

ResQ-ID: A Reproducible Face and Ear Retrieval Prototype for Disaster Identification Assistance

A. Parivazhagan, Mallu Sree Harika, Kota Koushik Reddy, Chala Venkata Srinivas, Chinthakayala Sriram

Department of Computer Science and Engineering, School of Computing, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, India

Published October 2026

Abstract

ResQ-ID is a reproducible local prototype for retrieving possible missing-person identities from face and ear photographs while retaining human decision authority. The implementation combines YuNet and pretrained SFace, a histogram of oriented gradients (HOG) ear descriptor, exact cosine retrieval, and a FastAPI service with SQLite persistence. Separate component experiments use 400 real Olivetti face images and 600 AMI right-ear images. Three image-disjoint gallery, validation, and test partitions include unseen identities; thresholds are selected only on validation probes. Mean rank-1 retrieval is 96.67% for SFace and 57.92% for HOG ears, while correct accepted identification is 81.11% and 35.83%. Fixed-threshold tests reveal substantial sensitivity to occlusion. The recorded software suite passes 31 checks; 30 loopback ear-query trials have median latency 18.68 ms in a ten-case registry. The datasets contain different subjects, so no paired fusion advantage is demonstrated. These findings establish a transparent component baseline; disaster-field reliability and deployment readiness remain unvalidated.

Keywords: biometric retrieval; face recognition; ear recognition; open-set identification; human verification; reproducible evaluation; disaster assistance

1 Introduction

Disaster identification assistance requires an operator to compare submitted photographs with references associated with missing-person cases. The registry may be incomplete, and image quality may vary. ResQ-ID returns possible identities for investigation; similarity scores alone cannot establish identity. The study evaluates a local research workflow using controlled public photographs rather than an operational disaster deployment.

Face and ear evidence can be available in different views. Ear recognition has an established literature [3], and face-ear fusion has already been studied [4]. The contribution here is a CPU-executable baseline with traceable references, modality-specific comparisons, validation-selected thresholds, and tested separation between algorithmic suggestions and recorded human decisions.

Closed-set rank-1 retrieval asks whether an enrolled identity leads a candidate list. Open-set identification must additionally reject people absent from the gallery. We report both quantities, test controlled changes to held-out real images, and examine software decision controls and small-registry latency. A new recognition architecture, measured fusion gain, or field effectiveness is outside the demonstrated contribution.

2 Related work

ArcFace learns discriminative face representations with an additive angular margin [10]; SFace uses sigmoid-constrained objectives [1]. AdaFace examines quality-adaptive margins [5]. ResQ-ID uses pretrained SFace without retraining and reports its own component measurements rather than importing published benchmark performance.

Detection and representation are distinct. RetinaFace addresses face localization [11], whereas the prototype uses OpenCV YuNet for single-face detection and alignment and SFace for representation [8]. The Olivetti benchmark uses existing crops, so its recognition results do not evaluate the detector.

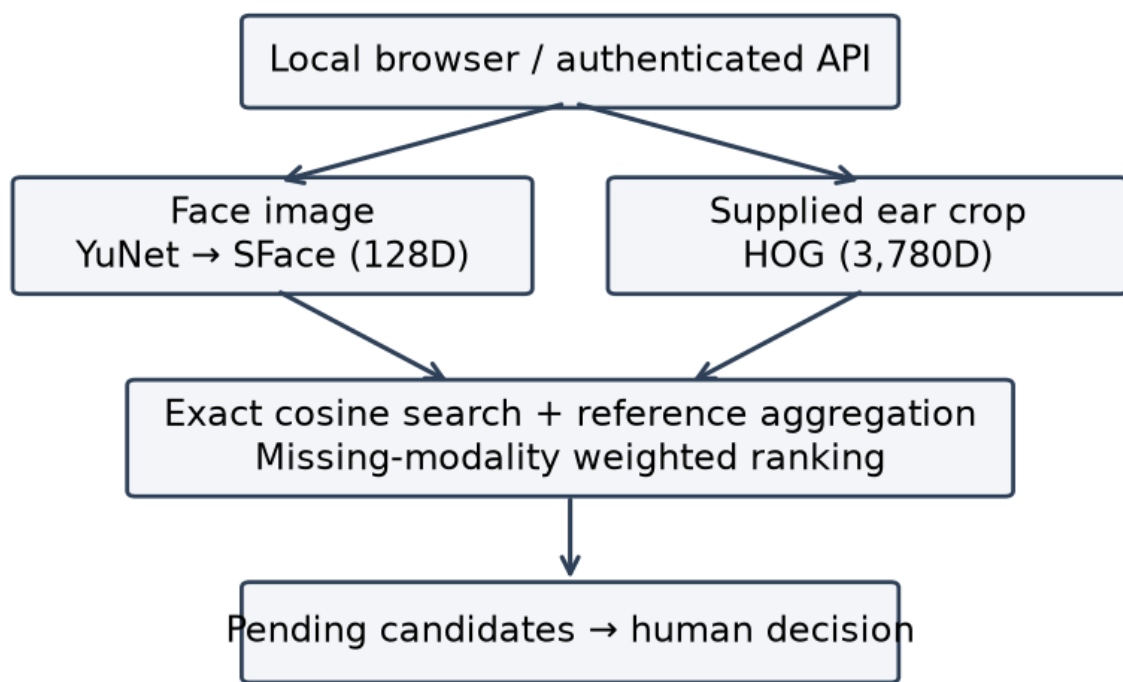
Ear-recognition research documents acquisition and descriptor challenges [3]. Our ear path adapts locally normalized HOG features [2] as an inspectable baseline. A gallery-fitted 64-dimensional PCA comparison tests compression of HOG; it differs from the original raw-image eigenface formulation [12]. Neither a trained ear network nor automatic ear localization is implemented.

Olivetti is a small controlled face collection [6], and AMI provides indoor ear views [7]. These resources enable repeatable comparisons but cannot establish performance on injured people, crowds, or rescue-site footage. NIST documents task-dependent demographic differences [9]; aggregate results here do not establish demographic parity.

3 System architecture

3.1 Implemented local workflow

Figure 1 summarizes the implemented architecture. A local browser interface provides case registration, reference enrollment, candidate search, and verification. FastAPI exposes these functions as authenticated endpoints. The face path supports either a single detected face or an explicitly pre-cropped face image. The ear path requires a user-supplied crop; automatic ear localization is outside this implementation. Reference representations are stored with modality and model identifiers, image checksums, and quality descriptors. Multiple references can be attached to the same case.



SQLite references, decisions and append-only audit triggers

Figure 1. Implemented local architecture. Solid arrows show the available processing path. Ear cropping is supplied by the operator; automatic ear detection and person detection are not implemented.

SQLite is used as the local persistence layer. At the experimental registry size, exact cosine retrieval is sufficiently simple to inspect, so the prototype does not require a separate vector index. This differs from the original PostgreSQL/pgvector design. The implementation also uses a browser client rather than a native mobile application. These choices make the baseline executable on the measured CPU-only environment while preserving the core register, compare, review, and decide workflow.

3.2 Candidate evidence and state transitions

A query may provide face evidence, ear evidence, or both. The matching engine computes component similarities, aggregates references per case, and returns a bounded ranked list. Returned records begin in the pending state. The response exposes which modalities contributed, their scores, the applied thresholds, and image-quality summaries. Overall scores are explicitly labeled as similarities rather than probabilities that the identity is correct.

The verification endpoint permits only verifier or administrator role keys to record a terminal verified or rejected decision. A reason is required, and subsequent terminal decisions on the same match are rejected. Match updates, decision records, and audit entries share a transaction. Database triggers prevent ordinary update or delete operations on audit rows. This is a research safeguard, not a claim of tamper-proof logging against a database administrator. Role keys represent roles rather than individually accountable production users.

3.3 Input handling and prototype boundaries

The API rejects undecodable image uploads, constrains file size and decoded pixel count, and does not trust supplied filenames for persistence. Reference records contain descriptors and checksums rather than public media URLs. Access keys are generated locally and the demo binds to localhost. Production account management, expiring tokens, encrypted storage, multi-person scene processing, video jobs, geospatial context, and alert delivery are not implemented. No legal-compliance certification or production security assessment is claimed.

4 Methodology

4.1 Representations and matching formulation

Equation (1) defines L2 normalization and cosine similarity for a descriptor x and query/reference representations. The centered raw-pixel baseline resizes an image to 32 by 32 grayscale pixels, subtracts its mean, and normalizes the resulting 1,024-dimensional vector. The HOG baseline resizes grayscale input to a 64 by 128 window and uses 16 by 16 blocks, an 8 by 8 block stride, 8 by 8 cells, and nine orientation bins, producing 3,780 dimensions. Zero-energy or non-finite vectors are rejected.

$$z = \frac{x}{\|x\|_2}, s(q, r) = z_q^T z_r \quad (1)$$

SFace produces a normalized 128-dimensional representation through the official ONNX model. In the controlled face experiment, existing Olivetti crops are converted to three channels and resized to 112 by 112 before feature extraction. No learned face weights are fitted on Olivetti. For the ear projection comparison, a 64-component, non-whitened randomized PCA is fitted only to gallery HOG vectors; projected gallery, validation, and test descriptors are subsequently normalized. The ear API uses full HOG, not the PCA comparison variant.

Equation (2) defines the reference score for candidate identity i and modality m , using the maximum similarity across that identity's compatible references. This maximum allows different views without creating duplicate candidate identities. The standalone component experiments rank identities by this component score. A candidate passes a modality threshold when the score reaches or exceeds that modality's validation-selected threshold; otherwise that modality does not establish candidate eligibility.

$$S_m(q, i) = \max_{r \in R_{m,i}} s(q_m, r) \quad (2)$$

Equation (3) defines joint ranking over the available compatible query/reference modalities in the set A . The provisional weights are 0.7 for face and 0.3 for ear. Eligibility requires at least one component to reach its own threshold. With one available modality, normalization reduces the ranking to that modality score. These weights are engineering defaults, not experimentally optimized fusion parameters, and the combination is not a calibrated identity probability. Without a verified paired dataset, the paper reports no fused identification accuracy.

(3)

4.2 Real datasets and split construction

The face evaluation includes all 400 real Olivetti observations from 40 labeled subjects, obtained through the official scikit-learn Figshare mirror [6]. The ear archives are downloaded directly from the AMI university host [7] and contain 700 images from 100 subjects. The ear experiment retains the six right-ear views per subject, giving 600 observations, and excludes the BACK left-ear view. This removes untested cross-side comparisons but does not create a new publicly redistributed image collection.

Three random partitions are generated with seeds 17, 29, and 43. For each face partition, 30 subjects are enrolled and ten are unseen. Five images per enrolled subject form the gallery, two form validation probes, and three form test probes. Five unseen subjects contribute all ten images each to validation; the remaining five contribute all ten images each to testing. Therefore the face gallery has 150 images, validation has 110 probes, and testing has 140 probes: 90 known and 50 unknown.

For each ear partition, 80 subjects are enrolled, ten unseen subjects are reserved for validation, and ten different unseen subjects are reserved for testing. Per enrolled subject, three right-ear observations form the gallery, one forms validation, and two form testing. Each unseen subject contributes all six retained views. The resulting ear gallery has 240 images, validation has 140 probes, and testing has 220 probes: 160 known and 60 unknown. Gallery, validation, and test images are disjoint; enrolled identities necessarily appear in gallery and probe sets, whereas unknown validation and test identities are disjoint.

4.3 Threshold selection and metrics

Thresholds are selected on validation scores by maximizing the balanced criterion in (4). DIR is the fraction of known probes that are correctly ranked first and accepted; FPIR is the fraction of unknown probes receiving an accepted candidate. Ties favor the larger threshold. This criterion balances known-correct identification and unknown rejection for the study; it is not an operational risk policy established with rescue personnel. All stress tests retain the clean-validation threshold without retuning.

$$B(\tau) = \frac{DIR_{val}(\tau) + 1 - FPIR_{val}(\tau)}{2} \quad (4)$$

Rank-1 and rank-5 rates are computed over known probes only. Accepted precision divides correctly accepted known probes by all accepted probes, counting wrong known identifications and accepted unknown people against precision. A known-versus-unknown score AUC is retained as a diagnostic, not as identity-verification AUC. Mean and sample standard deviation across the three random partitions are reported. The partitions reuse observations, so these deviations are sensitivity summaries rather than three independent population replications.

For a prespecified primary partition, seed 17, identity-level bootstrap intervals are also calculated for clean rank-1 retrieval. Two thousand resamples of known identities use bootstrap seed 101, retaining each identity's mean success across its probes. This avoids treating multiple images of the same subject as fully independent. The supplementary CSV stores the intervals, exact denominators, split manifests, and every per-probe outcome. No cross-paper numerical superiority claim is made.

4.4 Stress conditions and reproducibility

Four conditions are evaluated: unchanged images; Gaussian blur with a 7 by 7 kernel and standard deviation 1.5; reduced intensity by a factor of 0.35; and a central rectangle occupying one third of image width and one third of image height replaced with zero. These are deterministic transformations of real held-out observations, computed in memory. They are useful sensitivity probes but are not genuine smoke, dust, injury, or disaster-site datasets. Model parameters and thresholds remain fixed across conditions.

The recorded computer uses an Intel Core i5-1235U, approximately 16 GB RAM, Intel Iris Xe integrated graphics, and Windows 11. Inference is CPU-only. The environment uses Python 3.12, OpenCV 4.12.0.88, NumPy 2.2.6, and scikit-learn 1.9.1. OpenCV and numerical-library thread counts are bounded during descriptor experiments. Archive and model SHA-256 hashes, image-level split hashes, package versions, raw predictions, and scripts are retained. The hardware differs from the GTX 1650 configuration in the design document; no timings are attributed to that GPU.

5 Experimental results and evaluation

5.1 Direct representation comparisons

Table 1 reports direct comparisons under identical per-dataset partitions. SFace obtains mean face rank-1 retrieval of 96.67% with a partition standard deviation of 1.92 percentage points. HOG reaches 97.04%, while centered pixels reach 90.00%. HOG and SFace are close on these controlled crops; the experiment does not establish that a deep representation is universally superior. The difference is specific to this small acquisition setting and preprocessing choice.

Table 1. Clean retrieval and open-set identification.

Dataset / method	Rank 1 %	Rank 5 %	DIR %	FPIR %
Face / Raw pixels	90.00	95.56	60.37	10.00
Face / HOG	97.04	100.00	82.22	7.33
Face / SFace	96.67	100.00	81.11	10.67
Ear / Raw pixels	37.50	52.50	9.58	1.67
Ear / HOG	57.92	71.67	35.83	8.33
Ear / HOG + PCA64	52.08	67.29	23.13	7.22

For AMI right-ear images, HOG achieves 57.92% rank-1 retrieval, compared with 37.50% for centered pixels and 52.08% for HOG with PCA64. Full HOG exceeds the raw-pixel baseline by 20.42 percentage points in the partition means. The compact projection loses retrieval performance in this protocol. These observations support HOG as a lightweight starting point, while its error rate remains too large to support autonomous identification.

In the primary partition, SFace correctly retrieves 95.56% of known face probes; the identity-bootstrap 95% interval is 91.11% to 98.89%. For HOG ear retrieval, the corresponding estimate is 53.75% and the interval is 45.00% to 62.50%. These intervals quantify sampling variation within the fixed dataset and primary split. They do not quantify generalization to disaster populations or account for uncertainty about the pretrained face model's original training overlap.

5.2 Open-set acceptance and threshold sensitivity

Rank-based success is substantially higher than correct accepted identification. SFace's mean clean DIR is 81.11%, with FPIR 10.67%. HOG ear DIR is 35.83%, with FPIR 8.33%. The component thresholds from seed 17 are 0.793873 for face and 0.892733 for ear. These values are protocol-specific similarities. They are neither probabilities nor thresholds validated for arbitrary photographs or a deployed rescue registry.

The clean thresholds from the primary partition are fixed before test evaluation. Unknown samples are small, and the selection criterion weights known identification and unknown rejection equally. An operational study would need an independently specified false-alarm budget, larger galleries, and a human-review workload assessment.

5.3 Robustness on transformed real observations

With fixed thresholds, central occlusion changes SFace mean rank-1 retrieval from 96.67% to 91.48%, while DIR decreases from 81.11% to 19.63%. Blur produces rank-1 92.96% but DIR only 24.81%. Thus, a descriptor can retain useful candidate ordering while its scores fall below the clean-calibrated acceptance threshold. This is a material distinction for interfaces that show only high-score candidates.

For HOG ear retrieval, central occlusion yields rank-1 55.21% and DIR 0.21%, versus clean values of 57.92% and 35.83%. The near collapse of accepted correct candidates shows that the baseline threshold is not robust to this perturbation. A lower unknown acceptance rate under degradation cannot by itself be read as an improvement: it accompanies a large loss of known identifications. Figure 2 displays these differences.

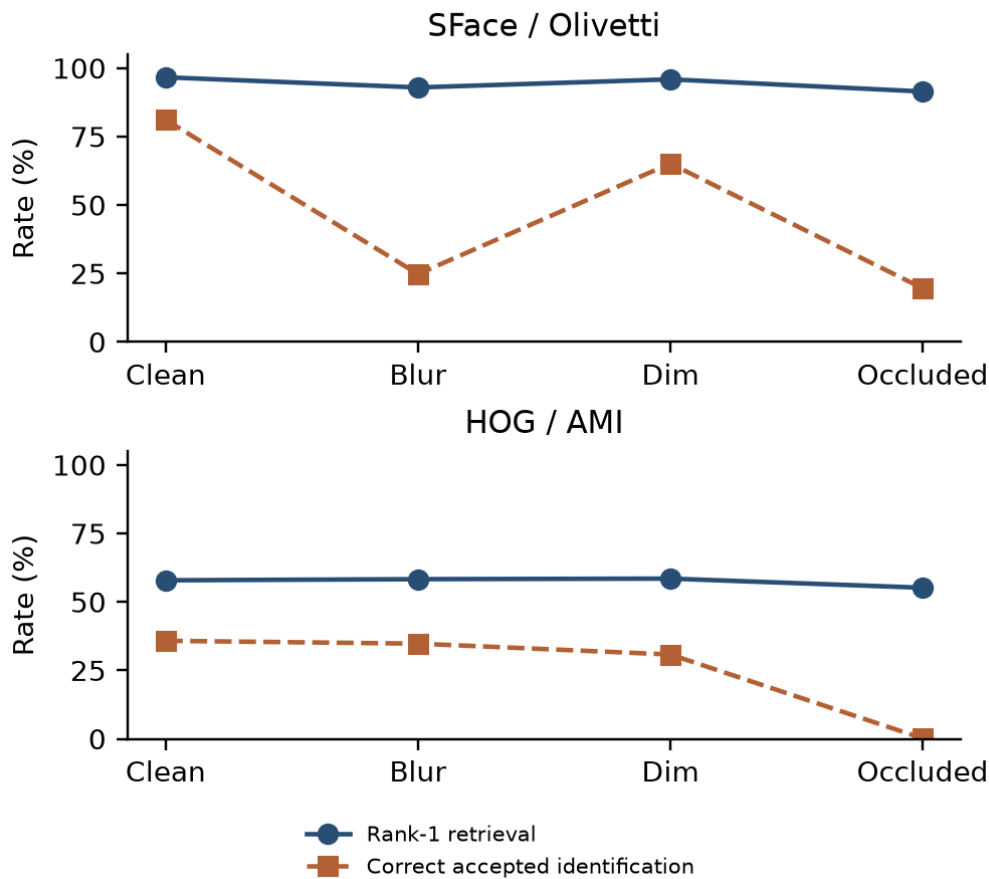


Figure 2. Rank-1 retrieval and correct accepted identification at fixed clean-validation thresholds. Perturbations transform real test images and do not constitute disaster-field validation.

Table 2. Fixed threshold stress results.

Input / condition	Rank 1 %	DIR %	FPIR %
Face clean	96.67	81.11	10.67
Face blur	92.96	24.81	2.67
Face dark	95.93	65.19	4.67
Face occlusion	91.48	19.63	1.33
Ear clean	57.92	35.83	8.33
Ear blur	58.33	34.79	7.22
Ear dark	58.54	30.83	5.56
Ear occlusion	55.21	0.21	0.00

5.4 Software correctness and measured latency

The final automated suite passes 31 checks. Covered cases include missing or invalid authentication, role-denied actions, required registration basis, real-image descriptor extraction, empty registries, invalid uploads, upload-size limits, candidate persistence, mandatory pending state, terminal decisions, duplicate-decision conflicts, audit access restrictions, incompatible models, multiple-reference aggregation, and dataset-split integrity. Passing this suite verifies the exercised contracts. It does not establish complete security, correctness under all failures, or human-verification accuracy.

Thirty actual loopback HTTP queries use held-out AMI UP views against a ten-case registry enrolled from FRONT views. All return HTTP 200. Median round-trip latency is 18.68 ms, mean is 19.96 ms, and the 95th percentile is 27.19 ms. The workflow control returns 403 for a rescuer's verification attempt, 200 for an authorized decision, and 409 for a duplicate terminal decision. A separately labeled exact-reference control is used for this decision-path check; it is not part of the recognition-accuracy benchmark.

Excluding the first descriptor call, SFace feature extraction has median 92.37 ms and 95th percentile 97.58 ms over 399 face crops. Ear HOG has median 0.73 ms and 95th percentile 1.23 ms over 599 retained ear images. These are descriptor-stage measurements. The HTTP ear-query measurement includes request handling, image decoding, descriptor calculation, reference retrieval, and persistence, but excludes mobile-network delay, human review, and full-scene detection. It cannot support a live multistream real-time claim.

6 Discussion and limitations

The benchmark uses real research observations with provider attribution. AMI specifies a noncommercial, no-derivatives license [7]; modified images are evaluated in memory and are not re-distributed as a new dataset. Reference descriptors, filenames, case records, and audit logs still require controlled handling. Localhost binding and role separation reduce exposure in the demonstration, but production notice, authorization basis, retention, individual accountability, encrypted storage, and misuse prevention require additional implementation and review. Human review reduces automatic-decision authority but does not by itself eliminate recognition error or reviewer bias.

Internal validity is limited by crop-based evaluation, maximum-reference aggregation, a fixed descriptor implementation, and the small number of partitions. The pretrained face artifact is not retrained here, and overlap between its original training subjects and the evaluation population is not independently audited. External validity is limited by controlled datasets, indoor ear imagery, small galleries, exclusion of left-ear views, and the absence of real disaster observations. Synthetic degradation of real images cannot reproduce injury, camera motion trajectories, smoke dynamics, or field registration conditions.

Construct validity is limited because retrieval success and correctly recorded decisions are not the same as operationally correct identity verification. The study includes no rescue-operator user trial or independent secondary-identification evidence. Statistical validity is limited by repeated use of the same finite population across splits and few unseen identities. Bootstrap intervals reflect within-dataset uncertainty only. Demographic fairness cannot be inferred from aggregate performance; task-specific subgroup evaluation would be required [9]. Most importantly, the two datasets do not share verified identities, so the implementation's multimodal ranking formula has no measured fusion-performance result.

7 Conclusion and future work

ResQ-ID provides a runnable and reproducible identification-assistance baseline with real-data component comparisons and tested human-decision controls. SFace face retrieval reaches mean rank-1 96.67%, and HOG ear retrieval reaches 57.92%, under the specified controlled protocols. The much lower thresholded identification rates, particularly under occlusion, show why rank accuracy

must not be presented as reliable automatic identification. The ear baseline also requires substantial improvement before field use.

The next research steps follow these limitations: obtain a consented paired face-ear collection; evaluate joint ranking against face-only and ear-only baselines on the same people; study quality-aware threshold calibration without test leakage; train and validate an ear-localization and representation model; and test complete scene processing and human review under realistic acquisition conditions. The present findings support continued development and careful evaluation. They do not establish disaster deployment readiness or superiority over published systems evaluated under different protocols.

References

- [1] Zhong Y, Deng W, Hu J, Zhao D, Li X, Wen D. SFace: Sigmoid-constrained hypersphere loss for robust face recognition. *IEEE Transactions on Image Processing*. 2021;30:2587–2598. <https://doi.org/10.1109/TIP.2020.3048632>
- [2] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2005. Vol. 1, p. 886–893. <https://doi.org/10.1109/CVPR.2005.177>
- [3] Emeršič Ž, Štruc V, Peer P. Ear recognition: More than a survey. *Neurocomputing*. 2017;255:26–39. <https://doi.org/10.1016/j.neucom.2016.08.139>
- [4] Boodoo NB, Subramanian RK. Robust multi biometric recognition using face and ear images. *International Journal of Computer Science and Information Security*. 2009;6(2):164–169. <https://arxiv.org/abs/0912.0955>
- [5] Kim M, Jain AK, Liu X. AdaFace: Quality adaptive margin for face recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2022. p. 18750–18759.
- [6] scikit-learn contributors. fetch_olivetti_faces: dataset documentation and official loader. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_olivetti_faces.html [Accessed 4 October 2026].
- [7] Gonzalez E, Alvarez L, Mazon L. *AMI Ear Database*. Universidad de Las Palmas de Gran Canaria. Available from: https://ctim.ulpgc.es/research_works/ami_ear_database/ [Accessed 4 October 2026].
- [8] OpenCV contributors. *OpenCV Zoo: Face detection YuNet and face recognition SFace* [model repository]. Available from: https://github.com/opencv/opencv_zoo [Accessed 4 October 2026].
- [9] Grother P, Ngan M, Hanaoka K. *Face Recognition Vendor Test Part 3: Demographic Effects*. NISTIR 8280. National Institute of Standards and Technology; 2019. <https://doi.org/10.6028/NIST.IR.8280>
- [10] Deng J, Guo J, Xue N, Zafeiriou S. ArcFace: Additive angular margin loss for deep face recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2019. p. 4690–4699.
- [11] Deng J, Guo J, Ververas E, Kotsia I, Zafeiriou S. RetinaFace: Single-shot multi-level face localisation in the wild. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2020. p. 5203–5212.
- [12] Turk M, Pentland A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*. 1991;3(1):71–86. <https://doi.org/10.1162/jocn.1991.3.1.71>