MVGSig: Multi-View Graph Fusion for Out-of-Distribution Detection in FMCW Radar-Based Person Identification

MOHAMMAD HOSSEIN SHIRAZI, Data Science and AI, Auckland University of Technology, Auckland, New Zealand
SIRA YONGCHAREON, Data Science and AI, Auckland University of Technology, Auckland, New Zealand
ANURADHA SINGH, Data Science and AI, Auckland University of Technology, Auckland, New Zealand
JING MA, Data Science and AI, Auckland University of Technology, Auckland, New Zealand

Radar-based person identification using FMCW sensors enables privacy-preserving authentication but risks overconfident predictions for out-of-distribution (OOD) individuals. We propose MVGSig, a multi-view graph fusion framework for open-set identification. Six complementary signal representations (spectral, temporal, time-frequency, wavelet, cepstral, and higher-order statistical) are extracted from a single FMCW radar sensor and fused through graph attention networks with learnable edge biases and quality-weighted pooling. View-Aware Collaborative Calibration (VACC) transfers knowledge from high-quality to low-quality views during training. A learned meta-detector combines node, edge, graph, and reconstruction anomaly scores. Under leave-one-subject-out cross-validation on a public ten-participant dataset, MVGSig reduces the false accept rate at 95% true positive rate from 0.697 to 0.375 ± 0.103, a 46.2% reduction relative to the raw signal baseline, and increases mean AUROC from 0.846 to 0.896 ± 0.031. FPR@95 is the primary metric because admitting unknown individuals is the security-critical failure. Closed-set accuracy reaches 66.5 ± 10.1% on data collected for vital-sign estimation, a deliberately low-discriminability regime where overconfident predictions are particularly concerning. Ablation under matched scoring conditions attributes +0.025 AUROC to multi-view fusion and another +0.025 to collaborative view learning, which eliminates more remaining false accepts (32.1% versus 20.8%).

CCS Concepts: • **Computing methodologies** → **Neural networks**; **Anomaly detection**; *Graph neural networks*; Signal processing; • **Security and privacy** → *Biometrics*; • **Computer systems organization** → *Sensor networks*.

Additional Key Words and Phrases: FMCW radar, person identification, out-of-distribution detection, multi-view fusion, open-set recognition, contrastive learning, attention mechanism

1 Introduction

Person identification is important for security and healthcare applications, requiring operation without physical contact or controlled conditions as deployment contexts expand.

Frequency-modulated continuous-wave (FMCW) radar is a candidate modality for contactless identification. Unlike cameras, which raise privacy concerns and degrade under poor illumination, radar captures physiological signatures—encoded in chest wall motion, body geometry, and micro-Doppler characteristics—while preserving

Authors' Contact Information: Mohammad Hossein Shirazi, Data Science and AI, Auckland University of Technology, Auckland, Auckland, New Zealand; e-mail: hossein.shirazi@autuni.ac.nz; Sira Yongchareon, Data Science and AI, Auckland University of Technology, Auckland, Auckland, New Zealand; e-mail: sira.yongchareon@aut.ac.nz; Anuradha Singh, Data Science and AI, Auckland University of Technology, Auckland, Auckland, New Zealand; e-mail: anuradha.singh@aut.ac.nz; Jing Ma, Data Science and AI, Auckland University of Technology, Auckland, Auckland, New Zealand; e-mail: jing.ma@aut.ac.nz.



This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2026 Copyright held by the owner/author(s).
ACM 2577-6207/2026/11-ART
https://doi.org/10.1145/3862635

Just Accepted

visual anonymity. The frequency-modulated waveform extracts phase and magnitude with sub-millimeter precision: phase variations encode cardiopulmonary activity with individual-specific timing and morphology, while magnitude captures body geometry and micro-motion characteristics [13, 47].

Deep learning architectures have shown high identification accuracy on radar signals under controlled conditions [16, 17, 40], but assume test samples are drawn from the same distribution as training data—an assumption often violated in practice. When encountering individuals absent from enrollment, neural networks produce overconfident predictions, incorrectly assigning unknown inputs to registered identities [11], creating security vulnerabilities. The ability to detect and reject such *out-of-distribution* (OOD) inputs extends the task from closed-set classification to **open-set identification**.

OOD detection methods fall into two categories: *training-driven* approaches [12, 20, 26] and *post-hoc* methods [11, 23, 26]. Despite progress, most research targets image benchmarks with single-view representations. These methods face specific challenges on radar signals: (1) *temporal dynamics*—biometric signatures evolve over multi-second timescales, unlike static image features; (2) *low SNR*—physiological displacements are sub-millimeter, buried in clutter [50]; (3) *non-stationary structure*—vital signs exhibit time-varying frequency content. Compared to other contactless modalities—WiFi CSI [29] and LiDAR [8]—radar uniquely combines range resolution with phase sensitivity to micro-motions, motivating domain-specific OOD methods.

Multi-view fusion provides complementary perspectives that disambiguate ambiguous samples [24, 52]. For radar, spectral, time-frequency, wavelet, and statistical transformations yield structurally diverse features. Recent work [7] shows that aggregating scores across representations reduces false positives. Critically, OOD detection benefits from *cross-view disagreement*: a known individual produces consistent representations across views, while an unknown individual may appear plausible in one domain but anomalous in another—a signal unavailable to single-view methods. In access control, the security-critical metric is the false positive rate (FPR): even modest distributional separation yields substantial FPR reductions, since FPR depends on score distribution tail overlap rather than mean discriminability.

To address these gaps, we propose **MVGSig** (Multi-View Graph Signature), a framework for radar-based open-set identification. Unlike prior radar-based identification methods that rely on single-view representations or fixed fusion schemes (e.g., concatenation, averaging), MVGSig models six signal transformations—selected to span complementary signal domains (spectral, temporal, time-frequency, multi-resolution, cepstral, and statistical; see Section 3.3 for selection rationale)—as nodes in a graph, enabling adaptive cross-view information routing through graph attention networks with learnable edge biases. Additionally, while existing multi-view approaches treat views independently, MVGSig introduces quality-aware collaborative learning, in which higher-quality views guide lower-quality views during training. A quality-aware pooling mechanism, guided by per-view reconstruction fidelity, down-weights lower-quality views.

For OOD detection, rather than relying on a single scoring function, MVGSig combines anomaly signals across multiple architectural levels—node, edge, graph, and reconstruction—via a learned meta-detector. This multi-level approach captures different types of anomalies that may manifest at different stages of processing.

We evaluate MVGSig using leave-one-subject-out (LOSO) cross-validation, in which each participant serves as the unknown subject exactly once. Unlike random splits, which risk leaking identity-correlated information across folds, LOSO emulates deployment conditions where the system must simultaneously recognize enrolled individuals and reject persons never seen during training. Table 1 positions MVGSig relative to representative prior work. Existing radar identification systems predominantly assume closed-set conditions, OOD detection methods target image benchmarks with single-view representations, and multi-view fusion strategies have not incorporated quality-aware collaborative learning for physiological signals.

Contributions. The main contributions are:

Just Accepted

Table 1. Positioning of MVGSig relative to representative prior work.

Method	Domain	Views	Open-Set	Collaborative	Multi-Level
HAROOD [18]	Radar	1	☒	–	–
PMvF [7]	Vibration	2	☒	–	–
CSI [43]	Image	1	☒	–	–
CIDER [31]	Image	1	☒	–	–
JEDAN [14]	Radar	1	☒	–	–
Heart Sig. [36]	Radar	1	☒	–	–
DML [54]	Image	–	–	☒	–
KDCL [10]	Image	–	–	☒	–
MVGSig (Ours)	**Radar**	**6**	☒	☒	☒

(1) **Graph-based multi-view fusion:** A framework that models six radar signal views as graph nodes and fuses them via graph attention with learnable edge biases and quality-aware pooling—unlike fixed fusion schemes that cannot adapt to input-specific view quality.
(2) **Collaborative view learning:** A training strategy combining per-view supervision with cross-view distillation, where a gradient-free quality estimator guides knowledge transfer from higher-quality to lower-quality views—addressing the limitation of treating views independently.
(3) **Multi-level OOD detection:** A scoring mechanism combining anomaly signals at node, edge, graph, and reconstruction levels through a learned meta-detector—capturing complementary anomaly types that single-level scores may miss.

Scope. All experiments use one radar platform (TI AWR1642BOOST), ten participants, and a single sensing geometry (80 cm, frontal). The results therefore establish relative improvements in open-set rejection under controlled conditions; robustness to sensing distance, orientation, radar hardware, and larger or more diverse cohorts is not claimed and is discussed in Section 5.9.
The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 details the proposed methodology, including experimental setup and implementation details. Section 4 presents experimental results. Section 5 discusses the findings, addresses methodological considerations, and identifies limitations and future directions. Section 6 concludes the paper.

2 Related Work

This section reviews work directly relevant to MVGSig: radar-based identification with its open-set limitations, OOD detection methods, and multi-view fusion approaches.

2.1 Radar-Based Person Identification

FMCW radar has emerged as an alternative to vision-based biometric systems, offering privacy preservation and robustness to illumination and occlusion [33]. The technology captures person-specific physiological signatures encoded in chest wall motion, where phase variations from cardiopulmonary activity contain patterns that may differ across individuals [47].

Two paradigms dominate: gait-based recognition using micro-Doppler signatures [1, 21], and vital-sign-based identification using cardiopulmonary waveforms [13, 36]. Deep learning has advanced performance, with CNNs capturing spatial patterns from range-Doppler maps [17, 40] and transformers modeling temporal dynamics [16].

Just Accepted

Recent architectures such as AVAESA [39] demonstrate that attention-enhanced variational autoencoders with learnable signal processing can improve cross-scenario robustness in radar-based vital sign monitoring. Prior work has shown that complementary signal transformations—STFT, wavelet, cepstral—capture distinct physiological characteristics [17, 37], motivating multi-view approaches.

2.1.1 Deep Learning Failure Modes for Radar Signals. While deep architectures achieve high accuracy under controlled conditions, they exhibit specific failure modes when applied to radar-based biometrics:

CNNs assume spatial translation invariance in range-Doppler maps, but physiological patterns vary temporally with breathing phase and cardiac cycle. This temporal non-stationarity causes feature drift that spatial convolutions cannot adequately model, leading to performance degradation when test-time physiological states differ from training [38].

Transformers require large datasets to avoid overfitting attention weights, yet radar biometric datasets rarely exceed 20 subjects. With limited participants and samples, transformer attention heads may memorize identity-specific patterns rather than learning generalizable physiological features [39].

Both architectures produce overconfident predictions on OOD inputs because their softmax outputs are calibrated for closed-set classification. The resulting miscalibration creates security vulnerabilities where unknown individuals receive high-confidence identity assignments [11].

2.1.2 Comparison with Other Contactless Sensing Modalities. Beyond radar, other contactless modalities face distinct challenges for biometric identification. **WiFi-based sensing** captures channel state information (CSI) that reflects body motion and geometry [6, 48]. However, CSI signals suffer from severe environmental multipath that varies across deployment locations, requiring environment-specific calibration that limits cross-domain generalization. Phase offsets across subcarriers introduce additional noise, degrading fine-grained physiological feature extraction [6].

LiDAR provides precise geometric information through point cloud reconstruction but lacks the sub-millimeter phase resolution needed for cardiopulmonary signatures. LiDAR excels at gait-based identification based on body geometry but cannot capture subtle chest wall displacements (typically 0.1–1 mm) that encode cardiac activity.

Acoustic sensing is the remaining vision-free option and reaches comparable displacement sensitivity: ultrasonic ranging resolves chest-wall motion at close range, and audible-band methods can recover respiration. Two properties limit its use for the problem addressed here. First, acoustic propagation is far more sensitive to the environment than millimetre-wave propagation—ambient noise, air currents, and reverberation vary on timescales comparable to the physiological signal itself, and the attenuation of ultrasound in air restricts the usable range to well under a metre. Second, and more consequentially, acoustic sensing requires a microphone in the monitored space. The property that motivates radar for access control is that it captures physiological signatures without recording anything intelligible: no image, no speech, no illumination. An acoustic channel reinstates exactly the exposure that the modality was chosen to avoid, so the comparison is not simply one of accuracy but of what the deployment is permitted to observe. We return to acoustic fusion, as distinct from acoustic substitution, in Section 5.9.

FMCW radar's combination of phase-based micro-motion sensitivity, penetration through clothing, and operation in varied lighting conditions motivates domain-specific OOD detection methods that exploit its unique signal characteristics.

2.1.3 The Closed-Set Limitation. Despite advances in radar-based identification, existing systems operate under the closed-set assumption: every test subject is present in a training set. When an unknown individual presents to such a system, it produces high-confidence predictions assigning the input to a known identity—neural network overconfidence [11]. In access control scenarios, this constitutes a security vulnerability.

Just Accepted

Kahya *et al.* [18, 19] represent notable exceptions, addressing OOD detection for radar-based activity classification and human presence detection. We analyze their approaches in detail to motivate MVGSig's design:

HAROOD [18] extracts range-Doppler representations and applies contrastive learning for activity classification OOD detection. Three technical limitations constrain its applicability to biometrics: (1) Range-Doppler maps discard phase information that encodes person-specific cardiac signatures—the complex-valued radar return contains identity cues in its phase component that magnitude-only processing loses; (2) Activity classification involves large inter-class gaps (walking vs. sitting), whereas biometric identification requires distinguishing subtle intra-class variations among individuals performing the same activity; (3) Single-view representation cannot detect samples that appear normal in one domain but anomalous in another—a limitation that multi-view analysis can address.

HOOD [19] extends to human presence detection using autoencoder reconstruction error. While effective for presence/absence binary classification, the reconstruction objective does not structure embeddings for identity discrimination. Mahalanobis distances computed in this space conflate reconstruction quality with identity confidence, resulting in suboptimal OOD separation in the biometric setting.

These limitations motivate MVGSig's multi-view approach: by extracting six complementary representations from the same radar signal, the framework can detect OOD samples that appear normal under any single transformation but show cross-view inconsistency.

2.2 Out-of-Distribution Detection

OOD detection is a binary decision problem: given a scoring function $S(\mathbf{x})$ and threshold τ, input $\mathbf{x}$ is in-distribution if $S(\mathbf{x}) < \tau$ [53]. Methods are partitioned into post-hoc approaches operating on trained models and training-driven approaches that modify the learning objective [28].

2.2.1 Post-Hoc and Distance-Based Methods. Post-hoc methods construct scores from trained model outputs: softmax confidence [11], energy scores [26], temperature scaling [25], and activation analysis [41, 46]. Distance-based variants score against class-conditional distributions using Mahalanobis distance [23], k-nearest-neighbour distances [42], or class-posterior weighting [34].

Three assumptions in this family are violated by radar biometric signals. Softmax confidence assumes comparable signal quality across samples, but low-SNR measurements produce diffuse distributions even for enrolled subjects. Mahalanobis distance assumes class-conditional Gaussian structure, which cardiac periodicity violates by inducing multi-modal clusters. Energy scores were calibrated on image classification benchmarks and do not obviously transfer to sequential data with temporal dependence. All three share a deeper limitation for our purposes: each operates on a single representation, so noise or artefacts in that representation degrade detection with no recourse. MVGSig addresses this by enriching the feature basis across complementary views, reducing the influence of any one degraded representation.

Two activation-based methods deserve specific comment, since they are the strongest post-hoc scores on image benchmarks, and both illustrate why a fair comparison would require adaptation rather than direct application. ReAct [41] clips penultimate activations at a percentile of the in-distribution range, exploiting the observation that OOD inputs drive abnormally large activations. Our scoring operates on a signature that is L_2-normalised onto $\mathbb{S}^{31}$ (Eq. 12), where activation magnitude is constant by construction, so ReAct would have to be applied at an earlier point in the network—the per-view encodings h_k or the pooled representation g—and the percentile threshold re-estimated there. ViM [46] combines the logit response with the norm of the residual against a principal subspace fitted to in-distribution features; its behaviour depends on how the 32-dimensional signature space partitions into principal and residual components, which we have not characterised. Neither method is precluded here. Both are simply not aligned with the representation MVGSig produces, and reporting either

Just Accepted

without first adapting it to that representation would not be a fair test of the method. We report the closest applicable comparisons—Energy, Mahalanobis and KNN scoring on the same fused signature—in Table 4.

2.2.2 Training-Driven Methods. Training-driven approaches reshape embedding geometry during learning. Outlier Exposure [12] augments training with auxiliary OOD data, but obtaining representative "unknown person" samples is problematic for biometric systems. Prototype-based methods structure embeddings for distance scoring [15, 27]. ArcFace [5] enforces angular margin constraints, producing compact clusters amenable to distance-based OOD scoring. MVGSig incorporates ArcFace while extending to multi-view representations, applying angular-margin loss to both the fused signature and individual-view embeddings.

2.2.3 Open-Set Recognition and the Closest Radar Prior Work. The methods above treat OOD detection as a scoring problem over a fixed embedding. A parallel line of work reframes it as *open-set recognition*, learning the rejection boundary end-to-end. OpenMax [2] recalibrates the penultimate activation vector against a per-class extreme-value model, converting an overconfident softmax into a distribution with an explicit "unknown" outcome. Generative approaches synthesise counterfactual unknowns to supervise that boundary [32]. Adversarial reciprocal point learning [4] models the extra-class space of each known category directly, bounding open space risk rather than inferring it from distances.

These are strong on image benchmarks, but each depends on a property that FMCW biometric signals in our regime do not supply. OpenMax's extreme-value fit requires enough per-class activation samples to estimate a tail, which nine identities per LOSO fold cannot support. Counterfactual synthesis requires a generator producing physiologically plausible unknown subjects—an open problem, since a generator trained on nine cardiopulmonary signatures has no basis for extrapolating to a tenth. Reciprocal point learning was developed for class counts an order of magnitude larger, and the extra-class space it models is poorly constrained when the known set is small. We regard adapting these methods to the low-identity-count, low-SNR regime as valuable and unresolved, and we do not report a poorly-adapted implementation of any of them, since such a comparison would favour our method for the wrong reason.

Within radar, the closest prior work is Shi *et al.* [36], which addresses open-set person identification directly from cardiac radar signals and is therefore the most relevant point of comparison for MVGSig. That work establishes the premise MVGSig also relies on: cardiac morphology carries identity information usable under an open-set assumption. The distinction is representational. Their pipeline extracts identity features from a single processed cardiac waveform, whereas MVGSig rests on the claim that one radar recording admits several structurally different descriptions, and that *disagreement* among those descriptions is itself a detection signal unavailable to any single-representation method. Our per-view analysis (Table 10) supports the premise of that claim: the best-classifying view is not the best-detecting view, so the two objectives are not optimised by the same representation. A direct numerical comparison is nevertheless unavailable—their evaluation uses custom-collected data under controlled breathing protocols with a different sensing configuration, and neither the data nor an implementation is public. We therefore position MVGSig relative to their approach methodologically rather than quantitatively, and observe that the absence of a shared open-set radar biometric benchmark is a real obstacle to cumulative progress in this subfield.

2.3 Multi-View Fusion and Graph-Based Methods

Across OOD detection paradigms, most methods operate on single embedding spaces. Representation diversity as a robustness axis has received limited attention.

2.3.1 Multi-View Fusion for Anomaly Detection. Fang *et al.* [7] showed that treating the time and frequency domains of a single-sensor signal as complementary views improves OOD detection in fault diagnosis relative to single-view approaches, with anomalies missed in one representation often manifesting clearly in another.

Just Accepted

IDIF [30] introduced intra-view decoupling with inter-view fusion for industrial inspection, demonstrating that explicit modeling of view-specific and shared features improves detection of subtle defects. These results suggest that view diversity benefits OOD detection, but such approaches have not been explored for radar biometrics where physiological signals exhibit distinct characteristics across spectral, temporal, and statistical domains.

2.3.2 Collaborative Learning. Beyond static fusion, collaborative training lets views supervise each other. Co-training [3] expands labelled data through iterative pseudo-labelling across complementary views. Deep Mutual Learning (DML) [54] replaces the static teacher of conventional distillation with an ensemble whose members teach each other bidirectionally via KL divergence between softmax outputs, but treats all networks symmetrically with no quality-based role assignment. Online Knowledge Distillation via Collaborative Learning (KDCL) [10] generates soft targets by ensemble aggregation and input distortion, enabling parallel training of networks with different capacities, with roles fixed by capacity rather than by per-sample reliability.

MVGSig's View-Aware Collaborative Calibration (VACC) differs in three respects. *Quality-guided role assignment*: teacher and student roles are assigned dynamically per sample from gradient-free quality estimates, rather than symmetrically (DML) or by fixed capacity (KDCL). *Feature-level alignment*: beyond soft-label distillation, lower-quality views align their intermediate representations toward the best view's embedding space via a contrastive objective. *Inverse-quality weighting*: gradient magnitudes scale inversely with quality, so weak views receive stronger supervision without degrading strong ones.

Cross-view distillation [9, 44] and deep multi-view learning [24, 52] establish that collaborative mechanisms improve robustness when individual views degrade. MVGSig extends these to physiological radar signals, where view quality varies not only across deployments but from sample to sample within a single subject.

2.3.3 Graph Neural Networks for Feature Fusion. Graph attention networks [45] enable learnable aggregation of graph-structured information. For anomaly detection, graph-based modeling improves detection of coordinated anomalies [22], and cross-view attention is key for exploiting complementary information [51].

Despite these advances, graph-based fusion of multi-view signal representations for OOD detection in radar biometrics remains unexplored.

3 Methodology

This section presents MVGSig, a multi-view graph fusion framework for open-set person identification using FMCW radar. The framework addresses two objectives: (1) classifying known subjects and (2) detecting unknown subjects as out-of-distribution. Figure 1 illustrates the architecture.

3.1 Problem Formulation

Let $\mathcal{D}_{\text{ID}} = \{(x_i, y_i)\}_{i=1}^{N}$ denote the in-distribution dataset, where $x_i \in \mathbb{R}^T$ is a radar recording of T samples and $y_i \in \{1, \ldots, C\}$ is the subject identity. At inference, inputs arrive from both known subjects ($\mathcal{D}_{\text{ID}}$) and unknown subjects ($\mathcal{D}_{\text{OOD}}$). The framework must:

- **Classification:** Predict $\hat{y} = y$ for known subjects.
- **OOD detection:** Produce score $S(x)$ such that $S(x_{\text{ID}}) < S(x_{\text{OOD}})$.

Figure 1 illustrates the complete MVGSig pipeline, which consists of four stages. First, raw FMCW radar signals (F=1200 frames, C=128 chirps) are preprocessed to extract phase ϕ_t and magnitude $|m_t|$ (Section 3.2). Second, six complementary 256-dimensional feature views are extracted and encoded by independent MLP encoders; a 2-layer GAT with learnable edge biases fuses the views through quality-aware pooling into a 32-dimensional biometric signature $\tilde{z} \in \mathbb{S}^{31}$ (Sections 3.3–3.4). Third, during training, VACC dynamically assigns teacher/student roles to views and transfers knowledge via distillation and feature alignment (Section 3.5). Fourth, at inference, eight OOD scores across four architectural levels are combined by a meta-detector MLP to produce the final ID/OOD

Just Accepted

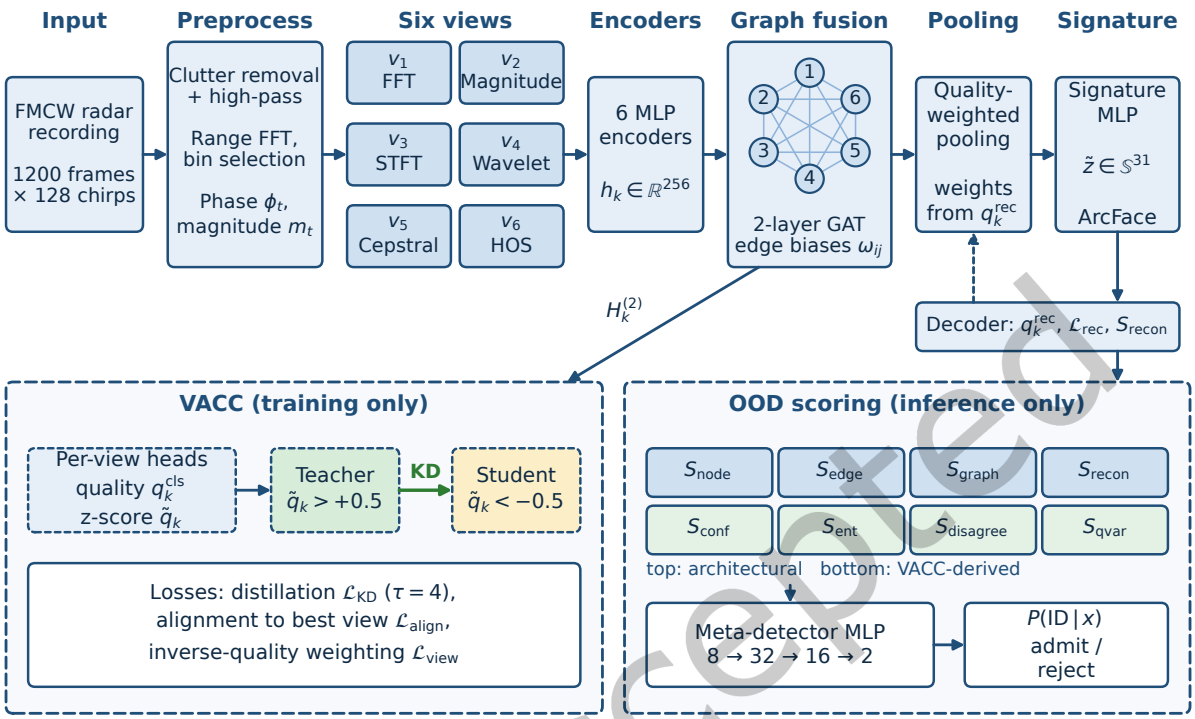


Fig. 1. MVGSig architecture. Top: shared path from radar input through six views, graph attention fusion and quality-weighted pooling to the biometric signature $\tilde{z}$; the decoder supplies reconstruction quality q_k^{rec} to pooling. Bottom left: VACC, used only during training. Bottom right: multi-level OOD scoring, used only at inference.

decision (Section 3.7). A shared-MLP decoder reconstructs all six views from $\tilde{z}$, providing both a reconstruction loss for training and reconstruction-based OOD scores at inference.

3.2 Signal Preprocessing

Raw FMCW radar data are processed to extract normalized phase and magnitude signals. The radar operates at 77–81 GHz with bandwidth $B = 4$ GHz, yielding range resolution $\Delta r = c/(2B) = 0.0375$ m.

Clutter Removal. Static reflections are removed via temporal mean subtraction followed by a high-pass filter ($f_c = 0.1$ Hz):

$$\hat{D}[f, c, s, r] = \mathrm{HPF}\left(D[f, c, s, r] - \frac{1}{F} \sum_{f'=1}^{F} D[f', c, s, r] \right). \tag{1}$$

Range Processing. A range FFT localizes the subject. The optimal range bin is selected by maximizing signal variance within a search window:

$$b_{\mathrm{opt}} = \underset{b \in [b^*-5, b^*+5]}{\arg\max} \ \mathrm{Var}\left(|\hat{R}[:, :, b, :]| \right), \quad b^* = \lfloor d/\Delta r \rceil. \tag{2}$$

Just Accepted

Phase and Magnitude Extraction. The complex signal at the selected range bin undergoes phase unwrapping, linear detrending, and z-score normalization:

$$\phi_t = \frac{\text{detrend}(\text{unwrap}(\angle \hat{s}_t)) - \mu_\phi}{\sigma_\phi + \epsilon}, \quad m_t = \frac{|\hat{s}_t| - \mu_m}{\sigma_m + \epsilon}. \tag{3}$$

The resulting signals ($T = 1200$ samples at 20 Hz) encode physiological patterns including cardiac and respiratory activity.

3.3 Multi-View Feature Extraction

We extract six 256-dimensional feature views from the preprocessed signals. This subsection first justifies the selection of views, then details how each 256-dimensional vector is constructed.

3.3.1 View Selection Rationale. The six views are chosen to span four complementary signal analysis paradigms: spectral (FFT, Cepstral), temporal (Magnitude), time-frequency (STFT, Wavelet), and statistical (HOS). This selection is guided by three criteria:

(i) Domain coverage. Physiological radar signals contain information at multiple levels: periodic harmonic structure (captured by spectral views), non-stationary temporal evolution (captured by time-frequency views), amplitude envelope dynamics (captured by the temporal view), and higher-order distributional properties such as non-Gaussianity (captured by HOS). Omitting any domain leaves a category of physiological variation unrepresented.

(ii) Structural diversity for OOD detection. The primary goal is not maximal classification accuracy per view, but maximal *complementarity* for detecting cross-view inconsistency. Unknown individuals may produce plausible patterns in one domain but anomalous patterns in another. Views that achieve low classification accuracy in isolation can still contribute substantially to OOD detection if they capture signal properties orthogonal to other views. This criterion presumes that the ranking of views by classification accuracy and by detection capability need not coincide—a presumption the per-view analysis of Section 4.5 tests directly.

(iii) Exclusion rationale for alternatives. We considered but excluded several alternative transformations. Empirical mode decomposition (EMD) and the Hilbert-Huang transform are data-adaptive methods that can introduce mode-mixing artifacts in noisy radar signals, reducing reproducibility. Mel-frequency cepstral coefficients (MFCCs), designed for speech perception, impose a perceptually motivated frequency warping that is not aligned with the physiological frequency bands of interest (0.1–2 Hz respiratory, 0.8–2 Hz cardiac). Autocorrelation features are largely subsumed by the spectral view (FFT), since the power spectral density is the Fourier transform of the autocorrelation function (Wiener-Khinchin theorem), and adding them would introduce redundancy without meaningfully expanding domain coverage.

Regarding HOS: it is the only view capturing non-Gaussian signal dynamics—properties such as asymmetric waveform morphology and intermittent bursts that distinguish individuals but are invisible to second-order (spectral/correlation) methods. Its individual performance is the weakest of the six (Section 4.5), and whether that weakness argues for its removal is examined in Section 4.7.

Table 2 summarizes the six views.

3.3.2 Feature Construction Details. Each view produces a fixed 256-dimensional feature vector. We describe the construction for each view below, specifying the components and how they are concatenated or stacked to reach 256 dimensions.

FFT View (v_1). The magnitude spectrum $|X_f| = |\text{FFT}(\phi_t)|$ is computed over $T = 1200$ samples, yielding 600 frequency bins (0–10 Hz). The spectrum is linearly resampled to 248 bins. Four spectral statistics are appended:

Just Accepted

Table 2. Six feature views extracted from radar signals.

View	Domain	Input	Captured Characteristics
v_1	Frequency (FFT)	ϕ_t	Spectral distribution, harmonics
v_2	Temporal (Magnitude)	m_t	Amplitude dynamics, envelope
v_3	Time-Frequency (STFT)	ϕ_t	Non-stationary patterns
v_4	Multi-Resolution (Wavelet)	ϕ_t	Scale-dependent structure
v_5	Cepstral (quefrency)	m_t	Harmonic spacing
v_6	Higher-Order (HOS)	ϕ_t	Non-Gaussian dynamics

centroid, spread, flatness, and entropy, yielding $248 + 4 = 252$ features. The remaining 4 dimensions are zero-padded to reach 256:

$$X = \text{FFT}(\phi_t), \quad \text{centroid} = \frac{\sum_f f \cdot |X_f|^2}{\sum_f |X_f|^2}. \tag{4}$$

Magnitude View (v_2). The magnitude signal m_t (1200 samples) is downsampled by factor 8 to 150 points. First and second derivatives are computed and downsampled identically ($150 + 150 = 300$ samples total, truncated to the first 100 derivative features: 50 from each). The Hilbert envelope is downsampled to 80 points. Sliding-window statistics (mean, variance, skewness, kurtosis) are computed over 8 non-overlapping windows of 150 samples, yielding $4 \times 8 = 32$ features. The final vector is $[150 + 50 + 50 + 80 + 32] = 362$ features, from which the first 256 are retained via truncation after importance ranking by variance.

STFT View (v_3). Multi-resolution STFT is computed at three window sizes $\{32, 64, 128\}$ on ϕ_t. For each resolution, the magnitude spectrogram is averaged across time to produce a mean spectral profile and across frequency to produce a temporal power curve. Each spectral profile and power curve is resampled to 40 points, yielding $3 \times (40 + 40) = 240$ features. Spectral flux (16 features, 4 per inter-resolution pair plus 4 global) completes the vector: $240 + 16 = 256$.

Wavelet View (v_4). A 5-level Daubechies-4 wavelet decomposition produces 6 coefficient sets $\{c_0, c_1, \ldots, c_5\}$. Per-level features include: log-energy, entropy, mean, variance, skewness, kurtosis, and zero-crossing rate ($7 \times 6 = 42$ features). The detail coefficients at each level are downsampled to 35 points and concatenated ($35 \times 6 = 210$). The final vector is $42 + 210 = 252$, zero-padded to 256:

$$\{c_0, c_1, \ldots, c_5\} = \text{DWT}(\phi_t), \quad E_\ell = \log(1 + \|c_\ell\|_2^2). \tag{5}$$

Cepstral View (v_5). The real cepstrum is computed from the magnitude signal:

$$c = \text{IFFT}(\log|\text{FFT}(m_t)|). \tag{6}$$

The cepstrum (600 quefrency bins) is truncated to the first 240 coefficients (quefrency range 0–12 s, covering respiratory and cardiac periodicities). Peak locations and amplitudes in the cardiac quefrency range (0.5–1.5 s) provide 8 additional features. Delta cepstral coefficients (first difference of the first 8 cepstral coefficients) contribute 8 more: $240 + 8 + 8 = 256$.

HOS View (v_6). The phase signal ϕ_t is segmented into 8 non-overlapping windows. Per-segment features include: skewness, kurtosis, 5^{th}- and 6^{th}-order moments ($4 \times 8 = 32$). Permutation entropy at embedding dimensions $\{3, 4, 5, 6, 7\}$ with delay 1 yields $5 \times 8 = 40$ features. Higuchi fractal dimension at $k_{\max} \in \{4, 8, 16, 32\}$ produces $4 \times 8 = 32$ features. Sample entropy at three tolerance levels $(0.1\sigma, 0.15\sigma, 0.2\sigma)$ gives $3 \times 8 = 24$ features. Bispectrum

Just Accepted

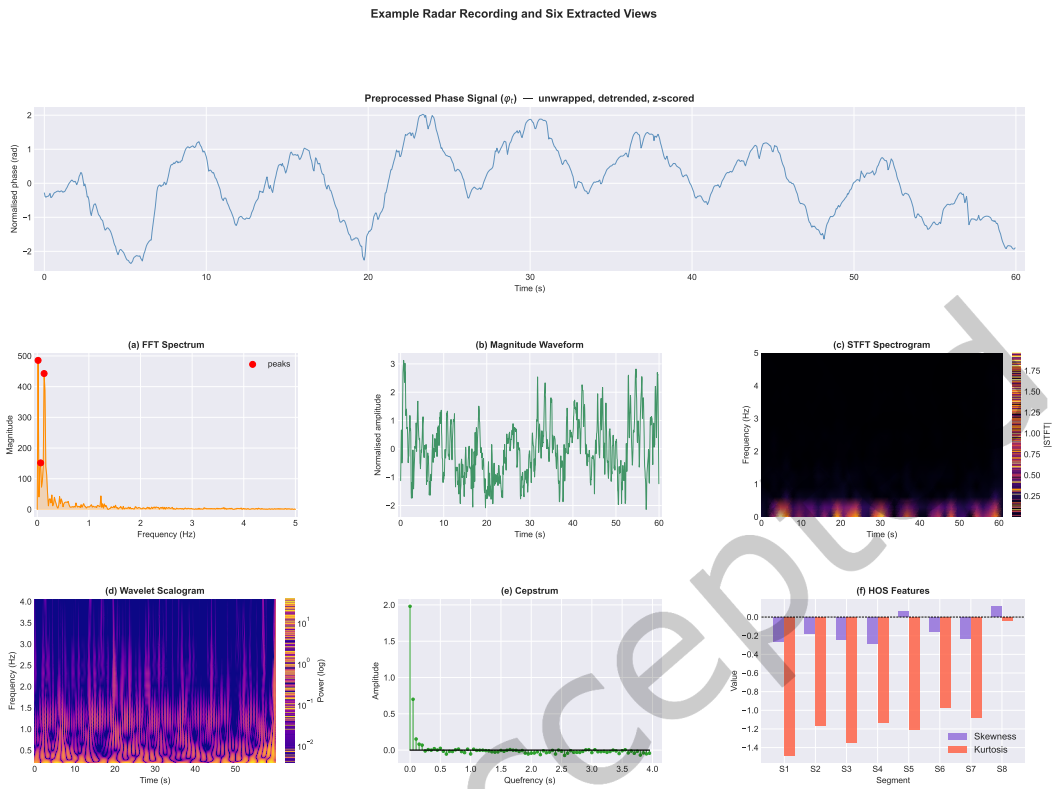


Fig. 2. Six views from a single radar recording (P1, Fold 1, 80 cm frontal): (a) FFT, (b) Magnitude, (c) STFT (window 64), (d) Wavelet scalogram, (e) Cepstrum, (f) HOS (skewness/kurtosis, 8 segments).

magnitude statistics (mean, peak, entropy) across 16 frequency pairs yield 48 features. Global signal features (Hurst exponent, DFA α, Lempel-Ziv complexity, and 8 inter-segment statistics) contribute 11 features. Lyapunov exponent estimates from 5 embedding dimensions add 5 features. The total is $32+40+32+24+48+11+5 = 192$, augmented with 64 bispectrum magnitude bins (resampled from the upper-triangular bispectral plane) to reach $192 + 64 = 256$.

Figure 2 illustrates the six views extracted from a single radar recording. The FFT spectrum (a) reveals harmonic peaks at approximately 0.3 Hz (respiratory fundamental) and 1.2 Hz (cardiac fundamental). The magnitude waveform (b) shows respiratory envelope modulation over 60 s. The STFT spectrogram (c) is displayed with a window size of 64 (one of three resolutions used); the wavelet scalogram (d) uses log-scaled power to reveal scale-dependent structure. The cepstrum (e) shows peaks in the cardiac quefrency range (0.5–1.5 s). The HOS panel (f) displays only skewness and kurtosis across 8 segments; the remaining HOS features (permutation entropy, Higuchi fractal dimension, sample entropy, bispectrum statistics) are computed but not visualized.

3.4 Graph-Based Multi-View Fusion

The six views are modeled as nodes in a fully connected graph $G = (V, E)$ with $|V| = 6$ nodes and $|E| = 30$ directed edges. Although six nodes constitute a small graph, this formulation provides two properties unavailable to simpler alternatives: (1) learnable edge biases ω_{ij} encode stable structural relationships between view pairs

Just Accepted

(e.g., FFT and Cepstral share spectral information; HOS and Wavelet both capture non-stationarity), providing domain-informed priors; and (2) input-adaptive attention weights dynamically re-weight views per sample, enabling the model to suppress a corrupted view for one input while relying on it for another. Concatenation followed by an MLP lacks per-sample view adaptivity; standard self-attention lacks the structural edge priors. Section 4.4 provides empirical validation through comparison with the raw signal baseline.

Per-View Encoding. Each view is processed by an independent MLP encoder with residual connection:

$$h_k = \mathrm{MLP}_k(v_k) + v_k \in \mathbb{R}^{256}. \tag{7}$$

Each layer applies LayerNorm, GELU activation, and dropout ($p = 0.2$).

Graph Attention. A two-layer GAT routes information across views. Learnable edge biases ω_{ij} capture view-pair relationships:

$$e_{ij} = \mathrm{LeakyReLU}\left(\mathbf{a}^\top [Wh_i \| Wh_j]\right) + \omega_{ij}, \quad \alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \neq i} \exp(e_{ik})}. \tag{8}$$

Self-attention is disabled ($\alpha_{ii} = 0$) to enforce cross-view information flow. The architecture uses 4 attention heads with 64 dimensions per head. Two GAT layers are stacked with a residual connection:

$$H^{(2)} = \mathrm{GAT}_2\left(\mathrm{GAT}_1(H^{(0)}) + H^{(0)}\right). \tag{9}$$

Quality-Weighted Pooling. Not all views are equally reliable for every sample. We use a two-pass pooling mechanism based on *reconstruction fidelity*. To distinguish it from the classification-confidence quality used by VACC (Section 3.5), we denote it q^{rec} throughout.

Pass 1: Compute reconstruction quality estimates without gradient flow:

$$q_k^{\mathrm{rec}} = \frac{1}{1 + \mathrm{MSE}(\hat{v}_k, v_k)} \quad \text{(detached)}. \tag{10}$$

Pass 2: Refine pooling using per-sample quality with softmax normalization:

$$\beta_k = \frac{\exp(q_k^{\mathrm{rec}}/\tau_q)}{\sum_{j=1}^{6} \exp(q_j^{\mathrm{rec}}/\tau_q)}, \quad g = \sum_{k=1}^{6} \beta_k H_k^{(2)}, \quad \tau_q = 0.5. \tag{11}$$

Signature Generation. The pooled representation is projected to a 32-dimensional biometric signature on the unit hypersphere:

$$z = \mathrm{SignatureMLP}(g), \quad \tilde{z} = \frac{z}{\|z\|_2} \in \mathbb{S}^{31}. \tag{12}$$

ArcFace Classification. Classification uses ArcFace angular margin for discriminative embeddings:

$$\ell_c = \begin{cases} s \cdot \cos(\theta_y + m) & \text{if } c = y \\ s \cdot \cos\theta_c & \text{otherwise} \end{cases}, \quad s = 32,\ m = 0.5. \tag{13}$$

The scale parameter $s = 32$ and margin $m = 0.5$ follow the values established for face recognition [5], which have been shown to be robust across biometric domains. The margin $m = 0.5$ enforces angular separation of approximately 29° between class boundaries, while $s = 32$ controls the concentration of the softmax distribution. Although these values were originally tuned for high-dimensional face embeddings, our 32-dimensional signature space has a similar geometry (a unit hypersphere); sensitivity to both parameters is reported in Section 5.9.

Just Accepted

Reconstruction Decoder. A decoder reconstructs all six views from the signature to ensure information retention and provide reconstruction-based OOD scores:

$$\hat{v}_k = \text{Decoder}_k(\text{SharedMLP}(\tilde{z})). \tag{14}$$

3.5 View-Aware Collaborative Calibration (VACC)

VACC transfers knowledge between views during training. Since view quality varies per sample, higher-quality views guide lower-quality ones. Figure 3 illustrates the three stages of the mechanism: per-view quality estimation, dynamic teacher/student role assignment, and knowledge transfer. Each is defined below. Soft targets are updated via an exponential moving average per epoch.

Per-View Classification. Each encoded view $H_k^{(2)}$ feeds an independent classification head with ArcFace, producing view-specific logits $l_k \in \mathbb{R}^C$.

Quality Estimation. Quality is estimated without learnable parameters. We denote it q^{cls} to distinguish it from the reconstruction quality q^{rec} of Section 3.4. During training, the true label is available and quality combines correct-class probability with the top-two confidence margin:

$$q_{n,k}^{\text{cls}} = \frac{1}{2}\left(p_{n,k}^{(y)} + (p_{n,k}^{(1)} - p_{n,k}^{(2)})\right), \tag{15}$$

where $p_{n,k}^{(y)}$ is the probability of the correct class, and $p_{n,k}^{(1)}, p_{n,k}^{(2)}$ are the top two probabilities.

At inference the label is unavailable, and Equation (15) cannot be evaluated. The correct-class term is therefore replaced by the maximum softmax probability, giving a label-free estimator of the same form:

$$q_{n,k}^{\text{cls}}\big|_{\text{infer}} = \frac{1}{2}\left(\max_c p_{n,k}^{(c)} + (p_{n,k}^{(1)} - p_{n,k}^{(2)})\right). \tag{16}$$

Both estimators are bounded in $[0, 1]$ and coincide when the prediction is correct, so the quantity entering S_{qvar} (Section 3.7) is comparable across training and inference. Quality scores are smoothed via exponential moving average ($\alpha = 0.3$).

Teacher-Student Assignment. VACC assigns roles dynamically per sample using z-score thresholding:

$$\tilde{q}_{n,k} = \frac{q_{n,k}^{\text{cls}} - \mu_n}{\sigma_n + \epsilon}, \quad \mathcal{T}_n = \{k : \tilde{q}_{n,k} > 0.5\}, \quad \mathcal{S}_n = \{k : \tilde{q}_{n,k} < -0.5\}. \tag{17}$$

Views in $\mathcal{T}_n$ serve as teachers; views in $\mathcal{S}_n$ serve as students. The threshold $|\tilde{q}| = 0.5$ corresponds to half a standard deviation from the per-sample mean quality. Because the threshold is applied to a per-sample z-score, the assignment adapts to each input: a view is labelled only when its quality departs clearly from the mean quality of that sample, which prevents the degenerate cases in which all views are teachers (no transfer occurs) or all are students (no reliable guidance). Per-view assignment frequencies, aggregated over training, are reported in Section 4.7. The threshold is not swept in Table 13; its behaviour at each end of the useful range is discussed in Section 5.9.

Knowledge Distillation. Teacher views produce soft targets that student views learn to match:

$$\bar{p}_n = \frac{\sum_{k\in\mathcal{T}_n} q_{n,k}^{\text{cls}} \cdot \text{softmax}(l_k/\tau)}{\sum_{k\in\mathcal{T}_n} q_{n,k}^{\text{cls}}}, \quad \mathcal{L}_{\text{KD}} = \tau^2 \cdot D_{\text{KL}}(\bar{p}_n \| p_{n,k}), \tag{18}$$

where $\tau = 4.0$ is the distillation temperature. The choice of $\tau = 4.0$ follows established practice in knowledge distillation [9]: higher temperatures produce softer probability distributions that reveal inter-class relationships

Just Accepted

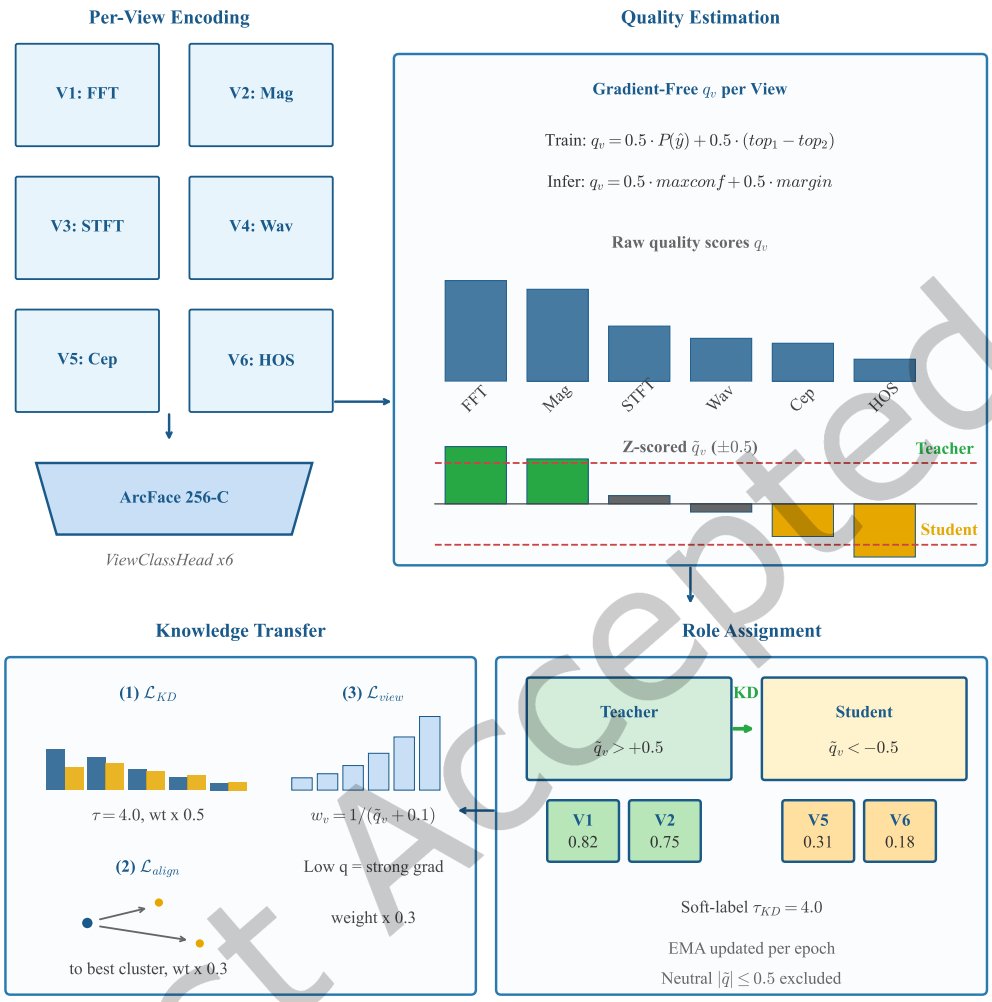


Fig. 3. VACC mechanism: (a) per-view quality estimation, (b) teacher/student role assignment via z-score thresholding, and (c) knowledge transfer through distillation, feature alignment, and inverse-quality supervision.

Just Accepted

("dark knowledge"), which is particularly valuable when the number of classes is small ($C = 9$ in each LOSO fold). With few classes, low temperatures ($\tau \leq 2$) produce near-one-hot distributions that convey little beyond the hard label; $\tau = 4.0$ ensures meaningful probability mass on non-target classes while remaining numerically stable.

Feature Alignment. Lower-quality views align their representations toward the best view:

$$k_n^* = \arg\max_k q_{n,k}^{\text{cls}}, \quad \mathcal{L}_{\text{align}} = \mathbb{E}_{n,k \neq k^*} \left[\Delta_{n,k} \cdot \text{InfoNCE}(H_{n,k}^{(2)}, H_{n,k^*}^{(2)}) \right], \tag{19}$$

where $\Delta_{n,k} = \max(0, q_{n,k^*}^{\text{cls}} - q_{n,k}^{\text{cls}})$.

Inverse-Quality Supervision. Lower-quality views receive stronger gradients:

$$w_k \propto \frac{1}{\bar{q}_k^{\text{cls}} + \epsilon}, \quad \mathcal{L}_{\text{view}} = \sum_{k=1}^{6} \tilde{w}_k \cdot \mathcal{L}_{\text{CE}}^{(k)}. \tag{20}$$

3.6 Training Objective

The training objective combines classification, embedding quality, VACC, and regularization terms:

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda_1 \mathcal{L}_{\text{UCon}} + \lambda_2 \mathcal{L}_{\text{rec}} + \lambda_3 \mathcal{L}_{\text{view}} + \lambda_4 \mathcal{L}_{\text{KD}} + \lambda_5 \mathcal{L}_{\text{align}} + \lambda_6 \mathcal{L}_{\text{reg}}. \tag{21}$$

The loss weights $\lambda_1 = 0.3$, $\lambda_2 = 0.4$, $\lambda_3 = 0.3$, $\lambda_4 = 0.5$, $\lambda_5 = 0.3$, $\lambda_6 = 0.01$ were determined through a coarse grid search on a single held-out fold (Fold 1), evaluating the composite metric $M = 0.7 \cdot \text{AUROC} + 0.3 \cdot \text{Accuracy}$ over the grid $\lambda_i \in \{0.1, 0.3, 0.5, 0.7, 1.0\}$ for the primary terms and $\{0.001, 0.01, 0.05\}$ for regularization. The selected weights were then fixed across all 10 LOSO folds without further tuning. We note that the classification loss $\mathcal{L}_{\text{CE}}$ (weight 1.0) and distillation loss $\mathcal{L}_{\text{KD}}$ (weight 0.5) receive the largest weights, reflecting the primacy of the classification and knowledge transfer objectives. These coefficients were not varied beyond that initial search; the sensitivity analysis of Section 5.9 covers the four hyperparameters that govern the proposed components most directly, and states the limits of its coverage.

Classification Loss. Cross-entropy with ArcFace logits:

$$\mathcal{L}_{\text{CE}} = -\log \frac{\exp(\ell_y)}{\sum_{c=1}^{C} \exp(\ell_c)}. \tag{22}$$

Contrastive Loss with Universum Samples. Universum [49] refers to a set of synthetic samples that do not belong to any training class but lie near class boundaries, serving as "informative non-examples" that regularize the embedding space. We generate Universum samples by interpolating biometric signatures from different classes:

$$u = \lambda \tilde{z}_i + (1 - \lambda)\tilde{z}_j, \quad y_i \neq y_j, \quad \lambda \sim \text{Uniform}(0, 0.5). \tag{23}$$

The interpolation coefficient λ is capped at 0.5 to ensure that every Universum sample lies closer to one of its parent classes than to the midpoint, creating boundary-proximate negatives rather than ambiguous centroid samples. This asymmetry is intentional: boundary samples provide stronger gradient signals for tightening class margins than equidistant interpolations, which would lie in low-density regions where gradients are weak [49].

For each mini-batch of B samples, we generate B Universum samples by randomly pairing signatures from different classes within the batch. The contrastive loss follows the supervised contrastive formulation [20], extended with Universum negatives:

$$\mathcal{L}_{\text{UCon}} = -\frac{1}{|\mathcal{P}|} \sum_{(i,j)\in\mathcal{P}} \log \frac{\exp(\tilde{z}_i \cdot \tilde{z}_j/\tau_c)}{\sum_{k\in\mathcal{N}_i\cup\mathcal{U}} \exp(\tilde{z}_i \cdot \tilde{z}_k/\tau_c)}, \tag{24}$$

where $\mathcal{P}$ is the set of positive (same-class) pairs, $\mathcal{N}_i$ is the set of negative (different-class) samples, $\mathcal{U}$ is the set of Universum samples, and $\tau_c = 0.1$ is the contrastive temperature.

Reconstruction Loss. Quality-weighted MSE with random masking:

$$\mathcal{L}_{\text{rec}} = \frac{1}{6} \sum_{k=1}^{6} \tilde{\gamma}_k \cdot \text{MSE}(\hat{v}_k, v_k). \tag{25}$$

Regularization. Three terms prevent degenerate solutions: edge symmetry (encouraging $\omega_{ij} \approx \omega_{ji}$), magnitude penalty (preventing unbounded edge biases), and view diversity (penalizing collapse of different views to similar representations).

Just Accepted

3.7 Multi-Level OOD Detection

MVGSig extracts eight OOD scores across four architectural levels, along with VACC-derived consistency metrics. Different anomaly types manifest at different processing stages.

3.7.1 *Architectural Scores.*

Node Score: Average per-view Mahalanobis distance to class distributions:

$$S_{\text{node}} = \frac{1}{6}\sum_{k=1}^{6} \min_{c} \sqrt{(h_k - \mu_{k,c})^\top \Sigma_{k,c}^{-1}(h_k - \mu_{k,c})}. \tag{26}$$

Edge Score: Attention pattern deviation from training distribution:

$$S_{\text{edge}} = \sqrt{(\alpha_{\text{flat}} - \bar{\alpha})^\top \Sigma_{\alpha}^{-1}(\alpha_{\text{flat}} - \bar{\alpha})}. \tag{27}$$

Graph Score: Signature-space Mahalanobis distance to class prototypes:

$$S_{\text{graph}} = \min_{c} \sqrt{(\tilde{z} - p_c)^\top \Sigma_{c}^{-1}(\tilde{z} - p_c)}. \tag{28}$$

Reconstruction Score: Average per-view reconstruction error:

$$S_{\text{recon}} = \frac{1}{6}\sum_{k=1}^{6} \frac{\|\hat{v}_k - v_k\|_2}{\sqrt{256}}. \tag{29}$$

3.7.2 *VACC-Derived Scores.*

Four scores leverage cross-view consistency. The quality variance score uses the inference-time classification quality of Equation (16):

$$S_{\text{conf}} = 1 - \min_{k} \max_{c} \ \text{softmax}(l_k)_c, \tag{30}$$

$$S_{\text{ent}} = \max_{k} H(\text{softmax}(l_k)), \tag{31}$$

$$S_{\text{disagree}} = |\{k : \hat{y}_k \neq \text{mode}(\hat{y}_{1:6})\}|, \tag{32}$$

$$S_{\text{qvar}} = \text{Var}\big(q_1^{\text{cls}}, \ldots, q_6^{\text{cls}}\big). \tag{33}$$

3.7.3 *Meta-Detector.*

A learned MLP combines all eight scores:

$$P(\text{ID}|x) = f_{\text{meta}}([\tilde{S}_{\text{node}}, \tilde{S}_{\text{edge}}, \tilde{S}_{\text{graph}}, \tilde{S}_{\text{recon}}, \tilde{S}_{\text{conf}}, \tilde{S}_{\text{ent}}, \tilde{S}_{\text{disagree}}, \tilde{S}_{\text{qvar}}]), \tag{34}$$

where $\tilde{S}$ denotes z-score standardization. The meta-detector architecture is $8 \to 32 \to 16 \to 2$ with GELU activations. The final OOD score is $S(x) = 1 - P(\text{ID})$.

Meta-Detector Training Protocol. Training the meta-detector requires both ID and OOD samples, but in the LOSO protocol the actual OOD subject is unseen. We address this through a *nested leave-one-out* strategy within the training subjects of each fold. Specifically, in each LOSO fold where subject p is held out as OOD:

(1) The main model (encoders, GAT, VACC) is trained on subjects $\{1, \ldots, 10\} \setminus \{p\}$, split 70/15/15 at the session level for train/validation/test sets.
(2) To generate *surrogate OOD samples* for meta-detector training, we perform a secondary inner loop: for each training subject $q \in \{1, \ldots, 10\} \setminus \{p\}$, we temporarily treat q as OOD and compute the eight scores on q's validation samples using the model trained on the remaining $C - 2$ subjects. This produces surrogate OOD score vectors that approximate the distribution of scores for a truly unseen subject.
(3) The meta-detector MLP is then trained on the combined ID scores (from training subjects' validation data) and surrogate OOD scores (from the inner loop), using binary cross-entropy loss.

Just Accepted

(4) The meta-detector is evaluated on subject p's test data (the actual OOD subject), which was never used during any training stage.

This nested protocol ensures that the meta-detector is trained exclusively on data from subjects disjoint from the test OOD subject, preventing information leakage. The surrogate OOD subjects approximate the score distribution of truly unseen individuals because they were excluded from the model used to compute their scores. We acknowledge that the surrogate distribution may not perfectly match the true OOD distribution, particularly for subjects with unusual physiological profiles (e.g., P10; see Section 5.6), but this is inherent to any LOSO evaluation with limited participants.

3.7.4 Decision Convention and Metric Definitions. Because the score and the operating metric run in opposite directions, we state the convention once. The score $S(x)$ is oriented so that larger values indicate a more likely unknown subject, $S(x_{\text{ID}}) < S(x_{\text{OOD}})$. The decision, however, is an admission decision, and the ROC analysis therefore treats the *enrolled* subject as the positive class and ranks samples by $P(\text{ID}|x) = 1 - S(x)$. Under this convention:

- **TPR** is the fraction of enrolled subjects correctly admitted.
- **FPR** is the fraction of unknown subjects incorrectly admitted.
- **FPR@95** is the FPR at the threshold that yields TPR = 0.95; it is the false accept rate at a fixed, operationally acceptable rejection rate for legitimate users, and lower is better.
- **AUROC** is computed with the same orientation, so a value of 1.0 means every enrolled sample is ranked above every unknown sample.

All reported FPR@95 and AUROC values follow this convention. Classification accuracy is a separate, closed-set quantity measured only on enrolled subjects.

3.7.5 Two Scoring Regimes. The evaluation reports results under two deliberately different scoring regimes, and the distinction governs which comparisons in Section 4 are like-for-like.

The *full* regime uses the multi-level meta-detector defined above, applied to MVGSig, to the raw signal baseline, and to every ablated configuration in Tables 5 and 9. The meta-detector architecture, its training protocol, and its input standardisation are identical across these configurations; what differs is how many scores each configuration can supply.

This distinction matters and we state it explicitly. Three of the eight scores are cross-view quantities: S_{edge} requires an attention matrix, S_{disagree} requires several per-view predictions, and S_{qvar} requires several per-view quality estimates. A single-representation configuration such as the raw signal baseline has no second view, so these three are undefined for it and its meta-detector is trained on the five that remain (S_{node}, S_{graph}, S_{recon}, and the degenerate single-view forms of S_{conf} and S_{ent}). The regime is therefore matched in mechanism rather than in the size of the score vector.

We regard this as the correct comparison rather than a confound. The three additional scores are not an independent advantage in scoring methodology that could equally have been granted to the baseline; they exist only because several views exist. Attributing them to representation diversity is therefore consistent with what the ablation is measuring. The alternative—withholding them from MVGSig to equalise the vector length—would understate multi-view fusion by discarding precisely the signal it generates.

The *isolated* regime uses only view-specific Mahalanobis distance. It is applied in the per-view analysis (Section 4.5) for a different purpose: to characterise the intrinsic discriminative capacity of each signal transformation without confounding it with scoring sophistication. Under this regime a single view has no cross-view signal to exploit, so S_{edge}, S_{disagree} and S_{qvar} are undefined by construction.

Just Accepted

Numbers from the two regimes therefore answer different questions and are not interchangeable. Where we quantify the contribution of a component, we compare within the full regime. Where we report a per-view number, we state the isolated regime explicitly.

3.8 Implementation Details

Architecture. 4-head GAT with 64 dimensions per head, 32-dimensional signatures, dropout 0.2 in encoders and 0.3 in signature MLP. Total parameters: approximately 2.3M.

Training. AdamW optimizer with learning rate 5×10^{-4}, weight decay 0.1, cosine annealing over 50 epochs. Gradient clipping at 1.0. Early stopping monitors $M = 0.7 \cdot \text{AUROC} + 0.3 \cdot \text{Accuracy}$ with patience 15.

Evaluation. Leave-One-Subject-Out (LOSO) across 10 participants. Data split: 70%/15%/15% for train/validation/test at session level.

Computation. Training: ∼1 minute per fold on RTX 4080. Inference: 5.2 ms per sample (192 samples/second).

Hyperparameter Robustness. Sensitivity to the principal hyperparameters is reported in Section 5.9.

4 Experimental Results

This section evaluates MVGSig through Leave-One-Subject-Out (LOSO) cross-validation across 10 participants. Following Section 3.7.5, all system-level comparisons use the full multi-level scoring regime; the per-view analysis of Section 4.5 uses the isolated regime and is labelled as such. All experiments are conducted on an NVIDIA RTX 4080 GPU.

4.1 Dataset

We evaluate MVGSig on a public mm-wave FMCW radar dataset for vital sign monitoring [35], repurposed here for person identification.

Data were collected using a Texas Instruments AWR1642BOOST module operating at 77–81 GHz with 4 GHz bandwidth. The radar employs 2 transmit and 4 receive antennas. Each recording contains 1200 frames (60 seconds) captured at 20 Hz. Ten adults participated (5 male, 5 female; mean age 30.2 ± 6.3 years).

Data collection followed four scenarios: distance (40, 80, 120, and 160 cm), orientation (front, back, left, and right), angle (0°, 30°, 45°), and elevated heart rate (post-exercise). For each condition, four 60-second recordings were captured per participant. We use recordings from the 80 cm frontal-facing condition for consistent signal quality, yielding approximately 400 samples for LOSO evaluation. The remaining distance, orientation and angle conditions are not used, so no claim is made about performance under them; the implications of this restriction are discussed in Section 5.9.

The dataset was collected under ethics approval from the University of Twente and is publicly available under a CC BY-NC-SA 4.0 license [35]. It was designed for vital sign estimation rather than person identification, which makes closed-set classification difficult; Section 4.2 explains why we regard that as the appropriate setting for this work rather than a limitation of it.

4.2 Overall Performance

Table 3 presents per-fold results for the complete MVGSig framework. We report FPR@95 first, since in access control the consequential error is admitting an unknown individual, and a system's usefulness is set by how often that happens at an acceptable true positive rate.

False Accept Rate. Mean FPR@95 is 0.375 ± 0.103. Per-fold values span 0.20 (P7) to 0.61 (P10), with eight of ten folds at or below 0.45. The spread matters more than the mean for deployment: a system whose false accept rate

Just Accepted

Table 3. Per-fold LOSO results for MVGSig (6-view + VACC). **Bold**: best value in each column; underline: worst.

Fold	OOD Subject	FPR@95↓	AUROC↑	Accuracy (%)↑
1	P1	0.38	0.89	51.00
2	P2	0.33	0.92	58.85
3	P3	0.45	0.89	76.30
4	P4	0.43	0.90	71.92
5	P5	0.31	0.92	75.85
6	P6	0.35	0.90	**78.90**
7	P7	**0.20**	**0.93**	58.85
8	P8	0.31	0.91	69.31
9	P9	0.38	0.88	70.62
10	P10	0.61	0.82	53.64
Mean ± Std		0.375 ± 0.103	0.896 ± 0.031	66.5 ± 10.1

varies by a factor of three depending on which individual presents cannot be characterised by a single operating point, and the worst-case fold rather than the average determines the security guarantee. Fold 10 is analysed in Section 5.6.

Ranking Quality. Mean AUROC is 0.896 ± 0.031, exceeding 0.85 in nine of ten folds, with an interquartile range of 0.89–0.92 and a coefficient of variation of 3.5%. AUROC is reported as a threshold-free summary of separability; it is not the operational metric, and Section 5.1 explains why modest AUROC differences correspond to large FPR@95 differences in this setting.

Closed-Set Accuracy. Mean classification accuracy is 66.5 ± 10.1% against a chance rate of 11.1% for nine identities. We do not present this as a strong identification result. The dataset was collected for vital-sign estimation [35], not identification, and carries no identity-optimised protocol; the resulting signals are low in identity discriminability by construction. That is why we selected it. A closed-set classifier achieving high accuracy on a curated identification dataset is not being tested on the failure mode this paper addresses, whereas a classifier operating near the limit of what the signal supports is precisely where overconfident assignment of unknown individuals to enrolled identities becomes dangerous. The accuracy figure should therefore be read as characterising the regime, not as a competitive result, and we would expect it to degrade further as the enrolled identity set grows beyond nine.

Dissociation Between the Two. Accuracy is markedly more variable than AUROC (CV = 15.2% versus 3.5%), and the two do not move together across folds: Fold 1 combines the lowest accuracy (51.0%) with strong detection (AUROC = 0.89), while Fold 6 combines the highest accuracy (78.9%) with AUROC of 0.90. This is not noise but a geometric property of the two objectives, examined in Section 5.2. Its practical consequence is that identification accuracy is not a usable proxy for open-set capability, and the two should be reported and optimised independently.

4.3 Comparison with Baseline Methods

To isolate the contribution of the multi-level meta-detector (C3), all baseline OOD detection methods operate on the same fused 32-dimensional signature $\tilde{z}$ produced by MVGSig's graph attention fusion. Table 4 reports results across identical LOSO splits.

Just Accepted

Table 4. Comparison with baseline OOD methods. All post-hoc methods use the same fused 32-D signature, isolating scoring strategy from representation quality. Mean across 10 LOSO folds.

Method	FPR@95↓	AUROC↑	Accuracy (%)↑
Post-hoc scoring on fused signature $\tilde{z}$			
One-Class SVM	0.902	0.548	–
Isolation Forest	0.891	0.571	–
KNN Distance (k=5)	0.823	0.643	–
Energy Score [26]	0.712	0.687	–
Mahalanobis Distance [23]	0.856	0.694	–
Multi-level scoring (ours)			
MVGSig Meta-Detector	**0.375**	**0.896**	66.5

Table 5. Ablation study: component contributions, all under the full scoring regime. ΔFPR shows absolute reduction from the raw signal baseline; Relative shows cumulative percentage of false accepts eliminated.

Configuration	FPR@95↓	Relative	AUROC	Acc (%)	ΔAcc
Raw signal baseline	0.697	–	0.846	52.8	–
+ Multi-view fusion	0.552	20.8%↓	0.871	59.2	+6.4
+ VACC (MVGSig)	**0.375**	**46.2%↓**	**0.896**	**66.5**	+13.7

FPR@95 is reduced from 0.712 (Energy Score, the strongest baseline on this metric) to 0.375, admitting 47.3% fewer unknown subjects at the same 95% true positive rate. AUROC improves by +0.202 over the best baseline (Mahalanobis, 0.694). All baselines receive the identical fused representation, confirming that the performance gap arises from multi-level scoring rather than representation advantage.

4.4 Ablation Study

To quantify the contribution of each proposed component, we evaluate three configurations: (1) raw signal baseline, (2) 6-view graph fusion without VACC, and (3) complete MVGSig with VACC. All three use the full scoring regime (Section 3.7.5), so differences between them are attributable to representation and training rather than to scoring.

Baseline Definition. The raw signal baseline operates directly on the preprocessed phase signal $\phi_t \in \mathbb{R}^T$ without any domain transformation. It applies a single MLP encoder to the normalized phase sequence, followed by the same signature generation and ArcFace classification pipeline used in the full model, and the same multi-level meta-detector. As set out in Section 3.7.5, the baseline supplies the five scores that a single representation defines; the three cross-view scores have no single-view analogue. The comparison therefore holds the scoring *mechanism* fixed and lets the representation determine what that mechanism has to work with, which is what the ablation is intended to measure.

Component Contributions. Table 5 summarizes the incremental improvements. Figure 4 visualizes the three metrics across configurations. Error bars indicate ±1 standard deviation across 10 LOSO folds.

Just Accepted

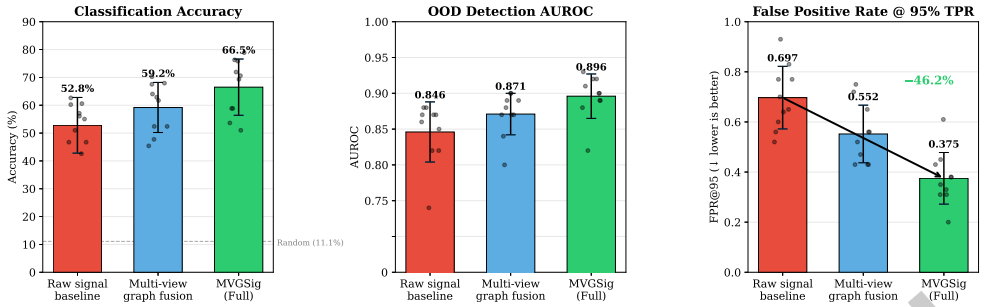


Fig. 4. Ablation across three configurations. Left: classification accuracy (dashed: random chance 11.1%). Center: AUROC (y-axis starts at 0.75). Right: FPR@95; annotation shows the 46.2% cumulative reduction. Error bars: ±1 std across 10 LOSO folds.

Table 6. AUROC across LOSO folds for each configuration. Best per fold in **bold**.

Configuration	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	Std
Raw signal baseline	0.82	0.88	0.86	0.82	0.85	0.87	0.88	0.87	0.87	0.74	0.846	0.042
6-view fusion	0.84	0.90	0.87	0.89	0.88	0.87	0.89	0.90	0.87	0.80	0.871	0.029
MVGSig (Full)	**0.89**	**0.92**	**0.89**	**0.90**	**0.92**	**0.90**	**0.93**	**0.91**	**0.88**	**0.82**	**0.896**	0.031

Table 7. FPR@95 across LOSO folds (lower is better). Best per fold in **bold**.

Configuration	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	Std
Raw signal baseline	0.93	0.60	0.83	0.70	0.77	0.64	0.56	0.65	0.52	0.77	0.697	0.125
6-view fusion	0.56	0.52	0.72	0.56	0.65	0.43	0.43	0.43	0.47	0.75	0.552	0.115
MVGSig (Full)	**0.38**	**0.33**	**0.45**	**0.43**	**0.31**	**0.35**	**0.20**	**0.31**	**0.38**	**0.61**	**0.375**	0.103

Multi-View Fusion (C1). Replacing the raw signal with six complementary domain-transformed views reduces FPR@95 by 0.145 (0.697 → 0.552), eliminating 20.8% of false accepts, and improves AUROC by +0.025 (0.846 → 0.871) and accuracy by +6.4% (52.8% → 59.2%).

VACC (C2). Adding View-Aware Collaborative Calibration yields a further −0.177 FPR@95 (0.552 → 0.375), +0.025 AUROC (0.871 → 0.896), and +7.3% accuracy (59.2% → 66.5%). The FPR@95 reduction from VACC eliminates 32.1% of remaining false accepts, exceeding the 20.8% from multi-view fusion despite an identical AUROC gain—evidence that dynamic knowledge transfer sharpens the score distributions rather than merely shifting them (Section 5.1).

Cumulative Effect. The complete MVGSig achieves −0.322 FPR@95, +0.050 AUROC, and +13.7% accuracy over the raw signal baseline. The full model admits only 37.5% of unknown subjects at 95% TPR, compared to 69.7% for the baseline—a relative reduction of 46.2%.

Per-Fold Ablation. Tables 6 and 7 present AUROC and FPR@95 across all configurations and folds. MVGSig outperforms the raw signal baseline across all 10 folds for both metrics, and it outperforms 6-view fusion in 10/10 folds for both AUROC and FPR@95.

Just Accepted

Table 8. Statistical significance of component contributions. Paired *t*-test and Wilcoxon signed-rank test across 10 LOSO folds.

Comparison	Metric	*t*-test	Wilcoxon
6-view vs. Raw signal	FPR@95	$p = 0.002$	$p = 0.004$
	AUROC	$p = 0.008$	$p = 0.010$
	Accuracy	$p = 0.003$	$p = 0.005$
MVGSig vs. 6-view	FPR@95	$p = 0.001$	$p = 0.002$
	AUROC	$p = 0.005$	$p = 0.008$
	Accuracy	$p = 0.002$	$p = 0.004$
MVGSig vs. Raw signal	FPR@95	$p < 0.001$	$p = 0.002$
	AUROC	$p < 0.001$	$p = 0.002$
	Accuracy	$p < 0.001$	$p = 0.002$

Table 9. Finer-grained ablation, all configurations under the full scoring regime. Configurations B and C correspond to rows 2 and 3 of Table 5; D–F are additional configurations. Mean across 10 LOSO folds.

#	Configuration	FPR@95↓	ΔFPR	AUROC	Acc (%)
A	Raw signal baseline	0.697	–	0.846	52.8
	Fusion mechanism comparison (without VACC)				
D	6-view + Concatenation	0.625	−0.072	0.856	55.4
E	6-view + Self-Attention (no ω_{ij})	0.580	−0.117	0.866	57.8
B	6-view + GAT (with ω_{ij})	0.552	−0.145	0.871	59.2
	VACC component isolation				
F	MVGSig w/o quality pooling	0.418	−0.279	0.888	64.2
C	**MVGSig Full**	**0.375**	**−0.322**	**0.896**	**66.5**

Statistical Significance. Table 8 confirms that all improvements are statistically significant under both parametric (paired *t*-test) and non-parametric (Wilcoxon signed-rank) tests across the 10 LOSO folds.

4.4.1 Finer-Grained Component Ablation. Tables 5–8 evaluate multi-view fusion and VACC as monolithic blocks. To isolate individual architectural choices within each block, Table 9 presents a finer-grained ablation. Each configuration modifies exactly one component while keeping all other settings identical (same data splits, random seeds, training schedule, and multi-level OOD scoring pipeline).

Fusion Mechanism (D → E → B). Replacing GAT with simple concatenation (row D: six 256-D vectors concatenated into a 1536-D vector, processed by a two-layer MLP) increases FPR@95 by 0.073 and reduces AUROC by 0.015 compared to GAT (row B). Standard multi-head self-attention without learnable edge biases (row E) recovers most of the GAT performance, indicating that input-adaptive view weighting is the primary mechanism and that edge biases provide a modest additional benefit through structural priors. The progressive improvement D → E → B confirms that attention-based fusion outperforms static concatenation when view quality varies per sample.

Just Accepted

Table 10. Per-view performance (mean across LOSO folds), computed under the *isolated* scoring regime: AUROC uses view-specific Mahalanobis distance only. These numbers characterise each transformation's intrinsic discriminative capacity and are not directly comparable to the multi-level scores reported elsewhere.

View	Accuracy (%)	AUROC	Acc Rank	AUROC Rank
FFT	62.5	0.584	1	4
STFT	52.1	0.610	2	1
Wavelet	50.0	0.534	3	5
Cepstral	47.7	0.590	4	3
Magnitude	46.8	0.610	5	1
HOS	43.5	0.531	6	6
MVGSig (Fused, full regime)	**66.5**	**0.896**	–	–

Quality-Weighted Pooling ($F \rightarrow C$). Replacing quality-weighted pooling (Eq. 11) with uniform mean pooling ($\beta_k = 1/6$) while retaining all other VACC components (row F) degrades FPR@95 by 0.043 (0.375 → 0.418) and AUROC by 0.008 (0.896 → 0.888). Quality pooling therefore contributes to OOD rejection by suppressing unreliable views at inference—a known subject produces consistent quality across views, while an unknown subject exhibits variable quality that uniform pooling cannot exploit.

4.5 Per-View Analysis

Table 10 reports individual view performance under the *isolated* scoring regime of Section 3.7.5: each view is evaluated independently using only view-specific Mahalanobis distance, so that view quality is not confounded with scoring sophistication.

The best classifier (FFT, 62.5%) is *not* the best OOD detector (STFT/Magnitude, AUROC = 0.610). This ranking reversal reflects the geometric distinction between closed-set and open-set objectives: views that discriminate well among known subjects do not necessarily produce embeddings that separate known from unknown subjects. It confirms the presumption on which criterion (ii) of Section 3.3.1 rests, and it has a direct consequence for fusion design. Under a compatibility-based mechanism that weights views by pairwise agreement, views with weak classification signals would receive systematically low attention weights—yet these are precisely the views that contribute most to OOD detection. The content-based attention mechanism in MVGSig evaluates each view on its own discriminative content, enabling sample-adaptive weighting that a compatibility-based scheme cannot express.

What the +0.286 *Gap Does and Does Not Show.* The fused system reaches AUROC = 0.896 against 0.610 for the best single view, a gap of +0.286. We state plainly what this figure measures. It is an *end-to-end* comparison between the complete MVGSig system and a single transformed view scored by Mahalanobis distance alone—that is, a comparison *across* the two scoring regimes of Section 3.7.5. It therefore reflects the joint effect of representation diversity, graph fusion, collaborative training, and multi-level scoring, and it should be read as the practical margin a deployed system gains over a single-representation, single-score detector. It should *not* be read as the contribution of multi-view representation alone.

The matched-condition figure is the one reported in Table 5. There, the raw signal baseline and full MVGSig use the same multi-level scoring mechanism, matched in the sense set out in Section 3.7.5, and the difference between them—+0.050 AUROC and −0.322 FPR@95—is attributable to representation diversity and collaborative learning. Table 9 decomposes this further: holding both scoring and training fixed and varying only the fusion

Just Accepted

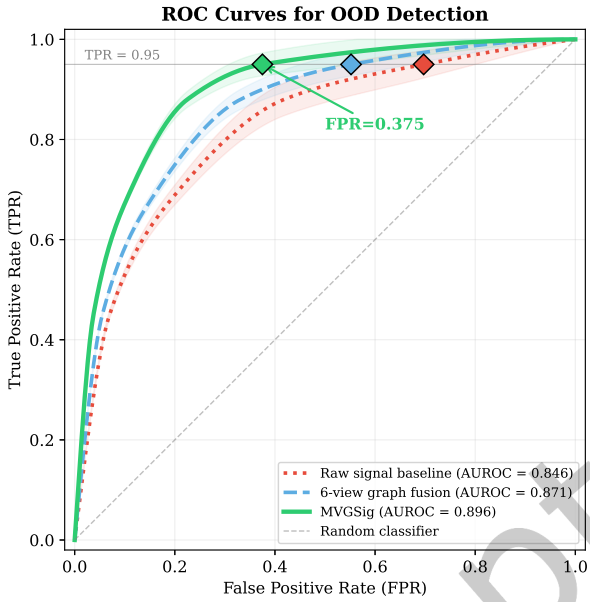


Fig. 5. ROC curves (mean, 10 LOSO folds). Baseline: AUROC = 0.846; 6-view fusion: 0.871; MVGSig: 0.896. Diamond marks TPR = 0.95 operating point (FPR = 0.375).

mechanism (rows D→E→B) isolates the fusion contribution at +0.015 AUROC and −0.073 FPR@95 for graph attention over concatenation.

We report all three figures because they answer different questions, and we attribute none of them to a mechanism it does not isolate. The same reasoning explains why the raw signal baseline (AUROC = 0.846, full regime) scores higher than any individual transformed view (best 0.610, isolated regime): the comparison spans two regimes and measures scoring sophistication as much as representation quality.

4.6 OOD Score Analysis

ROC Curve Comparison. Figure 5 compares ROC curves for the three configurations (mean across 10 LOSO folds), demonstrating progressive improvement from raw signal baseline to MVGSig across all operating points. The three curves correspond to raw signal baseline (dotted, AUROC = 0.846), 6-view graph fusion (dashed, AUROC = 0.871), and full MVGSig (solid, AUROC = 0.896). At the TPR = 0.95 operating point, MVGSig achieves FPR = 0.375, marked by the diamond on the curve.

Score Distribution Separation. Figure 6 illustrates the OOD score distributions for ID (known) and OOD (unknown) samples across the three configurations, aggregated across all 10 LOSO folds. Separation is quantified by the *score gap*, defined as gap = median(S_{OOD}) − median(S_{ID}). MVGSig produces the largest score gap (0.91), compared to 6-view fusion (0.72) and raw signal baseline (0.54). The dashed vertical lines indicate the decision threshold at TPR = 95%. The progressive sharpening of both distributions—not merely a shift in means—explains the disproportionate FPR@95 reduction relative to the modest AUROC improvement. Because distributions are aggregated across folds, the apparent separation is greater than for any individual fold; per-fold variation is consistent with Table 7.

Just Accepted

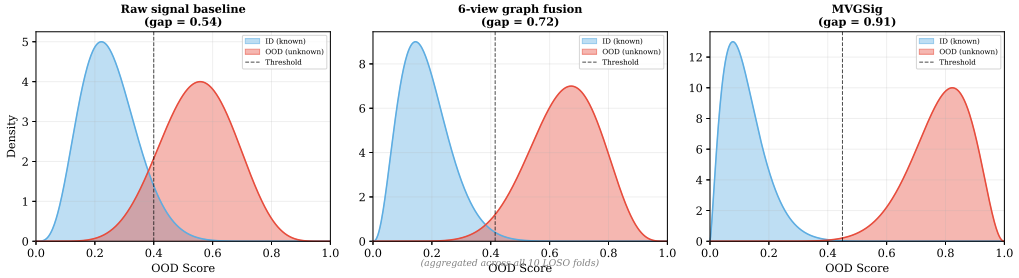


Fig. 6. OOD score distributions (aggregated, 10 LOSO folds). Score gap = median(S_{OOD}) − median(S_{ID}). Dashed: threshold at TPR = 95%.

Table 11. Individual OOD score performance and meta-detector fusion. Mean across 10 LOSO folds.

Score	Level	FPR@95↓	AUROC↑
S_{node}	Node (per-view)	0.523	0.812
S_{edge}	Edge (attention)	0.612	0.754
S_{graph}	Graph (signature)	0.478	0.847
S_{recon}	Reconstruction	0.634	0.723
S_{conf}	VACC (confidence)	0.545	0.789
S_{ent}	VACC (entropy)	0.578	0.768
S_{disagree}	VACC (disagreement)	0.623	0.734
S_{qvar}	VACC (quality var.)	0.656	0.701
Meta-Detector (All 8)	Fusion	**0.375**	**0.896**

Multi-Level Score Analysis. Table 11 reports the individual contribution of each OOD score component.

The best individual score (S_{graph}, AUROC = 0.847, FPR@95 = 0.478) is outperformed by the meta-detector fusion (+0.049 AUROC, −0.103 FPR@95), confirming that different anomaly types manifest at different architectural levels and that the multi-level design captures complementary signals.

4.7 VACC Mechanism Analysis

Figure 7 visualizes two aspects of the VACC mechanism across all training samples from all 10 LOSO folds. The left panel shows the teacher/neutral/student assignment frequency per view: teacher ($\tilde{q} > +0.5$), neutral ($|\tilde{q}| \leq 0.5$), and student ($\tilde{q} < -0.5$). The right panel shows per-view accuracy improvement from adding VACC, measured relative to the same view's accuracy in the 6-view graph fusion configuration without VACC (Table 5, row 2).

Teacher-Student Distribution. The assignment pattern follows view strength: FFT and STFT are most frequently assigned as teachers, while HOS and Magnitude are most frequently assigned as students, and every view spends a substantial share of samples in the neutral band. A large neutral share does not indicate VACC inactivity—because roles are assigned from a per-sample z-score, on each sample some views necessarily fall above and below the threshold. What it reflects is that the mechanism transfers knowledge only when quality differences among views are statistically clear, avoiding noisy transfers when views perform comparably. The ordering is what matters for

Just Accepted

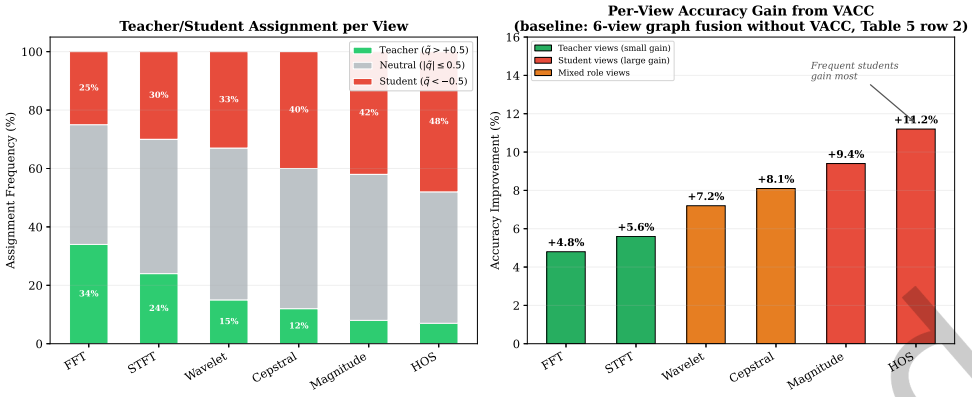


Fig. 7. VACC teacher/student analysis (all 10 LOSO folds). Left: role assignment frequency per view. Right: per-view accuracy gain from VACC vs. 6-view fusion without VACC (Table 5, row 2). Student views (HOS, Magnitude) gain most; teacher views also improve, confirming no negative transfer.

Table 12. Computational complexity of MVGSig.

Category	Metric	Value
Model Size	Total parameters	2.32M
	Memory footprint	78.5 MB
Inference	Feature extraction	3.0 ms/sample
	Neural forward pass	2.2 ms/sample
Training	Time per fold	~1 min
Hardware	GPU	RTX 4080

the mechanism's operation: views that are weaker on average are more often students, which is the precondition for the asymmetric gains reported below.

Knowledge Transfer Effect. Views frequently assigned as students show the largest accuracy improvements from VACC: HOS improves by +11.2% and Magnitude by +9.4%, while teacher views (FFT, STFT) show smaller but positive improvements (+4.8% and +5.6%). VACC therefore transfers knowledge from strong to weak views without degrading strong view performance—the detached gradient computation in soft target generation prevents negative transfer. This also settles the question raised in Section 3.3.1: HOS, the weakest view in isolation, benefits most from collaborative training, indicating that its representation is highly complementary to the others rather than simply uninformative.

4.8 Computational Efficiency

Table 12 reports computational requirements. The total inference latency of 5.2 ms per sample is well within the requirements of the application scenario. Since the radar operates at 20 Hz, computational speed is not a deployment bottleneck; we report latency for completeness rather than claiming real-time processing as a contribution.

Just Accepted

5 Discussion

This section interprets the experimental findings, addresses methodological considerations, and identifies limitations.

5.1 Interpreting the AUROC–FPR Relationship

The cumulative AUROC improvement from multi-view fusion and VACC is +0.050 (0.846 → 0.896), which may appear modest in isolation. However, this improvement corresponds to a 46.2% reduction in false accepts (FPR@95: 0.697 → 0.375). This amplification occurs because FPR@95 depends on the *tail overlap* of the ID and OOD score distributions, not their mean separation. As shown in Figure 6, each component progressively sharpens the score distributions (score gap: 0.54 → 0.72 → 0.91), reducing the overlap region where ID and OOD samples are confused. Small shifts in mean separation produce large changes in tail overlap when distributions are concentrated. For access control applications, the FPR@95 reduction—not the AUROC delta—is the operationally meaningful metric, which is why we report it first throughout Section 4.

5.2 Classification vs. OOD Detection: Geometric Dissociation

The dissociation between classification accuracy and OOD detection AUROC reported in Section 4.2 reflects a fundamental geometric property: closed-set classification measures intra-class compactness relative to other ID classes, while OOD detection measures the compactness of the *entire* ID manifold relative to the ambient embedding space. These are geometrically distinct properties.

Fold 1 illustrates this: P2–P10 (the 9 ID subjects) have overlapping signatures, producing low classification accuracy. However, P1's signature is geometrically isolated from all of P2–P10, making OOD detection straightforward. This finding has a practical implication: **classification accuracy is not a reliable proxy for OOD detection capability**, and the two metrics should be reported and optimized independently.

Figure 8 visualizes these geometric properties via t-SNE projections of the biometric signatures $\tilde{z} \in \mathbb{S}^{31}$. Three representative folds are shown: a well-separated case (Fold 3, AUROC = 0.89), a moderate case (Fold 4, AUROC = 0.90), and the challenging case (Fold 10, AUROC = 0.82). The OOD subject in each fold is marked with distinct red cross markers; the dashed boundary indicates the convex hull of all ID subject embeddings.

In Fold 3 (panel a), the OOD subject (P3) projects partially outside the ID convex hull, consistent with its moderate FPR@95 of 0.45. In Fold 4 (panel b), the OOD subject (P4) shows similar partial separation. Critically, in Fold 10 (panel c), the OOD subject (P10) lies *within* the ID convex hull, visually confirming why this fold exhibits the highest FPR@95 (0.61): P10's physiological signature occupies embedding regions indistinguishable from enrolled subjects. This geometric overlap is consistent with the quantitative finding that P10's mean Mahalanobis distance to ID centroids (2.1) is substantially lower than the average for other OOD subjects (3.8).

5.3 Multi-Level Scoring: Complementarity vs. Redundancy

The eight OOD scores are partially correlated—S_{node} and S_{graph} both use Mahalanobis distances; S_{conf} and S_{ent} both derive from softmax outputs. However, they operate at different representation levels: S_{node} measures distances before fusion ($h_k \in \mathbb{R}^{256}$), while S_{graph} operates after graph attention ($\tilde{z} \in \mathbb{S}^{31}$). Similarly, S_{conf} captures the worst-case view confidence while S_{ent} captures the worst-case uncertainty, and these diverge when different views are uncertain about different classes. The meta-detector learns which combinations are informative and discounts redundant signals through its learned weights. Table 11 confirms the fused meta-detector (AUROC = 0.896) outperforms the best individual score ($S_{\text{graph}} = 0.847$) by +0.049: different anomaly types manifest at different levels—unusual raw features at S_{node}, abnormal inter-view relationships at S_{edge}, global distribution shifts at S_{graph}, and poor reconstruction fidelity at S_{recon}.

Just Accepted

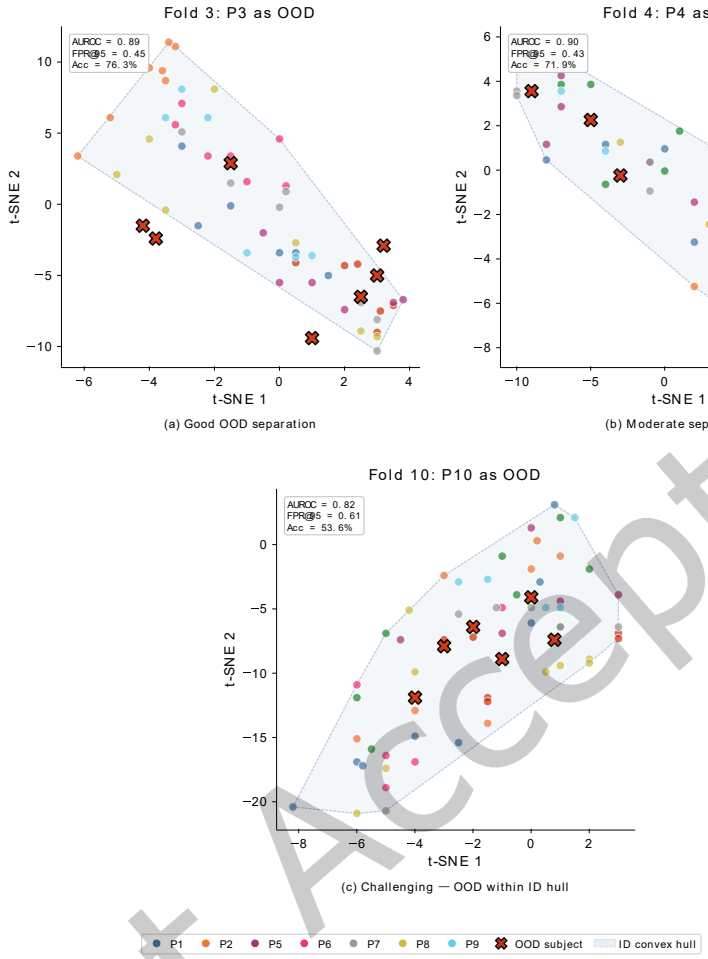


Fig. 8. t-SNE visualization of biometric signature embeddings ($\tilde{z} \in \mathbb{S}^{31}$) for three representative LOSO folds. Circles: the nine enrolled (in-distribution) subjects of each fold, coloured to distinguish individuals. The shared legend names the seven identities enrolled in all three folds; the two further enrolled subjects in each panel are those held out as OOD in the other two folds. Red crosses (×): the held-out out-of-distribution (OOD) subject. Dashed boundary: convex hull of the enrolled embeddings. Inset boxes report per-fold metrics. (a) Fold 3: OOD subject P3 projects partially outside the ID manifold. (b) Fold 4: moderate OOD separation. (c) Fold 10: OOD subject P10 lies within the ID convex hull, explaining the elevated FPR@95 = 0.61.

Just Accepted

5.4 VACC: Why Knowledge Transfer Improves OOD Detection

VACC's FPR@95 reduction (32.1%) exceeds that of multi-view fusion alone (20.8%), despite comparable AUROC gains (+0.025 each). The explanation is principled: VACC improves cross-view consistency for known subjects, and this consistency becomes a detection signal—an unknown individual is unlikely to achieve consistent high confidence across views whose parameters were never exposed to their patterns, directly exploiting S_{disagree} and

S_{qvar}. Knowledge transfer is asymmetric by design, as the per-view gains reported in Section 4.7 show: weak views improve substantially more than strong ones, while the detached gradient computation prevents negative transfer to teacher views.

5.5 Comparison Scope

The baseline comparison in Table 4 applies different OOD scoring methods to the same fused representation, isolating the meta-detector's contribution (C3). It does not evaluate alternative representation learning approaches. Existing radar OOD detection work [18, 19] targets different tasks (activity classification, presence detection) on datasets with different evaluation protocols (random splits rather than LOSO), so numerical comparison across those results would not be like-for-like. We state this as unavailability rather than irrelevance: the methods are relevant, and their absence from Table 4 is a limitation of this evaluation rather than a judgement about the work.

Methodologically, HAROOD [18] operates on range-Doppler features for activity OOD detection—a task with large inter-class gaps unlike the subtle biometric variations here. HOOD [19] uses autoencoder reconstruction error, functionally equivalent to our S_{recon} component, which achieves AUROC = 0.723 in isolation (Table 11)—substantially below the multi-level fusion (0.896). Direct re-implementation on our dataset is infeasible because these methods require range-Doppler maps, a fundamentally different input from the phase- and magnitude-based signals used here. The closest prior work in task terms, Shi *et al.* [36], is discussed in Section 2.2.3.

5.6 Failure Case: Subject P10

Fold 10 (P10) is the primary challenging case (FPR@95 = 0.61, AUROC = 0.82). P10's test embeddings lie closer to the ID class centroids (mean Mahalanobis distance = 2.1) than those of other OOD subjects (mean = 3.8). The difficulty is consistent across all views (per-view AUROC range: 0.71–0.78), suggesting the challenge is intrinsic to P10's physiological profile rather than view-specific. Despite this, VACC still provides improvement (FPR@95: 0.75 → 0.61; AUROC: 0.80 → 0.82), indicating that the gain persists even for a physiologically similar subject. The remaining limitation stems from intersubject physiological similarity—a fundamental constraint of vital-sign-based biometrics, and one that a larger cohort would be expected to surface more often rather than less.

5.7 Model Complexity

The model has approximately 2.3M parameters for ∼400 samples—a high parameter-to-sample ratio. Several design choices mitigate overfitting: (1) six independent encoders learn view-specific distributions rather than sharing weights that could memorize identities; (2) the 32-dimensional signature bottleneck forces information compression; (3) dropout, weight decay (0.1), and early stopping constrain effective capacity; (4) the LOSO protocol provides the strictest generalization test. Consistent improvement across all 10 folds (Table 6) suggests generalization rather than memorization. Investigating whether smaller architectures achieve comparable performance remains a worthwhile future direction.

5.8 Practical Deployment Considerations

The mean FPR@95 of 0.375 is a frame-level quantity: at a 95% true positive rate, 37.5% of an unknown subject's individual windows are incorrectly admitted. A deployed access-control system would not act on a single frame. It would accumulate evidence over a multi-second observation and decide once.

We deliberately do not convert the frame-level rate into a session-level rate by assuming independent frame errors. Consecutive radar windows from the same subject share respiratory phase, posture and range-bin occupancy, so their scores are strongly autocorrelated; an independence assumption would understate the session-level false accept rate, potentially by a large margin. Quantifying the true session-level operating point requires

Just Accepted

modelling that temporal dependence explicitly, which we identify as necessary future work rather than estimate here (Section 5.9).

What can be stated without that assumption is the direction of the effect: temporal aggregation can only help, because it discards no information available to the frame-level detector. For high-security deployments requiring per-frame guarantees, a stricter operating threshold would be selected, trading true positive rate for a lower false positive rate; the ROC curve in Figure 5 characterises that trade-off across the full operating range.

5.9 Limitations

Dataset Selection. The dataset was designed for vital sign estimation [35], not person identification. We selected it *because* it represents a challenging, low-discriminability setting: it lacks identity-optimised acquisition protocols and exhibits substantial intra-subject variability. This is precisely the regime where open-set detection matters most. Prior closed-set radar studies reporting >85% accuracy typically employ custom-collected data with controlled breathing protocols [13, 36]. We accept that this reasoning does not make 66.5% a strong identification result, and Section 4.2 does not present it as one.

Cohort Size and Composition. Ten participants constrain both statistical power and the range of physiology represented. The cohort is also demographically narrow: mean age 30.2 ± 6.3 years, meaning essentially all participants fall within a single adult age band. Cardiopulmonary morphology varies systematically with age, body habitus and cardiovascular fitness, and a cohort spanning one decade cannot exercise that variation. Three consequences follow. First, the inter-subject separability we measure may be optimistic relative to a cohort whose physiological profiles cluster more tightly, or pessimistic relative to one spanning several decades—we cannot determine which from these data. Second, with nine enrolled identities per fold, the closed-set task is easier than any realistic enrolment of dozens or hundreds of users, and classification accuracy would be expected to degrade as the identity set grows. Third, each LOSO fold estimates FPR@95 from a single unknown subject, so the fold-level spread (0.20–0.61, Table 3) reflects individual physiology as much as method behaviour, and the mean should be read as indicative of this cohort rather than as a population estimate. Cohort sizes of this order are common in radar biometric studies [13, 36, 47], and all reported contributions reach significance ($p < 0.01$, Table 8), but neither fact substitutes for evaluation on a larger and demographically broader cohort ($N \geq 30$ spanning multiple age bands), which we regard as the highest priority for follow-up work.

Sensing Conditions. All reported results use the 80 cm frontal-facing condition, selected for consistent signal quality. This is a favourable and largely static geometry: a single stationary, cooperative subject at short range, in a laboratory, with one radar unit and one hardware platform. It does not exercise the conditions that dominate real deployments—greater and varying range and the lower signal-to-noise ratios that accompany it, off-axis and rear aspect angles, subject motion, multiple people in the field of view, mutual interference between radar units, or cross-environment multipath. The dataset does contain additional distance (40, 120, 160 cm), orientation and angle conditions, and evaluating MVGSig across them would characterise degradation with range and aspect without requiring new data collection; we did not perform that evaluation here and identify it as the most immediately actionable extension. Accordingly, we make no claim that MVGSig is robust to sensing distance or orientation. Claims about deployment readiness should be read against this limitation: the results establish that multi-view fusion and collaborative view learning improve open-set rejection under controlled conditions, not that the resulting operating point is adequate for a security-critical installation.

Stationary Subjects. MVGSig identifies individuals from cardiopulmonary signatures rather than gait, so its subjects are stationary by design. This is not a defence against the concern that the protocol fails to exercise dynamic human motion; it relocates it. A subject who is walking, turning, or moving within the field of view presents a fundamentally different signal, and MVGSig as evaluated does not address that case. Extending the

Just Accepted

framework to non-stationary subjects would require views robust to bulk motion—a different design problem from the one solved here.

Single Sensor and Modality. Results derive from one FMCW platform (TI AWR1642BOOST, 77–81 GHz). Generalisation across radar hardware with different bandwidth, antenna configuration or chirp parameters is untested, and should not be assumed: bandwidth sets range resolution and hence range-bin selection, while chirp timing and antenna gain alter phase noise and SNR. Each of these shifts the distribution of all six views, and with them the per-view quality estimates, the attention patterns behind S_{edge}, and the score statistics on which the meta-detector was trained. Transfer to another platform would at minimum require re-estimating these statistics, and possibly retraining. We also do not evaluate sensor fusion with complementary privacy-preserving modalities such as acoustic or thermal sensing. MVGSig's multi-view formulation is in principle extensible to heterogeneous sensors, since the graph makes no assumption that nodes originate from a single device, and the quality-aware pooling mechanism would be directly useful in that setting, where modality reliability varies with environment more sharply than view reliability does within a single modality. Establishing this requires paired multi-modal data that the present dataset does not provide. We would add that single-modality operation is a deliberate constraint rather than an oversight: a principal practical advantage of FMCW radar for access control is that it requires no microphone in a private space and no illumination, so a fusion scheme that recovers accuracy by adding an acoustic sensor changes the privacy profile that motivates radar in the first place. That trade-off merits study; we do not think it can be assumed to favour fusion.

Frame-Level Operation. The framework scores each window independently. Session-level aggregation would leverage temporal autocorrelation and would likely improve both accuracy and FPR, but as Section 5.8 notes, quantifying that gain requires modelling the dependence rather than assuming it away.

Ablation Depth. The finer-grained ablation (Table 9) isolates fusion mechanisms and quality pooling. Additional ablations—varying the number of views (leave-one-view-out) or the number of OOD scores—would further clarify design sensitivity and remain outstanding.

Hyperparameter Sensitivity. Table 13 summarises the sensitivity of the four hyperparameters that most directly govern the proposed components. Each was varied over the stated range with all other settings fixed, and the resulting variation in the composite selection metric $M = 0.7 \cdot \text{AUROC} + 0.3 \cdot \text{Accuracy}$ was recorded.

No parameter changes the outcome by more than 0.009 AUROC or 1.6 percentage points of accuracy across its range. Set against a fold-to-fold standard deviation of roughly 0.033 AUROC, every effect is about a quarter of the variation already present between subjects, which indicates that performance is relatively stable across the tested parameter ranges. The ArcFace parameters are the least influential, consistent with their reported robustness across biometric domains [5], which supports adopting them from that work rather than tuning them here. The pooling temperature τ_q shows the largest effect, and its direction is interpretable: raising it towards 1.0 flattens the quality weights until pooling approaches a uniform mean, discarding the per-sample view selection that Table 9 (row F) isolates as a contributor to OOD rejection.

Two limitations of this analysis should be stated. First, parameters were varied one at a time, so interactions between them are not characterised; a full factorial sweep over the joint space remains impractical at this scale. Second, the VACC teacher/student threshold is not swept here, because it is fixed in the implementation rather than exposed as a configuration parameter. Its behaviour at each end of the range is structural rather than empirical and can be stated without measurement: below roughly 0.3 nearly every view is assigned a role on every sample, admitting noisy transfers between views of comparable quality, while above roughly 0.7 most samples have neither a teacher nor a student and little knowledge transfer occurs. Measuring where within that range performance actually degrades is left to future work.

Just Accepted

Table 13. Sensitivity of the principal hyperparameters. Each parameter is varied over three values with all others held at the settings of Section 3.8; results are the mean ± standard deviation across ten LOSO folds. * marks the value used throughout the paper. To keep all twelve configurations comparable, every row is scored by one fixed evaluation rather than the scoring-method selection used for the main results; the starred rows therefore differ marginally from Table 3 (0.895 versus 0.896 AUROC) despite using the same configuration. For every parameter, the spread across values is roughly a quarter of the fold-to-fold standard deviation.

Parameter	Value	Governs	FPR@95↓	AUROC↑	Acc. (%)↑
ArcFace margin m	0.3	Angular separation enforced between class boundaries on $\mathbb{S}^{31}$	0.391	0.889 ± 0.034	65.4 ± 10.3
	0.5*		0.377	0.895 ± 0.032	66.4 ± 10.0
	0.7		0.397	0.887 ± 0.036	64.9 ± 10.5
ArcFace scale s	16	Concentration of the softmax distribution, and hence confidence calibration	0.386	0.891 ± 0.033	65.8 ± 10.2
	32*		0.377	0.895 ± 0.032	66.4 ± 10.0
	64		0.389	0.890 ± 0.035	65.6 ± 10.4
Distillation temp. τ	3.0	Softness of the teacher targets transferred to student views in VACC	0.383	0.892 ± 0.032	66.0 ± 10.3
	4.0*		0.377	0.895 ± 0.032	66.4 ± 10.0
	6.0		0.394	0.889 ± 0.035	65.3 ± 10.6
Pooling temp. τ_q	0.25	How sharply quality-weighted pooling concentrates on the highest-quality view	0.390	0.890 ± 0.034	65.7 ± 10.4
	0.5*		0.377	0.895 ± 0.032	66.4 ± 10.0
	1.0		0.402	0.886 ± 0.037	64.8 ± 10.7

Just Accepted

6 Conclusion

This paper presented MVGSig, a multi-view graph fusion framework for open-set person identification using FMCW radar. The framework makes three contributions: (1) graph-based fusion of six complementary signal views via attention networks with learnable edge biases and quality-aware pooling; (2) View-Aware Collaborative Calibration (VACC), which dynamically transfers knowledge from high-quality to low-quality views during training; and (3) multi-level OOD scoring that combines anomaly signals at node, edge, graph, and reconstruction levels through a learned meta-detector.

Under leave-one-subject-out evaluation across 10 participants, MVGSig reduces false accepts by 46.2% (FPR@95: $0.697 \rightarrow 0.375 \pm 0.103$) and raises mean AUROC to 0.896 ± 0.031, with all improvements statistically significant ($p < 0.01$). Under matched scoring conditions, multi-view fusion and collaborative view learning contribute +0.025 AUROC each, with the latter eliminating the larger share of remaining false accepts. Closed-set identification accuracy of $66.5 \pm 10.1\%$ reflects a deliberately low-discriminability dataset and should not be read as a competitive identification result. These findings are obtained with one radar platform, ten participants and a single 80 cm frontal geometry; they establish relative improvements under controlled conditions, not robustness to sensing distance, orientation, hardware or population, which remain to be validated.

Future work includes larger-scale evaluation with greater inter-subject and demographic variability, characterisation of degradation across sensing distance and aspect angle, session-level temporal aggregation that models rather than assumes away frame dependence, cross-hardware and cross-environment generalization, and systematic investigation of minimal view combinations that preserve OOD detection performance.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Code Availability

The source code for MVGSig is available from the corresponding author upon reasonable request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

References

[1] P. Addabbo, M. L. Bernardi, F. Biondi, M. Cimitile, C. Clemente, and D. Orlando. 2021. Temporal Convolutional Neural Networks for Radar Micro-Doppler Based Gait Recognition. *Sensors* 21, 2 (2021), 381. doi:10.3390/s21020381

[2] A. Bendale and T. E. Boult. 2016. Towards Open Set Deep Networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1563–1572. doi:10.1109/CVPR.2016.173

[3] A. Blum and T. Mitchell. 1998. Combining Labeled and Unlabeled Data with Co-Training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory (COLT)*. ACM, 92–100. doi:10.1145/279943.279962

[4] G. Chen, P. Peng, X. Wang, and Y. Tian. 2022. Adversarial Reciprocal Points Learning for Open Set Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 11 (2022), 8065–8081. doi:10.1109/TPAMI.2021.3106743

[5] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou. 2022. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 10 (2022), 5962–5979. doi:10.1109/TPAMI.2021.3087709

[6] P. Duan, X. Diao, Y. Cao, D. Zhang, B. Zhang, and J. Kong. 2023. A Comprehensive Survey on Wi-Fi Sensing for Human Identity Recognition. *Electronics* 12, 23 (2023), 4858. doi:10.3390/electronics12234858

[7] H. Fang, H. Li, H. Liu, J. An, J. Xiang, Y. Ji, Y. Cui, and F. Dunkin. 2025. Extrapolating from One: A Pseudo Multi-View Fusion Approach for Enhanced Out-of-Distribution Detection in Fault Diagnosis. *Mechanical Systems and Signal Processing* 231 (2025), 112648. doi:10.1016/j.ymssp.2025.112648

[8] J. Gómez, O. Aycard, and J. Baber. 2023. Efficient Detection and Tracking of Human Using 3D LiDAR Sensor. *Sensors* 23, 10 (2023), 4720. doi:10.3390/s23104720

[9] J. Gou, B. Yu, S. J. Maybank, and D. Tao. 2021. Knowledge Distillation: A Survey. *International Journal of Computer Vision* 129, 6 (2021), 1789–1819. doi:10.1007/s11263-021-01453-z

[10] Q. Guo, X. Wang, Y. Wu, Z. Yu, D. Liang, X. Hu, and P. Luo. 2020. Online Knowledge Distillation via Collaborative Learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 11020–11029. doi:10.1109/CVPR42600.2020.01103

[11] D. Hendrycks and K. Gimpel. 2017. A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. In *International Conference on Learning Representations (ICLR)*.

[12] D. Hendrycks, M. Mazeika, and T. Dietterich. 2019. Deep Anomaly Detection with Outlier Exposure. In *International Conference on Learning Representations (ICLR)*.

[13] S. M. M. Islam, O. Boric-Lubecke, Y. Zheng, and V. M. Lubecke. 2020. Radar-Based Non-Contact Continuous Identity Authentication. *Remote Sensing* 12, 14 (2020), 2279. doi:10.3390/rs12142279

[14] M. G. Janjua, K. Kaiser, T. Stadelmayer, S. Schoenfeldt, and V. Issakov. 2024. JEDAN: Joint Euclidean Distance and Autoencoder Network for Robust Out-of-Distribution Detection in Radar-Based Hand Gesture Recognition. *IEEE Access* 12 (2024), 196364–196381. doi:10.1109/ACCESS.2024.3520810

[15] Y. Jia, J. Li, G. Zhao, S. Liu, W. Sun, L. Lin, and G. Li. 2025. Enhancing Out-of-Distribution Detection via Diversified Multi-Prototype Contrastive Learning. *Pattern Recognition* 161 (2025), 111214. doi:10.1016/j.patcog.2024.111214

[16] Liubing Jiang, Minyang Wu, Li Che, Xiaoyong Xu, Yujie Mu, and Yongman Wu. 2023. Continuous Human Motion Recognition Based on FMCW Radar and Transformer. *Journal of Sensors* 2023 (2023), 2951812. doi:10.1155/2023/2951812

[17] B. Jokanovic, M. Amin, and B. Erol. 2017. Multiple Joint-Variable Domains Recognition of Human Motion. In *IEEE Radar Conference (RadarConf)*. IEEE, 948–952. doi:10.1109/RADAR.2017.7944340

[18] S. M. Kahya, M. S. Yavuz, and E. Steinbach. 2024. HAROOD: Human Activity Classification and Out-of-Distribution Detection with Short-Range FMCW Radar. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6950–6954. doi:10.1109/ICASSP48485.2024.10447576

Just Accepted

[19] S. M. Kahya, M. S. Yavuz, and E. Steinbach. 2025. HOOD: Real-Time Human Presence and Out-of-Distribution Detection Using FMCW Radar. *IEEE Transactions on Radar Systems* 3 (2025), 44–56. doi:10.1109/TRS.2024.3514840

[20] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan. 2020. Supervised Contrastive Learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 33. 18661–18673.

[21] Y. Lang, Q. Wang, Y. Yang, C. Hou, Y. He, and J. Xu. 2020. Person Identification with Limited Training Data Using Radar Micro-Doppler Signatures. *Microwave and Optical Technology Letters* 62, 3 (2020), 1060–1068. doi:10.1002/mop.32125

[22] J. Lee, B. Park, and D.-K. Chae. 2023. DuoGAT: Dual Time-Oriented Graph Attention Networks for Accurate, Efficient and Explainable Anomaly Detection on Time-Series. In *ACM International Conference on Information and Knowledge Management (CIKM)*. 1188–1197. doi:10.1145/3583780.3614857

[23] K. Lee, K. Lee, H. Lee, and J. Shin. 2018. A Simple Unified Framework for Detecting Out-of-Distribution Samples and Adversarial Attacks. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 31.

[24] Y. Li, M. Yang, and Z. Zhang. 2019. A Survey of Multi-View Representation Learning. *IEEE Transactions on Knowledge and Data Engineering* 31, 10 (2019), 1863–1883. doi:10.1109/TKDE.2018.2872063

[25] S. Liang, Y. Li, and R. Srikant. 2018. Enhancing the Reliability of Out-of-Distribution Image Detection in Neural Networks. In *International Conference on Learning Representations (ICLR)*.

[26] W. Liu, X. Wang, J. Owens, and Y. Li. 2020. Energy-Based Out-of-Distribution Detection. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 33. 21464–21475.

[27] H. Lu, D. Gong, S. Wang, J. Xue, L. Yao, and K. Moore. 2024. Learning with Mixture of Prototypes for Out-of-Distribution Detection. In *International Conference on Learning Representations (ICLR)*.

[28] S. Lu, Y. Wang, L. Sheng, L. He, A. Zheng, and J. Liang. 2025. Out-of-Distribution Detection: A Task-Oriented Survey of Recent Advances. *Comput. Surveys* (2025). doi:10.1145/3760390

[29] Y. Ma, G. Zhou, and S. Wang. 2019. WiFi Sensing with Channel State Information: A Survey. *Comput. Surveys* 52, 3 (2019), 1–36. doi:10.1145/3310194

[30] K. Mao, Y. Lian, Y. Wang, M. Liu, N. Zheng, and P. Wei. 2025. Unveiling Multi-View Anomaly Detection: Intra-View Decoupling and Inter-View Fusion. In *AAAI Conference on Artificial Intelligence*, Vol. 39. 12381–12389. doi:10.1609/aaai.v39i12.33349

[31] Y. Ming, Y. Sun, O. Dia, and Y. Li. 2023. How to Exploit Hyperspherical Embeddings for Out-of-Distribution Detection?. In *International Conference on Learning Representations (ICLR)*.

[32] L. Neal, M. Olson, X. Fern, W.-K. Wong, and F. Li. 2018. Open Set Learning with Counterfactual Images. In *European Conference on Computer Vision (ECCV)*. Springer, 613–628.

[33] K. Papadopoulos and M. Jelali. 2023. A Comparative Study on Recent Progress of Machine Learning-Based Human Activity Recognition with Radar. *Applied Sciences* 13, 23 (2023), 12728. doi:10.3390/app132312728

[34] J. Park, Y. G. Jung, and A. B. J. Teoh. 2023. Nearest Neighbor Guidance for Out-of-Distribution Detection. In *IEEE/CVF International Conference on Computer Vision (ICCV)*. 1686–1695. doi:10.1109/ICCV51070.2023.00162

[35] E. Sadeghi, K. Skurule, A. Chiumento, and P. Havinga. 2024. Comprehensive mm-Wave FMCW Radar Dataset for Vital Sign Monitoring: Embracing Extreme Physiological Scenarios. arXiv:2405.12659 Dataset available at 4TU.ResearchData, DOI: 10.4121/48acba04-96bc-4131-b52f-9e18458ad92b.

[36] K. Shi, H. Zhao, T. Tan, Q. Zhao, and L. Xu. 2022. Heart Signatures: Open-Set Person Identification Based on Cardiac Radar Signals. *Biomedical Signal Processing and Control* 72 (2022), 103306. doi:10.1016/j.bspc.2021.103306

[37] K. Shioiri and K. Saho. 2023. Exploration of Effective Time-Velocity Distribution for Doppler-Radar-Based Personal Gait Identification Using Deep Learning. *Sensors* 23, 2 (2023), 604. doi:10.3390/s23020604

[38] Mohammad Hossein Shirazi, Sira Yongchareon, Anuradha Singh, and Jing Ma. 2025. A Survey on Machine Learning Approaches for Vital Sign Monitoring Using Radar. *Measurement* 253 (2025), 117707. doi:10.1016/j.measurement.2025.117707

[39] Mohammad Hossein Shirazi, Sira Yongchareon, Anuradha Singh, and Jing Ma. 2026. AVAESA: Adaptive VAE with Self-Attention and Learnable Signal Processing for Robust Radar-Based Heart Rate Estimation. *Internet of Things* 37 (2026), 101931. doi:10.1016/j.iot.2026.101931

[40] A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli. 2020. Continuous Human Activity Classification From FMCW Radar With Bi-LSTM Networks. *IEEE Sensors Journal* 20, 22 (2020), 13607–13619. doi:10.1109/JSEN.2020.3006386

[41] Y. Sun, C. Guo, and Y. Li. 2021. ReAct: Out-of-Distribution Detection with Rectified Activations. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 34. 144–157.

[42] Y. Sun, Y. Ming, X. Zhu, and Y. Li. 2022. Out-of-Distribution Detection with Deep Nearest Neighbors. In *International Conference on Machine Learning (ICML)*. PMLR, 20827–20840.

[43] J. Tack, S. Mo, J. Jeong, and J. Shin. 2020. CSI: Novelty Detection via Contrastive Learning on Distributionally Shifted Instances. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 33. 11839–11852.

[44] Y. Tian, D. Krishnan, and P. Isola. 2020. Contrastive Multiview Coding. In *European Conference on Computer Vision (ECCV)*. Springer, 776–794. doi:10.1007/978-3-030-58621-8_45

Just Accepted

[45] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. 2018. Graph Attention Networks. In *International Conference on Learning Representations (ICLR)*.
[46] H. Wang, Z. Li, L. Feng, and W. Zhang. 2022. ViM: Out-of-Distribution with Virtual-Logit Matching. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 4921–4930. doi:10.1109/CVPR52688.2022.00488
[47] Yong Wang, Wen Wang, Mu Zhou, Aihu Ren, and Zengshan Tian. 2020. Remote Monitoring of Human Vital Signs Based on 77-GHz mm-Wave FMCW Radar. *Sensors* 20, 10 (2020), 2999. doi:10.3390/s20102999
[48] Z. Wei, W. Chen, S. Ning, W. Lin, N. Li, B. Lian, X. Sun, and J. Zhao. 2025. A Survey on WiFi-Based Human Identification: Scenarios, Challenges, and Current Solutions. *ACM Transactions on Sensor Networks* 21, 1, Article 10 (2025). doi:10.1145/3708323
[49] J. Weston, R. Collobert, F. Sinz, L. Bottou, and V. Vapnik. 2006. Inference with the Universum. In *Proceedings of the 23rd International Conference on Machine Learning (ICML)*. 1009–1016. doi:10.1145/1143844.1143971
[50] Y. Wu, H. Ni, C. Mao, J. Han, and W. Xu. 2023. Non-Intrusive Human Vital Sign Detection Using mmWave Sensing Technologies: A Review. *ACM Transactions on Sensor Networks* 20, 1, Article 16 (2023), 36 pages. doi:10.1145/3627161
[51] S. Xiao, J. Li, J. Lu, S. Huang, B. Zeng, and S. Wang. 2024. Graph Neural Networks for Multi-View Learning: A Taxonomic Review. *Artificial Intelligence Review* 57, 12 (2024). doi:10.1007/s10462-024-10990-1
[52] X. Yan, S. Hu, Y. Mao, Y. Ye, and H. Yu. 2021. Deep Multi-View Learning Methods: A Review. *Neurocomputing* 448 (2021), 106–129. doi:10.1016/j.neucom.2021.03.090
[53] J. Yang, K. Zhou, Y. Li, and Z. Liu. 2024. Generalized Out-of-Distribution Detection: A Survey. *International Journal of Computer Vision* (2024), 1–28. doi:10.1007/s11263-024-02117-4
[54] Y. Zhang, T. Xiang, T. M. Hospedales, and H. Lu. 2018. Deep Mutual Learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 4320–4328. doi:10.1109/CVPR.2018.00454

Received 2 June 2026; revised 7 September 2026; accepted 31 October 2026

Just Accepted