independent researchers to reproduce the classification results without deploying the model locally.

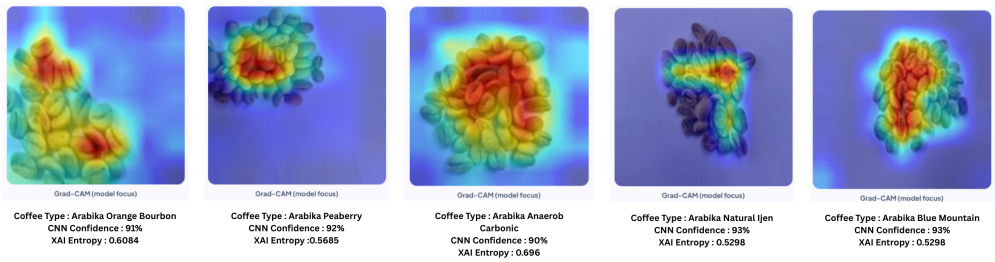


Figure 10. Grad-CAM saliency maps for one sample of each Arabica class.

Figure 10 shows Grad-CAM saliency maps for one sample of each Arabica class. In every case the strongest activations fall on the coffee beans rather than on the background, and the predictions are made with confidences of 90–93% and low predictive entropy (0.53–0.70 bits as displayed by the dApp, equivalent to 0.37–0.49 nats against the maximum $\ln 6 \approx 1.79$ nats of Eq. 11), indicating that the classifier relies on bean features when certifying a lot.

The certificate artefacts, namely the photograph, the Grad-CAM heatmap, and the ERC-721 metadata, are pinned on a single IPFS service (Pinata), so their availability depends on that provider [53]. Because IPFS is content-addressed, an identical copy can be re-pinned by another provider or a self-hosted node under the same CID without changing on-chain records, and the on-chain SHA-256 digest allows any restored copy to be verified; the core certificate fields (variety, identification label, farmer name, location, timestamp, confidence) remain on chain regardless. Cross-provider pinning and Filecoin or Arweave storage are future work.

4.4. Comparison with Related Systems

4.4.1. On-Chain Hash Registry vs. Application-Layer Duplicate Prevention

The clearest security difference from prior work concerns where duplicate prevention is enforced. Existing platforms typically check duplicates at the application layer, which can be bypassed by calling the contract directly. KopiArabikaNFT enforces hash uniqueness inside the contract, the guarantee applies to the submitted hash rather than to the photograph itself. Table 11 compares the system with representative agricultural NFT traceability approaches on design dimensions relevant to security, accessibility, and validation status.

Table 11. Feature comparison with existing agricultural NFT traceability systems.

Feature	**This Work**	**Existing Agri-NFT Systems**
CNN architecture	RepViT-M1.1 (CVPR 2024)	MobileNet / ResNet variants
OOD class (Non-Coffee)	Yes (95.67% raw; 98.00% deployed Non-Coffee recall)	None
On-chain hash registry	Yes (SHA-256)	No
Duplicate prevention	Blockchain consensus layer	Application layer only
Permissionless minting	Yes (V3 open mint)	Usually owner-only
Perceptual hash detection	Yes (aHash + Hamming)	No
Farmer interface	dApp, no installation	Mobile app / dedicated device
Transaction cost	$<$ \$0.01 USD (testnet estimate)	Variable (network-dependent)
Smart contract verified	Yes (Polygonscan)	Varies
Hash and image binding	Client-side; verifiable off-chain via CID	Rarely reported
Validation status	Testnet proof of concept; field trial (20 farmers)	Mostly prototype or simulation
Interface language	Indonesian-English	Varies (often English-only)

4.4.2. Permissionless Minting vs. Owner-Controlled Access

On most agricultural NFT platforms, only a single owner address or a pre-approved whitelist can mint, so farmers must register with the operator before they can certify. V3's open-mint design removes that dependency: any wallet holding sufficient POL can invoke mintKopiNFT() directly, which matters for smallholders facing poor connectivity, limited digital literacy, and geographic isolation. Routing issuance through an institutional wallet would instead bound throughput by that institution's availability and reintroduce the intermediary that blockchain traceability is meant to remove. Trust is thus relocated from the issuer to the evidence: each certificate carries the SHA-256 anchor of its photograph, the prediction confidence, and the petaniWallet address, so the contract guarantees record integrity and attributability rather than issuer legitimacy. Open minting does create a spam and false-certification surface. Three factors currently limit abuse: the gas cost of each mint, the OOD gate that rejects non-coffee submissions before minting, and the owner-restricted burnNFT() function, which revokes a fraudulent certificate after issuance and is the principal safeguard. Stronger measures such as decentralised-identity attestation, issuer staking, or a reputation layer are left to future work.

4.4.3. Farmer Accessibility and Interface Design

Existing agricultural blockchain deployments often require specific infrastructure, which can be challenging for smallholders. The proposed decentralized application (dApp) allows image capture and coffee classification via smartphones or computers, though NFT certificate minting currently necessitates a desktop browser with MetaMask. Future work aims to integrate mobile wallet solutions for complete smartphone functionality. The interface supports both Indonesian and English, catering to local users in Bondowoso, East Java, while remaining accessible to international stakeholders. To mitigate risks associated with key custody, farmers receive

comprehensive training to securely manage their recovery phrases, emphasizing handwritten copies and supervised recovery exercises. Though losing the phrase does not invalidate a certificate, it does hinder token transferability. Planned enhancements include account abstraction and social recovery wallets to further lower accessibility barriers.

MetaMask transfers the burden of key custody to the farmer, so a preventive training protocol was delivered before system use: farmers write the recovery phrase by hand, never photograph or type it into messaging or note-taking applications, keep two copies in separate locations, and are told that no legitimate party, including the research team, will ever request it. Training ends with a supervised recovery exercise on a second device. Losing the phrase does not invalidate a certificate, because verification resolves by token identifier and requires no wallet; only transferability of the token is lost. Account-abstraction or social-recovery wallets are planned to lower this barrier further.

5. Conclusion

This study presented a CNN and blockchain pipeline that classifies five Indonesian Arabica coffee classes, rejects non-coffee inputs, and issues NFT certificates through a permissionless ERC-721 smart contract (V3) with an on-chain SHA-256 hash registry on the Polygon Amoy testnet. Three findings stand out. First, training an explicit Non-Coffee class rejected far more out-of-distribution images than confidence or entropy thresholds alone, while separation among the five coffee classes was almost perfect on the controlled dataset. Second, the on-chain registry blocked every exact resubmission regardless of the submission path, and the perceptual-hash layer caught re-encoded copies that cryptographic hashing cannot detect, although not cropped or rotated ones. Third, certification time is dominated by blockchain minting in the laboratory and by image upload, endpoint cold start, and mobile-network latency in the field, whereas model inference is negligible. Because lighter backbones reached comparable accuracy, the contribution lies in the integrated, OOD-aware, anti-duplication workflow delivered through a bilingual dApp rather than in the choice of classifier. As a proof of concept, each certificate records a digital classification claim rather than physically verified quality. Future work will pursue background-augmented training for uncontrolled capture, IoT sensors for physical quality measurements, edge-computing inference for offline use, gasless meta-transaction relayers, and extension of the classifier to additional Indonesian coffee varieties.

6. Limitations and Future Directions

This system is presented as a proof of concept and is not a substitute for official certification. The NFT records a claim regarding digital classification rather than physically verified quality. In Indonesia, geographical origin is officially protected through Geographical Indication (GI) registration. Aligning the certificate with these schemes for instance, by including GI registration data and SNI quality grades issued by accredited bodies alongside the model outputs is an agenda for future development. The current design has several limitations. First, data and model: the dataset was captured under controlled lighting and neutral backgrounds in a single region, so accuracy on independent sites, devices, and seasons remains to be established, and the field trial recorded classification time on farmers' smartphones but not classification accuracy, cross-validation and repeated training runs were not performed, and the moderate overconfidence of the classifier could be reduced by post-hoc calibration such as temperature scaling. Second, The blockchain infrastructure relies on the Polygon Amoy testnet, which requires audits and cost management for mainnet deployment. Third, the inference infrastructure faces performance issues under high loads and requires an internet connection. Future development could incorporate offline inference using edge computing.

Declaration of Competing Interest

The authors declare no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability

The 12,000-image Arabica coffee dataset will be made available upon request to the corresponding author. Smart contract V3 is publicly accessible and verified at 0x5392C2F10d8Dea3e498726BcB8c806E8DA78834b on Polygon Amoy (Polygonscan). Web application: https://kopi-arabika-web3.vercel.app.

Acknowledgment

The authors gratefully acknowledge the support provided by the LP2M University of Jember, Department of Science Education-Faculty of Education, University of Jember, and Kemdiktisaintek for their valuable contributions, collaboration, and support.

REFERENCES

1. Amalia, F., Irifune, T., Takegami, T., Yusianto, Sumirat, U., Putri, S. P., & Fukusaki, E. (2023). Identification of Potential Quality Markers in Indonesia's Arabica Specialty Coffee Using GC/MS-Based Metabolomics Approach. Metabolomics, 19(11), 89. https://doi.org/10.1007/s11306-023-02051-5
2. Prawitasari, S. (2020). Assessment and Risk Mitigation of Arabica Ijen Coffee Supply Chains. Proceedings of the 1st Borobudur International Symposium on Humanities, Economics and Social Sciences (BIS-HESS 2019). https://doi.org/10.2991/assehr.k.200529.169
3. Munyendo, L., Njoroge, D., Zhang, Y., & Hitzmann, B. (2023). Novel Method for the Detection of Adulterants in Coffee and the Determination of a Coffee's Geographical Origin Using Near Infrared Spectroscopy Complemented by an Autoencoder. International Journal of Food Science & Technology, 58(3), 1284–1298. https://doi.org/10.1111/ijfs.16283
4. García, M., Candelo-Becerra, J. E., & Hoyos, F. (2019). Quality and Defect Inspection of Green Coffee Beans Using a Computer Vision System. Applied Sciences, 9(19), 4195. https://doi.org/10.3390/app9194195
5. Kuruppuarachchi, C. (2023). Grading of Green Coffee Beans for Specialty Coffee using Image Processing Techniques. https://doi.org/10.5281/zenodo.10464814
6. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A Survey of Deep Neural Network Architectures and Their Applications. Neurocomputing, 234, 11–26. https://doi.org/10.1016/j.neucom.2016.12.038
7. Gita, R. S. D., Ridlo, Z. R., Baihaki, R. I., Muttakin, F. N., Dafik, & Kristiana, A. I. (2026). STEM-Based Analysis of Avocados (Persea americana Mill.) Leaf Classification Using Deep Convolutional Neural Networks. Statistics, Optimization & Information Computing, 15(5), 4058–4082. https://doi.org/10.19139/soic-2310-5070-3124
8. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using Deep Learning for Image-Based Plant Disease Detection. Frontiers in Plant Science, 7, 1419. https://doi.org/10.3389/fpls.2016.01419
9. Zheng, R., & Kamruzzaman, M. (2023). Applications of hyperspectral imaging in the coffee industry: Current research and future outlook. Applied Spectroscopy Reviews. https://doi.org/10.1080/05704928.2023.2295351
10. Zimka, M., & Pentoś, K. (2025). Evaluation of the Feasibility of Coffee Variety Classification from Bean Images Using Convolutional Neural Networks. Acta Universitatis Cibiniensis. Series E: Food Technology, 29(2), 177–188. https://doi.org/10.2478/aucft-2025-0014
11. Lu, S., Wang, Y. P., Sheng, L., He, L., Zheng, A., & Liang, J. (2025). Out-of-Distribution Detection: A Task-Oriented Survey of Recent Advances. ACM Computing Surveys. https://doi.org/10.1145/3760390
12. Lu, S., Wang, S., Sheng, L., Zheng, A., He, L., & Liang, J. (2024). Recent Advances in OOD Detection: Problems and Approaches. https://doi.org/10.48550/arxiv.2409.11884
13. Ferentinos, K. P. (2018). Deep Learning Models for Plant Disease Detection and Diagnosis. Computers and Electronics in Agriculture, 145, 311–318. https://doi.org/10.1016/j.compag.2018.01.009
14. Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A Comparative Study of Fine-Tuning Deep Learning Models for Plant Disease Identification. Computers and Electronics in Agriculture, 161, 272–279. https://doi.org/10.1016/j.compag.2018.03.032
15. Korkmaz, A., Talan, T., Koşunalp, S., & Iliev, T. (2025). Comparison of deep learning models in automatic classification of coffee bean species. PeerJ Computer Science, 11, e2759. https://doi.org/10.7717/peerj-cs.2759
16. Arwatchananukul, S., Xu, D., Charoenkwan, P., Moon, S. A., & Saengrayap, R. (2024). Implementing a deep learning model for defect classification in Thai Arabica green coffee beans. Smart Agricultural Technology, 9, 100680. https://doi.org/10.1016/j.atech.2024.100680

17. Hassan, S. M., Maji, A. K., Jasiński, M., Leonowicz, Z., & Jasińska, E. (2021). Identification of Plant-Leaf Diseases Using CNN and Transfer-Learning Approach. Electronics, 10(12), 1388. https://doi.org/10.3390/electronics10121388
18. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674. https://doi.org/10.3390/s18082674
19. Kamilaris, A., & Prenafeta-Boldu, F. X. (2018). Deep Learning in Agriculture: A Survey. Computers and Electronics in Agriculture, 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016
20. Yang, J., Zhou, K., Li, Y., & Liu, Z. (2024). Generalized Out-of-Distribution Detection: A Survey. International Journal of Computer Vision, 132, 5635–5662. https://doi.org/10.1007/s11263-024-02117-4
21. Ridlo, Z. R., Dafik, Waluyo, J., Yushardi, & Venkatachalam, M. (2025). On RIDS Analysis for Shade Tree Placement and Its Application to STGNN Multi-Step Forecasting on RH and CO2 Concentration of Coffee Agroforestry. Statistics, Optimization & Information Computing, 14(4), 1889–1908. https://doi.org/10.19139/soic-2310-5070-2643
22. Meganingtyas, D. E. W., Dafik, Agustin, I. H., Baihaki, R. I., Ridlo, Z. R., Angrenani, A. B., & Nisviasari, R. (2026). The Implementation of Artificial Neural Networks and Resolving Efficient Dominating Set for Time Series Forecasting on Vertical Farming Soil Moisture. Statistics, Optimization & Information Computing, 15(6), 5148–5169. https://doi.org/10.19139/soic-2310-5070-3004
23. Ridlo, Z. R., Dafik, Waluyo, J., & Yushardi. (2024). Exploring the Impact of Makerspace-Based Learning Materials on Students' Computational Thinking Skills: Using Machine Learning to Address Challenges in Smart Coffee Agroforestry. Edelweiss Applied Science and Technology, 8(5), 52–73.
24. Ridlo, Z. R., Ningsih, S. P. A., & Anggraini, A. L. (2025). Computational Thinking and Deep Learning on Science Education Framework: A Systematic Review. Science Education International, 36(4), 480–489. https://doi.org/10.33828/sei.v36.i4.11
25. Tian, F. (2016). An Agri-Food Supply Chain Traceability System for China Based on RFID & Blockchain Technology. In 13th International Conference on Service Systems and Service Management (ICSSSM), IEEE, 1–6.
26. Kamilaris, A., Fonts, A., & Prenafeta-Boldú, F. X. (2019). The Rise of Blockchain Technology in Agriculture and Food Supply Chains. Trends in Food Science & Technology, 91, 640–652. https://doi.org/10.1016/j.tifs.2019.07.034
27. Pandey, V., Pant, M., & Snasel, V. (2022). Blockchain technology in food supply chains: Review and bibliometric analysis. Technology in Society, 69, 101954. https://doi.org/10.1016/j.techsoc.2022.101954
28. Ellahi, R. M., Wood, L. C., & Bekhit, A. E.-D. A. (2023). Blockchain-based frameworks for food traceability: A systematic review. Foods, 12(16), 3026. https://doi.org/10.3390/foods12163026
29. Zhao, G., Liu, S., Lopez, C., Lu, H., Elgueta, S., Chen, H., & Boshkoska, B. M. (2019). Blockchain Technology in Agri-Food Value Chain Management: A Synthesis of Applications, Challenges and Future Research Directions. Computers in Industry, 109, 83–99. https://doi.org/10.1016/j.compind.2019.04.002
30. Lezoche, M., Hernandez, J. E., Diaz, M. M. E. A., Panetto, H., & Kacprzyk, J. (2020). Agri-Food 4.0: A Survey of the Supply Chains and Technologies for the Future Agriculture. Computers in Industry, 117, 103187. https://doi.org/10.1016/j.compind.2020.103187
31. Queiroz, M. M., Telles, R., & Bonilla, S. H. (2020). Blockchain and Supply Chain Management Integration: A Systematic Review of the Literature. Supply Chain Management: An International Journal, 25(2), 241–254. https://doi.org/10.1108/SCM-03-2018-0143
32. Liu, Y., Ma, X., Shu, L., Hancke, G. P., & Abu-Mahfouz, A. M. (2021). From Industry 4.0 to Agriculture 4.0: Current Status, Enabling Technologies, and Research Challenges. IEEE Transactions on Industrial Informatics, 17(6), 4322–4334. https://doi.org/10.1109/TII.2020.3003910
33. Zou, W., Lo, D., Kochhar, P. S., Le, X. B. D., Xia, X., Feng, Y., Chen, Z., & Xu, B. (2021). Smart Contract Development: Challenges and Opportunities. IEEE Transactions on Software Engineering, 47(10), 2084–2106. https://doi.org/10.1109/TSE.2019.2942301
34. Khan, S. N., Loukil, F., Ghedira-Guegan, C., Benkhelifa, E., & Bani-Hani, A. (2021). Blockchain Smart Contracts: Applications, Challenges, and Future Trends. Peer-to-Peer Networking and Applications, 14(5), 2901–2925. https://doi.org/10.1007/s12083-021-01127-0
35. Nizamuddin, N., Salah, K., Azad, M. A., Arshad, J., & Rehman, M. H. (2019). Decentralized Document Version Control Using Ethereum Blockchain and IPFS. Computers & Electrical Engineering, 76, 183–197. https://doi.org/10.1016/j.compeleceng.2019.03.014
36. Feng, H., Wang, X., Duan, Y., Zhang, J., & Zhang, X. (2020). Applying Blockchain Technology to Improve Agri-Food Traceability: A Review of Development Methods, Benefits and Challenges. Journal of Cleaner Production, 260, 121031. https://doi.org/10.1016/j.jclepro.2020.121031
37. Antonucci, F., Figorilli, S., Costa, C., Pallottino, F., Raso, L., & Menesatti, P. (2019). A Review on Blockchain Applications in the Agri-Food Sector. Journal of the Science of Food and Agriculture, 99(14), 6129–6138. https://doi.org/10.1002/jsfa.9912
38. Zheng, Z., Xie, S., Dai, H. N., Chen, X., & Wang, H. (2018). Blockchain Challenges and Opportunities: A Survey. International Journal of Web and Grid Services, 14(4), 352–375. https://doi.org/10.1504/ijwgs.2018.095647
39. Bao, H., & Roubaud, D. (2022). Non-Fungible Token: A Systematic Review and Research Agenda. Journal of Risk and Financial Management, 15(5), 215. https://doi.org/10.3390/jrfm15050215
40. Wang, L., Sun, W., Zhao, J., Zhang, X., Lu, C., & Luo, H. (2024). A non-fungible token and blockchain-based cotton lint traceability solution. Applied Sciences, 14(4), 1610. https://doi.org/10.3390/app14041610
41. Oliva, G., Lazaro, A., Melero, F. J., & Miralles-Cuevas, S. (2021). NFT-Based Certification Framework for Sustainable Agriculture Products Using Smart Contracts. Sustainability, 13(24), 13862.
42. Tarazona, G., Chaparro-Beltran, R., & Guerrero, C. D. (2021). Advances of Emerging Technologies and Their Impact on Agri-Food Supply Chain Traceability. Sustainability, 13(24), 13948.
43. Sri Vigna Hema, V., & Manickavasagan, A. (2026). Blockchain non-fungible tokens for secure, sustainable, and traceable food systems in Web 3.0 era: A review of NFT marketplaces, sectoral applications, global adoption, potential benefits, and challenges. International Journal of System Assurance Engineering and Management. https://doi.org/10.1007/s13198-025-03129-0
44. Biørn-Hansen, A., Majchrzak, T. A., & Grønli, T.-M. (2017). Progressive Web Apps: The Possible Web-native Unifier for Mobile Development. In Proceedings of the 13th International Conference on Web Information Systems and Technologies (WEBIST 2017), pp. 344–351. SciTePress. https://doi.org/10.5220/0006353703440351

45. Malavolta, I., Procaccianti, G., Noorland, P., & Vukmirović, P. (2017). Assessing the Impact of Service Workers on the Energy Efficiency of Progressive Web Apps. In Proceedings of the 4th IEEE/ACM International Conference on Mobile Software Engineering and Systems (MOBILESoft 2017), pp. 35–45. IEEE. https://doi.org/10.1109/MOBILESoft.2017.7
46. Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Liu, S. (2024). RepViT: Revisiting Mobile CNN From ViT Perspective. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2024), pp. 15909–15920. IEEE.
47. Li, X., Xu, F., & Liu, Y. (2025). MS-RepViT: A Multi-Scale Feature Enhanced Reparameterized Vision Transformer for Breast Cancer Histopathology Image Classification. 2025 6th International Conference on Computer Vision, Image and Deep Learning (CVIDL), 01–05. https://doi.org/10.1109/cvidl65390.2025.11085561
48. Nguyen, A., Yosinski, J., & Clune, J. (2015). Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 427–436). https://doi.org/10.1109/CVPR.2015.7298640
49. Hendrycks, D., & Gimpel, K. (2017). A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. In International Conference on Learning Representations (ICLR 2017).
50. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 IEEE International Conference on Computer Vision (ICCV), 618–626. https://doi.org/10.1109/iccv.2017.74
51. Khan, A., Litchfield, A., Alabdulatif, A., & Khan, F. (2025). BlockPres IPFS: performance evaluation of blockchain based secure patients prescription record storage using IPFS for smart prescription management system. Cluster Computing, 28(4). https://doi.org/10.1007/s10586-024-05054-6
52. Jaya, A. (2025). Decentralized Data Storage Using IPFS for Sustainable Blockchain Availability Improvement. Blockchain Frontier Technology. https://doi.org/10.34306/bfront.v5i2.860
53. Daniel, E., & Tschorsch, F. (2022). IPFS and Friends: A Qualitative Comparison of Next Generation Peer-to-Peer Data Networks. IEEE Communications Surveys & Tutorials, 24(1), 31–52. https://doi.org/10.1109/COMST.2022.3143147
54. Trung, P. H. T., Bang, L. K., & Ngan, N. T. K. (2026). Enhancing Intellectual Property Rights Management Through Blockchain Technology, Smart Contracts, and IPFS: An Integrative Approach. Parallel and Distributed Computing, Applications and Technologies, 444–458. https://doi.org/10.1007/978-981-95-9846-5_36
55. Appel, A. W. (2015). Verification of a cryptographic primitive: SHA-256 (abstract). Proceedings of the 36th ACM SIGPLAN Conference on Programming Language Design and Implementation, 153. https://doi.org/10.1145/2737924.2774972
56. Li, A., Du, D., Zhu, M., & Yang, J. (2025). A lightweight blockchain computing method based on DAG parallel verification and improved SHA-256 optimization. 2025 China Automation Congress (CAC), 7349–7354. https://doi.org/10.1109/cac67268.2025.11486733
57. Shrivastava, P., Sinha, V., & Tak, D. (2024). A theoretic framework for NFT minting process. 2nd International Conference Series on Science, Engineering, and Technology (Icsset) 2022, 3167, 30006. https://doi.org/10.1063/5.0221603
58. Kumare, A. (2025). AI based NFT Minting using Blockchain. International Scientific Journal of Engineering and Management, 04(04), 1–7. https://doi.org/10.55041/isjem03299
59. Aishwarya, B., & Sriramya, P. (2025). Decentralized and Immutable Record-Keeping for Data Privacy-Preserving in Healthcare Using Homomorphic Encryption in Blockchain. 2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), 1–7. https://doi.org/10.1109/iacis65746.2025.11211455
60. Israni, D. K., & Shah, M. K. (2023). Blockchain: A Decentralized, Persistent, Immutable, Consensus, and Irrevocable System in Healthcare. Blockchain for Healthcare 4.0, 48–71. https://doi.org/10.1201/9781003408246-3
61. Hauck, R. (2021). Blockchain, Smart Contracts and Intellectual Property. Using distributed ledger technology to protect, license and enforce intellectual property rights. Legal Issues in the Digital Age, 2(1), 17–41. https://doi.org/10.17323/2713-2749.2021.1.17.41
62. Barz, B., & Denzler, J. (2020). Do We Train on Test Data? Purging CIFAR of Near-Duplicates. Journal of Imaging, 6(6), 41. https://doi.org/10.3390/jimaging6060041
63. Shrivastava, A., Gupta, A., & Girshick, R. (2016). Training Region-Based Object Detectors with Online Hard Example Mining. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 761–769). https://doi.org/10.1109/CVPR.2016.89
64. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. In Advances in Neural Information Processing Systems (NeurIPS) 30.
65. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & Houlsby, N. (2021). An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. In International Conference on Learning Representations (ICLR 2021).
66. Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., & Jégou, H. (2021). Training Data-Efficient Image Transformers & Distillation Through Attention. In Proceedings of the 38th International Conference on Machine Learning (ICML), PMLR 139, 10347–10357.
67. Mehta, S., & Rastegari, M. (2022). MobileViT: Light-Weight, General-Purpose, and Mobile-Friendly Vision Transformer. In International Conference on Learning Representations (ICLR 2022).
68. Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022). Transformers in Vision: A Survey. ACM Computing Surveys, 54(10s), 1–41. https://doi.org/10.1145/3505244
69. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4510–4520.
70. Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., & Sun, J. (2021). RepVGG: Making VGG-Style ConvNets Great Again. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 13733–13742.
71. Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191
72. Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. In International Conference on Learning Representations (ICLR).
73. Loshchilov, I., & Hutter, F. (2019). Decoupled Weight Decay Regularization. In International Conference on Learning Representations (ICLR).

74. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2818–2826.
75. Shorten, C., & Khoshgoftaar, T. M. (2019). A Survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6, 60. https://doi.org/10.1186/s40537-019-0197-0
76. Izza, R., Dafik, Agustin, I. H., Ridlo, Z. R., Baihaki, R. I., Fatahillah, A., & Muttakin, F. N. (2026). Cocoa Beans Roasting Classification using Hybrid Multi-Objective Golden Eagle Optimiser and Graph Convolutional Networks. Statistics, Optimization & Information Computing, 16(2). https://doi.org/10.19139/soic-2310-5070-3188
77. Liang, S., Li, Y., & Srikant, R. (2018). Enhancing the Reliability of Out-of-Distribution Image Detection in Neural Networks. In International Conference on Learning Representations (ICLR).
78. Szabo, N. (1997). Formalizing and Securing Relationships on Public Networks. First Monday, 2(9). https://doi.org/10.5210/fm.v2i9.548
79. Gilbert, H., & Handschuh, H. (2004). Security Analysis of SHA-256 and Sisters. In Selected Areas in Cryptography (SAC 2003), Lecture Notes in Computer Science, vol. 3006, pp. 175–193. Springer. https://doi.org/10.1007/978-3-540-24654-1_13
80. Wilson, E. B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. Journal of the American Statistical Association, 22(158), 209–212. https://doi.org/10.1080/01621459.1927.10502953
81. Lee, K., Lee, K., Lee, H., & Shin, J. (2018). A Simple Unified Framework for Detecting Out-of-Distribution Samples and Adversarial Attacks. In Advances in Neural Information Processing Systems (NeurIPS) 31, 7167–7177.
82. Liu, W., Wang, X., Owens, J., & Li, Y. (2020). Energy-Based Out-of-Distribution Detection. In Advances in Neural Information Processing Systems (NeurIPS) 33, 21464–21475.
83. Hendrycks, D., Mazeika, M., & Dietterich, T. (2019). Deep Anomaly Detection with Outlier Exposure. In International Conference on Learning Representations (ICLR 2019).
84. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On Calibration of Modern Neural Networks. In Proceedings of the 34th International Conference on Machine Learning (ICML), PMLR 70, 1321–1330.
85. Müller, R., Kornblith, S., & Hinton, G. E. (2019). When Does Label Smoothing Help? In Advances in Neural Information Processing Systems (NeurIPS), 32.
86. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90
87. Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., & Adam, H. (2019). Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2019), pp. 1314–1324. IEEE.
88. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning (ICML 2019), pp. 6105–6114.
89. Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 11966–11976). https://doi.org/10.1109/CVPR52688.2022.01167
90. Kornblith, S., Shlens, J., & Le, Q. V. (2019). Do Better ImageNet Models Transfer Better? In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019), pp. 2661–2671. IEEE. https://doi.org/10.1109/CVPR.2019.00277
91. Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How Transferable Are Features in Deep Neural Networks? In Advances in Neural Information Processing Systems (NeurIPS) 27, pp. 3320–3328.
92. Barbedo, J. G. A. (2018). Factors Influencing the Use of Deep Learning for Plant Disease Recognition. Biosystems Engineering, 172, 84–91. https://doi.org/10.1016/j.biosystemseng.2018.05.013
93. Caldarelli, G. (2020). Understanding the Blockchain Oracle Problem: A Call for Action. Information, 11(11), 509. https://doi.org/10.3390/info11110509
94. Khan, S., Haleem, A., Husain, Z., Samson, D., & Pathak, R. D. (2023). Barriers to blockchain technology adoption in supply chains: the case of India. Operations Management Research, 16(2), 668–683. https://doi.org/10.1007/s12063-023-00358-z
95. Rizq Moustafa, F., & Moustafa Aldabousi, A. (2026). Enhancing shareholder democracy through blockchain and decentralized autonomous organizations: A systematic review. Problems and Perspectives in Management, 24(1), 550–561. https://doi.org/10.21511/ppm.24(1).2026.36
96. Du, L., Ho, A. T. S., & Cong, R. (2020). Perceptual Hashing for Image Authentication: A Survey. Signal Processing: Image Communication, 81, 115713. https://doi.org/10.1016/j.image.2019.115713
97. Atzei, N., Bartoletti, M., & Cimoli, T. (2017). A Survey of Attacks on Ethereum Smart Contracts (SoK). In Principles of Security and Trust (POST 2017), Lecture Notes in Computer Science, 10204, 164–186. https://doi.org/10.1007/978-3-662-54455-6_8
98. Wüst, K., & Gervais, A. (2018). Do You Need a Blockchain? In 2018 Crypto Valley Conference on Blockchain Technology (CVCBT) (pp. 45–54). https://doi.org/10.1109/CVCBT.2018.00011
99. Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. International Journal of Human–Computer Interaction, 24(6), 574–594. https://doi.org/10.1080/10447310802205776
100. Daian, P., Goldfeder, S., Kell, T., Li, Y., Zhao, X., Bentov, I., Breidenbach, L., & Juels, A. (2020). Flash Boys 2.0: Frontrunning in Decentralized Exchanges, Miner Extractable Value, and Consensus Instability. In 2020 IEEE Symposium on Security and Privacy (SP) (pp. 910–927). https://doi.org/10.1109/SP40000.2020.00040
101. Finck, M. (2018). Blockchains and Data Protection in the European Union. European Data Protection Law Review, 4(1), 17–35. https://doi.org/10.21552/edpl/2018/1/6