Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identity Verification from Human Scent using Learned Representations of Two-Dimensional Gas Chromatography–Mass Spectrometry Data

Jan Hlavsa

hlavsja3@fel.cvut.cz

Czech Technical University in Prague

Radim Spetlik

Czech Technical University in Prague

Jana Čechová

University of Chemistry and Technology, Prague

Petra Pojmanová

University of Chemistry and Technology, Prague

Jiří Matas

Czech Technical University in Prague

Štěpán Urban

University of Chemistry and Technology, Prague

Research Article

Keywords: biometric identity verification, human scent, gas chromatography–mass spectrometry, metric learning, channel representation, cross-session generalisation

Posted Date: October 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10796400/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Identity Verification from Human Scent using Learned Representations of Two-Dimensional Gas Chromatography–Mass Spectrometry Data

Jan Hlavsa[1*], Radim Spetlik[1], Jana Čechová[2], Petra Pojmanová[2], Jiří Matas[1], Štěpán Urban[2]

[1]Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic.
[2]Department of Analytical Chemistry, University of Chemistry and Technology, Prague, Prague, Czech Republic.

*Corresponding author(s). E-mail(s): hlavsja3@fel.cvut.cz;

Abstract

This study examines the feasibility of employing raw two-dimensional gas chromatography/time-of-flight mass spectrometry (GC×GC–ToF-MS) data for the purpose of human scent identity verification. Unlike expert-driven techniques relying on compound identification, our framework transforms each GC×GC–ToF-MS sample into a multi-channel image. A comprehensive assessment is conducted on ten feature-embedding methods and five spatial-alignment strategies. The evaluation is performed on a newly assembled dataset of 252 individuals, comprising 2,528 raw samples and aggregating around 7.5 TB of data. Our original image-based method achieves **89.5%** area under the receiver operating characteristic curve (AUROC) and **53.3%** true-positive rate at 5% false-positive rate (TPR@5%FPR) without spatial registration. Extending the conference version of this work, we further analyse *in-session* versus *out-of-session* generalisation on the full dataset. We find that verification remains feasible for measurements from sessions absent from training, while AUROC drops by roughly **9** points when enrolment and query differ in session familiarity, consistently across both instruments. Because many out–out pairs share the same unseen session, the experiment indicates sensitivity to session familiarity rather than fully isolating cross-session generalisation. We further introduce a multi-view representation which, when trained with a cosine-softmax metric-learning objective, improves the performance to **92.8%** AUROC and **62.2%** TPR@5%FPR on the full **252**-identity dataset. Although the overall performance is below that of well-established biometrics (e.g., iris verification), the results demonstrate the feasibility of raw-odor verification for scenarios where direct line-of-sight or cooperation may be limited. Code and dataset access instructions are publicly available on github.com.

Keywords: biometric identity verification, human scent, gas chromatography–mass spectrometry, metric learning, channel representation, cross-session generalisation

1 Introduction

Identity verification is a fundamental problem in biometrics and a canonical computer vision decision problem, where the goal is to determine whether two samples originate from the same individual. Established biometric modalities, such as face, iris, and fingerprint recognition, provide robust performance in controlled scenarios. However, they also typically require the subject's active participation or direct line of sight and may not always be suitable for applications where direct physical or visual contact is not feasible. Consequently, the search for alternative biometrics that can operate passively—and without explicit subject cooperation—is a compelling research direction.

Body odor analysis has long been discussed as a biometric – forensic observations suggest that individuals emit distinctive scent signatures that remain relatively stable over time [21, 23] and can withstand environmental noise from various sources [16, 22, 24, 33], motivating previous attempts at odor-based person recognition using analytical chemistry tools [16, 21] and electronic noses [13, 28]. Existing studies often relied on small, private datasets and expert-driven compound identification [25, 26], limiting scalability and reproducibility. In this work, we explore whether modern representation learning can leverage raw outputs from two-dimensional gas chromatography with time-of-flight mass spectrometry (GC×GC–ToF-MS) to enable large-scale identity verification with minimal expert input.

GC×GC–ToF-MS separates a gas mixture with two capillary columns in series. Each column is a long, narrow tube whose inner wall is coated with a chemical layer called *stationary-phase*. The first column establishes a slow first-dimension retention-time axis t_1. A periodic modulator (period T_{mod}) slices the t_1 output and injects it into a faster second column, producing a fast axis $t_2 \in [0, T_{\text{mod}})$. The stationary phase chemistry of each column defines its separation pattern; using different coatings across the two columns yields different compound orderings and complementary—often orthogonal—separations. The resulting 2-D chromatogram (t_1, t_2) exhibits peaks that are narrow in t_1 and broad in t_2 (Fig. 1). At each (t_1, t_2),

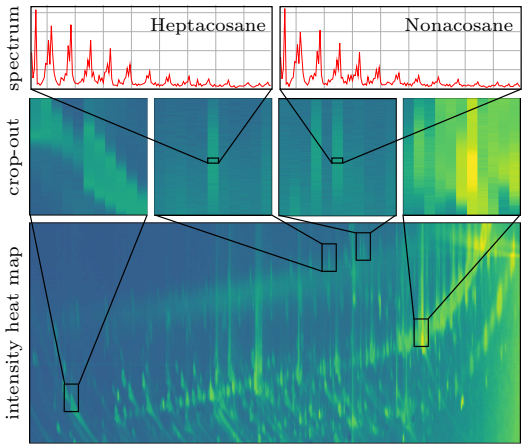


Fig. 1 Intensity heat map of the detector response in a 2D Gas Chromatography–Mass Spectrometry measurement, plotted over retention time with t_1 on the x-axis and t_2 on the y-axis (bottom). Mass spectra of Heptacosane and Nonacosane compounds (top). Notably, Heptacosane and Nonacosane spectra are almost identical, with spatial position (retention time) being the primary distinguishing factor. All visualizations are in arbitrary units. Reproduced from the conference version [32]

an electron-impact time-of-flight mass spectrometer (ToF-MS) records an intensity spectrum over mass-to-charge (m/z). The measurement is thus a set of spectra indexed by (t_1, t_2).

While GC×GC–ToF-MS measurements can capture highly detailed chemical profiles, they historically face two practical barriers: (i) long measurement times (often exceeding 2.5 hours per sample), and (ii) specific sampling procedures (e.g., requiring a subject to touch collecting beads). A recent non-contact sampling device [14] offers a promising direction toward circumventing these limitations: by collecting airborne volatiles passively at a distance, one could potentially avoid visual or physical contact. Such capabilities distinguish scent-based approaches from other biometrics and could eventually enable remote identification where other modalities are not available.

In this work, we represent GC×GC–ToF-MS data as multi-channel images and apply modern feature embedding techniques to the problem of identity verification. In our newly curated dataset of 252 individuals, the best-performing approach yields $\approx 53\%$ true positive rate at a 5% false positive rate. Although this performance is below that of established modalities (e.g., iris recognition), it demonstrates the feasibility

of odor-based verification from raw measurements while revealing considerable room for improvement. Our aim is to introduce this problem to the community and encourage further research by publishing our complete dataset (≈ 7.5 TB) and all supporting code.

A preliminary version of this work, entitled *Identity Verification from Human Scent using Channel Representation of 2D Gas Chromatography–Mass Spectrometry Data*, was presented at the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) 2026 [32]. This article substantially extends the conference paper in two key ways.

First, we add a dedicated study of *session generalisation*. A scent measurement is acquired in a session (a measurement campaign on a given date); instrument state, column condition, and ambient factors drift between sessions. A deployed verifier must therefore enrol and query individuals whose acquisition session may or may not be represented in the data used to build the model. We make this dependence explicit: for each evaluated pair we record whether the enrolled and the queried measurement come from sessions seen during training, yielding four operating regimes (in–in, in–out, out–in, out–out). Retraining the best benchmark on the full dataset and evaluating it across these regimes quantifies its dependence on session familiarity. The results show that measurements from sessions absent from training retain verification signal and that the cross-familiarity penalty persists on both the higher- and lower-resolution instruments. Second, we strengthen the verifier itself. We introduce a multi-view representation that preserves complementary mass-spectral and retention-time information and also investigate training objectives beyond the triplet loss used in the conference version. Specifically, we evaluated cosine-softmax metric learning and direct pairwise binary classification. We further perform a pair-level comparison of the original channel-based and proposed multi-view models, analyzing where their verification decisions agree and where they diverge. The best configuration, combining multi-view representation with cosine-softmax objective, reaches 92.8% AUROC and 62.2% TPR@5%FPR on the full 252-identity dataset. Controlled ablations show that the multi-view representation consistently improves global ranking, while the gain at the strict operating point emerges specifically when this representation is trained with the cosine-softmax objective.

In summary, our contributions are as follows. We

(i) propose a new identity verification paradigm that treats raw GC×GC–ToF-MS outputs as multi-channel images, eliminating the need for expert-driven compound identification;
(ii) introduce a newly assembled, large-scale, open dataset of human-scent GC×GC–ToF-MS measurements (≈ 7.5 TB), covering 252 unique identities;
(iii) systematically evaluate ten feature-embedding methods and five data-alignment strategies;
(iv) demonstrate a prototype identity verification system that achieves a TPR of ≈ 53% at a 5% FPR, highlighting the feasibility of the approach and the substantial opportunities for further improvement;
(v) *(new in this extension)* retrain and dissect the best model into four in/out-session regimes on the full 252-identity dataset, measuring the effect of session familiarity—and of detector resolution—on verification accuracy; and
(vi) *(new in this extension)* improve the best verification result to 92.8% AUROC / 62.2% TPR@5%FPR on the full 252-identity dataset with a multi-view representation and a cosine-softmax metric-learning recipe.

2 Related Work

Human body odor has emerged as a distinctive biometric identifier, rooted in the notion that the scent of each individual is unique and relatively stable over time. This idea is supported by long-standing forensic practice: trained dogs can reliably track or identify people by smell, even identical twins who share the same DNA [23]. Scientific analyses have confirmed that individuals emit distinct profiles of volatile organic compounds (VOCs) through sweat and other secretions. For example, Penn et al. conducted a large-scale gas chromatography-mass spectrometry (GC-MS) study of axillary sweat, finding "individually distinct and reproducible GC-MS

fingerprints" for each of 197 subjects [21]. In particular, they also observed consistent differences between male and female odor profiles, identifying numerous sex-specific compounds in sweat. This evidence suggests that body odor can serve as a unique "chemical fingerprint" for identity, analogous to a visual fingerprint.

Based on these findings, researchers have developed biometric systems to automate identity verification via scent. Early work applied pattern recognition to personal odor data, demonstrating that people could be identified by their odor with precision of 85% [28]. In that study, an electronic nose (an array of chemical sensors) captured each person's scent repeatedly (13 subjects over 28 sessions) while learning the odor patterns. The results showed that the odor profile of each subject remained consistent enough for identification, with only an error rate of 15%. This proof-of-concept established odor biometrics as feasible and highlighted its practical appeal: odor-based screening could happen passively and non-invasively. Unlike fingerprint or iris scans that require active cooperation, odor sensors can unobtrusively sample human scent as one walks by. In fact, body odor biometrics have been proposed as a "less aggressive" identification method for settings such as airport security, since the smell of a person could be verified in real time without physical contact or stigmatizing procedures.

2.1 Advancements in Scent Detection Technologies

Recent studies have refined both the sampling sites for human scent and the sensing technologies. Determining which part of the body provides the most reliable odor signal remains a critical area of research: breath has high humidity that can interfere with sensors, and armpit odor is often masked by deodorants [2]. In the same study, it is hypothesized that the ear canal could be an effective, relatively uncontaminated source of human odor. Using a portable electronic nose, they collected ear volatiles from volunteers and attempted identification. Remarkably, their system could identify a person by ear odor with $\sim$ 87% accuracy (10 out of 12 correct) when comparing scent samples taken on the same day. (This greatly exceeds random-chance accuracy of 8.3% for 12 individuals.) However, the study also observed a drop in performance for samples collected on different days—accuracy fell to $\sim$ 22% for day-to-day matching—suggesting that while the odor is unique, its temporal stability and sensor calibration remain challenges for long-term biometric applications. Beyond skin emanations, other work has investigated human breath as an identifier. For example, a study combined a 16-channel chemiresistive sensor array with machine learning to analyze breath samples. With sufficient sensors and data, the authors achieved over 97% identification accuracy on a set of 20 individuals using only their breath odor signatures [13]. These results underscore that, given a robust "digital nose" and algorithms, chemical biometrics can approach the reliability of traditional modalities. In summary, a range of odor-capturing approaches—from GC-MS laboratory analysis to hand-held electronic noses—have proven capable of distinguishing individuals by scent. However, all these findings are limited by small datasets consisting only from dozens of samples. In [16], human scent residues from fired cartridge cases were reliably characterized using GC×GC–ToF-MS, with chemical profiles remaining stable even after the extreme conditions of a gunshot. In particular, the experiments were carried out on a relatively small dataset involving 20 volunteers, producing statistically robust results (e.g., a Random Forest classifier with an F1 score of 1 and 100% success in Spearman's correlation similarity comparisons). This not only underscores the forensic potential of the approach but also supports the underlying stability of the human odor signature as a reliable identifier.

2.2 Scent-Based Sex Classification

In addition to identifying who a person is, biological sex can often be inferred from scent alone—a subtopic of great interest in biometrics and forensic science. Early investigations mainly used one-dimensional gas chromatography mass spectrometry (1D GC-MS), using targeted approaches to pinpoint predefined marker compounds indicative of sex, ethnicity, or disease states [3, 5, 10, 15, 21, 27, 30]. The reported accuracies ranged from around 70% to 96.67%, but these findings were based on private datasets and were therefore difficult to replicate [5,

10, 21]. More recently, two-dimensional gas chromatography-time-of-flight mass spectrometry (GC×GC–ToF-MS) has been explored, promising superior compound separation albeit at the cost of more complex data handling. In the first such study, 504 samples collected from 20 women and 20 men were analyzed with a targeted approach, yielding up to 90% accuracy using an SVM classifier [25, 26]. However, the inclusion of samples from the same individuals in both training and test sets limits the generalizability of these results to previously unseen subjects.

2.3 Image-Based Representations

Recent work shows that treating GC×GC–ToF-MS outputs as 2D images rather than isolated peaks can substantially improve performance [1, 9]. These image-based approaches capture subtle chemical differences without manual compound identification, achieving near 100% sensitivity and high specificity in detecting adulterated vs. authentic samples. They also hold promise for identity verification, as the full 2D fingerprint can reveal robust class signatures. Further reviews confirm the suitability of machine learning algorithms for these high-dimensional datasets, with neural networks effectively "smelling" raw chromatograms [4].

2.4 Discussion

Despite these promising results, previous research suffers from several constraints that limit broader adoption and reproducibility. First, almost all studies relied on private or relatively small datasets (often fewer than a few dozen subjects), complicating efforts to validate and compare different methods in larger or more diverse populations [16, 21, 25, 26, 28]. Second, the vast majority of chemical analyses followed a pipeline intensive with human labor, which involved the manual identification and quantification of compounds in 1D or 2D GC-MS experiments [3, 5, 6, 15, 27, 30]. Such manual workflows are prone to operator bias, require specialized chemistry expertise, and often employ targeted search for predefined compounds rather than unbiased, high-throughput discovery. Finally, because of their proprietary nature and specialized hardware, published studies typically do not release raw chemical data in a way that allows the broader scientific community to replicate or extend the findings. A related but distinct concern, which we address quantitatively in Sec. 5.1, is whether verification performance depends on acquisition-session familiarity; we therefore stratify evaluation pairs according to whether their sessions are represented in training.

In contrast, we advocate for a fully automated, end-to-end methodology for large-scale odor-based recognition. Rather than relying on manual compound extraction or proprietary pipelines, we employ neural networks to learn robust features, enabling efficient and scalable inference with minimal human intervention. Following a recently published computer vision approach to sex classification [12], we make our dataset publicly available.

3 Methods

To assess GC×GC–ToF-MS measurements for identity verification, we use a standard verification pipeline. For each input sample, a channel-encoded GC×GC–ToF-MS scan, we compute its embedding. We then compare this embedding to all other embeddings in the dataset. Ideally, embeddings from the same identity form tight clusters, while embeddings from different identities stay well separated.

We extract embeddings from GC×GC–ToF-MS measurements with four approaches: (i) a published baseline method [6] developed by analytical chemists, (ii) hand-engineered descriptors HOG [7], LBP [19], and SIFT [18], operating on a single-channel image representation, (iii) five off-the-shelf convolutional neural network (CNN) backbones followed by a learned projection layer, and (iv) a custom lightweight CNN that we train from scratch. Following the sex-classification pipeline introduced in [12], we evaluate five spatial-alignment strategies: the four schemes proposed in the original work plus an unaligned variant.

Let $\Omega \coloneqq \{1, \dots, H\} \times \{1, \dots, W\}$ be the pixel grid and $\mathbf{u} = (u_1, u_2) \in \Omega$. A GC×GC–ToF-MS measurement is $\mathcal{S} \coloneqq \{s(\mathbf{u})\}_{\mathbf{u}\in\Omega}$, where the local spectrum at $\mathbf{u}$ is $s(\mathbf{u}) = \{(m_{n,\mathbf{u}}, I_{n,\mathbf{u}})\}_{n=1}^{N(\mathbf{u})}$, with $m_{n,\mathbf{u}}$ the mass-to-charge ratio (m/z) and $I_{n,\mathbf{u}}$ its intensity. Our Human Scent Dataset includes higher-resolution (HR) and lower-resolution (LR)

scans. High-resolution produces an irregular, per-pixel set with variable length $N(\mathbf{u})$. Low-resolution uses a fixed m/z grid $\{m_n\}_{n=1}^{M}$ ($M \approx 800$), yielding $s(\mathbf{u}) = \{(m_n, I_{n,\mathbf{u}})\}_{n=1}^{M}$ with m_n independent of $\mathbf{u}$.

3.1 Canonical Frame Registration

Instrument factors, such as temperature changes or hardware drift, can shift retention times across a GC×GC–ToF-MS measurement, leading to misaligned compounds and unreliable comparisons [12]. To assess alignment explicitly, we evaluate both aligned and unaligned regimes. For the aligned baseline, we reimplement the canonical-frame registration exactly as in [12], aligning all samples so that the same compounds appear at the same retention times across measurements. For the unaligned setting, we process the raw data directly without any retention-time correction. The registration pipeline is fully automated and requires no manual tuning.

Let $\mathcal{R}^{\text{ref}} = \{c_1, c_2, \ldots\}$ denote the set of reference compounds. For each $c \in \mathcal{R}^{\text{ref}}$, let $\mathbf{u}_c \in \mathbb{R}^2$ be the reference location and s_c^{ref} the corresponding mass spectrum. Following [12], we evaluate two subsets reported to occur in all samples of that dataset: $\mathcal{R}^{(22)}$ and $\mathcal{R}^{(24)}$ (22- and 24-compound sets). The registration procedure has four steps and is performed for each measurement separately: (1) **Peak detection**: the peak location $\hat{\mathbf{u}}_c^{(\mathcal{S})}$ of every compound c is estimated using one of the compound peak detectors from [12]; (2) **Delaunay triangulation** over the estimated locations $\{\hat{\mathbf{u}}_c^{(\mathcal{S})}\}$; (3) **Coordinate transformation**: new coordinates for enclosed points are obtained via barycentric interpolation, mapping $\hat{\mathbf{u}}_c^{(\mathcal{S})}$ to the reference location $\mathbf{u}_c$; and (4) **Piecewise-linear interpolation** of the original measurement onto the transformed mesh to produce a canonical grid.

3.2 Channel Representation of Time-of-Flight Mass Spectrometry Measurements

We compress the raw mass spectra $s(\mathbf{u})$ with a bank of C *overlapping, learnable kernels*

$$\phi_k(m_{n,\mathbf{u}}) = \alpha_k \kappa(m_{n,\mathbf{u}}; \theta_k), \quad k = 1, \ldots, C, \tag{1}$$

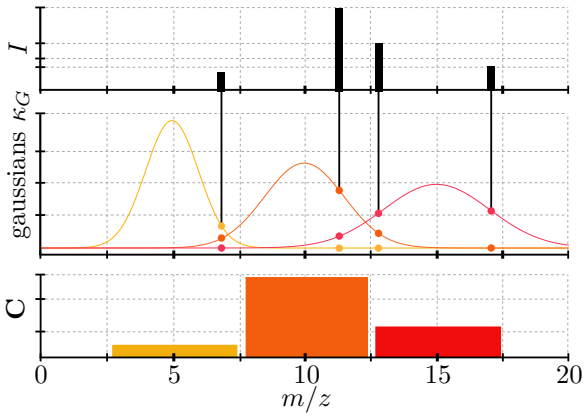


Fig. 2 Channel representation of a toy time-of-flight mass spectrometry measurement for $C = 3$ (see Sec. 3.2). The high-resolution spectrum, comprising m/z values and intensities I, is transformed into a compact channel representation $\mathbf{C}$ by applying three basis functions (here Gaussians) with learnable parameters θ and scale factors α. All values are in arbitrary units. Reproduced from the conference version [32]

where $\alpha_k > 0$ is a learnable amplitude and $\kappa : \mathbb{R} \to \mathbb{R}$ is a parametric basis function. We experiment with the Gaussian kernel

$$\kappa_G(m_{n,\mathbf{u}}; \theta_k) = \frac{\exp\left(-\frac{(m_{n,\mathbf{u}} - \mu_k)^2}{2\sigma_k^2}\right)}{\sqrt{2\pi}\sigma_k}, \tag{2}$$

where $\theta_k := (\mu_k, \sigma_k)$, with learnable width $\sigma_k > 0$ (Fig. 2). We also experimented with triangular kernels κ_Δ, which performed worse than Gaussians.

Input Measurement Encoding.

We convert $\mathcal{S}$ to a *channel-encoded image* $\mathbf{C} \in \mathbb{R}^{C \times H \times W}$ by applying the trained kernel bank to each spectrum: $\mathbf{C}_{k,\mathbf{u}} = \sum_{n=1}^{N(\mathbf{u})} \phi_k(m_{n,\mathbf{u}})\, I_{n,\mathbf{u}}$.

3.3 Feature Extraction Approaches

3.3.1 Compound-Centric Baseline

We reimplement the compound-centric baseline [6]. Let $\mathcal{R}^{\text{q}} = \{c_1, \ldots, c_Q\}$ be the set of compounds with estimated quantities, $Q := |\mathcal{R}^q|$. A sample $\mathcal{S}$ is represented by $\mathbf{x} \in \mathbb{R}^Q$, where the k-th element of $\mathbf{x}$ is the measured quantity of c_k; we apply ℓ_1 normalization $\mathbf{x} \leftarrow \mathbf{x}/\|\mathbf{x}\|_1$. For feature selection, for each compound c we compare absolute quantity differences in within-identity pairs to those in across-identity pairs using a nonparametric two-sample rank test, yielding a p-value

p_c, and retain c if $-\log_{10}(p_c) > \tau_{\text{feat}}$. For samples i and j we define a dissimilarity d_{ij} using one of (i) $\|\mathbf{x}-\mathbf{y}\|_2$, (ii) $1-\rho_{\text{Pearson}}$, or (iii) $1-\rho_{\text{Spearman}}$; we fit separate univariate Gaussian mixture models to the positive (same-identity) and negative (different-identity) distance distributions, obtaining $p_{\text{pos}}(d)$ and $p_{\text{neg}}(d)$, and declare a match if the likelihood ratio $\Lambda(d_{ij}) = p_{\text{pos}}(d_{ij})/p_{\text{neg}}(d_{ij}) > \tau_\Lambda$.

3.3.2 Hand-engineered Descriptors

We form a single-channel image by summing all m/z intensities at each spatial location. From this image, we extract three classic grayscale descriptors: (i) local binary patterns (LBP) [19], (ii) histogram of oriented gradients (HOG) [7], and (iii) scale-invariant feature transform (SIFT) [18]. Each descriptor yields a feature vector, which is ℓ_2-normalized and used for verification.

3.3.3 Convolutional Neural Networks

We implement the embedding function f using six CNNs—five pretrained and one custom-designed network. We use five pretrained backbones: *ResNet-18* [11], *ResNet-101* [11], *VGG-16* [31], *Inception-v3* [34], and *GoogLeNet* [34], followed by a lightweight projection layer. The custom network comprises three modules: (i) a convolutional backbone, (ii) a downsampling stack, and (iii) a residual multi-layer perceptron head. We choose an anisotropic 2×8 kernel in (i), reflecting the structure of the input data (497×2000) and the fact that peaks in GC×GC–ToF-MS typically appear as elongated blobs stretched along the second (retention-time) dimension (Fig. 1), so a wider kernel in this direction provides a better inductive bias for capturing peak shapes.

3.4 Identity Verification

Let $f : \mathbb{R}^{C\times H\times W} \to \mathbb{R}^D$ be an embedding function and $d : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}_{\geq 0}$ a dissimilarity (e.g., Euclidean or the cosine distance $d_{\cos}(\mathbf{x},\mathbf{y}) := 1 - \frac{\mathbf{x}^\top \mathbf{y}}{\|\mathbf{x}\|_2\,\|\mathbf{y}\|_2}$). For a pair $(\mathbf{C}_i, \mathbf{C}_j)$, $d_{ij} := d(f(\mathbf{C}_i), f(\mathbf{C}_j))$; at test time we declare "same identity" if $d_{ij} < \tau$.

Metric Learning with Triplet Loss.

We adopt the metric-learning setting with the triplet loss [29]. Given an anchor a, a matched positive p from the same identity, and a negative n from a different identity, we minimize the margin-based triplet loss

$$\mathcal{L} = \frac{1}{|\mathcal{T}|} \sum_{(a,p,n)\in\mathcal{T}} \max\{d_{ap} - d_{an} + \gamma, 0\}, \quad (3)$$

where $\gamma > 0$ is a predefined margin ($\gamma = 1$ in our experiments). The hinge term vanishes when $d_{ap} + \gamma \leq d_{an}$, encouraging tight same-identity clusters and well-separated different-identity clusters.

The single-image channel representation (Sec. 3.2) and the triplet objective are the two ingredients of the benchmark embedding, and each can be strengthened; the *representation* governs how well identities can be ranked (AUROC) and the *objective* how well they separate at a strict false-positive rate (TPR@5%FPR). We revisit both below.

Multi-View Representation.

A GC×GC–ToF-MS measurement is a long sequence of mass spectra: scan u records points with masses $m_{n,u}$ and intensities $I_{n,u}$, which we bin onto a fixed m/z grid, $b_j(u) = \sum_{n:\, m_{n,u}\in\beta_j} I_{n,u}$, where β_j is the j-th m/z bin. Two-dimensional gas chromatography folds the scan index into a first- and a second-dimension retention coordinate through the modulation period P (estimated per file from the total-ion signal): $r_1(u) = \lfloor u/P \rfloor$ and $r_2(u) = u \bmod P$. The channel representation of Sec. 3.2 collapses this structure onto a few Gaussian m/z channels and treats the result as one image; this is compact but discards most of the fine m/z detail. We instead form three *m/z-resolved* views of the same measurement, each retaining the full binned m/z axis as feature channels and differing only in how the retention structure is exposed,

$$\begin{aligned} \mathbf{X}^{\text{ms}} &\in \mathbb{R}^{S\times M_s}, \\ \mathbf{X}^{\text{rt}} &\in \mathbb{R}^{R\times M}, \\ \mathbf{X}^{\text{cube}} &\in \mathbb{R}^{R_1\times R_2\times M}. \end{aligned} \quad (4)$$

where: (i) $\mathbf{X}^{\text{ms}}$ is the *mass-spectral scan sequence* — the scans are grouped into S=256 contiguous blocks and each row holds that block's mean binned spectrum on an M_s=1174-bin (0.5-Da) grid, capturing which ions are present and how

they evolve along elution; (ii) $\mathbf{X}^{\mathrm{rt}}$ is the *first-dimension retention* field, intensity on R=512 first-dimension retention steps $\times$ M=587 m/z bins (1-Da, aggregated over the second dimension); and (iii) $\mathbf{X}^{\mathrm{cube}}$ is the *joint two-dimensional retention* field, the full R_1=64 (first-dimension) $\times\, R_2$=32 (second-dimension) retention plane $\times$ M=587 m/z bins. The three thus expose the same m/z-resolved intensity at increasing retention detail — folded into a scan sequence, as the first-dimension profile, and as the complete GC×GC plane. Each view $v \in \{\mathrm{ms}, \mathrm{rt}, \mathrm{cube}\}$ is encoded by its own convolutional branch g_v: a one-dimensional branch convolving *along the scan* (resp. first-dimension-retention) axis for $\mathbf{X}^{\mathrm{ms}}$ and $\mathbf{X}^{\mathrm{rt}}$, with the M_s (resp. M) m/z bins as input channels; and a two-dimensional branch convolving *over the retention plane* for $\mathbf{X}^{\mathrm{cube}}$, with the M m/z bins as channels. Each branch g_v ends in global pooling followed by a linear projection, mapping its view to a fixed-width descriptor $\mathbf{p}_v = g_v(\mathbf{X}^v) \in \mathbb{R}^D$ regardless of the view's length. Writing the three views jointly as $\mathbf{X} = (\mathbf{X}^{\mathrm{ms}}, \mathbf{X}^{\mathrm{rt}}, \mathbf{X}^{\mathrm{cube}})$, a learned fusion h combines the descriptors into the single D=256-dimensional embedding

$$\mathbf{z} = f_{\mathrm{mv}}(\mathbf{X}) = h\big(\mathbf{p}_{\mathrm{ms}} \oplus \mathbf{p}_{\mathrm{rt}} \oplus \mathbf{p}_{\mathrm{cube}}\big) \in \mathbb{R}^D, \quad (5)$$

where $\oplus$ denotes vector concatenation. Unlike the single-image benchmark encoder f of Sec. 3.4, f_{mv} consumes the three views jointly, yet its output plays the same role: one embedding, compared at test time by the same cosine distance $d_{\cos}$. This is not an ensemble: the model retains one embedding and one decision, and the construction serves only to avoid forcing complementary m/z and retention cues through one lossy projection.

Multi-View Architecture and Training.

The branch depths follow from the modest number of measurements per identity. The mass-spectral branch is a stack of four convolution–BatchNorm–ReLU blocks (width 256, kernel 9); the first-dimension retention branch, three such blocks (width 128, kernel 7); and the joint-retention branch, three 3×3 blocks (width 128). Each branch concatenates global max- and mean-pooling, is linearly projected to D=256 with a ReLU and dropout, and the three projections are concatenated and fused by a linear layer with BatchNorm into the shared 256-d embedding, which feeds the cosine-softmax head (κ=24). To keep every view discriminative on its own and prevent a single branch from dominating, we add *deep supervision*: the shared identity prototypes also score each branch's projection $\mathbf{p}_v$, and this auxiliary loss (weight 0.5) is added during training and removed at test, so a single head remains at inference. We regularise strongly (weight decay 3×10^{-3}, dropout 0.45, label smoothing 0.1, and BatchNorm) and augment the two retention-bearing views ($\mathbf{X}^{\mathrm{rt}}$, $\mathbf{X}^{\mathrm{cube}}$) with smooth warps and shifts along the retention axis, the direction of the observed instrument drift; optimisation uses AdamW (learning rate 3×10^{-4}). The improved embedding is trained and evaluated on the same data, identity-disjoint folds, and all-pairs protocol as the benchmark (Sec. 5).

Cosine-Softmax Objective.

Verification is judged at a fixed, small false-positive rate, at which the relevant quantity is how *tightly* an identity's measurements cluster and how rarely impostors fall within that cluster. Triplet loss addresses this only indirectly: it enforces a margin on individually sampled (anchor, positive, negative) triplets, which is effective for ranking but constrains the global cluster geometry only loosely and remains sensitive to the mining strategy. We instead train the embedding as open-set classification with a *cosine-softmax* head. Each training identity y is given a learnable prototype $\mathbf{w}_y \in \mathbb{R}^D$, and a measurement i with embedding $\mathbf{z}_i = f_{\mathrm{mv}}(\cdot)$ (Eq. 5) and label y_i is scored against every prototype by their cosine similarity, scaled by a factor κ:

$$\mathcal{L} = -\frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \log \frac{\exp\big(\kappa\, \hat{c}(\mathbf{z}_i, \mathbf{w}_{y_i})\big)}{\sum_{y'} \exp\big(\kappa\, \hat{c}(\mathbf{z}_i, \mathbf{w}_{y'})\big)}, \qquad \hat{c}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}^\top \mathbf{b}}{\|\mathbf{a}\|_2\, \|\mathbf{b}\|_2}, \quad (6)$$

where $\mathcal{B}$ is the minibatch and y' ranges over the training identities of the fold. Because the normalised scores lie in $[-1, 1]$, the scale κ (a temperature) is what allows the posterior to become confident; we use $\kappa = 24$. Minimising Eq. 6 draws each measurement towards its own prototype and repels it from all others, which concentrates

each identity on the unit sphere, the property that a low false-positive operating point requires, and follows the same progression from triplet to angular-margin objectives that advanced deep face recognition [8, 35]; we retain the margin-free form, having found an additive angular margin to give no improvement at our number of training identities. At test time identities are compared, as before, by the cosine similarity of their embeddings (Sec. 3.4); only the training signal changes.

Pairwise Binary Classification Objective.

The metric-learning objectives learn an embedding space in which verification is subsequently performed using a predefined similarity measure, such as cosine similarity. As a complementary formulation, we therefore consider learning the verification decision directly as a binary classification problem. This formulation allows us to compare the channel and multi-view representations when verification is learned directly, rather than performed using a predefined similarity metric.

Each input measurement is represented either by a channel-encoded image $\mathbf{C} \in \mathbb{R}^{C\times H\times W}$, as introduced in Sec. 3.2, or by the joint multi-view representation $\mathbf{X} = (\mathbf{X}^{\mathrm{ms}}, \mathbf{X}^{\mathrm{rt}}, \mathbf{X}^{\mathrm{cube}})$. We denote input representation generically by $\mathbf{U}$.

Given two measurements $\mathbf{U}_i$ and $\mathbf{U}_j$, the corresponding encoder produces embeddings

$$\mathbf{z}_i = f(\mathbf{U}_i), \qquad \mathbf{z}_j = f(\mathbf{U}_j), \tag{7}$$

We construct a symmetric pair representation by concatenating the absolute difference of the two embeddings and their element-wise product,

$$\mathbf{q}_{ij} = |\mathbf{z}_i - \mathbf{z}_j| \oplus (\mathbf{z}_i \odot \mathbf{z}_j)\,, \tag{8}$$

where $\odot$ denotes element-wise multiplication and $\oplus$ denotes vector concatenation. A learned pairwise classification head g_{pair} maps this representation to a scalar logit,

$$\ell_{ij} = g_{\mathrm{pair}}(\mathbf{q}_{ij})\,. \tag{9}$$

For a pair label $y_{ij} \in \{0, 1\}$, where $y_{ij} = 1$ denotes measurements from the same identity and $y_{ij} = 0$ denotes measurements from different identities, the model is optimized using binary cross-entropy with logits,

$$\begin{aligned}\mathcal{L}_{\mathrm{BCE}} = -\frac{1}{|\mathcal{P}|}\sum_{(i,j)\in\mathcal{P}} &[y_{ij}\log\sigma(\ell_{ij}) \\ &+(1-y_{ij})\log(1-\sigma(\ell_{ij}))]\,.\end{aligned} \tag{10}$$

where σ denotes the logistic sigmoid and $\mathcal{P}$ is the set of training pairs. Positive and negative pairs are sampled in balanced proportions during training.

Unlike triplet and cosine-softmax training, the pairwise formulation does not require the verification relation to be expressed by a fixed distance in the learned embedding space. Instead, the encoder and comparison head are trained jointly to discriminate between genuine and impostor pairs. It also avoids explicit triplet mining and provides direct supervision at the level at which verification is evaluated. We use it as a complementary objective to determine whether performance differences between the channel and multi-view representations arise from the representation itself or from the metric-learning formulation used to train it.

Implementation Details.

We implemented all methods in PyTorch [20] and trained neural networks using AdamW [17] with a fixed learning rate of $3 \cdot 10^{-4}$ or $3 \cdot 10^{-5}$. Training was performed on a cluster of NVIDIA A100 GPUs with 40 GB of RAM and on a cluster of AMD MI250x GPUs with 64 GB of RAM.

Use of generative AI

OpenAI Codex, Google Gemini, and Anthropic Claude were used to assist with code development, manuscript drafting, and language revision. The scientific ideas, methodology, experiments, analysis, and conclusions were developed by the authors. All AI-assisted outputs were critically reviewed, edited, tested where applicable, and verified by the authors, who take full responsibility for the final manuscript.

4 The Human Scent Dataset

The Human Scent Dataset (HSD) is a curated dataset consisting of 504 samples from 40 identities from a previously published 4D-C dataset [12],

Dataset	Total	n	Samples per identity			
			min	max	mean	median
4D-C	504	40	10	37	12.6	10
HSD	2528	252	1	76	10.0	10

Table 1 Dataset statistics: dataset name, total number of samples, number of identities n, and the minimum, maximum, average, and median number of measurements per identity. Reproduced from the conference version [32]

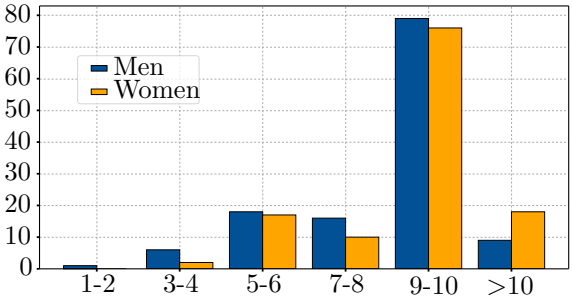


Fig. 3 Histogram of measurement counts per identity in the HSD dataset, grouped by biological sex. X-axis: measurements per identity, y-axis: number of subjects. Reproduced from the conference version [32]

and 2,024 newly collected samples from 212 identities. In total, HSD contains 129 male identities contributing 1,178 samples and 123 female identities contributing 1,350 samples (Fig. 3). Data were acquired using the Pegasus® 4D-C and a higher-resolution Pegasus® BT 4D, which provides enhanced mass-to-charge ratio precision. Collecting each human scent sample—including all preparation steps—takes $\approx$ 20 minutes; sensitivity and specificity enhancement takes $\approx$ 4 hours; the measurement and cleaning phase lasts about 2.5 hours, with data export requiring up to 4 hours. GC×GC–ToF-MS data files average $\approx$ 3 GB each, so that the dataset requires $\approx$ 7.5 TB of storage (Table 1). All participants provided their informed consent.

5 Experiments

In all our experiments, we use identity- and sex-aware 10-fold cross-validation on the HSD dataset with fixed splits. We report two metrics: the area under the ROC curve (AUROC) and the true positive rate at a fixed false positive rate TPR@c-FPR. Given positive (same-identity) pairs $\mathcal{P}$ and negative pairs $\mathcal{N}$, $\mathrm{TPR}(\tau) = |\{(i,j) \in \mathcal{P} : d_{ij} < \tau\}|/|\mathcal{P}|$ and $\mathrm{FPR}(\tau) = |\{(i,j) \in \mathcal{N} : d_{ij} < \tau\}|/|\mathcal{N}|$. The AUROC integrates TPR over FPR; TPR@c-FPR is $\mathrm{TPR}(\tau_c)$ at $\tau_c = \sup\{\tau : \mathrm{FPR}(\tau) \le c\}$, with $c = 5\%$ throughout.

Channel representation learning.

We train the Gaussian kernel bank (Sec. 3.2) by minimizing a weighted reconstruction loss with stochastic gradient descent. Kernel centers $\{\mu_k\}$ are fixed by a simple m/z-aware schedule: place $(C-2)$ centers uniformly at interval midpoints over $[0, 300]$, and place the remaining two at the midpoints of an equal two-way partition of $[300, 800]$. Kernel widths σ_k and amplitudes α_k are learned. Both kernels and weighting schemes achieve a reconstruction MSE below $\varepsilon = 3 \cdot 10^{-4}$ for all C, indicating high-fidelity reconstructions. Further details of the training procedure are provided in Online Resource 1 (Sec. S1.4).

Main results.

Table 2 reports the best cross-validation AUROC and TPR@5%FPR for the four feature-extraction families and the five alignment strategies. The custom CNN with $C = 5$ channels and *no* registration attains the best overall result, AUROC $= 89.5 \pm 3.4$ and TPR@5% $= 53.3 \pm 8.3$. The compound-centric chemistry baseline reaches AUROC $=$ 79.3 and TPR@5% $=$ 42.0 on the 4D-C subset (which carries the manual compound annotations it requires), and the best hand-engineered descriptor (LBP) reaches AUROC $=$ 78.1. Notably, on HSD the best results are obtained *without* spatial alignment, indicating that modern learned embeddings tolerate the observed retention-time drift; we attribute this to (i) marginalisation over run-specific drift instead of reliance on noisy alignment, (ii) alignment acting as a lossy bottleneck that discards discriminative retention-time/mass-spectral coupling cues, and (iii) estimated warps amplifying noise and injecting artifacts.

5.1 In-Session and Out-of-Session Generalisation

The cross-validation protocol above holds out *identities* but not acquisition *sessions*: a validation identity may have been measured on a date

				$c_{\cos}(\overline{\mathbf{u}}_c, \overline{\mathbf{s}}_c)$				$\mathrm{FCN}(\mathbf{u}, \mathbf{s})^{+}_{\cos}$			
		no registration		$\mathcal{R}^{(22)}$		$\mathcal{R}^{(24)}$		$\mathcal{R}^{(22)}$		$\mathcal{R}^{(24)}$	
	C	AUROC	TPR@5%	AUROC	TPR@5%	AUROC	TPR@5%	AUROC	TPR@5%	AUROC	TPR@5%
baseline	–	79.3±6.2	42.0±12.6	–	–	–	–	–	–	–	–
custom CNN	1	75.5±7.2	29.2±14.0	75.2±7.6	29.7±16.2	75.5±7.6	26.6±13.1	76.8±7.7	33.2±15.4	76.4±7.9	30.8±14.2
HOG	1	73.0±6.6	39.1± 7.8	74.3±7.2	39.1± 7.8	74.4±7.3	39.3± 7.7	76.0±4.9	33.4± 7.2	76.7±5.0	33.8± 7.1
LBP	1	78.1±3.2	15.8± 2.2	77.1±2.8	15.8± 2.2	77.1±2.8	15.5± 2.1	76.8±4.0	22.4± 4.3	78.1±4.3	22.9± 4.9
SIFT	1	66.6±6.0	18.6± 5.3	63.8±4.1	15.5± 4.0	63.4±3.9	15.1± 3.8	59.0±6.7	13.4± 4.0	63.8±6.5	17.4± 4.9
custom CNN	1	85.8±5.1	39.0±11.8	88.0±3.1	48.3± 9.2	88.0±3.6	50.5±10.5	87.9±3.5	49.1± 8.9	87.4±2.9	50.2± 9.1
VGG16	3	77.0±4.6	40.5± 6.6	79.5±9.6	42.9±10.4	79.8±9.4	42.2± 9.4	74.6±7.2	26.2± 8.3	74.5±6.9	26.6± 6.6
ResNet-18	3	87.5±3.9	44.9± 8.8	87.3±3.5	42.9± 5.2	86.9±3.0	41.1± 5.9	85.2±3.8	36.1± 7.1	85.5±4.3	36.2± 8.2
ResNet-101	3	86.3±4.1	41.5±11.4	86.0±4.2	38.5± 8.7	86.5±4.1	40.6± 7.7	84.5±4.7	35.3± 8.7	84.9±4.4	35.2± 9.1
GoogLeNet	3	87.9±3.5	46.8± 7.7	86.8±3.9	44.3± 8.4	86.7±3.4	44.7± 7.5	85.0±3.8	35.2± 8.9	84.9±4.1	36.1±10.3
Inception-v3	3	86.0±3.5	40.2± 5.6	85.5±3.3	39.2± 5.0	85.5±3.5	40.2± 5.1	83.5±3.9	32.8± 6.8	84.0±4.0	34.3± 9.1
custom CNN	3	88.3±3.8	52.4± 9.3	88.2±2.9	50.8± 5.1	87.3±3.5	48.6± 7.1	84.6±4.5	42.9± 6.8	83.8±5.1	40.0± 7.8
custom CNN	5	**89.5±3.4**	**53.3± 8.3**	88.3±3.4	51.3± 6.5	88.4±3.4	50.7± 6.7	85.8±3.4	46.5± 5.5	85.4±3.8	43.9± 6.2
custom CNN	10	88.7±3.9	53.7± 9.7	87.0±4.1	43.2± 8.4	86.3±3.9	50.0± 6.8	83.0±4.7	36.8± 9.8	83.3±4.9	39.5± 8.3
custom CNN	20	89.0±3.7	53.1± 7.1	87.5±3.9	49.0± 6.6	87.9±3.0	50.3± 6.6	84.9±3.9	43.6± 5.6	84.2±4.8	43.4± 7.1

Table 2 Best cross-validation Area Under ROC Curve and TPR@5%FPR for five spatial-alignment strategies on channel representation (Sec. 3.2) of 2D Gas Chromatography Mass Spectrometry measurements. Top: 4D-C dataset [12] with manual annotations of chemical compounds, bottom: Human Scent Dataset (Sec. 4). Results reported as $\mu \pm \sigma$ across 10 folds. Identity verification performed with: (i) analyst chemists *baseline* [6], (ii) deterministic descriptors LBP [19], HOG [7], and SIFT [18], (iii) custom-designed Convolutional Neural Network (CNN), and (iv) five pretrained neural network backbones *ResNet-18* [11], *ResNet-101* [11], *VGG-16* [31], *Inception-v3* [34], and *GoogLeNet* [34] each with one appended and trained projection layer. Number of channel representation kernels $C \in \{1, 3, 5, 10, 20\}$ (not applicable to baseline chemistry pipeline). Analyst chemistry baseline uses all available mass spectra $\mathcal{S}$ instead of channel representation. We perform registration using the weighted-cosine $c_{\cos}(\overline{\mathbf{u}}_c, \overline{\mathbf{s}}_c)$ and fully convolutional network-based $\mathrm{FCN}(\mathbf{u}, \mathbf{s})^{+}_{\cos}$ compound-peak detectors introduced in [12], evaluated on the $\mathcal{R}^{(22)}$ and $\mathcal{R}^{(24)}$ sets (22 and 24 compounds). Reproduced from the conference version [32]

that also appears in training through other identities. Because instrument state and ambient conditions drift between sessions, the session is a confounder that a deployed verifier cannot control, since at enrolment and at query the relevant session may be one the model never saw during training. We quantify this dependence on the full 252-identity HSD using the best benchmark configuration identified on the full dataset in Table 2 (custom CNN, Gaussian $C = 5$, no registration). For the primary analysis, we retrain this configuration on each fold's full training partition, including both detector populations, so that session membership is defined relative to the model's own training data. As a complementary single-detector analysis, we reuse the saved fold-specific checkpoints from the HR-only experiments reported in the supplementary material of our WACV paper [32] and stratify their original held-out HR pairs by session familiarity.

5.1.1 Definitions and Protocol

We define a *session* as the acquisition date of a measurement. For cross-validation fold k with training files $\mathcal{T}_k$, let $\mathcal{D}_k = \{\mathrm{date}(f) : f \in \mathcal{T}_k\}$ be the set of training sessions. A validation measurement x is *in-session* if $\mathrm{date}(x) \in \mathcal{D}_k$ and *out-of-session* otherwise. Verification compares an enrolled measurement i against a query measurement j; we therefore classify each ordered validation pair (i, j) into one of four regimes according to the session membership of its two members:

$$(\text{enrolled, query}) \in \{\textsf{in–in}, \textsf{in–out}, \textsf{out–in}, \textsf{out–out}\}.$$

The in/out-session label is *orthogonal* to the identity split and introduces no training leakage. The folds are partitioned by *identity*: every measurement of a validation identity is held out, so the network never observes a validation file, or its identity, during training, regardless of that file's session. Crucially, the *in-session* label does *not* mean the model saw the validation file; it means the model saw *other* identities measured on the same date, and was thereby exposed to that session's instrument and ambient

nuisance distribution. Both the full-HSD and HR-only analyses remain identity-disjoint because no validation identity is used for training.

In both settings we re-use the exact pairing of the main protocol (all ordered same-fold validation pairs, excluding self-pairs; a pair is genuine iff the two measurements share identity) and the exact metric (cosine similarity of the model's 4096-d embeddings, AUROC and TPR@5%FPR). Because per-fold out-of-session pairs are scarce, and entirely absent for some folds, a per-fold metric cannot be formed for every regime. We therefore *pool* pairs across the ten folds within each regime: the (similarity, genuine/impostor) pairs contributed by all ten folds are concatenated and scored by a *single* ROC per regime, rather than averaging the folds' individual curves. Because cosine similarity is bounded and shares the same scale across the folds' common architecture and training recipe, the concatenated scores are used to compute a single ROC for each regime.

5.1.2 Results on the Full HSD

For the primary analysis we retrain the benchmark architecture (custom CNN, $C = 5$, no registration) on the full training partition of each of the ten identity-disjoint folds, with the identical optimisation recipe, and apply the stratified pairing to all 2,528 held-out measurements of the 252 identities. Training on the full HSD avoids conflating session familiarity with detector-domain shift: applying the HR-only models to lower-resolution measurements would introduce a detector domain absent from their training data. We therefore use the full-HSD models for the primary analysis in Tables 3 and 4 and reserve the saved HR-only checkpoints for the within-detector analysis in Sec. 5.1.3.

Table 3 reports the four regimes pooled across folds. The model attains a per-fold mean AUROC of 89.3 ± 3.4 and TPR@5% of 47.9 ± 9.0. Session membership dominates: the same-familiarity in–in regime holds at 85.5/35.1, the cross-familiarity in–out and out–in regimes fall symmetrically to 76.3/28.7 — a 9-point AUROC loss and the hardest operating point — and the fully novel out–out regime is highest (90.6/76.4; accounted for in the discussion).

enrolled	query	AUROC	TPR@5%	#genuine
in	in	85.5	35.1	23314
in	out	76.3	28.7	3395
out	in	76.3	28.7	3395
out	out	90.6	76.4	3218
all pairs (pooled)		84.4	37.8	33322

Table 3 In/out-session verification on the *full* 252-identity HSD: the benchmark architecture (custom CNN, $C = 5$, no registration) *retrained* on each fold's full training partition, pooled across the ten folds (all-pairs AUROC / TPR@5%FPR, %). The per-fold mean is 89.3 ± 3.4 AUROC and 47.9 ± 9.0 TPR@5%. Cross-session mismatch (in–out/out–in) is the hardest regime, as on the new-detector subset (Table 5)

The two instruments.

HSD pools a higher-resolution (new-detector, exact-mass) and a lower-resolution (older-detector, unit-mass) instrument, contributing 1,187 and 1,341 of the evaluation measurements (Table 6). Because the two populations are almost identity-disjoint, the same full model exposes each instrument's in/out behaviour when the evaluation pairs are restricted to one at a time (Table 4). The higher-resolution half is substantially easier (90.7 vs. 74.2 mean AUROC), consistent with finer mass resolution preserving discriminative spectral detail that is lost in unit-mass acquisitions — the same type of information that appears important for our improved model at the strict TPR@5% operating point (Sec. 5.2). The cross-session penalty is nonetheless present on *both* instruments, and slightly larger on the harder low-resolution data (-10.7 vs. -8.2 AUROC from in–in to in–out), so the session-familiarity gap appears on both instruments rather than being specific to one of them.

5.1.3 Analysis with the Original HR-Only Models

The detector-stratified full-HSD analysis shows that the session-familiarity gap appears within both instrument subsets: AUROC decreases from 88.7 to 80.5 on the higher-resolution subset and from 67.1 to 56.4 on the lower-resolution subset (Table 4). As a complementary single-detector analysis, we reuse the saved fold-specific checkpoints underlying the HR-only experiments summarised in Table S5 of Online Resource 1 and originally reported in the supplementary material

Evaluation set	overall	in–in	in–out	out–out
higher-res (new)	90.7 ± 3.6	88.7	80.5	93.1
lower-res (older)	74.2 ± 6.8	67.1	56.4	77.4
full (HR + LR)	89.3 ± 3.4	85.5	76.3	90.6

Table 4 In/out-session AUROC (%) of the full-HSD model, evaluated separately on the higher- and lower-resolution instrument subsets and on the full dataset. All fold-specific models are trained on the full-HSD training partitions containing both instrument populations; only the evaluation pairs are restricted for the instrument-specific rows, while the full row additionally includes cross-instrument comparisons. "Overall" reports the mean ± standard deviation of the ten fold-level AUROCs over all pairs, whereas the session-regime columns report AUROCs computed after pooling the corresponding pairs across folds. The higher-resolution subset is easier, but the in–in→in–out decrease associated with a mismatch in session familiarity appears within both instrument subsets (−8.2 and −10.7 points)

enrolled	query	AUROC	TPR@5%	#genuine
in	in	90.3	56.3	9068
in	out	74.5	19.6	550
out	in	74.5	19.6	550
out	out	94.4	84.5	892
all pairs (pooled)		89.3	55.4	11060

Table 5 In/out-session verification on the higher-resolution new-detector subset using the saved fold-specific models underlying the HR-only experiments summarised in Table S5 of Online Resource 1 and originally reported in the supplementary material of our WACV paper [32]. The models (custom CNN, $C = 5$, no registration) are evaluated without retraining on their original held-out measurements, with scores pooled across the ten folds within each session regime. The pooled all-pairs result (89.3 AUROC and 55.4 TPR@5%) is consistent with the original per-fold mean AUROC of 88.5 ± 3.5 reported in Table S5

of our WACV paper [32]. The models are evaluated without retraining on their original held-out subset (1,187 measurements from 139 identities over 145 sessions, of which 1,060 are in-session and 127 out-of-session; Table 6). Their pooled all-pairs result (89.3/55.4 AUROC/TPR@5%) is consistent with the original per-fold mean AUROC of 88.5 ± 3.5 reported in Table S5. Performance decreases from 90.3/56.3 for in–in pairs to 74.5/19.6 for in–out pairs (Table 5). Thus, the same qualitative degradation appears for both full-HSD and HR-only training, with the latter containing no cross-instrument comparisons.

5.1.4 Discussion

First, measurements from sessions absent from training retain identity-discriminative signal: the out–out AUROC of 90.6 indicates that identities remain separable under the observed pairing structure when neither measurement comes from a training session. This is partly structural — out-of-session measurements concentrate in a few held-out sessions, so many out–out pairs compare two measurements from the *same* novel session, where shared context eases genuine acceptance, and the regime holds relatively few impostor pairs. It suggests that identity-discriminative signal persists in novel sessions, not that cross-session matching is easy.

Second, the cross-familiarity in–out/out–in regimes quantify the performance penalty associated with session-familiarity mismatch: with one side coming from a training-represented session and the other not, AUROC falls by ≈ 9 points on the full set — 8 and 11 points on the higher- and lower-resolution instruments respectively (Table 4), and 16 for the original HR-only models on their subset (Table 5). As expected for a symmetric similarity function evaluated on ordered pairs, the in–out and out–in regimes yield identical results; we therefore report them jointly as the cross-familiarity condition. The in–out direction corresponds to an enrol-now/query-later setting, in which a training-represented-session gallery is queried with novel-session probes. Session variability is therefore one important source of error on the full HSD, and session-invariant training (e.g., session-adversarial objectives or per-session calibration) is a natural next step. The pooled results should accordingly not be read as cross-session deployment accuracy.

5.2 Evaluating the Improved Verification Framework

We evaluate the multi-view representation and the training objectives introduced in Sec. 3.4 on the full HSD 10-fold protocol (all 252 identities, 2,528 measurements). To separate the effect of the input representation from that of the training objective, we evaluate both the original channel representation ($C = 5$) and the proposed multi-view representation. We compare three

	HR	LR	Full
Measurements			
in-session	1,060	1,177	2,237
out-of-session	127	164	291
total	1,187	1,341	2,528
Pairs (Gen. / Imp.)			
in → in	$9k/109k$	$14k/130k$	$23k/490k$
in → out	$0.6k/13k$	$2.7k/18k$	$3.4k/66k$
out → in	$0.6k/13k$	$2.7k/18k$	$3.4k/66k$
out → out	$0.9k/1k$	$2.3k/3k$	$3.2k/8k$

Table 6 Session composition of the evaluation data for each instrument population, pooled over the ten folds. *Top:* validation measurements that are in- vs. out-of-session relative to their fold's training dates. *Bottom:* genuine/impostor ordered pairs in each enrolled×query session regime. Measurement counts are exact, whereas pair counts are rounded and expressed in thousands (k). **HR**: new-detector subset (Sec. 5.1.3); **LR**: older-detector subset. The **Full** measurement count equals HR+LR, whereas the Full pair counts additionally include cross-instrument pairs and may therefore exceed the sum of the two subsets

Objective	Rep.	AUROC	TPR@5%
Triplet	channel	89.5±3.4	53.3±8.3
	multi	90.3±1.8	40.3±8.7
Cos.-softmax	channel	88.5±4.4	50.9±8.5
	multi	**92.8±2.1**	**62.2±8.0**
Pairwise BCE	channel	90.3±3.5	55.6±11.1
	multi	92.3±1.8	53.0±7.7

Table 7 Verification performance of the channel ($C = 5$) and multi-view representations with different objectives under identity-disjoint 10-fold cross-validation. This comparison is separate from the session-stratified analysis in Sec. 5.1; the channel-triplet row corresponds to the benchmark configuration in Table 2. Values are mean ± standard deviation across folds. Across all three objectives, the multi-view representation improves AUROC, whereas its effect on TPR@5% depends on the training objective. Cosine-softmax is the only objective for which the multi-view representation improves both metrics and yields the best overall result

objectives: triplet loss, cosine-softmax, and direct pairwise binary classification.

Across all three objectives (Table 7), the multi-view representation improves AUROC over the channel representation. With triplet loss, AUROC increases from 89.5% to 90.3%; with cosine-softmax, from 88.5% to 92.8%; and with pairwise binary classification, from 90.3% to 92.3%. This consistent improvement indicates that retaining complementary mass-spectral and retention-time information provides a stronger representation for globally ranking genuine and impostor pairs.

The behavior at the strict TPR@5%FPR operating point is less consistent. With the cosine-softmax objective, the multi-view representation improves TPR@5%FPR by approximately 11 points, yielding the best overall model. However, under the other two objectives, the multi-view representation reduces TPR@5%FPR. The decrease is particularly pronounced for triplet loss, where TPR@5%FPR drops by approximately 13 points. Thus, a multi-view representation alone does not guarantee better separation in the low-FPR regime.

Pairwise binary classification provides a complementary evaluation, as the similarity between two measurements is learned directly from their embeddings rather than defined by a fixed distance function. With this objective, the multi-view representation improves AUROC from 90.3% to 92.3%, while TPR@5%FPR remains similar (55.6% vs. 53.0%). This suggests that the AUROC benefit of the multi-view representation is not specific to metric-learning objectives. However, this improvement does not extend to the strict low-FPR regime, where cosine-softmax achieves substantially higher TPR@5%FPR.

With the multi-view representation fixed, replacing triplet loss with cosine-softmax increases AUROC from 90.3% to 92.8% and TPR@5%FPR from 40.3% to 62.2%, demonstrating the strong influence of the training objective, particularly in the low-FPR regime. Overall, the multi-view representation consistently improves global genuine–impostor ranking, whereas performance at a strict operating point depends strongly on the training objective.

Because a GC×GC–ToF-MS measurement bundles distinct mass-spectral and retention cues, a representation that keeps them separable sets the ceiling on how well identities can be distinguished, and the training objective determines how much of that ceiling is realised at a usable false-positive rate. On the full HSD the objective is decisive, and the gain it contributes falls almost entirely on TPR@5%FPR, the strict

operating point most relevant to deployment. These gains are, however, evaluated separately from session shift: the multi-view representation and cosine-softmax objective improve both AUROC and TPR@5%FPR under the standard protocol, while their robustness to the cross-familiarity gap identified in Sec. 5.1 remains to be evaluated. Combining them with a session-invariant criterion is a natural direction for future work.

5.3 Error Overlap and Complementarity of Single- and Multi-View Verification

Global verification metrics indicate which representation performs better on average, but do not reveal whether the corresponding models fail on the same measurement pairs. We therefore compare the channel and multi-view representations while keeping the training objective fixed to the best-performing cosine-softmax objective. This allows us to determine whether the channel and multi-view representations lead to the same verification errors under the same cosine-softmax training objective. Limited error overlap would indicate that the two representations capture complementary identity information and could potentially benefit from fusion.

Binary error agreement.

For each fold, both approaches are evaluated on exactly the same pairs of measurements from the full HSD. Because the similarity functions are symmetric, this analysis retains one unordered instance of each pair; the session-stratified analysis in Sec. 5.1 reports ordered pairs. A pair is genuine if both measurements share an identity and impostor otherwise. Each model uses its own fold-specific threshold at the 5%-FPR operating point. The resulting correct and incorrect decisions are summarized using 2×2 agreement tables (Tab. 8).

Across all pairs, the two models agree on 92.8% of decisions. This high agreement is largely driven by the 315,059 impostor pairs, for which the models agree in 93.8% of cases. Agreement is substantially lower for the 16,661 genuine pairs, reaching 73.8%. The remaining genuine-pair disagreements are nearly symmetric: 13.6% of genuine pairs are classified correctly only by the multi-view model, whereas 12.6% are classified correctly only by the channel-representation model. Thus, although the multi-view model achieves the stronger overall verification result, its improvement does not arise from uniformly correcting the errors of the channel-based model; the two representations still fail on partly different genuine pairs.

The disagreement pattern indicates some complementarity between the representations, particularly for genuine pairs. However, the two disagreement categories are of similar size, with only a small excess in favor of the multi-view model. Whether this complementarity can be exploited through model fusion remains an open question.

Decision-margin analysis.

To examine not only the binary decisions but also their confidence, we define the decision margin for model m, fold k, and pair (i, j) as

$$d_{m,k}(i,j) = s_{m,k}(i,j) - \tau_{m,k}, \tag{11}$$

where $s_{m,k}(i,j)$ is the similarity score and $\tau_{m,k}$ is the corresponding 5%-FPR threshold. Positive margins indicate a genuine-pair decision, negative margins indicate an impostor-pair decision, and their magnitude measures the distance from the decision boundary.

Figure 4 shows the joint margin distributions pooled across folds. Diagonal quadrants contain matching decisions, whereas off-diagonal quadrants contain disagreements.

The impostor-pair distribution reveals a notable difference between the two models. While many pairs receive channel-representation margins relatively close to the decision boundary, the multi-view model produces a pronounced tail towards strongly negative margins. Thus, for a substantial subset of correctly rejected impostor pairs, the multi-view representation places them considerably farther from the decision boundary. This suggests stronger separation of some difficult impostor pairs, even when the final binary decision agrees between the two models.

	(a) All pairs $n = 331{,}720$		(b) Genuine $n = 16{,}661$		(c) Impostor $n = 315{,}059$	
	Multi-view		Multi-view		Multi-view	
Channel repre.	Correct	Wrong	Correct	Wrong	Correct	Wrong
Correct	89.1%	3.6%	36.3%	12.6%	91.9%	3.1%
Wrong	3.6%	3.7%	13.6%	37.5%	3.1%	1.9%

Table 8 Correctness agreement between verifiers using the channel representation ($C = 5$) and the proposed multi-view representation, both trained with the cosine-softmax objective. Rows denote decisions of the channel-representation model and columns denote decisions of the multi-view model. Results are reported separately for all, genuine, and impostor pairs at the fold-specific 5%-FPR operating point. Percentages are relative to the number of pairs in each category. Since the table pools pair-level decisions across folds, its marginal correctness rates are pair-weighted and are therefore not directly comparable to the fold-averaged TPR@5%FPR values reported in Tab. 7

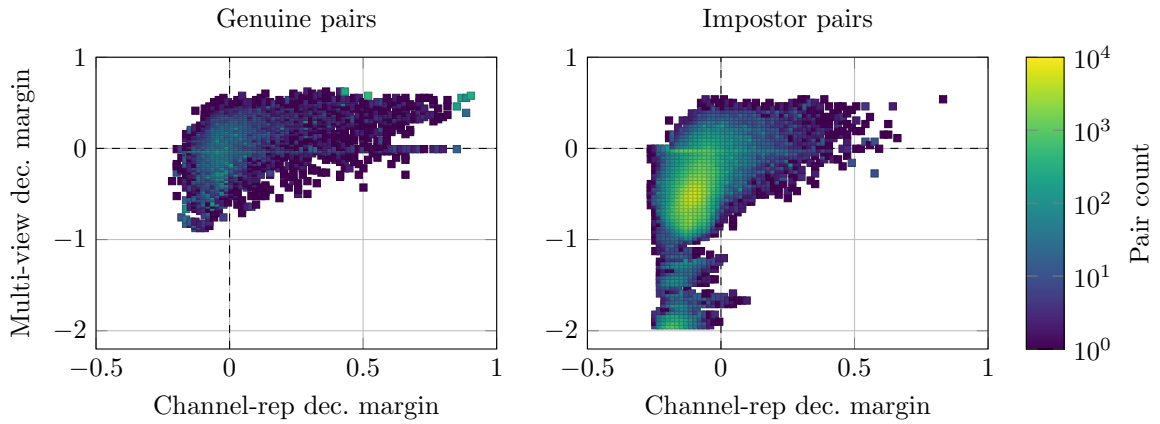


Fig. 4 Joint distributions of decision margins for the channel-rep and multi-view verification systems. Dashed lines indicate the respective decision boundaries. Diagonal quadrants contain pairs for which both systems make the same decision, while off-diagonal quadrants represent method-specific decisions. Colour indicates the number of pairs on a logarithmic scale

6 Limitations

Although our results establish feasibility, the following limitations apply: (i) accuracy (best ≈ 62% TPR@5%FPR; AUROC ≈ 93%) remains below that of established modalities; (ii) external validity across labs, climates, diets, hygiene, medications, and environments remains to be investigated; (iii) long-term permanence and drift are unquantified; (iv) the observation that explicit canonical registration is unnecessary is currently demonstrated only on Human Scent Dataset, with robustness under stronger retention-time drift and in settings requiring chemical interpretability still to be evaluated; and (v) throughout our verification experiments, impostor pairs are not restricted to same-sex comparisons. Consequently, sex-associated cues may contribute to the reported performance; a same-sex evaluation is needed to quantify the identity-specific signal independently of biological sex.

7 Conclusion

We systematically evaluated ten approaches for odor-based identity verification using raw GC×GC–ToF-MS data, represented as multi-channel images. In contrast to common analytical pipelines, the best results on our dataset are achieved without spatial alignment, indicating that modern learned embeddings can tolerate the observed retention-time drift. Our experiments on a newly proposed dataset of 252 individuals demonstrate the feasibility of scent-based recognition: the original channel-based benchmark achieves an ≈ 53% true positive rate at a 5% false positive rate with an 89.5% AUROC. Extending the analysis to in/out-session generalisation, we found that measurements from sessions absent

from training retain identity-discriminative signal under the observed pairing structure, with an out–out AUROC of 90.6 on the full HSD and 94.4 on the new-detector subset. Because many out–out pairs originate from the same unseen session, these values should not be interpreted as matching performance between two different unseen sessions. Nevertheless, session-familiarity mismatch remains substantially harder: AUROC drops by approximately 9 points on the full dataset (85.5 to 76.3) and by approximately 16 points on the new-detector subset (90.3 to 74.5) when a known-session gallery is queried with novel-session probes. These results identify acquisition-session variability as one important source of error to address next. Finally, a multi-view representation (mass-spectral, retention, and joint $m/z\times$retention views) trained with a cosine-softmax metric-learning recipe improves the best verification result to 92.8% AUROC and 62.2% TPR@5%FPR on the full 252-identity dataset; controlled ablations indicate that this low-FPR gain depends on matching the multi-view representation with an objective that produces sufficiently compact identity clusters. Importantly, these two extensions expose complementary aspects of the verification problem. The improved representation and training objective increase discriminative performance, whereas the session analysis shows that acquisition variability remains a separate source of error. Further progress therefore requires both stronger identity representations and greater robustness to session shift. These findings confirm that odor signatures can serve as a biometric trait when processed with modern feature embedding techniques, while underscoring clear opportunities for further improvement and more efficient measurement protocols.

Acknowledgements. The research reported in this paper has been partly funded by BMK, BMAW, and the State of Upper Austria in the frame of the SCCH competence center INTEGRATE [Project 1.6 TFI (Transferable Intelligence)] part of the FFG COMET Competence Centers for Excellent Technologies Programme, by the Czech Technical University in Prague grant No. SGS23/173/OHK3/3T/13, by the Ministry of the Interior of the Czech Republic grant No. VJ01010123, and by the Ministry of Education, Youth and Sports of the Czech Republic through the e-INFRA CZ (ID:90254).

Statements and Declarations. Author contributions. J.H.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Visualization, Writing—original draft, Writing—review and editing; R.S.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Visualization, Writing—original draft, Writing—review and editing; J.C.: Methodology, Investigation, Data curation, Resources; P.P.: Methodology, Investigation, Data curation, Resources; J.M.: Conceptualization, Methodology, Supervision; S.U.: Resources, Supervision.

Funding. The research reported in this paper has been partly funded by BMK, BMAW, and the State of Upper Austria in the frame of the SCCH competence center INTEGRATE [Project 1.6 TFI (Transferable Intelligence)] part of the FFG COMET Competence Centers for Excellent Technologies Programme, by the Czech Technical University in Prague grant No. SGS23/173/OHK3/3T/13, by the Ministry of the Interior of the Czech Republic grant No. VJ01010123, and by the Ministry of Education, Youth and Sports of the Czech Republic through the e-INFRA CZ (ID:90254).

Competing interests. The authors have no relevant financial or non-financial interests to disclose.

Ethics approval. This study was approved by the Institutional Review Board of the University of Chemistry and Technology, Prague (approval number EK/8/2020) and complies with the Declaration of Helsinki for Medical Research involving Human Subjects.

Consent to participate. All participants provided informed consent prior to participation.

Data availability. The Human Scent Dataset used in this study is available upon request. Access instructions and the required data-access request form are provided on https://github.com/radimspetlik/identity_verification_from_human_scent.

Code availability. The source code supporting this study is publicly available on https://github.com/radimspetlik/identity_verification_from_human_scent.

Supplementary Information. Online Resource 1. Additional methodological details, instrumentation specifications, registration procedures, complete benchmark results, and supplementary ablation studies.

References

[1] Ballén Castiblanco JE, Calvacanti Ferreira VH, Teixeira CA, et al (2024) Classification of produced water samples using class-oriented chemometrics and comprehensive two-dimensional gas chromatography coupled to mass spectrometry. Talanta 268:125343. https://doi.org/10.1016/j.talanta.2023.125343

[2] Brener S, Snitz K, Sobel N (2024) An electronic nose can identify humans by the smell of their ear. Chemical Senses 49. https://doi.org/10.1093/chemse/bjad053

[3] Buszewski B, Grzywinski D, Ligor T, et al (2013) Detection of Volatile Organic Compounds as Biomarkers in Breath Analysis by Different Analytical Techniques. Bioanalysis 5(18):2287–2306. https://doi.org/10.4155/bio.13.183

[4] Caratti A, Squara S, Bicchi C, et al (2024) Boosting comprehensive two-dimensional chromatography with artificial intelligence: Application to food-omics. TrAC Trends in Analytical Chemistry 174:117669. https://doi.org/10.1016/j.trac.2024.117669

[5] Colón-Crespo LJ, Herrera-Hernández D, Holness H, et al (2017) Determination of VOC marker combinations for the classification of individuals by gender and race/ethnicity. Forensic Science International 270:193–199. https://doi.org/10.1016/j.forsciint.2016.09.011

[6] Cuzuel V, Leconte R, Cognon G, et al (2018) Human odor and forensics: Towards Bayesian suspect identification using GC×GC–MS characterization of hand odor. Journal of Chromatography B 1092:379–385. https://doi.org/10.1016/j.jchromb.2018.06.018

[7] Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), Ieee, pp 886–893

[8] Deng J, Guo J, Xue N, et al (2019) Arcface: Additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 4690–4699

[9] Ferreira VHC, Hantao LW, Poppi RJ (2021) Use of color based chromatographic images obtained from comprehensive two-dimensional gas chromatography in authentication analyses. Talanta 234:122616. https://doi.org/10.1016/j.talanta.2021.122616

[10] Frazier CJG, Gokool VA, Holness HK, et al (2023) Multivariate regression modelling for gender prediction using volatile organic compounds from hand odor profiles via HS-SPME-GC-MS. PLOS ONE 18(7). https://doi.org/10.1371/journal.pone.0286452

[11] He K, Zhang X, Ren S, et al (2016) Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 770–778, https://doi.org/10.1109/CVPR.2016.90

[12] Hlavsa J, Spetlik R, Čechová J, et al (2025) Sex Classification from Human Scent Using Image Interpretation of 2D Gas Chromatography-Mass Spectrometry Data. In: Petersen J, Dahl VA (eds) Image Analysis. Springer Nature Switzerland, pp 457–470, https://doi.org/10.1007/978-3-031-95911-0_32

[13] Jirayupat C, Nagashima K, Hosomi T, et al (2022) Breath odor-based individual authentication by an artificial olfactory sensor system and machine learning. Chemical Communications 58(44):6377–6380. https://doi.org/10.1039/d1cc06384g

[14] Kaminskyi O, Čechová J, Malá U, et al (2023) Device for Non-contact Odor Sampling. Forenzni vedy, pravo, kriminalistika 8(1):33–41

[15] Kataoka H, Saito K, Kato H, et al (2013) Noninvasive analysis of volatile biomarkers in human emanations for health and early disease diagnosis. Bioanalysis 5(11):1443–1459. https://doi.org/10.4155/bio.13.85

[16] Ladislavová N, Pojmanová P, Vrbka P, et al (2023) Human scent signature on cartridge case survives gun being fired: A preliminary study on a potential of scent residues as an identification tool. PLOS ONE 18(3):e0283259

[17] Loshchilov I, Hutter F (2019) Decoupled weight decay regularization. In: 7th International Conference on Learning Representations

[18] Lowe DG (2004) Distinctive image features from scale-invariant keypoints. International journal of computer vision 60(2):91–110

[19] Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence 24(7):971–987

[20] Paszke A, Gross S, Massa F, et al (2019) PyTorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems

[21] Penn DJ, Oberzaucher E, Grammer K, et al (2006) Individual and gender fingerprints in human body odour. Journal of The Royal Society Interface 4(13):331–340. https://doi.org/10.1098/rsif.2006.0182

[22] Peters R, Veenstra R, Heutinck K, et al (2023) Human scent characterization: A review. Forensic Science International 349:111743

[23] Pinc L, Bartoš L, Reslová A, et al (2011) Dogs Discriminate Identical Twins. PLoS ONE 6(6):e20704

[24] Pinc L, Bartos L, Vyplelova P (2016) Resistance of human odours to extremely high temperature as revealed by trained dogs. Czech Journal of Animal Science 61(4):172–176. https://doi.org/10.17221/8848-CJAS

[25] Pojmanová P (2020) Molekulová skladba pachové signatury. PhD thesis, VŠCHT

[26] Pojmanová P, Ladislavová N, Urban Š (2022) Human Skin Scent Class and Individual Identification

[27] Prokop-Prigge KA, Greene K, Varallo L, et al (2016) The effect of ethnicity on human axillary odorant production. Journal of Chemical Ecology 42(1):33–39. https://doi.org/10.1007/s10886-015-0657-8

[28] Rodriguez-Lujan I, Bailador G, Sanchez-Avila C, et al (2013) Analysis of pattern recognition and dimensionality reduction techniques for odor biometrics. Knowledge-Based Systems 52:279–289. https://doi.org/10.1016/j.knosys.2013.08.002

[29] Schroff F, Kalenichenko D, Philbin J (2015) FaceNet: A unified embedding for face recognition and clustering. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 815–823, https://doi.org/10.1109/CVPR.2015.7298682

[30] Shirasu M, Touhara K (2011) The scent of disease: volatile organic compounds of the human body related to disease and disorder. The Journal of Biochemistry 150(3):257–266. https://doi.org/10.1093/jb/mvr090

[31] Simonyan K, Zisserman A (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. In: Proceedings of 3rd International Conference on Learning Representations

[32] Spetlik R, Hlavsa J, Čechová J, et al (2026) Identity verification from human scent using channel representation of 2d gas

chromatography-mass spectrometry data. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp 1471–1480

[33] Stockham RA, Slavin DL, Kift W (2004) Survivability of human scent. Forensic Science Communications 6(4). Publisher: Federal Bureau of Investigation

[34] Szegedy C, Liu W, Jia Y, et al (2015) Going deeper with convolutions. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 1–9, https://doi.org/10.1109/CVPR.2015.7298594

[35] Wang H, Wang Y, Zhou Z, et al (2018) Cosface: Large margin cosine loss for deep face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 5265–5274

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf