# The Unconstrained Ear Recognition Challenge 2026
# Size Matters: Ear Biometrics on Edge*

D. Sušanj[1], J. Jin[2], M. Tokumasu[2], T. Ohki[2], A. Kamboj[3], A. Rai[4],S. Aggarwal[4],S. Ahluwalia[4],M. Chipade[4],A. Sharma[4], D. Pinčić[5], M. Banov[6], K. Lenac[6], K. Shigematsu[7], K. Inoue[7], A. Shin[7], H. Shirono[8], A. Hazra[9], T. Batabyal[9], S. Pal[9], D. Freire-Obregón[10], G. Gupta[11], A. Gupta[11], D. Arun[12], P. Flynn[12], J. L. Godoy[13], D. D. A. Ruiz[13], J. N. Khiarak[14], F. N. Khiarak[14], T. A. Saeed[14], L. D. M. S. S. Teja[15], D. P. Chowdhury[15], K. Griparić[1], R. Lajić[18], D. Tomašević[18], B. Meden[18], H. K. Ekenel[16], G. Cámara-Chávez[17], V. Štruc[19], P. Peer[18], Ž. Emeršič[18]

[1]Juraj Dobrila University of Pula, Faculty of Engineering, Croatia, [2]Shizuoka University, Japan, [3]Michelin, AI-COE Pune, India, [4]National Institute of Technology Kurukshetra, India, [5]University of Rijeka, Faculty of Law, Croatia, [6]University of Rijeka, Faculty of Engineering, Croatia, [7]National Institute of Technology, Oita College, Japan, [8]Kyushu Institute of Technology, Japan, [9]Department of CST & CSE-AIML, JIS College of Engineering, India, [10]Universidad de Las Palmas de Gran Canaria, Spain, [11]Arogyapandit Private Limited, India, [12]University of Notre Dame, USA, [13]Pontifícia Universidade Católica do Rio Grande do Sul, Brazil, [14]Kartal Ol NGO, Poland, [15]National Institute of Technology Silchar, India, [16]Istanbul Technical University, Turkey, [17]Federal University of Ouro Preto, Brazil, [18]University of Ljubljana, Faculty of Computer and Information Sciences, Slovenia, [19]University of Ljubljana, Faculty of Electrical Engineering, Slovenia

## Abstract

*The 2026 Unconstrained Ear Recognition Challenge (UERC 2026), a benchmarking competition focused on lightweight and efficient ear recognition from images captured in uncontrolled environments. Building on the previous editions of the UERC series, the 2026 challenge introduced a novel evaluation paradigm that jointly assesses recognition accuracy and computational efficiency, with the goal of promoting the development of compact ear recognition models suitable for deployment on resource-constrained and edge platforms. The competition was organized into two tracks: Track 1: Lightweight Model Design, which focused on balancing recognition performance with model efficiency through a composite evaluation criterion combining verification accuracy, parameter count, and model size, and Track 2: Edge Deployment which assessed recognition accuracy together with end-to-end inference time on a standardized Raspberry Pi 4 platform. A total of 11 research groups participated in Track 1 and 8 groups competed in Track 2, submitting solutions spanning a wide range of architectures including lightweight convolutional networks, vision transformers, ensemble-distillation pipelines, and ONNX-accelerated models. The results demonstrate that carefully designed compact architectures, such as knowledge-distilled MobileNetV4 and INT8-quantized TinyViT networks, can achieve competitive verification performance while drastically reducing parameter counts and inference time. The competition's starter kit, baseline code, and dataset are publicly available at* http://uerc.fri.uni-lj.si/.

*This research was supported in parts by the ARRS Research Programmes P2-0250(B) "Metrology and Biometric Systems" and P2-0214 "Computer Vision", by the EU's Horizon Europe research and innovation programme under Grant Agreement No 101225635 – OnMoveID, and EU NextGeneration under the project number IIP_UNIPU_A010148

## 1. Introduction

Ear recognition is an important research area within the field of biometrics [6,7]. Unlike face recognition, which benefits from dedicated hardware (e.g., structured-light depth sensors) and abundant large-scale training data, ear recognition remains a comparably less explored modality, yet it holds practical appeal due to the non-intrusiveness of the acquisition process and the distinctiveness of ear morphology. Existing work in this field has predominantly focused on maximizing raw recognition performance, while aspects critical for real-world deployment have largely been overlooked, including model efficiency.

Modern biometric recognition systems are increasingly expected not only to deliver high accuracy, but also to operate under the constraints imposed by resource-limited hardware, such as smartphones, wearables, and edge sensors. In these settings, model size, parameter count, and inference latency are key concerns alongside recognition performance [11,13]. Despite significant progress in efficient deep learning, including model pruning, quantization [8,15,18], knowledge distillation [10] and the design of inherently compact architectures [26,29,32], these advances have, to the best of our knowldge, now been systematically benchmarked in the context of ear recognition.

To address the growing need for deployable biometric systems, the 2026 edition of the Unconstrained Ear Recognition Challenge (UERC 2026) was organized as part of the 2026 IEEE/IAPR International Joint Conference on Biometrics. In contrast to traditional biometric benchmarks that primarily emphasize recognition performance, UERC 2026 introduced an evaluation paradigm that jointly considers recognition accuracy, model complexity, and practical inference efficiency. The competition was specifically designed to encourage the development of lightweight ear recognition systems suitable for resource-constrained and edge deployment scenarios.

The challenge was organized into two complementary tracks. *Track 1: Lightweight Model Design* focused on the development of compact ear recognition models and ranked submissions using a composite criterion combining verification performance, parameter count, and model size. *Track 2: Edge Deployment* focused on practical runtime efficiency and evaluated submissions based on recognition performance and end-to-end inference latency measured on a standardized Raspberry Pi 4 under identical CPU only conditions.

A total of 11 research groups participated in Track 1 and 8 groups competed in Track 2. The submitted solutions covered a diverse range of architectural and deployment strategies, including lightweight convolutional networks (EfficientNet-B0, MobileNetV3-Large, ConvNeXt variants), compact vision transformers (TinyViT, DINOv2-Small), ensemble distillation frameworks (GIN-Ear, DEAR), and ONNX Runtime optimized inference pipelines. The results demonstrate that carefully designed lightweight architectures, particularly those leveraging distillation, quantization, and deployment aware optimization, can achieve strong verification performance while maintaining low computational and memory requirements suitable for edge deployment.

The main contributions of this paper are summarized as follows:

- We introduce the first large-scale benchmark in ear recognition that jointly evaluates biometric verification performance, model complexity, and practical deployment efficiency under a unified experimental framework.
- We present a comprehensive comparative analysis of 11 lightweight ear recognition approaches and 8 edge deployment solutions, providing detailed insights into the trade-offs between recognition accuracy, parameter efficiency, and inference latency.
- We evaluate a broad range of modern efficiency-oriented design strategies, including lightweight convolutional networks, compact vision transformers, knowledge distillation, quantization, and ONNX Runtime optimized deployment pipelines.
- We publicly release the competition toolkit, including the evaluation protocol, baseline implementation, and dataset, to support reproducible research and future benchmarking efforts in efficient ear recognition.

## 2. Related Work

This section briefly reviews prior editions of the UERC competition series together with recent advances in efficient deep learning and modern ear recognition methodologies that form the technical foundation of UERC 2026.

**UERC Competition Series.** UERC 2026 is the fourth in the Unconstrained Ear Recognition Challenge series. The first UERC [6] was organized at IJCB 2017 and introduced one of the first large-scale in-the-wild ear recognition benchmarks, focusing on open-set identification in the presence of many distractor identities. UERC 2019 [5], held at ICB 2019, again addressed identification, but additionally investigated robustness to image resolution, head rotation, and occlusion. UERC 2023 [4] shifted the focus to verification performance and demographic fairness, introducing a joint criterion that penalized verification errors and performance differentials due to gender and ethnicity. All three competitions attracted multiple international research groups and produced widely referenced summary papers. UERC 2026 builds on these foundations but introduces an entirely different challenge dimension: lightweight design and edge deployability.

**Efficient Deep Learning.** The rapid growth of mobile and IoT devices has driven significant interest in efficient neural network design. Foundational lightweight architectures such as MobileNet [11], EfficientNet [29], and ConvNeXt [20] have demonstrated that competitive accuracy can be achieved with far fewer parameters than their larger counterparts. More recently, compact vision transformers like TinyViT [32] and self-supervised models such as DINOv2 [24] have extended this trend to attention-based architectures. Knowledge distillation [10] and quantization [15] further compress models for embedded deployment, while ONNX Runtime and hardware-aware inference frameworks enable efficient execution on CPUs. Despite this rich body of work, systematic benchmarking of these methods specifically for ear recognition has not yet been considered.

**Ear Recognition.** Ear recognition has been studied as a biometric modality for several decades [7], with recent work adopting deep convolutional networks and, increasingly, vision transformers for feature extraction. Margin-based learning objectives, including ArcFace [3], CosFace [30], and AdaFace [17], have proven particularly effective for learning discriminative ear embeddings. However, prior to UERC 2026, the field lacked a benchmark that explicitly targets the accuracy-efficiency trade-off under controlled conditions.

## 3. Methodology

This section describes the organization of UERC 2026, including the competition tracks, datasets, evaluation protocol, performance indicators, and baseline resources provided to participants.

### 3.1. Competition Organization

UERC 2026 was organized as a two-track competition focused on lightweight ear recognition and edge deployment. Participants were free to compete in one or both tracks.

**Track 1: Lightweight Model Design.** The first track collected compact ear recognition models and scored their performance on images captured in unconstrained environments.

The performance indicators included both recognition accuracy and quantitative measures of model complexity (number of parameters and model size). These components jointly contributed to the overall ranking, encouraging participants to design efficient architectures without imposing strict hard limits. Participants were free to develop any type of model and optimize it through architectural design, pruning, quantization, distillation, or other model compression strategies. Final submissions consisted of a working solution (source code or compiled binary) that the organizers executed to evaluate recognition performance and compute complexity measures on a sequestered test dataset.

**Track 2: Edge Deployment.** The second track focused on practical deployment efficiency under fixed hardware conditions. All submitted solutions were executed on a standardized Raspberry Pi 4 Model B (Revision 1.2) with 4 GB of RAM, equipped with a Broadcom BCM2711 quad-core Cortex-A72 CPU running at 1.5 GHz, and running a 64-bit Raspberry Pi OS (Debian Trixie port). All inference was CPU-only, with no GPU or NPU acceleration. The evaluation combined recognition accuracy with measured inference time per image, reflecting real-world edge constraints. Batch size was fixed to 1, timing measurements followed a standardized protocol with thermal throttling disabled and non-essential background processes minimized to ensure fair comparison, and each submission was warmed up with 1000 runs of model inference/feature extraction on random pairs of images before timed evaluation on 3500 image pairs.

### 3.2. Training and Testing Data

The training and testing data for UERC 2026 are based on the UERC 2023 dataset [4]. The training set consists of:

- A dataset of 14,004 images from 650 distinct subjects, collected in-the-wild from the web (including images from UERC 2017 and UERC 2019 training sets).
- Over 234,651 images from 660 subjects from the VGGFace-Ear dataset [33], generated by cropping and normalizing the ear region from VGGFace images.

The training data was made available along with identity labels to allow participants to train their models and select suitable hyperparameters. Additional external data was permitted at participants' discretion.

The test data was sequestered and not made accessible to participants. Final evaluation was conducted exclusively by the organizers on the hidden (sequestered) test set, and the same sequestered set was used for both Track 1 and Track 2 evaluations. The sequestered test set consists of 1,670 images from 70 distinct subjects across four ethnicity categories (Asian, Black, White, Other) and two gender categories (female, male). The dataset is gender-balanced (35 female, 35 male subjects). Asian, Black, and White groups each contain 10 subjects per gender (10 female + 10 male), while the Other group contains 5 subjects per gender (5 female + 5 male). Like the training data, the test images were collected in-the-wild and exhibit substantial appearance variability, including lateral and frontal head pose, partial occlusion from hair, earrings, and accessories, varying illumination (from controlled indoor to challenging outdoor conditions), and a range of image resolutions and compression levels, as illustrated in Fig. 1. Track 1 is evaluated on all pairwise comparisons within the sequestered set, yielding a total of 1,393,615 pairs (20,000 mated and 1,373,615 non-mated). Track 2 evaluation uses a fixed subset of 3,500 image pairs on the Raspberry Pi 4 platform.

Figure 1. **Example images from two subjects (rows) from the sequestered test dataset of UERC 2026.** Images are captured in-the-wild and exhibit considerable appearance variability due to pose, lighting, and occlusion.

### 3.3. Experimental Protocol

For all submissions, participants were required to submit a working solution (source code or binary) that the organizers ran on the sequestered test data. Track 1 solutions were executed on the organizers' GPU server. Track 2 solutions were executed exclusively on the Raspberry Pi 4 platform described above. Verification scores were computed using cosine similarity between L2-normalized feature embeddings. Track 1 performance was evaluated using all pairwise comparisons within the test set (mated and non-mated pairs), following the standard biometric verification protocol [14]. For Track 2, each repeated run consisted of 1000 warmup executions of model inference/feature extraction on random image pairs followed by evaluation on 3500 image pairs on the Raspberry Pi 4 platform. Final rankings for both tracks were computed using averages over repeated evaluation runs

to reduce the influence of stochastic variation.

### 3.4. Performance Indicators

**Track 1: Lightweight Model Design.** The primary ranking criterion combines recognition accuracy with model efficiency:

$$S_1 = 0.5\,v + 0.2\,p + 0.3\,s, \tag{1}$$
$$v = \mathcal{N}(\mathrm{VER}_{0.1\%}),$$
$$p = 1 - \mathcal{N}(\log P),$$
$$s = 1 - \mathcal{N}(\log M),$$

where $\mathrm{VER}_{0.1\%}$ is the Verification Rate at a False Acceptance Rate (FAR) of 0.1%, $P$ is the number of model parameters, and $M$ is the model size. $\mathcal{N}(\cdot)$ denotes min–max normalization applied across valid submissions, mapping values to $[0, 1]$. Following the scoring code, the parameter count and model size are log-transformed before normalization, and both complexity terms are inverted so that smaller models receive larger contributions to the final score. In addition, the starter-kit baseline performance is used as a Track 1 reference limit in the normalization step: the lower normalization bound for $\mathrm{VER}_{0.1\%}$ is clipped to at least the baseline value, so submissions falling below the baseline are penalized accordingly.

**Track 2: Edge Deployment.** The second track evaluates practical deployment efficiency on the fixed edge platform:

$$S_2 = 0.7\,v + 0.3\,t, \tag{2}$$
$$v = \mathcal{N}(\mathrm{VER}_{0.1\%}),$$
$$t = 1 - \mathcal{N}(T),$$

where $\mathrm{VER}_{0.1\%}$ is the Verification Rate at FAR of 0.1%, and $T$ is the total measured inference time over the 3500 evaluated image pairs (in seconds). Again, min–max normalization is applied across valid submissions, and the latency term is inverted so that faster methods receive a larger contribution to the final score.

**Scoring rationale.** The weighting in $S_1$ assigns the largest weight to verification accuracy ($w_v$=0.5) to ensure that biometric utility remains the primary objective, while the complementary weight is split between parameter count ($w_p$=0.2) and model size ($w_s$=0.3). The slightly higher weight on model size reflects that on-disk storage is a harder constraint than theoretical parameter count for edge deployment. For Track 2, accuracy carries a dominant weight ($w_v$=0.7) because a biometric system must first satisfy a minimum accuracy threshold before deployment speed becomes the deciding factor; inference time receives weight $w_t$=0.3 to meaningfully reward faster edge execution.

To assess whether the official rankings are sensitive to these choices, we re-ranked all submissions under three alternative weighting schemes per track. For **Track 1**, the tested schemes were: accuracy-heavy ($0.7v+0.15p+0.15s$), equal ($1/3$ each), and efficiency-heavy ($0.3v + 0.35p + 0.35s$). GIN-Ear remains the top-ranked submission in all four Track 1 scenarios. The top-four set (GIN-Ear, Cascaded-PruneEar, DEAR, EfficientNet-B0 AdaFace) is stable in three of four weightings; only in the two efficiency-heavy variants do DEAR and Cascaded-PruneEar swap positions 2 and 3. For **Track 2**, we tested accuracy-heavy ($0.9v + 0.1t$), equal ($0.5v + 0.5t$), and speed-heavy ($0.3v + 0.7t$) schemes. GIN-Ear leads in all four Track 2 scenarios; the same four approaches (GIN-Ear, Cascaded-PruneEar, EfficientNet-B0 AdaFace, DEAR) occupy the top four positions in all weightings, with positions 2–4 varying by accuracy/speed emphasis. These results confirm that the overall competition outcome is robust to the specific choice of weighting coefficients.

In addition to the composite scores, the following standard biometric performance indicators are reported for each submission: Area Under the ROC Curve (AUC), Equal Error Rate (EER), and the Verification Rate (VER) at FAR values of 0.1% and 1%.

### 3.5. Starter Kit and Baseline

To provide a quick entry point for participants, a starter kit was distributed that included the UERC dataset, evaluation code, and a baseline model based on a compact convolutional architecture. The baseline followed the UERC 2023 setup [4] and was provided with pre-trained weights. All evaluation scripts and competition materials are publicly available at http://uerc.fri.uni-lj.si/.

## 4. Summary of Participating Approaches

UERC 2026 received submissions from 11 research groups in Track 1 and 8 groups in Track 2. Table 1 provides a high-level overview of all participating approaches. Brief descriptions of each approach follow below.

**Baseline.** The starter kit included a ConvNeXt-Base baseline model, distributed with pre-trained weights and the standard UERC verification pipeline. The baseline was intended as a reference implementation for feature extraction and cosine-based scoring rather than as a ranked UERC entry.

**GIN-Ear.** The approach builds a compact **G**uided **I**dentity-balanced **N**etwork for **Ear** recognition [26, 28]. Before training, an Ear Identity-Balanced Augmentation procedure (EIB-Aug) balances the UERC training set by augmenting under-represented identities using 55 estimated ear landmarks [9] to generate occlusion samples (hair, earrings, earphones) [22] and pseudo-rotated images at yaw angles of 15°, 30°, 45°, and 60° [1]. A two-stage knowledge distillation pipeline then transfers knowledge from a DINOv3 ViT-B teacher to a DINOv3 ViT-S intermediate model, and subsequently to a MobileNetV4 Conv-Small student. The student produces a 512-dimensional normalized embedding optimized with

Table 1. **High-level summary of the approaches submitted to UERC 2026.** Eleven groups participated in Track 1 and eight in Track 2. The submitted solutions span a range of deep learning architectures, including lightweight CNNs, compact vision transformers, and ensemble distillation pipelines.

| Approach | Institution | Backbone | Track 2 |
|---|---|---|---|
| GIN-Ear | Shizuoka University, Japan | MobileNetV4 Conv-Small | ✓ |
| Cascaded-PruneEar | Michelin, AI-COE Pune / National Institute of Technology Kurukshetra, India | TinyViT-11M / TinyViT-5M (pruned) | ✓ |
| DEAR | University of Rijeka, Faculty of Law / University of Rijeka, Faculty of Engineering, Croatia | TinyViT-5M | ✓ |
| EfficientNet-B0 AdaFace | National Institute of Technology, Oita College / Kyushu Institute of Technology, Japan | EfficientNet-B0 | ✓ |
| TinyViT-Ear / EdgeNeXt-Ear | Department of CST & CSE-AIML, JIS College of Engineering, India | TinyViT-5M / EdgeNeXt-Small | ✓ |
| AurisPico | Universidad de Las Palmas de Gran Canaria, Spain | ConvNeXtV2-Pico | – |
| ConvNeXt ArcFace | Arogyapandit Private Limited, India | ConvNeXt-Zepto | ✓ |
| MobileNetEar | University of Notre Dame, USA | MobileNetV3-Large | ✓ |
| Stripe | Pontifícia Universidade Católica do Rio Grande do Sul, Brazil | ConvNeXt-Tiny | – |
| DINOv2-Ear | Kartal OI NGO, Poland | DINOv2 ViT-Small | – |
| Dino-Ear-NITS | National Institute of Technology Silchar, India | DINOv2-B/14 / S/14 | ✓ |

a combination of ArcFace loss, KL-divergence loss, and embedding-level distillation (MSE in stage 1; cosine loss in stage 2). For Track 1, only the MobileNetV4 feature extractor is retained. For Track 2, the model is converted to FP32 ONNX and executed with ONNX Runtime [23] for efficient CPU-only inference on the Raspberry Pi.

**Cascaded-PruneEar.** The submission uses cascaded knowledge distillation and structured pruning for compact ear verification. Ear images are left-aligned with a CLIP-based model, resized to $224 \times 224$, stored as lossless WebP, and augmented with CLAHE, random unsharp masking, ColorJitter, and RandomGrayscale. For Track 1, a SwinV2-Base teacher [19] is distilled into TinyViT-21M and then TinyViT-11M [32], after which the final model is pruned by $45\%$ using first-order Taylor importance to obtain an approximately 6M-parameter model. For Track 2, TinyViT-21M is distilled into TinyViT-5M and then pruned to approximately 1M parameters for edge deployment. The network uses a 512-dimensional $L_2$-normalized embedding for verification and is trained with AdaFace loss [17] together with an MSE-based embedding distillation loss.

**DEAR.** The approach employs a compact ensemble-distillation strategy. In the first stage, three teacher models (TinyViT-11M, SwinV2-Small, and SE-ResNet152D) [12, 19, 32] are trained independently on the UERC 2026 training data using a supervised classification objective. A student TinyViT-11M network is then trained to absorb the ensemble knowledge via a multi-term distillation objective [10] that combines cross-entropy loss, cosine embedding alignment with each teacher, relational knowledge distillation [25], and supervised contrastive loss [16]. A second compression stage further distills TinyViT-11M into a smaller TinyViT-5M architecture. The final model is quantized to INT8 precision to reduce model size and accelerate inference. The same model is used for the evaluation in both tracks.

**EfficientNet-B0 AdaFace.** The approach uses an EfficientNet-B0 [29] backbone as a compact ear embedding extractor, initialized from ImageNet pre-trained weights. The model is fine-tuned with an AdaFace-style [17] angular margin loss (margin $m$=0.30, quality scale $h$=0.45) using AdamW optimization. Checkpoint selection is based on validation AUC to avoid overfitting on low-FAR metrics alone. For Track 1, model weights are stored in FP16 to reduce the on-disk footprint. For Track 2, an ONNX Runtime [23] inference path is used with qnnpack quantization backend, and repeated-image caching eliminates redundant forward passes during pair-wise evaluation.

**TinyViT-Ear / EdgeNeXt-Ear.** The submission uses different backbone architectures for the two tracks [21, 32]. **TinyViT-Ear** is the Track 1 model and uses TinyViT-5M (`tiny_vit_5m_224` [32]), initialized from ImageNet pre-trained weights via the `timm` framework [31]. It is trained with the standard CrossEntropy loss and AdamW optimizer with a CosineAnnealingLR schedule. During inference, the classification head is removed and the resulting embeddings are L2-normalized before cosine similarity scoring. **EdgeNeXt-Ear** is the Track 2 model and uses EdgeNeXt-Small (`edgenext_small_rw` [21]), again initialized from ImageNet pre-trained weights via the `timm` framework [31]. It follows the same training recipe based on CrossEntropy loss, AdamW, and CosineAnnealingLR. As with Track 1, the classification head is removed during inference and the output embeddings are L2-normalized before scoring.

**AurisPico.** The approach targets ear verification under explicit model size constraints using a ConvNeXtV2-Pico backbone with an ArcFace [3] metric learning objective. Training follows a staged fine-tuning schedule with progressively increasing angular margin. Prior to submission, the checkpoint is converted to FP16 precision, reducing the on-disk footprint to approximately 17 MB. Inference uses a lightweight test-time augmentation (TTA) strategy that averages embeddings from several deterministic crops. Additionally, batch normalization recalibration is applied using the unlabeled sequestered images as a domain adaptation mechanism. This approach participated in Track 1 only.

**ConvNeXt ArcFace.** The approach is based on a compact ConvNeXt backbone [20, 31] with a lightweight projection head outputting 256-dimensional L2-normalized em-

Table 2. **Track 1 results – Lightweight Model Design.** Results are sorted by the composite score $S_1$ (higher is better). VER@0.1% and VER@1% are verification rates (higher is better); EER and AUC are the Equal Error Rate (lower is better) and Area Under ROC Curve (higher is better). The last row shows the starter-kit baseline for reference only and it is not part of the official ranking.

| Rank | Approach | VER@0.1% ↑ | VER@1% ↑ | EER ↓ | AUC ↑ | P (M) ↓ | Size (MB) ↓ | Score ↑ |
|---|---|---|---|---|---|---|---|---|
| 1 | GIN-Ear | **0.508** | **0.700** | **0.095** | **0.967** | **1.78** | **7.1** | **1.000** |
| 2 | Cascaded-PruneEar | 0.471 | 0.689 | 0.096 | 0.965 | 7.54 | 30.1 | 0.740 |
| 3 | DEAR | 0.362 | 0.591 | 0.119 | 0.951 | 5.35 | 7.3 | 0.644 |
| 4 | EfficientNet-B0 AdaFace | 0.339 | 0.575 | 0.126 | 0.947 | 4.05 | 8.1 | 0.605 |
| 5 | TinyViT-Ear | 0.282 | 0.506 | 0.149 | 0.928 | 5.08 | 20.3 | 0.405 |
| 6 | AurisPico | 0.248 | 0.502 | 0.145 | 0.930 | 8.82 | 17.6 | 0.320 |
| 7 | ConvNeXt ArcFace | 0.161 | 0.348 | 0.222 | 0.860 | 1.97 | 7.9 | 0.280 |
| 8 | MobileNetEar | 0.190 | 0.376 | 0.199 | 0.882 | 4.88 | 19.5 | 0.223 |
| 9 | Stripe | 0.213 | 0.382 | 0.234 | 0.849 | 29.6 | 118.4 | 0.039 |
| 10 | DINOv2-Ear | 0.173 | 0.435 | 0.166 | 0.913 | 22.3 | 89.0 | -0.006 |
| 11 | Dino-Ear-NITS | 0.157 | 0.351 | 0.215 | 0.863 | 87.3 | 349.3 | -0.215 |
| Ref. | Baseline (ConvNeXt-Base) | 0.262 | 0.463 | 0.157 | 0.921 | 87.6 | 350.3 | – |

beddings. The model is trained for identity classification using an ArcFace-style [3] angular margin loss, with AdamW, mixed precision, and a cosine learning-rate schedule. Model selection is based on full validation-pair evaluation using VER@0.1%, VER@1%, EER, and AUC. For Track 1, compact ConvNeXt-Zepto-style is used. For Track 2, horizontal-flip TTA is applied to ConvNeXt-Tiny variant by averaging embeddings from original and mirrored images.

**MobileNetEar.** The approach is designed as a lightweight ear verification model based on ImageNet-pretrained MobileNetV3-Large [2, 11]. The original classification head is replaced with an embedding head consisting of a linear projection, Hardswish activation, dropout, a bias-free projection layer, and batch normalization, producing a 512-dimensional embedding. The model is trained exclusively on the official UERC training data with ArcFace loss [3] (scale 30.0, margin 0.50) and AdamW optimization. Same model is used in both tracks.

**Stripe.** The approach uses a ConvNeXt-Tiny [20] backbone with a dual-branch pooling head that combines global average pooling and stripe pooling (dividing the feature map horizontally into three stripes), projecting to a 384-dimensional embedding. Ear images are pre-aligned using the offline normalization method from [9] and converted to grayscale. The training loss combines CosFace [30] and triplet loss with cosine-distance mining. A staged training strategy gradually shifts from semi-hard to hard mining and adjusts loss weights over 100 epochs. This approach participated in Track 1 only.

**DINOv2-Ear.** The approach adapts a self-supervised DINOv2 ViT-Small backbone [24] for ear recognition. The model processes $518\times518$ input images and produces 512-dimensional L2-normalized embeddings. Training combines ArcFace loss [3], triplet margin loss [27], and supervised contrastive loss [16], with PK sampling (16 identities, 2 images per identity). The DINOv2 backbone is frozen for the first epoch to stabilize the embedding head, after which the full network is fine-tuned end-to-end. This approach participated in Track 1 only.

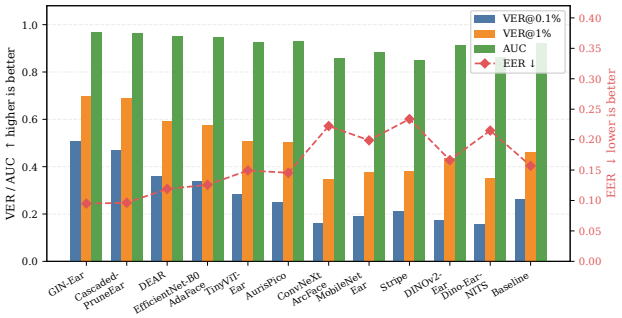


Figure 2. **Track 1 – verification performance metrics per submission.** VER@0.1%, VER@1%, and AUC are shown for each submission, sorted by composite score $S_1$. Lower EER (dashed, right axis) and higher rates indicate better performance. The baseline is shown for reference only and is not ranked.

**Dino-Ear-NITS.** The approach fuses DINOv2-B/14 [24] transformer features with lightweight handcrafted descriptors (RGB mean/variance, grayscale texture, Sobel edge responses, local contrast, Laplacian detail maps), projected and concatenated before a final fusion layer [3, 16]. Training combines an ArcFace-style additive angular margin loss and supervised contrastive learning ($\mathcal{L} = \mathcal{L}_{\text{ArcFace}} + 0.05\,\mathcal{L}_{\text{SupCon}}$). For Track 2, the DINOv2-S/14 backbone is used for improved runtime efficiency, while Track 1 uses the larger DINOv2-B/14 backbone.

## 5. Experiments and Results

This section presents the experimental results and comparative analysis of all submissions across both competition tracks, with a particular focus on the trade-offs between

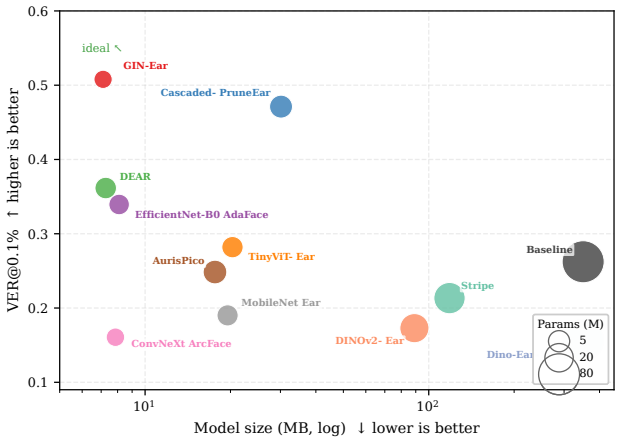


Figure 3. **Track 1 – trade-off between verification performance and model complexity.** Verification performance is measured with EER (lower is better) and VER@0.1% FAR (higher is better). Bubble size is proportional to the number of model parameters. An ideal submission is located in the upper-left with a small bubble. The baseline is shown for reference only and is not ranked.

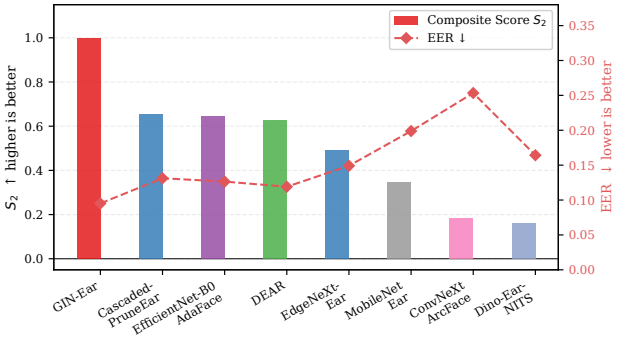


Figure 4. **Track 2 – composite scores $S_2$ and EER per submission.** Submissions are sorted by $S_2$ (higher is better).

recognition accuracy, model complexity, and practical deployment efficiency.

## 5.1. Track 1: Lightweight Model Design

**Overall Comparison.** Table 2 reports the full set of performance indicators for all Track 1 submissions, sorted by the composite score $S_1$ (Eq. (1)). Graphical results are presented in Figure 2. The results show considerable diversity in the efficiency–accuracy trade-off across the submitted approaches. We note that biometric metrics (VER, EER, AUC) are computed from a single evaluation pass over the fixed sequestered test set.

**Top Performers.** GIN-Ear achieved the top overall rank in Track 1 with a perfect composite score of 1.000. This result stems from GIN-Ear simultaneously achieving the highest VER@0.1% (50.8%), the lowest EER (9.5%), the best AUC (0.967), the fewest parameters (1.78 M), and one of the smallest model sizes (7.1 MB). The combination of identity-balanced augmentation and two-stage knowledge distillation into a MobileNetV4 student proved highly effective at compressing recognition capability from a large teacher into a very compact model.

Cascaded-PruneEar ranked second with a score of 0.740, achieving the second-highest VER@0.1% (47.1%) and second-lowest EER (9.6%) while maintaining a relatively small model size. DEAR (0.644) and EfficientNet-B0 AdaFace (0.605) rounded out the top four, both achieving strong recognition performance (EER of 0.119 and 0.126, respectively) while keeping model sizes below 10 MB.

**Efficiency–Accuracy Trade-off.** Figure 3 illustrates the trade-off between verification performance (EER, x-axis), VER@0.1% (y-axis), and model complexity (bubble size proportional to number of parameters). An ideal submission would be located in the upper-left region of the plot (low EER, high VER, small bubble). The figure highlights that GIN-Ear and DEAR achieve the best trade-off, while approaches relying on large pretrained vision transformers (Dino-Ear-NITS, DINOv2-Ear) achieve competitive recognition results but at a significantly higher parameter count, resulting in lower overall scores. Approaches in the lower-right quadrant, i.e., Stripe, ConvNeXt ArcFace, and MobileNetEar, demonstrate that small model size alone is insufficient if the verifiction performance is also low. Notably, despite having the smallest parameter count (1.97 M), ConvNeXt ArcFace ranked 7th in Track 1 due to lower verification accuracy compared to the top competitors.

**Track 1 Ranking.** The full ranking based on the composite score $S_1$ is provided in Table 2. A clear trend among the highest ranked submissions (ranks 1–4) is the combination of lightweight backbone architectures with discriminative embedding learning objectives such as ArcFace and AdaFace. In addition, the top performing approaches increasingly rely on advanced compression and transfer strategies, particularly knowledge distillation, to effectively transfer discriminative capability from larger pretrained models into substantially smaller and more deployment efficient architectures. This trend is especially evident in GIN-Ear and DEAR, which achieve strong verification performance while maintaining low parameter counts and compact model sizes.

## 5.2. Track 2: Edge Deployment

**Overall Comparison.** Table 3 reports the results for all Track 2 submissions, evaluated on the Raspberry Pi 4 platform. The composite score $S_2$ (Eq. (2)) jointly reflects recognition quality at FAR=0.1% and total inference time (Fig. 4). Final rankings were averaged over repeated evaluation runs; standard deviations of per-pair inference latency range from 0.0004 s (GIN-Ear, highly stable ONNX pipeline) to 0.071 s (Cascaded-PruneEar), indicating that timing measurements are consistent and rank-order conclusions are not affected by

Table 3. **Track 2 results – Edge Deployment on Raspberry Pi 4.** Results are sorted by the composite score $S_2$ (higher is better). VER@0.1% and VER@1% are reported together with EER, AUC, the total measured inference time (seconds) over the full test set, and its per-pair standard deviation $\sigma$ (ms).

| Rank | Approach | VER@0.1% ↑ | VER@1% ↑ | EER ↓ | AUC ↑ | Time (s) ↓ | $\sigma$ (ms) | Score ↑ |
|---|---|---|---|---|---|---|---|---|
| 1 | GIN-Ear | **0.512** | **0.700** | **0.095** | **0.967** | **267** | **0.4** | **1.000** |
| 2 | Cascaded-PruneEar | 0.377 | 0.615 | 0.131 | 0.934 | 3,151 | 70.8 | 0.656 |
| 3 | EfficientNet-B0 AdaFace | 0.334 | 0.566 | 0.126 | 0.946 | 710 | 3.0 | 0.644 |
| 4 | DEAR | 0.362 | 0.591 | 0.119 | 0.951 | 3,067 | 10.3 | 0.628 |
| 5 | EdgeNeXt-Ear | 0.271 | 0.482 | 0.149 | 0.928 | 1,754 | 4.9 | 0.490 |
| 6 | MobileNetEar | 0.190 | 0.376 | 0.199 | 0.882 | 1,340 | 33.8 | 0.346 |
| 7 | ConvNeXt ArcFace | 0.150 | 0.305 | 0.253 | 0.829 | 4,325 | 19.3 | 0.182 |
| 8 | Dino-Ear-NITS | 0.234 | 0.437 | 0.164 | 0.913 | 10,559 | 16.8 | 0.162 |

run-to-run variability.

**Top Performers.** GIN-Ear again achieved the top position in Track 2 with a score of 1.000, driven by an exceptional combination of high VER@0.1% (51.2%), high VER@1% (70.0%), low EER (9.5%), and by far the fastest inference time (267 seconds total over the 3500-pair evaluation). This substantial speed advantage – roughly 2.7× faster than the next-fastest submission (EfficientNet-B0 AdaFace, 710 seconds) – is primarily attributable to the ONNX Runtime-based deployment of the lightweight MobileNetV4 model, demonstrating the practical effectiveness of the two-stage distillation and ONNX export pipeline.

Cascaded-PruneEar (2nd, score 0.656) and EfficientNet-B0 AdaFace (3rd, score 0.644) achieved closely competitive scores. Despite Cascaded-PruneEar's significantly longer total inference time (3,151 seconds), its comparatively strong VER@0.1% (37.7%) and VER@1% (61.5%) kept it competitive. EfficientNet-B0 AdaFace benefited from FP16 weights and an ONNX Runtime-based inference path, achieving a favorable speed-accuracy trade-off. DEAR (4th, score 0.628) similarly demonstrated competitive recognition (VER@0.1% 36.2%, VER@1% 59.1%) but was penalized by a longer inference time due to the TinyViT architecture.

**Accuracy–Latency Trade-off.** Fig. 5 plots each submission in the space of total inference time (x-axis, log scale) versus VER@1% (y-axis). An ideal submission would be in the upper-left region. The figure illustrates a clear spread: GIN-Ear is isolated in the upper-left corner, representing the gold standard of both fast and accurate edge inference. EfficientNet-B0 AdaFace clusters nearby with a good speed-accuracy balance. Dino-Ear-NITS, despite a reasonable VER@1% (43.7%), suffers from the highest inference time (10,559 seconds) despite switching to the smaller DINOv2-S/14 backbone for Track 2, illustrating the cost of transformer-based hybrid inference on edge hardware.

**Comparison of Tracks 1 and 2.** A comparison of the Track 1 and Track 2 rankings reveals that the top performers are consistent across both tracks: GIN-Ear leads both, and approaches with efficient, compact models (EfficientNet-B0 AdaFace, DEAR, Cascaded-PruneEar) perform well in both settings. This consistency indicates that architectures optimized for lightweight design (via distillation, quantization, ONNX export) also translate effectively to the edge deployment scenario. In contrast, submissions relying on high-capacity pretrained transformers or high-resolution DINOv2 inference (e.g., DINOv2-B/14 and DINOv2-ViT-Small at 518×518) are penalized both for parameter count in Track 1 and for inference time in Track 2, illustrating the practical challenges of deploying such architectures on embedded hardware.

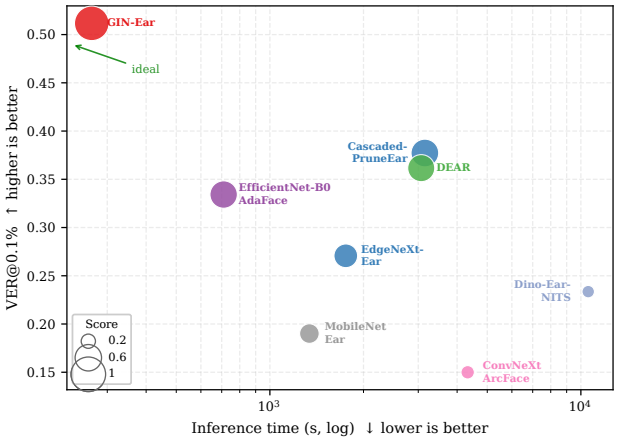


Figure 5. **Track 2 – accuracy versus inference time on Raspberry Pi 4.** Accuracy is measured with VER@1% (higher is better), while total inference time is shown on a logarithmic x-axis (lower is better). An ideal submission is in the upper-left corner. Bubble size is proportional to the composite score $S_2$.

## 6. Conclusion

This paper presented the results of the 2026 Unconstrained Ear Recognition Challenge (UERC 2026), the first competition in the UERC series to jointly evaluate recognition accuracy and computational efficiency. A total of 11 research groups participated in Track 1 (Lightweight Model Design) and 8 competed in Track 2 (Edge Deployment on

Raspberry Pi 4), submitting a diverse range of architectures spanning compact CNNs, lightweight vision transformers, and ensemble-distillation pipelines.

The competition results yield several important observations. First, knowledge distillation from large pretrained models into compact student architectures (e.g., MobileNetV4, TinyViT-5M) is highly effective: the top-ranked submission (GIN-Ear) achieved the lowest EER (9.5%) with only 1.78 M parameters, outperforming methods with 10–50× more parameters in terms of the composite efficiency-accuracy score. Second, model deployment choices matter significantly for edge performance: ONNX Runtime-based inference consistently enabled faster latency than standard PyTorch paths, and INT8 quantization enabled smaller model sizes with negligible accuracy loss. Third, there remains a clear accuracy gap between the top approaches and larger pretrained models (e.g., DINOv2-B/14), suggesting that further research into efficient transfer learning for ear recognition is warranted.

Compared to UERC 2023, where the best model achieved an EER of 0.146 on the same sequestered dataset and under the same protocol, but without any efficiency constraints [4], UERC 2026's top performer achieves an EER of 0.095, a notable improvement, while simultaneously satisfying strict efficiency requirements. This suggests that the field has matured considerably in its ability to build compact and highly accurate ear recognition systems. We anticipate that the benchmark datasets, evaluation toolkit, and competition outcomes will serve as a valuable resource for continued progress in efficient and deployable biometric recognition.

**Limitations and generalization.** Several limitations of the current benchmark should be acknowledged. First, the evaluation is conducted on a single in-the-wild dataset derived from UERC 2023, whose training portion consists largely of ear regions cropped and normalised from VGGFace images [33] rather than dedicated ear captures. While the sequestered test set was collected independently from the web and covers four ethnicity categories (Asian, Black, White, Other) with balanced gender representation, results may not transfer directly to acquisition conditions outside this web-image domain, and the extent to which the distribution of pose, lighting, and occlusion severity mirrors target deployment scenarios remains an open question. Results on this benchmark may therefore not transfer directly to specialized settings, such as surveillance video frames, near-infrared imagery, or ear data collected from specific demographic groups underrepresented in web-sourced collections. Second, the hardware platform for Track 2 (Raspberry Pi 4 Model B, CPU-only) represents one point in the space of edge devices; rankings could shift on platforms with different memory hierarchies, NPUs, or DSPs. Third, model compression strategies evaluated here (distillation, INT8 quantization, ONNX export) may interact differently with other backbone families not explored in this competition, and absolute performance levels are expected to increase as more capable lightweight architectures emerge.

**Fairness and failure analysis.** The sequestered test set covers four ethnicity categories (Asian, Black, White, Other) combined with two genders, with Asian, Black, and White groups containing 10 subjects per gender and the Other group containing 5 subjects per gender (70 subjects total), but a systematic per-group performance breakdown was not conducted in the current evaluation, as UERC 2026 prioritized efficiency benchmarking over demographic fairness analysis. This is a recognized limitation: differential accuracy across demographic groups is an important fairness consideration for biometric deployment, and the absence of such analysis means that potential performance disparities in UERC 2026 submissions remain unquantified. We plan to address this in future editions by integrating the demographic bias metrics established in UERC 2023 [4] with the efficiency-aware evaluation framework of UERC 2026, enabling a joint assessment of recognition accuracy, computational efficiency, and demographic fairness.

## References


[1] T. Arakawa, T. Ohki, T. Maeda, et al. Development of a pseudo-rotated ear image generation method using planar approximation. In *Proc. IPSJ Annual Conference*, 2023.

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[3] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[4] Ž. Emeršič, T. Ohki, M. Akasaka, T. Arakawa, S. Maeda, M. Okano, Y. Sato, A. George, S. Marcel, I. I. Ganapathi, S. S. Ali, S. Javed, N. Werghi, S. G. Işık, E. Sarıtaş, H. K. Ekenel, V. Hudovernik, J. N. Kolf, F. Boutros, N. Damer, G. Sharma, A. Kamboj, A. Nigam, D. K. Jain, G. Cámara-Chávez, P. Peer, and V. Štruc. The unconstrained ear recognition challenge 2023: Maximizing performance and minimizing bias. In *Proc. IEEE International Joint Conference on Biometrics (IJCB)*, 2023.

[5] Ž. Emeršič, A. K. S. V., B. S. Harish, W. Gutfeter, J. N. Khiarak, A. Pacut, E. Hansley, M. P. Segundo, S. Sarkar, H. Park, G. P. Nam, I.-J. Kim, S. G. Sangodkar, Kaçar, M. Kirci, L. Yuan, J. Yuan, H. Zhao, F. Lu, J. Mao, X. Zhang, D. Yaman, F. I. Eyiokur, K. B. Özler, H. K. Ekenel, D. P. Chowdhury, S. Bakshi, P. K. Sa, B. Majhi, P. Peer, and V. Štruc. The unconstrained ear recognition challenge 2019. In *Proc. IAPR International Conference on Biometrics (ICB)*, 2019.

[6] Ž. Emeršič, D. Štepec, V. Štruc, P. Peer, A. George, A. Ahmad, E. Omar, T. E. Boult, R. Safdari, Y. Zhou, S. Zafeiriou, D. Yaman, F. I. Eyiokur, and H. K. Ekenel. The unconstrained

ear recognition challenge. In *Proc. IEEE International Joint Conference on Biometrics (IJCB)*, pages 715–724, 2017.

[7] Ž. Emeršič, V. Štruc, and P. Peer. Ear recognition: More than a survey. *Neurocomputing*, 255:26–45, 2017.

[8] S. Han, H. Mao, and W. J. Dally. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. In *Proc. International Conference on Learning Representations (ICLR)*, 2016.

[9] E. A. Hansley, M. P. Segundo, and S. Sarkar. Employing fusion of learned and handcrafted features for unconstrained ear recognition. In *Proc. IET Biometrics*, 2018.

[10] G. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. In *Proc. NIPS Workshop on Deep Learning and Representation Learning*, 2015.

[11] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam. Searching for MobileNetV3. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

[12] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[13] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and less than 0.5MB model size. 2016.

[14] ISO/IEC. ISO/IEC 19795-1: Biometric performance testing and reporting. *International Organization for Standardization*, 2021.

[15] B. Jacob, S. Kligys, B. Chen, M. Zhu, M. Tang, A. Howard, H. Adam, and D. Kalenichenko. Quantization and training of neural networks for efficient integer-arithmetic-only inference. 2018.

[16] P. Khosla, P. Tian, C. Wang, A. Ramanujan, Y. Tian, T. Collins, A. Norouzi, P. Isola, A. Krishnamurthy, A. Efros, and T. Salimans. Supervised contrastive learning. In *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[17] M. Kim, A. K. Jain, and X. Liu. AdaFace: Quality adaptive margin for face recognition. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[18] R. Lajić, P. Peer, V. Štruc, D. S. Han, B. Meden, and Ž. Emeršič. Faces: Facial analysis with compressed efficient systems. *ICT express*, 2026.

[19] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Wei, Z. Guo, X. Zhang, L. Yang, K. Guo, M. Zhang, Z. Wei, and B. Guo. Swin Transformer V2: Scaling up capacity and resolution. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[20] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie. A ConvNet for the 2020s. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[21] M. Maaz, A. Shaker, H. Cholakkal, S. Khan, S. W. Zamir, R. M. Anwer, and F. S. Khan. EdgeNeXt: Efficiently amalgamated CNN-Transformer architecture for mobile vision applications. In *Proc. European Conference on Computer Vision (ECCV) Workshops*, 2022.

[22] S. Maeda, T. Ohki, et al. Occlusion synthesis for ear recognition augmentation. In *Proc. IPSJ Annual Conference*, 2022.

[23] Microsoft. ONNX Runtime: Cross-platform, high performance ML inferencing and training accelerator. https://onnxruntime.ai, 2018.

[24] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Bhatt, L. Szlam, P. Labatut, A. Joulin, J. Mairal, G. Synnaeve, H. Jégou, and P. Bojanowski. DINOv2: Learning robust visual features without supervision. 2023.

[25] W. Park, D. Kim, Y. Lu, and M. Cho. Relational knowledge distillation. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[26] D. Qin, C. Leichner, M. Delakis, M. Fornoni, S. Luo, F. Yang, W. Wang, C. Banbury, C. Ye, B. Akin, O. Danaci, B. Karaman, V. Vasudevan, B. Chen, and M. Sandler. MobileNetV4 – universal models for the mobile ecosystem. *arXiv preprint arXiv:2404.10518*, 2024.

[27] F. Schroff, D. Kalenichenko, and J. Philbin. FaceNet: A unified embedding for face recognition and clustering. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[28] O. Siméoni, E. Bautista, J. Revaud, and P. Weinzaepfel. DINOv3: Revisiting self-supervised pretraining with curriculum. *arXiv preprint arXiv:2501.09678*, 2025.

[29] M. Tan and Q. V. Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proc. International Conference on Machine Learning (ICML)*, 2019.

[30] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu. CosFace: Large margin cosine loss for deep face recognition. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[31] R. Wightman. PyTorch image models. https://github.com/huggingface/pytorch-image-models, 2019.

[32] K. Wu, J. Zhang, H. Peng, M. Liu, B. Xiao, J. Fu, and Z. Yuan. TinyViT: Fast pretraining distillation for small vision transformers. In *Proc. European Conference on Computer Vision (ECCV)*, 2022.

[33] X. Zhou, E. Sariyanidi, and H. K. Ekenel. VGGFace-ear: Ear recognition using VGGFace. *IET Biometrics*, 2021.