International Journal of Safety and Security Engineering
Vol. 16, No. 9, September, 2026, pp. 2133-2141
Journal homepage: http://iieta.org/journals/ijsse

Evidence-Gated Security and Performance Evaluation of Hash-Chain-Audited Biometric Verification for Smart Parking

Adli A. Nababan[1*], Miftahul Jannah[2], Jefri J. Pangaribuan[1]

[1] Information Systems Department, School of Information Systems, Bina Nusantara University, Jakarta 11480, Indonesia
[2] Department of Computer Engineering and Informatics, Politeknik Negeri Medan, Medan 20155, Indonesia

Corresponding Author Email: adli.nababan@binus.ac.id

Copyright: ©2026 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/ijsse.160916

Received: 31 July 2026
Revised: 12 September 2026
Accepted: 20 September 2026
Available online: 30 September 2026

Keywords:
biometric verification, evidence gating, hash-chain auditing, security-performance evaluation, smart parking, verification consistency

ABSTRACT

Smart parking services must verify authorized field officers and preserve auditable records, yet small biometric samples and isolated ledger tests can be overstated. This study presents an evidence-gated evaluation of a MobileFaceNet-AdaFace cascade (quality-aware calibrated open-set face-identification workload (QAC-OSFI)) and permissioned hash-chain logging. Recognition evaluation comprised 1,675 repeated placements from 554 unique images (516 known and 38 unknown), representing 21 known and 20 unknown identities; five seeds repartitioned the same collection. QAC-OSFI used the fast route for 94.21% of placements and the expert rejection route for 5.79%, yielding 1,672/1,675 correct placement-level decisions. The isolated ledger benchmark comprised 96 configurations and 2,976 observations. Composite latency combined stored recognition timings with separately measured ledger stages and was not a live camera-to-audit measurement. On a reused 19-configuration validation subset, main-effects and interaction models yielded append-latency R^2 values of 0.585 and 0.629 and Mean Absolute Percentage Error (MAPE) values of 45.33% and 40.07%, respectively. Because this subset informed interaction selection, these scores are exploratory, not unbiased independent-test estimates. A provenance audit excluded validation-latency, storage, and throughput optimization because materialized chain length was unverified, storage was cumulative, and throughput was derived. Five unrehashed record/linkage inconsistencies were detected by a trusted verifier; two structural attempts were prevented by database constraints, with one trial per scenario. These checks do not establish protection against coordinated record rewriting and subsequent hash recomputation. The findings support bounded consistency-check and append-contention characterization, not population-wide biometric validity, general tamper-evident security, or deployment optimality.

1. INTRODUCTION

Urban parking management is increasingly digitized through mobile applications, sensing, electronic payment, and data-driven platforms. However, many systems optimize space availability, payment, or traffic flow without treating field-officer identity and verification-record integrity as linked security-engineering requirements. Recent smart-parking reviews describe the transition toward integrated services [1], while Indonesian implementations demonstrate the operational relevance of roadside supervision and public accountability [2]. In the Medan use case, a mobile face-verification service is intended to help distinguish registered parking officers from unknown persons, but the resulting decision must also remain auditable after the transaction.

Biometric verification and audit logging are treated here as complementary controls with different responsibilities. The recognition subsystem estimates identity membership, whereas the audit subsystem checks the internal consistency and linkage of stored decisions under a trusted verifier. The ledger does not improve recognition accuracy, and the recognition model does not prevent later record modification. Unkeyed hash linkage alone does not authenticate the stored history against a writer who can replace records and recompute subsequent hashes. The security interpretation therefore depends on the stated adversary, trust assumptions, and evidence boundary.

This distinction is necessary because the available evidence is heterogeneous. Recognition and ledger timing were collected in separate experiments; the biometric data contain repeated placements of a small local identity set; and the ledger benchmark is a sparse fractional allocation rather than a balanced factorial experiment. Moreover, several archived variables were later found to be unsuitable as independent optimization outcomes. A safety- and security-oriented analysis should expose these limitations rather than convert them into deployment claims.

Accordingly, this study addresses three questions: (1) what within-dataset information about route utilization, errors, and recognition latency is supported by the repeated biometric placements; (2) how do the configured ledger mechanism and worker count relate to directly measured append latency,

including interaction sensitivity and exploratory scores on a reused validation subset; and (3) which isolated record-integrity inconsistencies were detected by the trusted verifier, which structural operations were prevented by database constraints, and which adversarial conditions remain untested?

The paper contributes an explicit evidence and threat boundary, a reproducible specification of the quality-aware calibrated open-set face-identification workload (QAC-OSFI), an audit of measurement provenance and experimental allocation, an exploratory interaction-aware append-latency analysis, and a mechanism-specific account of integrity outcomes. The scope is limited to observed consistency checks and measured append-performance associations. Neither independent predictive validation nor resistance to a consistently rewritten hash-chain history is established.

2. RELATED WORK

2.1 Smart parking and auditable identity

Contemporary smart-parking research combines sensing, prediction, routing, payment, and management services [1]. The Surabaya cashless-parking study shows that digitalization also changes institutional control and service accountability [2]. For public-facing parking operations, the verification question is not only whether a parking space is available, but whether the field actor is an enrolled and authorized person. Blockchain-enhanced vision reviews identify image provenance and tamper-resistant audit as underdeveloped integration opportunities [3], while blockchain-secured mobility records demonstrate that a limited permissioned record can be more appropriate than a public cryptocurrency architecture [4].

2.2 Quality-aware face recognition

AdaFace adapts the angular margin to sample quality and provides a strong always-on workload [5]. QMagFace [6], CR-FIQA [7], CLIB-FIQA [8], eDifFIQA [9], and transformer-based face-image quality assessment [10] show that quality can inform confidence or route selection. Edge-deployment analysis further confirms that accuracy must be considered jointly with resource cost [11]. A survey of face-image quality [12], ISO/IEC 29794-5:2025 [13], and NIST defect-vector evaluation [14] motivate explicit quality definitions and cautious interpretation. The present study uses quality-aware routing as a workload component; it does not infer population-wide biometric validity from repeated local placements.

2.3 Tamper-evident logging and performance modelling

Tamper-evident logging can use forward-secure authentication, chained digests, ratcheting, or storage-assisted mechanisms. LAVA formalizes authenticated log verification [15], SealFSv2 combines storage protection with ratcheting [16], and hash-chain data tracking illustrates linked-record verification [17]. Permissioned-ledger performance is configuration dependent: prior studies model Hyperledger Fabric scalability [18], characterize latency for Internet-of-Things workloads [19, 20], and compare platform throughput [21]. These findings support component-level measurement, but they do not justify treating derived throughput, cumulative storage snapshots, or unverified chain-size labels as independent deployment metrics.

2.4 Security-performance evidence in related systems

Blockchain-based integrity-auditing literature emphasizes that security guarantees and operational efficiency must be evaluated together [22], while identity-management research emphasizes privacy, access control, and scalability [23-25]. Blockchain and biometric hashing have been integrated for lightweight IoT authentication [26], while blockchain-based controls have been used to protect digital-evidence workflows [27]. The unresolved issue addressed here is methodological: integrated security systems require explicit attribution of each outcome to its control mechanism and measurement source, particularly when recognition, audit processing, and threat trials are not collected as one live transaction.

3. SYSTEM AND SECURITY MODEL

3.1 Evidence and threat boundary

Figure 1 depicts the actual evidence boundary. Recognition replay and audit benchmarking were conducted as separate experiments. Stored recognition timings were combined offline with isolated ledger timings to form a composite benchmark value; this value is not a synchronously measured mobile-camera-to-audit transaction. The evidence-admissibility gate retains only outcomes whose provenance supports the stated analysis. For record integrity, the evaluated evidence is detection of unrehashed inconsistencies by a trusted verifier, not authentication of history against a database writer who recomputes the affected hash chain. Table 1 summarizes the evaluated threats, implemented controls, observed evidence, and explicit boundaries.

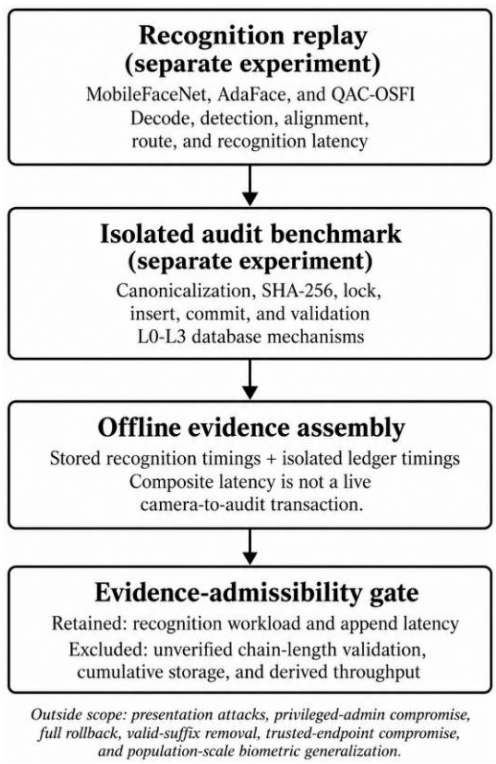


Figure 1. Measurement provenance, security boundary, and evidence-admissibility gate

Table 1. Threat model, implemented controls, and tested boundaries

Threat/Security Objective	Implemented Control	Observed Evidence	Boundary
Unknown or incorrect identity acceptance	Quality-aware fast route with rejection-oriented expert fallback	Repeated local placements; route and error audit	No population, camera/session-disjoint, liveness, or presentation-attack claim
Protected-field modification	Canonical payload, SHA-256 data hash, and block hash	One isolated trial each for payload, data-hash, block-hash, and previous-hash changes	Unrehashed inconsistencies only; no resistance to coordinated record rewriting and hash recomputation
Chain-continuity disruption	Previous-hash linkage and block-number continuity	One deletion trial	No full rollback, valid-suffix removal, or deletion followed by consistent suffix rehashing
Invalid insertion or order exchange	Database uniqueness and referential constraints	One attempted insertion and one order swap were prevented	Prevention is not attributed to hash-chain verification

Note: Secure Hash Algorithm 256-bit (SHA-256).

3.2 Frozen quality-aware calibrated open-set face-identification workload specification

MobileFaceNet is the fast backbone and AdaFace IR-50 is the expert backbone. All embeddings are L2-normalized. For each aligned enrollment image, the scalar quality score combines YuNet detection confidence d, normalized Laplacian sharpness s, and normalized face-size ratio r. The result is clipped to $[0,1]$ and rounded to four decimal places; a detected face is rejected before embedding when either bounding-box side is below 60 pixels.

$$q = clip(0.45d + 0.35s + 0.20r, 0,1) \quad (1)$$

Enrollment images are combined using quality-weighted prototypes with the frozen temperature $\beta = 2$. Only ENROLLMENT images contribute to a prototype; VALIDATION and UNKNOWN_VALIDATION data are used for calibration and thresholds, whereas PROBE and UNKNOWN_PROBE data are reserved for reporting.

$$w_{ij} = \frac{exp(2q_{ij})}{\sum_k exp(2q_{ik})}, p_i = normalize(\sum_j w_{ij} e_{ij}) \quad (2)$$

The fast-stage calibrator is deterministic L2-regularized logistic regression over the standardized feature vector $z_F = [s_{(1)}, \Delta, q, \log(1 + n_{i^*})]$, where $s_{(1)}$ is the top cosine score, Δ is the top-1/top-2 margin, and n_{i^*} is the enrollment-support count of the leading candidate. Standardization uses the population means and standard deviation with a 10^{-6} floor. Full-batch gradient descent is run for 800 iterations with a learning rate of 0.05 and an L2 coefficient of 0.001.

$$\pi_F = \sigma(w_F{}^T z_F + b_F) \quad (3)$$

For seeds 42, 101, 202, 303, and 404, the fast thresholds are 0.156261, 0.236002, 0.126669, 0.188091, and 0.122805, respectively. Each threshold equals the maximum unknown-validation probability plus 10^{-8} and is therefore marked LOW_SUPPORT. A query is accepted on the fast route only when $q \geq 0.40$, $\pi_F \geq \tau_F$, and $\Delta \geq 0.02$; $q < 0.40$ yields RETRY, and other cases proceed to AdaFace.

The final calibrator uses standardized fast/expert top scores and margins plus quality. It follows the same optimization procedure, with comparison operator $\geq$. The stored final threshold is 1.00000001 for every seed, so an expert-routed probe cannot satisfy the acceptance condition and the expert branch is rejection-only in this frozen evaluation. No gallery-dispersion or cross-model-agreement feature is used. No expert-assisted correction, degradation, or identity change occurred. The measured expert-fallback probability is

$$p_f = \frac{97}{1{,}675} = 0.0579$$

3.3 Permissioned audit record

The audit service recursively sorts associative payload keys before canonical JavaScript Object Notation (JSON) serialization. The data hash is computed as the SHA-256 digest of the canonical payload:

$$h_{data} = SHA256(canonical(payload)) \quad (4)$$

The block hash is computed from the length-prefixed block number, previous hash, data hash, and creation time. Length-prefix encoding prevents delimiter ambiguity; it does not authenticate the fields against an authorized or unauthorized database writer:

$$h_{block} = SHA256(enc(b) \parallel enc(h_{prev}) \parallel enc(h_{data}) \parallel enc(t)) \quad (5)$$

L0 is a conventional database log, L1 stores an independent record hash, L2 links sequential records, and L3 adds stored checkpoints. Full-chain validation recomputes the data and block hashes and verifies previous-hash linkage, continuity, schema version, and the verification-record link. These operations check the internal consistency of the presented records. SHA-256 is unkeyed: a writer able to modify records and recompute all affected subsequent hashes can construct a self-consistent replacement history without finding a hash collision. The archived evaluation establishes neither an independently authenticated chain head nor protection of such a reference outside the writable database. The source audit found checkpoint creation but no dedicated production checkpoint-validation command; stored checkpoints are not treated as authenticated trust anchors. Accordingly, the experiments support detection of the tested unrehashed inconsistencies only. They do not establish general tamper-evident security against coordinated rewriting, and checkpoint-validation and analytical break-even claims remain excluded.

4. EXPERIMENTAL DESIGN AND ANALYSIS

4.1 Recognition protocol and statistical dependence

Five deterministic seeds (42, 101, 202, 303, and 404) repartitioned the same local identity collection. Each seed used 21 enrolled identities, 315 known probes, and 20 unknown probes, but the placements are not independent populations. Across all seeds, 1,675 primary placements reduce to 554 unique source-image hashes: 516 known images and 38 unknown images representing 20 unknown identities. No verified camera, capture-session, time, or location field is available. Recognition outcomes are therefore reported as within-dataset counts, with repeated source-image errors identified explicitly, not as population estimates. Table 2 summarizes the identity, image, placement, and seed support underlying this interpretation.

4.2 Measurement provenance

The audit benchmark ran against the isolated schema parkirverify_benchmark rather than production records. The environment used an AMD64 platform with 16 logical processors, Windows build 10.0.26200, Python 3.11.9, PyMySQL 2.2.8, and the monotonic time.perf_counter_ns timer. The frozen application source revision was 4ac83848b0dd6d050a890a61bc14bc0b614081d8.
Recognition replay and audit-stage benchmarking were separate experiments; the composite response-time field is therefore an offline analytical combination rather than a live camera-to-audit measurement. Table 3 distinguishes direct measurements from derived or non-admissible fields.

Table 2. Recognition evidence supports and dependence structure

Item	Observed Support	Interpretation
Known identities	21	Same enrolled identity set across seeds
Unknown identities	20	Local unknown population
Unique known images	516	Repeated across deterministic placements
Unique unknown images	38	One unknown-image outcome changes the image-level rate by 2.63 percentage points
Primary placements	1,675	1,575 known + 100 unknown; not 1,675 independent subjects
Five seeds	42, 101, 202, 303, 404	Repeated partition placements, not external validation

Table 3. Measurement provenance and admissible use

Metric	Provenance	Admissible Use
Recognition latency	Stored recognition replay/snapshot timings	Recognition-service workload only
Composite benchmark latency	Offline sum of stored recognition and isolated ledger stages	Descriptive only; not a live mobile transaction
Append latency	Direct monotonic-timer measurement in isolated database	Primary quantitative ledger outcome
Validation latency	Benchmark timing with nominal chain_length field	Diagnostic only; excluded from optimization
Throughput_rps	Exactly 10^9/composite_latency_ns in all 2,880 warm rows	Not an independent throughput measurement
Storage_bytes	Fourteen cumulative database-size snapshots	Excluded from configuration comparison

Note: Requests per second (rps); nanoseconds (ns). The suffixes _rps, _ns, and _bytes indicate the measurement units of the corresponding variables.

4.3 Frozen fractional allocation and reused validation subset

The reference universe contains 11,520 possible factor combinations, of which 96 configuration keys were executed. The archive preserves the complete allocation and historical "holdout" flags but not the original generator code or randomization seed; the study therefore does not claim an optimal, random, orthogonal, or balanced selection algorithm. The realized 96-key allocation is published as configuration_allocation.csv and summarized in Appendix. The historical flag is an archival label, not evidence of an independent test set.

The archived dataset assigns 77 configuration keys to coefficient estimation and 19 to the historical "holdout" subset. Although the 19 keys were excluded from coefficient fitting, their results were inspected when deciding whether to retain interaction terms. They therefore served model selection and are termed the reused validation subset in this paper. They cannot provide unbiased independent-test performance estimates. No additional untouched test set is established by the archive. The allocation-generator code, randomization seed, and time-stamped preregistration are unavailable. Every key has one first-run observation and 30 warm repetitions, giving 2,976 successful rows. The full-factor categorical allocation matrix is full rank (21/21) with condition number 10.89, but selected two-factor cells contain only 2-16 configurations. Unequal factor counts limit the coefficients to adjusted associations rather than isolated causal effects.

4.4 Explanatory models and interaction sensitivity

Warm observations were summarized by the configuration median. The archived main-effects specification used categorical recognition and ledger terms together with log-scaled engineering factors. HC3 heteroscedasticity-consistent intervals were used. The primary outcome is directly measured append latency Y; composite and validation-latency outcomes are retained only as diagnostic records of poor model fit on the comparison subset. Seed was not entered into the append model because it labels the separate recognition replay rather than the independently timed audit append stage; its unequal allocation is nevertheless disclosed in Appendix.

$$\begin{aligned} log_{10}Y &= \theta_0 + \theta_R + \theta_L \\ &+\theta_n log_{10}n + \theta_c log_{10}c \\ &+\theta_P log_{10}P + \theta_k log_{10}k + \varepsilon \end{aligned} \tag{6}$$

Because contention may depend on ledger architecture and payload, a sensitivity model adds ledger-by-concurrency and ledger-by-payload interactions. In Eq. (7), $I(\cdot)$ denotes the dummy indicator for L1-L3 relative to L0; "main effects" denotes the systematic terms of Eq. (6), excluding its residual:

$$\log_{10} Y = \text{main effects} + \sum_{\ell \in \{L1,L2,L3\}} \eta_{\ell c} I(L = \ell) \log_{10} c + \sum_{\ell \in \{L1,L2,L3\}} \eta_{\ell P} I(L = \ell) \log_{10} P + \varepsilon \quad (7)$$

MAE, RMSE, MAPE, and R^2 were calculated on the original millisecond scale for the reused 19-configuration validation subset. This subset informed the decision to retain interaction terms; the resulting scores are therefore exploratory model-comparison quantities, not unbiased out-of-sample or independent-test estimates. Both specifications are retained in the results for transparent comparison, without claiming that the observed difference confirms generalization. HC3 and fitted-mean intervals are conditional on the fitted specification and do not adjust for this model-selection step. Independent evaluation after the model is frozen is required before a confirmatory predictive claim can be made.

4.5 Integrity-trial protocol

Seven prescribed actions were executed once each on a copied benchmark database. Payload, data-hash, block-hash, previous-hash, and deletion actions were assessed through the trusted verifier. These trials introduced record or linkage inconsistencies; they did not evaluate a coordinated rewrite followed by consistent recomputation of every affected subsequent hash. Insertion and order-exchange attempts raised database integrity errors before a hash-chain verification decision was produced. The observed outcomes are single-scenario consistency detections or database preventions, not a detection-rate estimate or proof of history authenticity. Rehashing attacks by database writers, full rollback, valid-suffix removal, malicious administrators with database and key access, trusted-endpoint compromise, and presentation attacks were not tested.

5. RESULTS

5.1 Recognition workload characterization

Figure 2 visualizes the corresponding seed-level latency and Table 4 summarizes the recognition-workload statistics. QAC-OSFI reduced mean recognition-service latency by 49.26% relative to the reported AdaFace workload, but its P95 was similar and its P99 was higher. It used the fast route for 1,578/1,675 placements (94.21%) and the expert rejection route for 97/1,675 (5.79%). All 1,575 known placements were accepted through the fast route. Among 100 unknown placements, 97 were routed to the expert branch and rejected; three remained on the fast route and were falsely accepted.

The three placement-level false accepts were the same source image from one unknown subject repeated in seeds 101, 303, and 404. Thus, 1,672/1,675 correct placement decisions and 100% known rank-1 accuracy describe this dataset only. The evidence does not establish low-FPIR performance, independent-population accuracy, or external generalization.

5.2 Append-latency model and interaction sensitivity

Table 5 retains the numerical results of both specifications on the same reused validation subset. The append interaction variant had lower comparison-subset MAPE (40.07% versus 45.33%) and higher R^2 (0.629 versus 0.585). Because those results informed interaction selection, this apparent improvement is exploratory and cannot establish superior independent-test performance. Composite and validation-latency scores also remained poor on this subset. Neither specification has an untouched-test estimate in this study, and neither is presented as a deployment predictor.

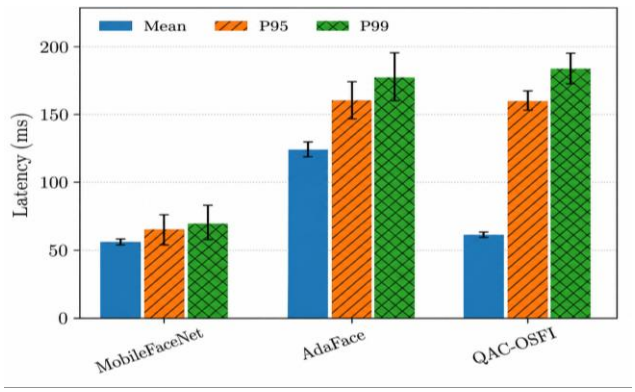


Figure 2. Seed-level recognition latency; error bars denote standard deviations across five repeated seeds

Table 4. Recognition workload summary across repeated seeds (latency: Mean ± standard deviation)

Workload	Mean (ms)	P95 (ms)	P99 (ms)	OSA	R1	UDR
MobileFaceNet	55.48 ± 0.84	64.80 ± 12.25	70.10 ± 16.60	99.76%	100.0%	96/100
AdaFace	122.93 ± 5.32	160.57 ± 19.48	179.38 ± 24.22	99.82%	100.0%	97/100
QAC-OSFI	62.38 ± 0.94	160.19 ± 6.30	185.49 ± 9.88	99.82%	100.0%	97/100

Note: Quality-Aware Calibrated Open-Set Face Identification (QAC-OSFI); milliseconds (ms); 95th percentile latency (P95); 99th percentile latency (P99); Open-Set Accuracy (OSA); Rank-1 Accuracy (R1); Unknown Detection Rate (UDR).

Table 5. Exploratory model-comparison scores on the reused validation subset (19 configuration keys)

Outcome	Model	MAE (ms)	RMSE (ms)	MAPE	R^2
Append	Main effects	34.23	64.12	45.33%	0.585
Append	+ $L \times c$, $L \times P$	33.20	60.66	40.07%	0.629
Composite	Main effects	129.37	185.40	49.22%	-0.966
Composite	+ $L \times c$, $L \times P$	133.08	203.08	46.56%	-1.359
Validation*	Main effects	184.92	368.09	494.00%	-8.115
Validation*	+ $L \times c$, $L \times P$	180.02	340.35	480.23%	-6.793

Notes: The same 19 keys informed interaction selection. All scores are exploratory, not unbiased independent-test estimates. "Validation*" denotes the recorded chain-validation latency, shown only as a diagnostic and excluded from configuration selection. Mean Absolute Error (MAE); Root Mean Squared Error (RMSE); Mean Absolute Percentage Error (MAPE)

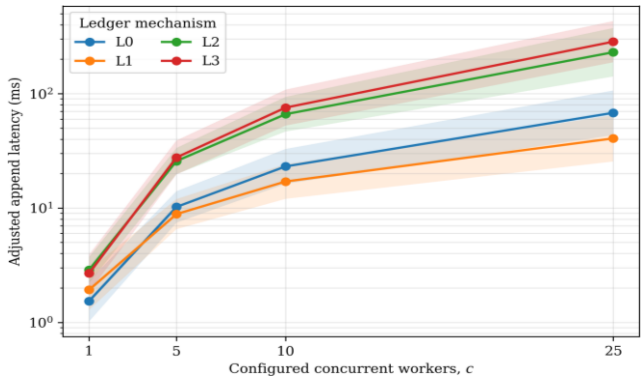


Figure 3. Interaction-aware fitted append latency; bands are nominal 95% fitted-mean intervals conditional on the model specification

At the reference workload, the fitted interaction specification permits different concurrency slopes by ledger mechanism; its 25-worker/one-worker ratios ranged from 21.1 for L1 to 106.0 for L3. Figure 3 illustrates these interaction-aware fitted associations and their model-conditional confidence bands. The fitted-mean bands widen substantially at 25 workers. These model-conditional associations, and the 40.07% MAPE on the reused validation subset, do not account for selection on that subset and must not be treated as independently validated capacity-planning constants.

5.3 Provenance audit and removal of unsupported optimization

The provenance audit identified three metrics that were not admissible for operational optimization. First, throughput_rps is algebraically identical to 10^9 divided by composite latency for every warm row, so it is not an independently observed throughput window. Second, storage_bytes contains only 14 stepwise values and correlates with execution order (Spearman $\rho = 0.996$), indicating cumulative database state rather than configuration-specific storage growth. Third, the archived evidence does not verify that nominal chain_length equals the number of materialized linked records. The audit disposition is summarized in Table 6.

An archived configuration carrying nominal $n = 10^5$ illustrates the final issue: its database-size snapshot was only 98,304 bytes and its 30-repeat validation mean was 1.95 ms. These values cannot support a full-100,000-block validation claim. Consequently, validation-based Pareto points, checkpoint break-even, storage optimization, and global-optimum statements are excluded. This study reports no operational Pareto set.

5.4 Integrity-control outcomes

Five isolated record/linkage inconsistencies were detected by the trusted verifier, and two structural actions were prevented by database constraints. Table 7 reports the mechanism and single-trial latency for transparency. No trial tested a coordinated history rewrite with consistent recomputation of subsequent hashes. These outcomes demonstrate only the specified consistency checks and structural preventions; they establish neither $D = 1$ as a general detection probability nor tamper-evident protection against an attacker with sufficient database write access.

Table 6. Audit-based metric disposition

Candidate Metric	Audit Result	Disposition
Append latency	Directly timed; 30 warm repeats per key	Retained as primary performance outcome
Composite latency	Offline composition of separate experiments	Descriptive only
Validation latency	Materialized chain length not verified	Excluded from operational comparison
Throughput	Exact reciprocal of composite latency	Excluded as non-independent
Storage	Cumulative snapshots tied to run order	Excluded from optimization

Table 7. Single-trial consistency-check and database-prevention outcomes in the copied benchmark database

Action	Control Response	Code	*n*	Latency (ms)
Payload change	Verifier detection	INVALID_DATA_HASH	1	5.47
Data-hash change	Verifier detection	INVALID_DATA_HASH	1	4.87
Block-hash change	Verifier detection	INVALID_PREVIOUS_HASH	1	5.77
Previous-hash change	Verifier detection	INVALID_PREVIOUS_HASH	1	5.02
Block deletion	Verifier detection	MISSING_OR_REORDERED_BLOCK	1	5.14
Block insertion	Database prevention	IntegrityError	1	18.33
Order exchange	Database prevention	IntegrityError	1	1.40

6. DISCUSSION

The recognition results are useful for workload construction but are not an independent biometric validation. The five seeds reuse the same local identity collection, and the 100 unknown placements represent only 38 unique images from 20 identities. The QAC-OSFI specification is now reproducible, including quality weights, calibrator features, thresholds, route guards, and the rejection-only expert branch. Nevertheless, the observed 99.82% placement accuracy must remain dataset-specific.

The distinction between recognition-service timing and ledger timing is equally important. Composite latency was assembled offline from separate evidence sources; it should not be interpreted as a live mobile-camera-to-audit service-level measurement. Direct append timing is the strongest performance evidence. The interaction analysis describes possible ledger-dependent concurrency associations, but the 19-key validation subset also informed model selection. Its errors and apparent improvements are therefore exploratory, not independent generalization evidence. A prospective evaluation on untouched configurations is required after the

model specification is frozen.

The sparse realized allocation covers all factor levels and has a full-rank main-effect matrix, but factor counts and interaction cells are unequal. Publishing the 96-key allocation improves transparency, yet it cannot replace a prospective balanced or D-optimal design with an archived generator and independent validation block. A new experiment should materialize the declared chain length, reset or snapshot the database for every configuration, measure completed requests over fixed throughput windows, and implement a dedicated checkpoint validator.

Security interpretation also depends on the responsible control and adversary. The five verifier outcomes concern broken record/hash/linkage consistency, while database constraints prevented two structurally invalid writes. An attacker with sufficient database write permissions may modify records and recompute the affected suffix, producing a self-consistent chain that passes the same internal checks; compromise of a secret key is not required for unkeyed hashing. The study does not establish an authenticated reference beyond the attacker's rewrite capability, and no such protection is inferred from stored checkpoints. A separately protected and authenticated reference would need implementation and adversarial evaluation before stronger history-authenticity claims could be made. Coordinated rehashing, full rollback, valid-suffix removal, endpoint compromise, credential theft, and presentation attacks remain untested. The current single trials establish neither a general detection probability nor resistance to a consistently rewritten history.

Relative to studies that integrate blockchain with biometric authentication [26] or protect digital-evidence workflows [27], the contribution here is not another authentication protocol or application workflow. It is an evidence-admissibility method that attributes each claim to a threat, control, measurement source, statistical support, and excluded condition, thereby preventing an isolated proof-of-concept benchmark from being presented as deployment readiness.

7. CONCLUSIONS

This study evaluated hash-chain-audited biometric verification within an explicit evidence-admissibility boundary. QAC-OSFI is specified as a MobileFaceNet fast route with quality-aware prototypes, validation-only logistic calibration, and an AdaFace rejection route. Its placement-level results are limited to 554 unique local images and 20 unknown identities. The ledger benchmark contributes 2,976 observations and supports descriptive append-contention analysis. On the reused 19-configuration validation subset, the interaction model yielded $R^2 = 0.629$ and MAPE = 40.07%; because this subset informed model selection, these are exploratory scores rather than unbiased independent-test estimates. A provenance audit excluded unsupported validation-latency, storage, throughput, checkpoint, and Pareto claims. Five unrehashed record/linkage inconsistencies were detected by a trusted verifier, and two structural attempts were prevented by database constraints, each in one trial. These outcomes do not establish tamper-evident protection against coordinated database rewriting and hash recomputation. Broader claims require untouched predictive evaluation after model freezing, independent biometric populations, verified measurement provenance, and repeated adversarial tests with explicitly protected trust references.

AUTHOR CONTRIBUTIONS

Miftahul Jannah: Conceptualization, Methodology, Investigation, Resources, Data curation, Supervision, Funding acquisition, Project administration, Writing – review and editing. Adli A. Nababan: Conceptualization, Methodology, Formal analysis, Software, Validation, Visualization, Project administration, Writing – original draft, Writing – review and editing. Jefri J. Pangaribuan: Software, Investigation, Data curation, Validation, Resources, Visualization, Writing – review and editing.

DATA AVAILABILITY

The anonymized performance measurements, model specifications, archived prediction files, integrity outcomes, environment metadata, and reproducibility documentation are available on Zenodo under DOI 10.5281/zenodo.21615345. The allocation matrix, interaction-sensitivity results, metric-admissibility audit, and tamper-outcome reclassification are supplied as supplementary material. Historical "holdout" labels in archived files denote the 19-key subset subsequently reused for model comparison and selection; they do not identify an untouched test set. Facial images, embeddings, names, identification numbers, and precise locations are not publicly released.

ACKNOWLEDGMENT

The authors acknowledge Bina Nusantara University and Politeknik Negeri Medan for their institutional and in-kind research support. The authors also thank the North Sumatra Provincial Transportation Agency for facilitating field coordination and providing operational context for the smart-parking officer verification study. The authors further thank the participating parking officers for their voluntary participation and cooperation.

REFERENCES

[1] Channamallu, S.S., Kermanshachi, S., Rosenberger, J.M., Pamidimukkala, A. (2026). Smart parking systems: A comprehensive review of digitalization of parking services. Green Energy and Intelligent Transportation, 5(1): 100293. https://doi.org/10.1016/j.geits.2025.100293

[2] Nugroho, B.L., Supeno, E., Asmorowati, S., Hapsari, W. (2024). Implementation of the smart parking system (SPS): A study on cashless payment for public roadside parking fees in Surabaya City. Journal of Governance and Administrative Reform, 5(2): 119-136. https://doi.org/10.20473/jgar.v5i2.61650

[3] Ottakath, N., Al-Ali, A., Al-Maadeed, S., Elharrouss, O., Mohamed, A. (2023). Enhanced computer vision applications with blockchain: A review of applications and opportunities. Journal of King Saud University - Computer and Information Sciences, 35(10): 101801. https://doi.org/10.1016/j.jksuci.2023.101801

[4] Kurniawan, F., Putra, D.P., Hammad, J., Satria Prabuwono, A. (2023). A blockchain-secure mobility data in smart campus. International Journal of Interactive Mobile Technologies, 17(18): 55-66. https://doi.org/10.3991/ijim.v17i18.41823
[5] Kim, M., Jain, A.K., Liu, X. (2022). AdaFace: Quality adaptive margin for face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, USA, pp. 18750-18759. https://doi.org/10.1109/CVPR52688.2022.01808
[6] Terhörst, P., Ihlefeld, M., Huber, M., et al. (2023). QMagFace: Simple and accurate quality-aware face recognition. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, USA, pp. 3473-3483. https://doi.org/10.1109/WACV56688.2023.00348
[7] Boutros, F., Fang, M., Klemt, M., Fu, B., Damer, N. (2023). CR-FIQA: Face image quality assessment by learning sample relative classifiability. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, pp. 5836-5845. https://doi.org/10.1109/CVPR52729.2023.00565
[8] Ou, F.Z., Li, C., Wang, S., Kwong, S. (2024). CLIB-FIQA: Face image quality assessment with confidence calibration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, pp. 1694-1704. https://doi.org/10.1109/CVPR52733.2024.00167
[9] Babnik, Ž., Peer, P., Štruc, V. (2024). eDifFIQA: Towards efficient face image quality assessment based on denoising diffusion probabilistic models. IEEE Transactions on Biometrics, Behavior, and Identity Science, 6(4): 458-474. https://doi.org/10.1109/TBIOM.2024.3376236
[10] Liu, T., Li, S., Xu, M., Yang, L., Wang, X. (2024). Assessing face image quality: A large-scale database and a transformer method. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(5): 3981-4000. https://doi.org/10.1109/TPAMI.2024.3350049
[11] Biswas, A., Patnaik, S.A., Hafez, A.H.A., Namboodiri, A.M. (2023). Characterizing face recognition for resource-efficient deployment on edge. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Paris, France, pp. 1312-1321. https://doi.org/10.1109/ICCVW60793.2023.00141
[12] Schlett, T., Rathgeb, C., Henniger, O., Galbally, J., Fierrez, J., Busch, C. (2022). Face image quality assessment: A literature survey. ACM Computing Surveys, 54(10s): 210. https://doi.org/10.1145/3507901
[13] International Organization for Standardization & International Electrotechnical Commission (ISO/IEC). (2025). ISO/IEC 29794-5:2025 Information Technology — Biometric Sample Quality-Part 5: Face Image Data. Geneva, Switzerland. https://www.iso.org/standard/81005.html.
[14] Yang, J., Grother, P.J., Ngan, M.L., Hanaoka, K., Hom, A. (2023). Face Analysis Technology Evaluation (FATE) Part 11: Face image quality vector assessment: Specific image defect detection. National Institute of Standards and Technology Internal Report 8485. https://doi.org/10.6028/NIST.IR.8485
[15] Bajramovic, E., Fein, C., Frinken, M., Rösler, P., Freiling, F. (2023). LAVA: Log authentication and verification algorithm. Digital Threats: Research and Practice, 4(3): 35. https://doi.org/10.1145/3609233
[16] Guardiola-Múzquiz, G., Soriano-Salvador, E. (2023). SealFSv2: Combining storage-based and ratcheting for tamper-evident logging. International Journal of Information Security, 22: 447-466. https://doi.org/10.1007/s10207-022-00643-1
[17] Kim, S., Kim, D. (2024). Data-tracking in blockchain utilizing hash chain: A study of structured and adaptive process. Symmetry, 16(1): 62. https://doi.org/10.3390/sym16010062
[18] Ke, Z., Park, N. (2023). Performance modeling and analysis of Hyperledger Fabric. Cluster Computing, 26: 2681-2699. https://doi.org/10.1007/s10586-022-03800-2
[19] Lee, S., Kim, M., Lee, J., Hsu, R.H., Kim, M.S., Quek, T.Q.S. (2023). Facing latency of Hyperledger Fabric for blockchain-enabled IoT: Modeling and analysis. IEEE Network, 37(6): 232-239. https://doi.org/10.1109/MNET.120.2200064
[20] Abang, J.E., Takruri, H., Al-Zaidi, R., Al-Khalidi, M. (2024). Latency performance modelling in Hyperledger Fabric blockchain: Challenges and directions with an IoT perspective. Internet of Things, 26: 101217. https://doi.org/10.1016/j.iot.2024.101217
[21] Ucbas, Y., Eleyan, A., Hammoudeh, M., Alohaly, M. (2023). Performance and scalability analysis of Ethereum and Hyperledger Fabric. IEEE Access, 11: 67156-67167. https://doi.org/10.1109/ACCESS.2023.3291618
[22] Han, H., Fei, S., Yan, Z., Zhou, X. (2022). A survey on blockchain-based integrity auditing for cloud data. Digital Communications and Networks, 8(5): 591-603. https://doi.org/10.1016/j.dcan.2022.04.036
[23] Sutradhar, S., Karforma, S., Bose, R., Roy, S., Djebali, S., Bhattacharyya, D. (2024). Enhancing identity and access management using Hyperledger Fabric and OAuth 2.0: A blockchain-based approach for security and scalability for healthcare industry. Internet of Things and Cyber-Physical Systems, 4: 49-67. https://doi.org/10.1016/j.iotcps.2023.07.004
[24] Alanzi, H.M., Alkhatib, M. (2025). Blockchain-based identity management system prototype for enhanced privacy and security. Electronics, 14(13): 2605. https://doi.org/10.3390/electronics14132605
[25] Ahmed, M.R., Islam, A.K.M.M., Shatabda, S., Islam, S. (2022). Blockchain-based identity management system and self-sovereign identity ecosystem: A comprehensive survey. IEEE Access, 10: 113436-113481. https://doi.org/10.1109/ACCESS.2022.3216643
[26] Gudivada, D., Manchiraju, K.R. (2025). A blockchain-biometric lightweight mutual authentication protocol for IoT security. International Journal of Safety and Security Engineering, 15(12): 2589-2602. https://doi.org/10.18280/ijsse.151215
[27] Sunardi, Kusuma, R.S. (2023). Digital evidence security system design using blockchain technology. International Journal of Safety and Security Engineering, 13(1): 159-165. https://doi.org/10.18280/ijsse.130118

NOMENCLATURE

c	configured concurrent workers, dimensionless
d, s, r	detection confidence, normalized sharpness,

	and normalized face-size ratio, dimensionless
q	enrollment-image quality score, dimensionless
e_{ij}, w_{ij}, p_i	normalized embedding, quality weight, and identity prototype, dimensionless
$s_{(1)}$, Δ, n_{i^*}	top cosine score, top-1/top-2 margin, and leading-candidate support count
z_F, w_F, b_F	standardized fast-stage feature vector, coefficient vector, and intercept
p_f, π_F, τ_F	expert-fallback probability, fast knownness probability, and fast threshold, dimensionless
L0-L3	conventional log, independent hash, linked hash chain, and checkpoint-recording chain
n	nominal chain-length factor; not accepted as a verified materialized count
P	canonical payload size, bytes
h_{data}, h_{prev}, h_{block}	data hash, previous-block hash, and block hash
b, t, enc(·)	block number, creation time, and length-prefix encoding operation
Y	directly measured append latency, ms
θ, η, ε	main-effect coefficients, interaction coefficients, and residual term
$I(\cdot)$	dummy indicator function, dimensionless
OSA, R1, UDR	open-set accuracy, rank-1 accuracy, and unknown detection rate
QAC-OSFI	quality-aware calibrated open-set face-identification workload

APPENDIX

Realized configuration allocation

The complete 96-key allocation is provided in configuration_allocation.csv. The marginal factor counts are reported in Table A1 and demonstrate coverage, but not orthogonality. Additional supplementary files report pairwise allocation counts, interaction-model scores on the reused validation subset, metric admissibility, the anomalous validation audit, and tamper-outcome reclassification. Historical "holdout" filenames and flags are retained for provenance and must be interpreted using Section 4.3.

Table A1. Configuration counts by factor level

Factor	Level: Count
Recognition workload	AdaFace: 39; MobileFaceNet: 18; QAC-OSFI: 39
Ledger mechanism	L0: 21; L1: 28; L2: 24; L3: 23
Seed	42: 23; 101: 21; 202: 13; 303: 21; 404: 18
Nominal chain length	10^2: 29; 10^3: 21; 10^4: 22; 10^5: 24
Concurrent workers	1: 30; 5: 21; 10: 22; 25: 23
Payload bytes	512: 33; 2,048: 27; 4,096: 36
Checkpoint interval	100: 24; 500: 20; 1,000: 26; 5,000: 26
Analysis roles	Coefficient estimation: 77; reused validation: 19