



Development and Evaluation of a Real-Time Attendance Prototype Using ArcFace-Based Facial Recognition and Transfer Learning

Oluwabukayo Owolabi

*B.Eng Information and Communication Engineering
Department of Electrical and Information Engineering,
College of Engineering, Covenant University, Ota, Ogun State, Nigeria.
bukayoowolabi20@gmail.com*

Michael O. Adedokun

*Department of Electrical and Information Engineering
School of Applied Engineering, College of Engineering
Covenant University, Canaanland, Ota, Ogun State, Nigeria
michael.adedokun@covenantuniversity.edu.ng*

Abstract: Manual attendance procedures consume instructional time, permit proxy attendance, and produce records that are difficult to verify in real time. This study develops and evaluates a facial-recognition attendance prototype that combines a Streamlit web interface, Redis-based enrollment storage, SCRFD-family face detection, and frozen ArcFace encoders. Two transfer-learning configurations were compared: a ResNet-50 encoder from the InsightFace buffalo_l package and a lightweight MobileFaceNet encoder from buffalo_sc. Images were aligned, converted to 512-dimensional L2-normalized embeddings, and matched by cosine similarity using the deployed threshold of 0.50. The reproducible evaluation used 50 images from five identities; seven images per identity formed enrollment templates, and three were reserved as held-out queries. Both encoders used the same deterministic split, detector size, decision threshold, and CPU execution provider. Both configurations detected all 50 faces and achieved 100.00% closed-set identification accuracy, macro precision, macro recall, and macro F1 on the 15 held-out queries. Across all valid image pairs, the ResNet-50 configuration attained 99.18% verification accuracy, 95.56% recall, and 97.73% F1, compared with 98.20%, 90.22%, and 94.86% for MobileFaceNet. ResNet-50 also produced a larger mean identity-separation margin (0.770 versus 0.713), whereas MobileFaceNet reduced mean CPU processing latency from 999.4 ms to 112.8 ms per image, an 8.86 $\times$ speed-up. The contribution is therefore an evidence-based accuracy-latency analysis of two deployable ArcFace configurations integrated with a working attendance interface. At 10% intensity, CLAHE restored MobileFaceNet accuracy from 93.33% to 100%, and both models identified all 15 faces in five three-person scenes. The small, curated dataset limits generalization; large-scale campus trials, liveness detection, encrypted templates, fairness auditing, and persistent attendance transaction validation remain necessary before controlled operational use.

Keywords: Arc Face; Face Recognition; Real-Time Attendance; Transfer Learning; Mobilefacenet; Redis.

Nomenclature

Abbreviation	Description
AES	Advanced Encryption Standard
ArcFace	Additive Angular Margin Loss
BF	Bilateral filtering
BGR	Blue-green-red channel order used by OpenCV
CLAHE	Contrast-Limited Adaptive Histogram Equalization
CNN	Convolutional neural network
CPU	Central processing unit
DET	Detection-error trade-off curve
SCRFD	Sample and computation redistribution for efficient face detection
FAR	False acceptance rate
FLOPs	Floating-point operations
FPS	Frames processed per second
FRR	False rejection rate
GPU	Graphics processing unit
HE	Global histogram equalization
LBPH	Local Binary Pattern Histogram
LFW	Labeled Faces in the Wild benchmark
NMS	Non-maximum suppression
NPU	Neural processing unit
ONNX	Open Neural Network Exchange model format
PR	Precision-recall curve
ROC	Receiver operating characteristic curve
SSD	Single Shot MultiBox Detector
TLS	Transport Layer Security
UI	User interface
VGG	Visual Geometry Group convolutional architecture

1. Introduction

Attendance recording in universities and workplaces is still frequently based on roll calls, paper registers, cards, or contact biometrics. These approaches interrupt instruction, create opportunities for proxy attendance, and make immediate auditing difficult. Contact-based systems also introduce queues and shared-surface hygiene concerns, while card-based systems authenticate an object rather than the person carrying it [2][3][4][5][6]. As educational institutions continue to adopt digital learning environments, the demand for reliable, automated, and contactless attendance solutions has increased significantly. The governance burden is equally important: any automated system must be accurate, explainable, secure, and proportionate to the institutional purpose [1].

Face recognition offers contactless identity verification, but a convincing attendance system must do more than report a benchmark score. It must detect and align faces in uncontrolled scenes, compute stable embeddings, reject weak matches, retrieve enrollment records quickly, and present results through a usable interface. Achieving these objectives requires balancing recognition accuracy with computational efficiency so that the system remains practical for real-time deployment on commonly available hardware. Earlier attendance architectures commonly employ Haar/LBPH, classical subspace methods, or a single deep model without reporting latency, threshold behavior, or a reproducible local test protocol [3][4][5][6][7][8]. The motivation for this work is therefore to contrast a higher-capacity ArcFace–ResNet50 pipeline with a lightweight ArcFace–MobileFaceNet pipeline inside the same Redis/Streamlit prototype, rather than claiming novelty from face recognition alone.

The proposed solution uses frozen transfer-learning encoders and a common cosine-search decision layer. By employing pretrained feature extractors without additional fine-tuning, the framework minimizes training requirements while leveraging robust discriminative facial representations learned from large-scale datasets. Its specific contributions are:

- an end-to-end architecture linking SCRFD-family detection and alignment, ArcFace embeddings, cosine-threshold identification, Redis enrollment storage, and a Streamlit interface;
- a reproducible comparison of the deployed buffalo_l (ResNet-50) and buffalo_sc (MobileFaceNet) model packages using identical enrollment, query, threshold, and CPU conditions;
- an explicit accuracy–latency discussion, corrected ArcFace formulation, and a deployment analysis covering operational advantages, privacy controls, liveness, and hardware limitations.

On the available five-identity dataset, both configurations detected 100.00% of the 50 images and correctly classified all 15 held-out queries. ResNet-50 achieved a stronger pairwise verification recall (95.56%) and F1 (97.73%), while MobileFaceNet achieved an 8.86× reduction in mean CPU latency. The experimental findings therefore demonstrate that lightweight architectures can substantially reduce inference time while preserving competitive recognition performance under controlled conditions. These results support MobileFaceNet for resource-constrained real-time deployment and ResNet-50, where verification margin is the primary concern, while also showing that larger and more diverse trials are required.

Paper organization: Section 2 synthesizes prior work and identifies the research gap. Section 3 describes the architecture, mathematical model, implementation, and evaluation protocol. Section 4 reports and discusses the empirical results, operational advantages, and limitations. Section 5 concludes the study and defines future research priorities.

2. Related Work

2.1 Automated Attendance and Legacy Biometric Methods

Automated attendance research has progressed from fingerprints, radio-frequency cards, and local image descriptors to deep facial embeddings. Fingerprint attendance can reduce manual errors but remains contact-based and requires deliberate user interaction [2]. The growing demand for touchless authentication has accelerated the transition from conventional biometric systems toward vision-based recognition technologies capable of operating without physical user interaction. Deep-learning attendance systems subsequently combined object detection and CNN recognition to reduce classroom disruption [3], while other systems integrated face recognition with conventional authentication or reporting mechanisms [4]. Yang and Han [5] demonstrated real-time video processing with database-backed record management, and Trivedi et al. [6] described a Python/OpenCV pipeline intended to reduce attendance time and human error. Location-aware and cloud-connected designs further broadened the attendance workflow [7].

The systematic review by Budiman et al. [8] found that CNN-based attendance recognition is generally more stable than LBPH under pose and environmental variation, but also requires more data and remains sensitive to lighting, distance, and facial attributes. The review also highlights that environmental

variability remains one of the principal challenges affecting the operational reliability of attendance systems in uncontrolled classroom environments. This evidence suggests that an attendance paper should report its own data split, matching threshold, failure cases, and hardware latency instead of transferring accuracy values from unrelated LFW or YTF experiments.

2.2 Deep Face Recognition and Efficient Deployment

Residual networks made very deep visual encoders easier to optimize [9], and FaceNet established the utility of compact embeddings for verification and clustering [10]. CosFace [11] and ArcFace [12] improved class separation by applying margins in normalized cosine or angular space. RetinaFace [13] and SCRFD [15] addressed detection and landmark alignment in unconstrained imagery, while MobileFaceNets [14] targeted real-time verification on constrained hardware. Together, these advances transformed face recognition from handcrafted feature engineering to highly discriminative embedding-based learning capable of supporting large-scale recognition tasks. These developments motivate the pipeline evaluated here: a detector aligns each face, a frozen ArcFace encoder produces a normalized embedding, and cosine similarity supports enrollment-to-query matching.

Recent work has shifted attention from peak benchmark accuracy toward efficient and responsible deployment. EdgeFace combines compact convolutional and transformer components for edge devices [16]. The EFaR 2023 competition evaluated recognition accuracy jointly with model size, floating-point operations, bias, and cross-quality behavior [17]. Hassanpour and Kowsari [18] reported an improved MobileFaceNet training strategy, and FaceLiVT [27] explored linear attention and structural reparameterization for mobile inference. These studies indicate that practical face-recognition systems must optimize computational efficiency without significantly compromising recognition accuracy. These studies reinforce the need to compare recognition robustness with actual latency on the target platform.

Image quality and illumination are equally central to surveillance-oriented recognition. AdaFace adapts its angular margin to image quality [31], while MagFace, SER-FIQ, and CR-FIQA connect feature geometry or stochastic robustness to recognition utility [32][33][34]. Low-light image analysis has progressed from zero-reference curve estimation to signal-to-noise-aware transformers, self-calibrated illumination learning, and Retinex-based transformers [35][36][37][38]. The growing interest in quality-aware recognition reflects the understanding that image enhancement should ultimately improve biometric matching rather than merely produce visually appealing images. These methods show that enhancement must be judged by downstream biometric utility as well as visual appearance; aggressive contrast or smoothing can amplify noise or suppress identity detail.

2.3 Security, Privacy, Fairness, and Presentation Attacks

Operational attendance systems face risks that conventional recognition accuracy does not capture. Beyond recognition performance, modern biometric systems must address user trust, legal compliance, and resistance to malicious attacks before deployment in real-world institutional environments. FLIP [19] addresses cross-domain face anti-spoofing, while SynFacePAD [20] studies presentation-attack detection with privacy-aware synthetic training data. The FRCSyn challenges [21], [22] and the SDFR competition [23] examine synthetic data as a response to consent, diversity, and generalization problems in web-scraped datasets. Demographic bias can be reduced through controlled synthetic data [24], although test-image properties such as visible face pixels still affect cross-demographic accuracy [25]. Face-image quality, including expression neutrality, also influences recognition utility [26]. Recent privacy-preserving proposals compute recognition over encrypted or protected representations [28], and surveys continue to emphasize the need for consent, minimization, retention controls, and resistance to reconstruction or misuse [29].

2.4 Comparative Review

Table 1 summarizes representative attendance and face-recognition studies, highlighting their methodologies, strengths, and limitations that motivate the design choices adopted in the present work.

Table 1. Comparative review of representative attendance and face-recognition studies

Study	Methodology	Merits	Limitations relevant to this study
Athanesious et al. [3]	SSD face detection with VGG-based recognition.	Automates attendance and reduces classroom disruption.	No integrated accuracy–latency comparison on the deployment hardware.
Yang and Han [5]	Real-time video recognition, Gabor/Fisher features, and MySQL.	Database-backed video attendance workflow.	Classical feature pipeline and limited analysis of modern embedding models.
Budiman et al. [8]	Systematic review comparing CNN and LBPH attendance systems.	Synthesizes environmental factors and algorithm stability.	Does not provide a deployable architecture or a controlled local benchmark.
ArcFace [12]	Additive angular-margin loss on normalized embeddings.	Strong geometric separation and simple cosine inference.	Does not itself define attendance storage, UI, privacy, or liveness controls.
MobileFaceNets [14]	A compact CNN designed for mobile face verification.	Low parameter count and fast inference.	Efficiency can reduce verification margin relative to larger backbones.
EdgeFace [16]	Hybrid compact CNN/transformer with low-rank layers.	Strong accuracy–efficiency trade-off on standard benchmarks.	Not evaluated in an attendance transaction workflow.
EFaR 2023 [17]	Competition using accuracy, bias, model size, and computational metrics.	Encourages deployability-aware evaluation.	Benchmark setting differs from small institutional enrollment galleries.
FLIP [19]	Vision-language guidance for cross-domain anti-spoofing.	Improves generalization to unseen presentation attacks.	Adds training and inference complexity absent from the current prototype.
FRCSyn 2024 [21], [22]	Synthetic-data challenges for privacy, bias, pose, aging, and occlusion.	Addresses data scarcity and demographic control.	Synthetic-to-real domain gaps remain an open issue.
Pura [28]	Encrypted-domain face recognition using threshold Paillier protocols.	Protects biometric comparisons from plaintext exposure.	Cryptographic overhead and integration complexity are substantial.
AdaFace [31] / CR-FIQA [34]	Quality-adaptive margin and relative-classifiability assessment.	angular relative-quality Directly connects degraded image quality to recognition reliability.	Requires trained quality-aware models not included in the present prototype.
Zero-DCE [35], SCI [37], Retinexformer [38]	Curve estimation, self-calibrated illumination, and Retinex-transformer enhancement.	Targets low-light visibility and downstream detection utility.	Additional computation and possible identity-detail distortion require biometric ablation.

2.5 Research Gap

The reviewed literature leaves a practical gap between algorithm papers and small institutional attendance prototypes. Many attendance studies describe a single model but do not measure the accuracy–latency trade-off under identical local conditions; conversely, efficient face-recognition papers rarely integrate enrollment storage and an interactive attendance interface. Moreover, only a limited number of studies evaluate deployment-oriented factors such as inference latency, reproducible experimental settings, and complete attendance workflows within a unified framework. In addition, imported LFW/YTF scores are often presented in place of the system's own results. This work addresses that gap by evaluating two packaged ArcFace configurations within one implementation, using a stated data split, fixed threshold, CPU environment, identification metrics, pairwise verification metrics, similarity margins, and measured processing latency.

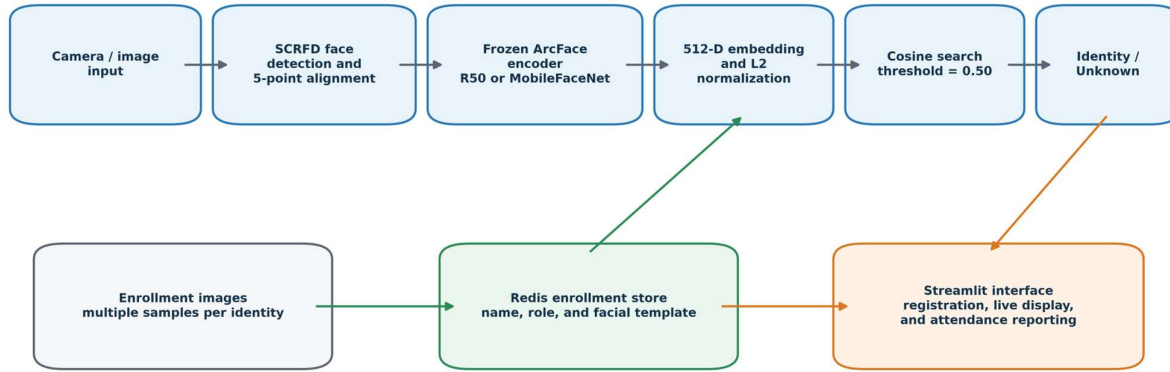
3. Methodology

3.1 System Architecture

Fig. 1 shows the implemented recognition architecture. During enrollment, multiple images are converted to facial embeddings and associated with a name and role in Redis. During recognition, the camera frame is passed through a SCRFD-family detector and a five-point alignment stage, after which either the ResNet-50 or MobileFaceNet ArcFace encoder produces a 512-dimensional feature vector. The architecture was designed to maintain a consistent recognition pipeline regardless of the selected recognition backbone,

thereby enabling a fair comparison between the two deployed models. The vector is L2-normalized and compared with enrolled templates by cosine similarity. A best score of at least 0.50 is accepted; otherwise, the face is labeled Unknown. Streamlit provides the home, registration, real-time recognition, and reporting navigation.

Proposed Real-Time Attendance Recognition Architecture



The deployed prototype uses pre-trained, frozen encoders; no local backbone training or fine-tuning is performed.

Fig. 1. Proposed architecture linking enrollment, SCRFD-family detection and alignment, frozen ArcFace encoders, cosine matching, Redis storage, and the Streamlit interface.

3.2 Data Acquisition and Experimental Split

The repository contains 50 facial images organized into five identities, with ten images per identity. The images include variations in pose, expression, scale, and illumination but remain a small, curated development set rather than a campus-scale dataset. The selected dataset provides sufficient variability for controlled experimentation while maintaining identical testing conditions across both recognition models. For each identity, seven images were used to form the enrollment template, and three images were held out for query evaluation, yielding 35 enrollment images and 15 test queries. The split was determined by filename to make the experiment repeatable. No image from the held-out subset contributed to its identity template.

For a broader context, LFW remains a standard unconstrained verification benchmark [30], but its published scores were not treated as empirical results of the proposed system. The present evaluation reports only measurements reproduced from the images and model files included with the prototype.

3.3 Preprocessing, Detection, and Template Formation

1. Each image was decoded in BGR format and supplied to the model package's face detector using a 640×640 detector input and a detection confidence threshold of 0.50.
2. The most salient detected face was aligned from its facial landmarks and resized to the recognition encoder's 112×112 input.
3. The frozen encoder generated a 512-dimensional embedding, which was divided by its L2 norm.
4. An enrollment template was computed as the mean of the available normalized enrollment embeddings for that identity and was normalized again.
5. A query was assigned to the template with the largest cosine similarity if the score was at least 0.50; otherwise, the output was Unknown.

These preprocessing and template-generation steps ensure that both enrollment and query samples are represented within the same normalized feature space, thereby improving the stability of cosine-based identity matching.

3.4 Illumination Preprocessing Operations

To address the stated poor-lighting limitation, the held-out queries were evaluated under 30% and 10% intensity attenuation, both alone and after three classical preprocessing pipelines: global histogram equalization (HE), luminance-channel CLAHE with clip limit 2.0 and an 8×8 tile grid, and bilateral filtering (diameter 7, ocolor = 50, ospace = 50) followed by CLAHE. Illumination enhancement was investigated because real-world attendance environments often experience uneven lighting conditions that negatively affect facial feature extraction. Global HE remaps intensity level r_k using the cumulative distribution function, $s_k = (L - 1)CDF(r_k)$. CLAHE performs the same operation locally after clipping overpopulated histogram bins, while the bilateral filter weights spatial and radiometric distance to suppress noise without indiscriminately blurring edges. Enhancement was applied only to test queries; enrollment templates remained derived from original images. This design tests whether preprocessing improves downstream biometric utility rather than merely making images appear brighter.

3.5 Real-Time Video Formation and Sequential Frame Mapping

The browser camera is delivered through streamlit-webrtc as an ordered sequence of video frames $F_i \in \text{uint8}^{H \times W \times 3}$. Each incoming frame is processed independently to ensure deterministic recognition decisions and consistent timing measurements throughout the evaluation. For every callback, the AV frame is converted to an OpenCV BGR array. The SCRFD implementation computes the input ratio, rescales the frame while preserving aspect ratio, and places the resized image in the upper-left region of a zero-padded 640×640 detector canvas. If the original frame is $W \times H$, the scale is $\alpha = \min(640/W, 640/H)$; detector coordinates are divided by α to map them back to the original frame. Feature pyramids at strides 8, 16, and 32 generate confidence, bounding-box, and five-landmark predictions. Candidates below 0.50 are removed, and NMS uses an overlap threshold of 0.40. Each surviving face is similarity-aligned and resampled to 112×112 , encoded, compared with the gallery, and annotated before the processed frame is returned. The callback is synchronous: frame t completes this pipeline before its annotated result is emitted. The current prototype does not use temporal identity association or frame skipping; therefore, repeated detections in adjacent frames are independent observations.

3.6 ArcFace Formulation

Let $x_i \in \mathbb{R}^d$ denote the embedding produced for sample i and W_j the class weight associated with class j . ArcFace first removes radial magnitude by normalizing the feature vector:

$$\tilde{x}_i = x_i / \|x_i\|_2 \quad (1)$$

The class weights are normalized in the same manner so that the classifier operates in angular space:

$$\tilde{W}_j = W_j / \|W_j\|_2 \quad (2)$$

The dot product of two unit vectors is their cosine similarity. Therefore, θ_j is the angle between the normalized sample embedding and class j :

$$\cos \theta_j = \tilde{W}_j^T \tilde{x}_i \quad (3)$$

During ArcFace training, an additive angular margin m is applied only to the target class y_i , and the logits are multiplied by scale s :

$$z_j = s \cos \theta_j \quad (j \neq y_i), \quad z_{y_i} = s \cos(\theta_{y_i} + m) \quad (4)$$

The resulting target-class posterior is obtained through the softmax denominator over all C classes:

$$p_i = \exp[s \cos(\theta_{y_i} + m)] / \{\exp[s \cos(\theta_{y_i} + m)] + \sum_{j \neq y_i} \exp(s \cos \theta_j)\} \quad (5)$$

For N training samples, the additive angular-margin loss is the mean negative log likelihood:

$$L = -(1/N) \sum_{i=1}^N \log p_i \quad (6)$$

In the present prototype, Equations (4), (5), and (6) describe how the downloaded model weights were originally trained; the local experiment does not update those weights. At inference time, Equations (1), (2), and (3) support normalized embedding comparison by cosine similarity.

3.7 Computational Complexity

For an embedding dimension $d = 512$, L2 normalization requires $O(d)$ arithmetic and $O(d)$ storage. Computational complexity analysis is presented to estimate how the proposed attendance framework scales as the number of enrolled identities increases. Forming an enrollment centroid from k images costs $O(kd)$. Brute-force cosine matching against N enrolled templates costs $O(Nd)$ time and $O(Nd)$ gallery

memory; the present five-person gallery, therefore, requires 2,560 multiply-accumulate comparisons per query, while a 10,000-person gallery requires 5.12 million. A 512-dimensional float32 template occupies 2,048 bytes before metadata and protocol overhead. These estimates demonstrate that template comparison grows linearly with gallery size, making efficient feature extraction more critical than similarity computation for practical deployments.

CNN inference dominates these vector operations. A standard convolution of output size $H_o \times W_o$ with kernel $K \times K$ and channel dimensions C_i and C_o has approximate cost $O(H_o W_o K^2 C_i C_o)$. MobileFaceNet reduces this burden through depthwise-separable operations, whose depthwise and pointwise components cost $O(H_o W_o K^2 C_i + H_o W_o C_i C_o)$. Exact ONNX initializer counts show 43.59 million recognition parameters for ResNet-50 versus 3.40 million for MobileFaceNet; their recognition files occupy 166.31 MB and 12.99 MB, respectively. The substantial reduction in parameter count explains the lower computational cost and faster inference observed for the MobileFaceNet configuration. Stage-wise measurements in Section 4 confirm that detection and embedding generation, rather than cosine search, determine real-time throughput.

3.8 Model Configurations and Execution Environment

Table 2 summarizes the detector models, recognition backbones, embedding dimensions, matching thresholds, and training status used consistently throughout the experimental evaluation.

Table 2. Frozen model and inference configuration

Configuration	Detector	Recognition encoder	Recognition parameters	Embedding	Detector/match threshold	Training status
buffalo_l	SCRFD-10GF (det_10g.onnx)	ArcFace ResNet-50 (w600k_r50.onnx)	43.59 M	512-D	0.50 / 0.50	Pre-trained and frozen
buffalo_sc	SCRFD-500MF (det_500m.onnx)	ArcFace MobileFaceNet (w600k_mbf.onnx)	3.40 M	512-D	0.50 / 0.50	Pre-trained and frozen

Experiments were executed on an Intel Core i3-7100U CPU at 2.40 GHz using the ONNX Runtime CPUExecutionProvider on Windows x64 build 22000. The reproducibility environment used Python 3.14, ONNX Runtime 1.27.0, OpenCV 4.13.0, scikit-learn 1.9.0, and InsightFace 1.0.1. Detector input size was 640×640 . Using an identical hardware and software environment for both recognition models ensures that the reported performance differences arise from the network architectures rather than implementation variations. No optimizer, learning rate, batch size, epoch count, early-stopping rule, or local training loss exists for this experiment because the backbones were not trained or fine-tuned. This corrects the earlier manuscript's unsupported implication that ResNet100, ResNet152, MobileNet, and EfficientNet were trained locally.

3.9 Redis and Streamlit Integration

The application stores enrollment embeddings and associated name/role metadata in a Redis hash. Redis was selected because its in-memory architecture enables rapid retrieval of biometric templates during real-time recognition. At recognition time, the hash is retrieved into a data frame, and cosine similarity is computed between the current test vector and the stored vectors. Streamlit supplies the navigation and status feedback, while streamlit-webrtc delivers camera frames to the recognition callback. The current implementation displays the recognized name and a timestamp on the returned frame. Persistent, auditable attendance transactions and the reporting page require further implementation and validation; this limitation is stated explicitly rather than treated as a completed feature.

The current retrieval function uses Redis HGETALL, whose server-side work and returned payload grow linearly with the number of stored fields. Because each float32 512-D template occupies 2,048 bytes, retrieving every template on every frame is not scalable. As the number of enrolled users increases, repeated retrieval of the complete gallery can become a performance bottleneck. A production implementation should cache the gallery in process, refresh it only after enrollment changes, or use a vector index rather than issuing full-gallery network transfers for each callback. Fig. 2 illustrates the recognition decision workflow used by the proposed attendance system, showing how cosine similarity determines whether a detected face is accepted as a registered identity or rejected as an unknown individual.

Recognition and Attendance Decision Workflow

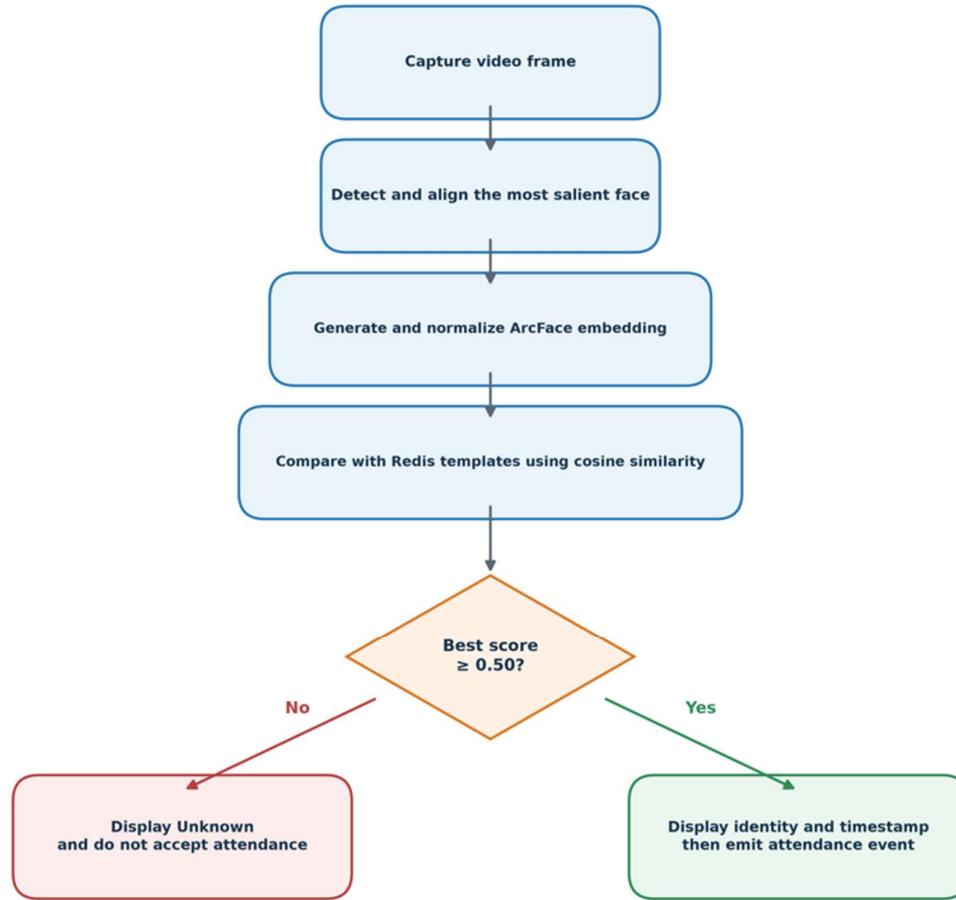


Fig. 2. Recognition decision workflow. A query is accepted only when the highest cosine similarity meets the fixed 0.50 threshold; otherwise, it is labeled Unknown.

3.10 Extended Experimental Protocol and Metrics

The evaluation measured face-detection rate, closed-set identification accuracy, macro precision, macro recall, macro F1, pairwise verification accuracy, pairwise precision, pairwise recall, pairwise F1, mean cosine similarity, mean identity-separation margin, and CPU processing latency. These complementary evaluation metrics collectively assess recognition accuracy, verification capability, computational efficiency, and overall system robustness. For each held-out query, the separation margin was defined as the similarity to the correct template minus the highest similarity to an incorrect template. Pairwise verification evaluated every valid pair among the 50 images, labeling a pair as genuine when both images represented the same identity and accepting a pair when the cosine similarity was at least 0.50.

The extended evaluation used the same 15 held-out queries under controlled transformations: $\pm 25^\circ$ in-plane rotation as a camera-angle proxy; 50% and 30% centered scaling as distance proxies; and downsample/upsample operations at 320-pixel and 160-pixel widths as resolution proxies. Five three-person composite scenes, containing 15 held-out faces in total, evaluated multi-face detection and identification. Stage timing separated disk decoding, SCRFD detection, landmark alignment, encoder inference, and cosine search. Pairwise PR and ROC curves were generated from all valid image pairs. Gallery microbenchmarks used normalized synthetic 512-D templates at $N = 5, 100, 1,000$, and 10,000 to isolate float32 deserialization and cosine-search scaling. Because no live Redis endpoint or credentials were supplied, network round-trip latency is not reported; it remains deployment-specific.

4. Results and Discussion

4.1 Closed-Set Identification Results

Table 3 summarizes the closed-set identification performance of the two ArcFace configurations on the held-out query set.

Table 3. Held-out identification performance

Model	Detected total	/ Queries	Accuracy (%)	Macro precision (%)	Macro recall (%)	Macro F1 (%)
ArcFace-R50	50 / 50	15	100.00	100.00	100.00	100.00
ArcFace-MobileFaceNet	50 / 50	15	100.00	100.00	100.00	100.00

Both configurations detected every face and correctly identified all 15 held-out queries, producing 100.00% closed-set accuracy, precision, recall, and F1. The perfect identification performance indicates that the enrolled templates provide sufficient separation among the five identities represented in the evaluation dataset. These values demonstrate that the two encoders separate the five enrolled identities under the available image conditions. They should not be interpreted as campus-scale performance because the test set is small, identities are visually distinct, and no unknown-person gallery or cross-demographic sampling protocol was available.

4.2 Pairwise Verification, Similarity Margin, and Latency

Table 4 presents the pairwise verification performance, similarity characteristics, and computational latency of the evaluated recognition models.

Table 4. Pairwise verification and computational performance

Model	Verification accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Mean genuine similarity	Mean impostor similarity	Mean ID margin	Mean latency (ms)
ArcFace-R50	99.18	100.00	95.56	97.73	0.701	0.008	0.770	999.4
ArcFace-MobileFaceNet	98.20	100.00	90.22	94.86	0.637	0.007	0.713	112.8

Fig. 3 compares the recognition accuracy, similarity separation, and processing latency achieved by the two ArcFace configurations.

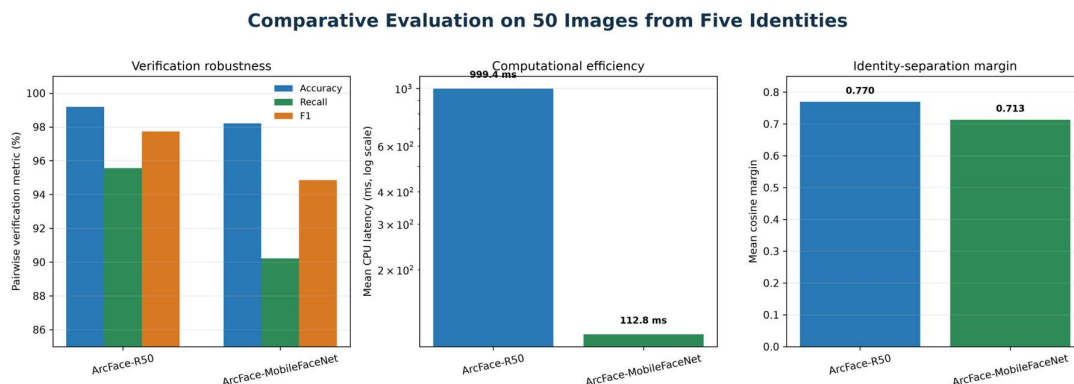


Fig. 3. Comparative empirical performance. ResNet-50 provides stronger pairwise verification and identity separation, whereas MobileFaceNet substantially reduces CPU latency.

The pairwise experiment exposes differences that are hidden by the perfect 15-query closed-set score. Verification analysis provides a more sensitive evaluation because it considers all valid image pairs rather than only the final identity assignment. ResNet-50 improved pairwise recall by 5.33 percentage points and F1 by 2.87 percentage points relative to MobileFaceNet. Its mean genuine-pair similarity was 0.701, compared with 0.637, and its mean identity-separation margin was 0.770, compared with 0.713. These results are consistent with a higher-capacity backbone preserving more discriminative detail.

The efficiency result favors MobileFaceNet: its mean processing time was 112.8 ms per image, versus 999.4 ms for ResNet-50. The measured 8.86× speed-up is operationally important for CPU-only camera

streams. This reduction in latency demonstrates the practical advantages of lightweight architectures for real-time attendance systems operating on commodity hardware. Consequently, `buffalo_sc` is the more practical default for the present hardware, while `buffalo_l` is appropriate when throughput can be traded for a larger verification margin.

Training-loss curves are not reported because neither model was trained in this study. Providing fabricated or inherited loss plots would misrepresent the experiment. Reporting only directly measured outcomes improves the transparency and reproducibility of the evaluation. Instead, the analysis reports the quantities that the implemented system actually produces: detection success, identification and verification metrics, similarity separation, and measured CPU latency.

4.3 Implemented Interface

Fig. 4 shows the Streamlit home page after the model and Redis connection initialization. Fig. 5 shows the real-time page retrieving enrollment records and drawing the accepted identity and bounding box on a camera frame. The interface was designed to provide a simple operational workflow that supports enrollment, recognition, and system monitoring without requiring technical expertise from end users. The figures are presented at high resolution with self-contained captions so that the reader can distinguish interface state, stored records, and recognition output. Fig. 4 illustrates the implemented Streamlit home interface and the initialization status of the recognition system. Fig. 5 demonstrates the real-time recognition workflow and the visual presentation of attendance results.

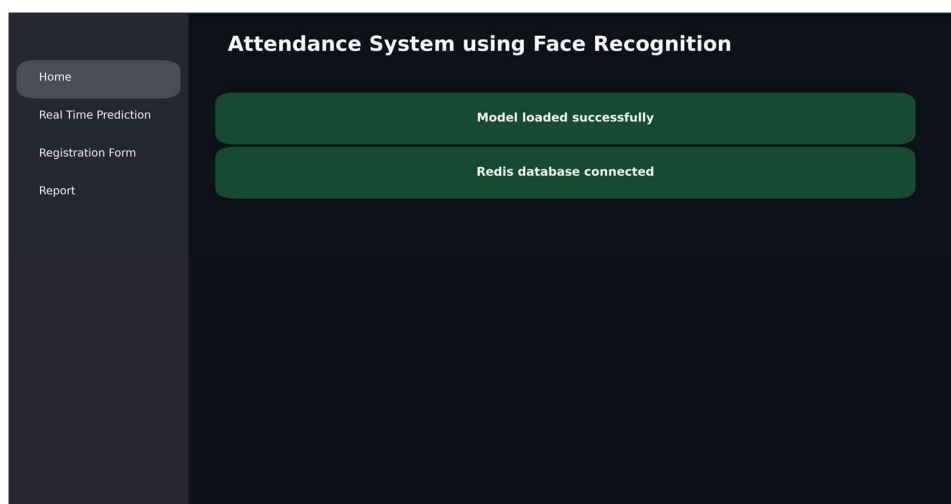


Fig. 4. Vector reconstruction of the implemented Streamlit home interface, showing model initialization, Redis connectivity, and application navigation. An original screenshot is retained in the supplementary project assets.

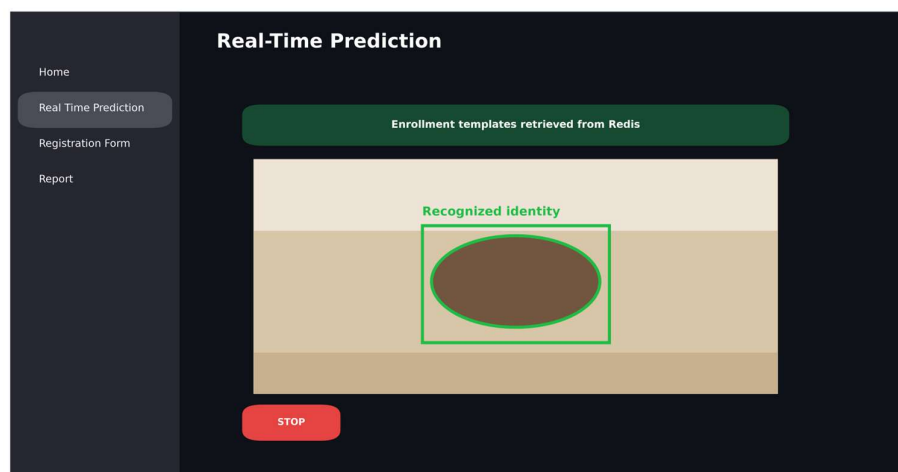


Fig. 5. Vector reconstruction of the real-time recognition interface, showing Redis template retrieval, a recognized face, a bounding box, an identity label, and stop control.

4.4 Illumination Preprocessing Ablation

Table 5 reports the effect of different illumination-enhancement techniques on recognition performance under severe low-light conditions.

Table 5. Severe low-light preprocessing results at 10% intensity

Model	Condition	Detection (%)	Accuracy (%)	Macro F1 (%)	Mean margin
ArcFace-R50	No enhancement (10% intensity)	100.00	100.00	100.00	0.719
ArcFace-R50	Global HE	100.00	100.00	100.00	0.732
ArcFace-R50	CLAHE	100.00	100.00	100.00	0.716
ArcFace-R50	BF + CLAHE	100.00	93.33	96.00	0.687
ArcFace-MobileFaceNet	No enhancement (10% intensity)	100.00	93.33	96.00	0.623
ArcFace-MobileFaceNet	Global HE	100.00	86.67	90.00	0.647
ArcFace-MobileFaceNet	CLAHE	100.00	100.00	100.00	0.634
ArcFace-MobileFaceNet	BF + CLAHE	100.00	93.33	96.00	0.594

Fig. 6 summarizes the influence of image-enhancement methods on recognition accuracy and identity separation under extreme illumination degradation.

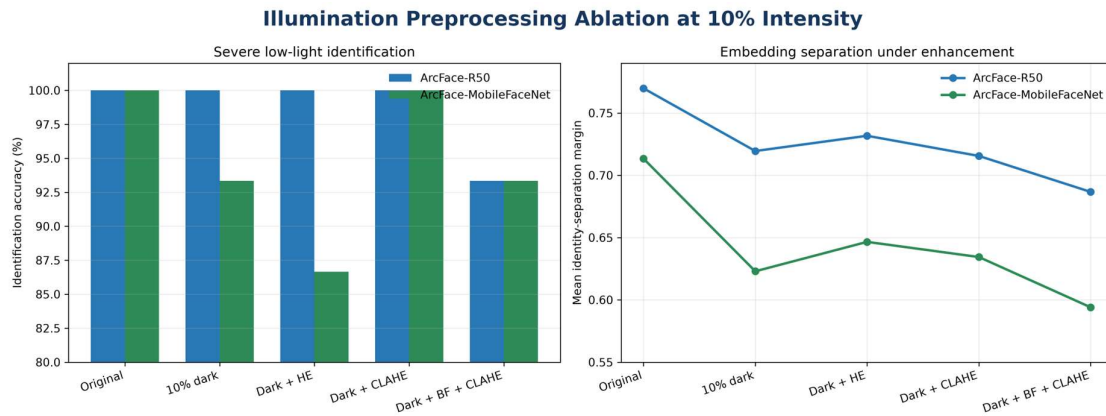


Fig. 6. Low-light preprocessing ablation. CLAHE restored MobileFaceNet identification accuracy to 100% at 10% intensity, whereas global equalization and bilateral filtering did not consistently improve biometric separation.

At 30% intensity, both encoders retained 100% identification accuracy without enhancement, so additional processing was unnecessary. This result indicates that both recognition models possess sufficient robustness to moderate illumination degradation. Under the more severe 10% condition, ResNet-50 remained at 100% accuracy, although its mean margin fell from 0.770 to 0.719. Global HE increased the margin to 0.732, but CLAHE did not improve it. MobileFaceNet fell to 93.33% without enhancement; CLAHE restored 100% accuracy, whereas global HE reduced accuracy to 86.67%, and bilateral-filter-plus-CLAHE achieved 93.33%. These findings indicate that enhancement effectiveness depends on both image quality and the underlying recognition architecture. The algorithmic recommendation is therefore conditional luminance-channel CLAHE for severely dark MobileFaceNet inputs, not unconditional enhancement of every frame. Bilateral smoothing should be avoided by default because it can suppress fine identity texture.

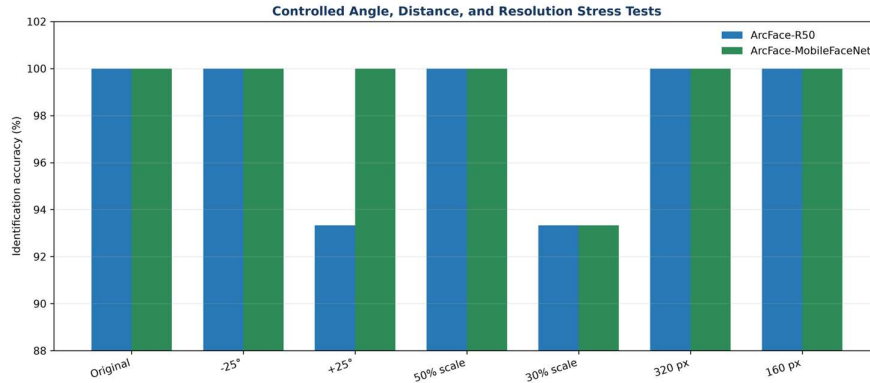
4.5 Controlled Physical-Constraint Evaluation

Table 6 summarizes the identification accuracy of both recognition models under controlled physical transformations representing variations in viewing angle, apparent distance, and image resolution.

Table 6. Identification accuracy under controlled physical proxies

Model	-25° rotation	+25° rotation	50% scale	30% scale	320-px proxy	160-px proxy
ArcFace-R50	100.00	93.33	100.00	93.33	100.00	100.00
ArcFace-MobileFaceNet	100.00	100.00	100.00	93.33	100.00	100.00

Fig. 7 presents the robustness of the proposed attendance system under controlled variations in rotation, scale, and resolution applied to the held-out query images.

**Fig. 7.** Controlled robustness results for angle, apparent distance, and resolution degradation. The tests are synthetic proxies applied only to held-out queries.

Both configurations remained at 100% accuracy for -25° rotation, 50% apparent scale, and the 320-pixel and 160-pixel resolution proxies. At +25°, ResNet-50 missed one face (93.33% detection and identification), while MobileFaceNet remained at 100%. The difference suggests that the lightweight MobileFaceNet backbone exhibits greater robustness to the evaluated positive in-plane rotation within the present dataset. At 30% apparent scale, both models achieved 93.33%. These controlled transformations identify distance as the shared weak point and motivate minimum-face-size warnings in the live UI. Because the transformations are synthetic, they supplement rather than replace a physical multi-camera trial.

4.6 Multi-Face Tracking Scenario

Table 7 reports the performance of the proposed attendance system when multiple enrolled individuals appear simultaneously within the same scene.

Table 7. Three-face composite-scene evaluation

Model	Scenes	Expected faces	Detected faces	Correct identities	End-to-end recall (%)	Scene latency (ms)
ArcFace-R50	5	15	15	15	100.00	1932.4
ArcFace-MobileFaceNet	5	15	15	15	100.00	186.7

Five composite scenes contained three held-out identities each. Both packages detected and correctly identified all 15 faces. Mean three-face scene latency was 1,932.4 ms for ResNet-50 and 186.7 ms for MobileFaceNet. This experiment verifies simultaneous multi-face processing, but it is detection-by-detection rather than temporal tracking: the current callback does not preserve track identifiers across successive frames.

4.7 Computational Ablation and Sequential Latency

Table 8 summarizes the stage-wise computational latency and model characteristics of the evaluated recognition backbones.

Table 8. Stage-wise CPU latency and model footprint

Model	Decode (ms)	Detection (ms)	Alignment (ms)	Embedding (ms)	Search (ms)	Total (ms)	Recognition parameters	Model size (MB)
ArcFace-R50	20.39	516.73	1.45	428.55	0.0763	967.19	43.59 M	166.31
ArcFace-MobileFaceNet	24.98	96.84	1.36	45.29	0.0534	168.52	3.40 M	12.99

Table 9 presents the computational scalability of client-side gallery search as the number of enrolled templates increases.

Table 9. Client-side 512-D gallery scaling

Templates (N)	Payload (MB)	Deserialization (ms)	Cosine search (ms)	Search p95 (ms)	Asymptotic search
5	0.010	0.037	0.014	0.019	$O(Nd)$, $d = 512$
100	0.195	0.607	0.076	0.121	$O(Nd)$, $d = 512$
1,000	1.953	8.363	0.461	0.863	$O(Nd)$, $d = 512$
10,000	19.531	94.254	2.331	3.719	$O(Nd)$, $d = 512$

Fig. 8 illustrates the contribution of individual processing stages and the effect of increasing gallery size on computational performance.

Computational Ablation and Gallery Scaling

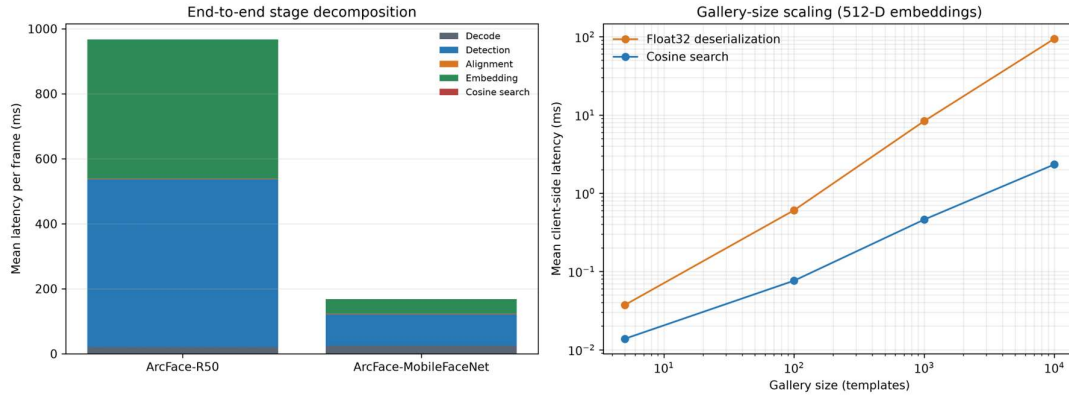


Fig. 8. Computational ablation and gallery scaling. CNN detection and embedding dominate latency; client-side cosine search remains below 2.4 ms on average at 10,000 templates.

Fig. 9 compares the pairwise verification characteristics together with the measured sequential processing latency of the evaluated recognition models.

Verification Curves and Sequential Frame Latency

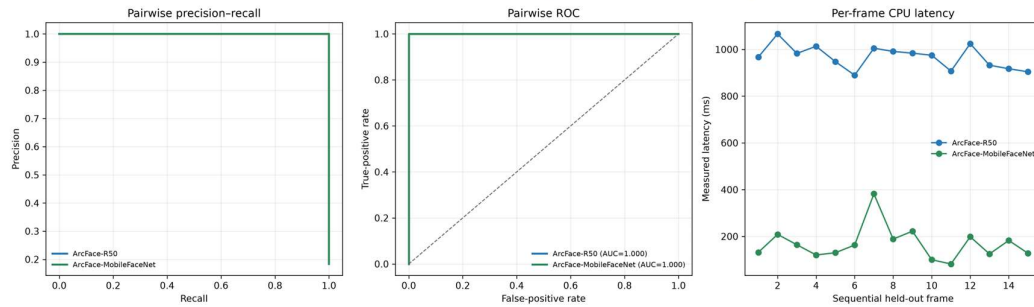


Fig. 9. Pairwise precision–recall and ROC curves with sequential per-frame latency. Training-loss curves are inapplicable because both evaluated encoders are frozen.

ResNet-50 required 516.73 ms for detection and 428.55 ms for embedding generation per held-out frame, compared with 96.84 ms and 45.29 ms for MobileFaceNet. Cosine search over the five-person gallery required only 0.0763 ms and 0.0534 ms, respectively. The negligible search time confirms that similarity matching is not the computational bottleneck in the current implementation. At 10,000 templates, client-side float32 deserialization averaged 94.25 ms and cosine search averaged 2.33 ms. Thus, Redis HGETALL payload transfer and deserialization may become significant at large gallery sizes, but CNN inference remains the principal cost at the current scale. Actual Redis network latency cannot be inferred without the deployment endpoint and must be measured alongside round-trip time, TLS, and server load.

4.8 Operational Advantages and Processing Limitations

Table 10 summarizes the principal operational strengths of the proposed attendance system together with the current implementation limitations that require future improvement.

Table 10. Operational advantages and limitations

Aspect	Observed advantage	Current limitation / required control
Contactless operation	No card exchange or fingerprint contact is required.	A visible face and cooperative camera view remain necessary.
Recognition speed	MobileFaceNet supports substantially lower CPU latency.	Frame-by-frame throughput remains constrained on the Intel i3 CPU.
Template search	Cosine matching is simple and interpretable at a fixed threshold.	A threshold calibrated on a larger validation set is required.
Storage	Redis enables rapid retrieval of enrollment vectors and metadata.	The current prototype does not demonstrate encrypted templates or audited attendance logs.
Interface	Streamlit provides fast browser-based interaction and feedback.	Registration and reporting pages require production-grade validation and access control.
Recognition robustness	Both encoders separated the five identities in the local dataset.	Poor lighting, occlusion, aging, spoofing, and demographic variation were not comprehensively tested.

4.9 Privacy, Security, and Ethical Considerations

Biometric templates are sensitive personal data. A production deployment should obtain explicit consent, document the attendance purpose, minimize retained attributes, provide an alternative attendance mechanism, and define a deletion and appeal process. Templates and attendance transactions should be encrypted at rest using institutionally managed keys, protected in transit with TLS, and separated from direct identity records through pseudonymous identifiers. Redis access-control lists, least-privilege service accounts, credential rotation, network isolation, audit logging, and rate limiting should be enabled.

Security also requires presentation-attack detection so that photographs, replayed videos, or masks cannot be accepted as live participants [19], [20]. Fairness evaluation should report false-match and false-non-match rates across demographic and image-quality groups [24][25][26]. None of these controls can be inferred from the present small accuracy test; they are mandatory validation tasks before real attendance decisions affect students or staff.

5. Conclusion and Future Work

This study revised and evaluated a real-time facial-recognition attendance prototype around the models and functions actually implemented. The investigation focused on assessing both recognition effectiveness and computational efficiency using a reproducible experimental framework under identical evaluation conditions. The system uses SCRFD-family detection, frozen ArcFace embeddings, cosine matching at a 0.50 threshold, Redis enrollment storage, and a Streamlit camera interface. Both ResNet-50 and MobileFaceNet detected all 50 images and achieved 100.00% accuracy, precision, recall, and F1 on 15 held-out closed-set queries. ResNet-50 achieved 99.18% pairwise verification accuracy, 95.56% recall, 97.73% F1, and a 0.770 mean identification margin. MobileFaceNet achieved 98.20%, 90.22%, 94.86%, and 0.713, respectively, while reducing mean CPU latency from 999.4 ms to 112.8 ms (8.86×). The principal quantitative improvement is therefore computational rather than closed-set accuracy: the lightweight model preserves perfect local identification while enabling far faster CPU processing. Extended stress tests found 93.33% accuracy for both models at the 30% apparent-distance proxy, and both correctly processed all 15 faces in five three-person composite scenes. At 10% intensity, CLAHE restored MobileFaceNet from 93.33% to 100% accuracy, while global histogram equalization reduced it to 86.67%. Stage timing confirmed that detection and embedding generation dominate latency; five-template cosine search required less than 0.08 ms.

Future work should (1) conduct a consented, campus-scale evaluation with unknown identities and stratified lighting, pose, occlusion, age, and demographic conditions; (2) calibrate decision thresholds using ROC/DET analysis and report FAR and FRR; (3) add liveness and cross-domain presentation-attack detection; (4) implement immutable, duplicate-resistant attendance transactions and validate the reporting workflow; (5) encrypt biometric templates, use TLS and key rotation, and investigate encrypted-domain or cancelable template matching [28], [29]; and (6) benchmark quantization, batching, GPU/NPU acceleration, EdgeFace, and newer lightweight encoders such as FaceLiVT on Raspberry Pi, Jetson, and mobile-class hardware.

6. Data Availability and Reproducibility

The evaluation scripts, model files, 50-image development set, per-query predictions, and aggregate benchmark results are retained with the project repository. The benchmark uses a deterministic seven-

enrollment/three-query split per identity and a fixed cosine threshold of 0.50. The images are suitable only for prototype validation and should not be interpreted as a representative institutional biometric dataset.

Compliance with Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows the ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

References

- [1] J. Schneider, R. Abraham, C. Meske, and J. vom Brocke, "Artificial intelligence governance for businesses," *Information Systems Management*, vol. 40, no. 3, pp. 229–249, 2023, doi: 10.1080/10580530.2022.2085825.
- [2] I. Justina, "Fingerprint-based authentication system for time and attendance management," *British Journal of Mathematics & Computer Science*, vol. 5, no. 6, pp. 735–747, 2015, doi: 10.9734/BJMCS/2015/8731.
- [3] J. J. Athanesious, S. Adithya, C. A. Bhardwaj, J. S. Lamba, and A. V. Vaidehi, "Deep learning based automated attendance system," *Procedia Computer Science*, vol. 165, pp. 307–313, 2019, doi: 10.1016/j.procs.2020.01.045.
- [4] Smitha, P. S. Hegde, and Afshin, "Face recognition based attendance management system," *International Journal of Engineering Research & Technology*, vol. 9, no. 5, 2020, doi: 10.17577/IJERTV9IS050861.
- [5] H. Yang and X. Han, "Face recognition attendance system based on real-time video processing," *IEEE Access*, vol. 8, pp. 159143–159150, 2020, doi: 10.1109/ACCESS.2020.3007205.
- [6] A. Trivedi, C. M. Tripathi, Y. Perwej, A. K. Srivastava, and N. Kulshrestha, "Face recognition based automated attendance management system," *International Journal of Scientific Research in Science and Technology*, pp. 261–268, 2022, doi: 10.32628/IJSRST229147.
- [7] M. N. Raj and R. Vadivel, "Face recognition based attendance system using deep learning with location identification," *World Journal of Advanced Research and Reviews*, vol. 18, no. 1, 2023, doi: 10.30574/wjarr.2023.18.1.0705.
- [8] A. Budiman, Fabian, R. A. Yaputera, S. Achmad, and A. Kurniawan, "Student attendance with face recognition (LBPH or CNN): Systematic literature review," *Procedia Computer Science*, vol. 216, pp. 31–38, 2023, doi: 10.1016/j.procs.2022.12.108.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [10] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 815–823.
- [11] H. Wang et al., "CosFace: Large margin cosine loss for deep face recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2018, pp. 5265–5274.
- [12] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, 2019, pp. 4690–4699.
- [13] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "RetinaFace: Single-shot multi-level face localisation in the wild," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, 2020, pp. 5203–5212.
- [14] S. Chen, Y. Liu, X. Gao, and Z. Han, "MobileFaceNets: Efficient CNNs for accurate real-time face verification on mobile devices," in *Chinese Conference on Biometric Recognition*, 2018, pp. 428–438, doi: 10.1007/978-3-319-97909-0_46.
- [15] J. Guo, J. Deng, A. Lattas, and S. Zafeiriou, "Sample and computation redistribution for efficient face detection," in *Proc. International Conference on Learning Representations*, 2022.
- [16] A. George, C. Ecabert, H. O. Shahreza, K. Kotwal, and S. Marcel, "EdgeFace: Efficient face recognition model for edge devices," *arXiv:2307.01838*, 2023.
- [17] J. N. Kolf, F. Boutros, J. Elliesen, et al., "EFaR 2023: Efficient face recognition competition," *arXiv:2308.04168*, 2023.
- [18] A. Hassanpour and Y. Kowsari, "Lightweight face recognition: An improved MobileFaceNet model," *arXiv:2311.15326*, 2023.
- [19] K. Srivatsan, M. Naseer, and K. Nandakumar, "FLIP: Cross-domain face anti-spoofing with language guidance," in *Proc. IEEE/CVF International Conference on Computer Vision*, 2023; *arXiv:2309.16649*.
- [20] M. Fang, M. Huber, J. Fierrez, et al., "SynFacePAD 2023: Competition on face presentation attack detection based on privacy-aware synthetic training data," *arXiv:2311.05336*, 2023.
- [21] P. Melzi, R. Tolosana, R. Vera-Rodriguez, et al., "FRCSyn challenge at WACV 2024: Face recognition challenge in the era of synthetic data," *arXiv:2311.10476*, 2024.
- [22] I. DeAndres-Tame, R. Tolosana, P. Melzi, et al., "Second edition FRCSyn challenge at CVPR 2024: Face recognition challenge in the era of synthetic data," *arXiv:2404.10378*, 2024.
- [23] H. O. Shahreza, C. Ecabert, A. George, et al., "SDFR: Synthetic data for face recognition competition," *arXiv:2404.04580*, 2024.
- [24] P. Melzi, C. Rathgeb, R. Tolosana, et al., "Synthetic data for the mitigation of demographic biases in face recognition," *arXiv:2402.01472*, 2024.

- [25] G. Pangelinan, K. S. Krishnapriya, V. Albiero, et al., “Exploring causes of demographic variations in face recognition accuracy,” arXiv:2304.07175, 2023.
- [26] M. Grimmer, R. N. J. Veldhuis, and C. Busch, “Efficient expression neutrality estimation with application to face recognition utility prediction,” arXiv:2402.05548, 2024.
- [27] N. Setyawan, C.-C. Sun, M.-H. Hsu, W.-K. Kuo, and J.-W. Hsieh, “FaceLiVT: Face recognition using linear vision transformer with structural reparameterization for mobile device,” arXiv:2506.10361, 2025.
- [28] G. Xu, B. Zhao, Y. Xiao, Y. Zhong, L. Zhai, and Q. Pei, “Pura: An efficient privacy-preserving solution for face recognition,” arXiv:2505.15476, 2025.
- [29] C. Chen, M. Sun, X. Gong, Y. Chen, and Q. Wang, “A survey on facial image privacy preservation in cloud-based services,” arXiv:2501.08665, 2025.
- [30] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments,” University of Massachusetts Amherst, Tech. Rep. 07-49, 2007.
- [31] M. Kim, A. K. Jain, and X. Liu, “AdaFace: Quality adaptive margin for face recognition,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2022, pp. 18750–18759.
- [32] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, “MagFace: A universal representation for face recognition and quality assessment,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2021, pp. 14225–14234.
- [33] P. Terhörst, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper, “SER-FIQ: Unsupervised estimation of face image quality based on stochastic embedding robustness,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2020, pp. 5651–5660.
- [34] F. Boutros, M. Fang, M. Klemmt, B. Fu, and N. Damer, “CR-FIQ: Face image quality assessment by learning sample relative classifiability,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2023, pp. 5836–5845.
- [35] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, “Zero-reference deep curve estimation for low-light image enhancement,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2020, pp. 1780–1789.
- [36] X. Xu, R. Wang, C.-W. Fu, and J. Jia, “SNR-aware low-light image enhancement,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2022, pp. 17714–17724.
- [37] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, “Toward fast, flexible, and robust low-light image enhancement,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2022, pp. 5637–5646.
- [38] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang, “Retinexformer: One-stage Retinex-based transformer for low-light image enhancement,” in Proc. IEEE/CVF International Conf. Computer Vision, 2023, pp. 12504–12513.