This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2026.3731584

SAFE: Spatially-Aligned Feature Extraction for Metal Detection via Top-View UWB Radar

Dokyun Ryoo, Jihwan Suh, Jeongyeup Paek, and Saewoong Bahk

***Abstract*—Securing high-traffic public venues requires detecting concealed threats without disrupting pedestrian flow or compromising privacy. While ceiling-mounted Ultra-Wideband (UWB) radar offers a promising non-intrusive solution, it faces severe challenges such as signal distortion from the overhead view and dominant floor clutter that masks faint metallic signatures. To address these limitations, we propose *SAFE*, a lightweight pass-through detection framework utilizing a single UWB sensor. Instead of relying on computationally expensive deep learning models, *SAFE* adopts a physics-aware approach: it aligns radar signals with the pedestrian's trajectory to rectify geometric distortions and extracts distinct spatial and kinematic features to isolate metallic objects from body motion. Extensive experiments demonstrate that *SAFE* achieves robust discriminative performance with an Area Under the Curve (AUC) of 0.921 and over 3,800× greater computational efficiency compared to state-of-the-art models, and a benchmark-based latency projection indicates its potential for real-time, privacy-preserving edge deployment.**

***Index Terms*—Ultra-Wideband (UWB) Radar, Walk-through Security, Metal Detection, Physics-aware Learning**

I. Introduction

In the contemporary security landscape, a significant shift has emerged with *"soft targets"* (i.e., public environments characterized by unrestricted access and high civilian density) becoming focal points for mass casualty incidents [1], [2]. Unlike *"hard targets"* (e.g., airports) that enforce strict perimeter control, high-traffic environments remain inherently vulnerable due to unrestricted accessibility. The proliferation of concealed threats, ranging from firearms to improvised explosive devices (IEDs), has created an urgent imperative to extend security screening beyond traditional checkpoints. However, securing high-traffic venues presents a fundamental paradox: detecting lethal threats without impeding high-volume pedestrian flow.

Security protocols for high-traffic environments diverge fundamentally from those of controlled facilities. While airports rely on time-consuming "stop-and-divest" inspections, transposing such static protocols to public venues is operationally infeasible as the sheer volume of pedestrians inevitably triggers severe bottlenecks. Such congestion hinders daily operations and paradoxically compromises safety by creating high-density bottlenecks inherently vulnerable to attacks [3]. Consequently, effective countermeasures must operate continuously and unobtrusively, balancing robust threat detection with the preservation of public convenience [4].

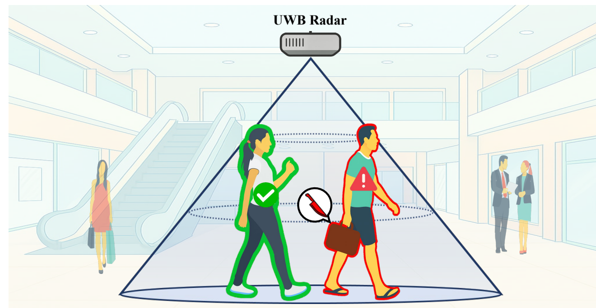


Fig. 1. Pass-through screening scenario for metal detection.

Current sensing modalities, however, face structural and functional limitations in supporting efficient, non-stop walk-through screening. Standard Walk-Through Metal Detectors (WTMD) mandate single-file queuing that inevitably causes congestion [5], while the reliance on strong magnetic fields poses safety risks to medical implants [5], [6]. Millimeter-wave (mmWave) imaging provides high resolution but typically requires subjects to pause and pose for scanning [7]. Compounding this operational bottleneck, prohibitive hardware costs and intrusive privacy concerns further preclude widespread use of mmWave in public venues [8], [9]. In contrast, Wi-Fi based sensing offers a cost-effective alternative but suffers from low spatial resolution, preventing reliable distinction between concealed weapons and normal body movements [10].

To balance security with efficiency, we propose a novel system utilizing ceiling-mounted Ultra-Wideband (UWB) radar. This modality effectively bridges the gap between existing sensors: it penetrates heavy clothing better than mmWave and resolves fine details sharper than Wi-Fi. By monitoring from overhead, the system enables non-stop screening at normal walking speeds as shown in Fig. 1. This configuration affords three key operational advantages:

1) **Non-Stop Screening:** Unobtrusive overhead monitoring allows pedestrians to be screened during natural walking without requiring a stop-and-scan procedure.
2) **Universal Safety:** Ultra-low transmission power safeguards medical implants and sensitive devices [11].

D. Ryoo and J. Suh are with the Department of Electrical and Computer Engineering and INMC, Seoul National University, Seoul, Republic of Korea (e-mail: {dkryoo, jhsuh}@netlab.snu.ac.kr); J. Paek is with the Department of Computer Science and Engineering, Chung-Ang University, Seoul, Republic of Korea (e-mail: jpaek@cau.ac.kr); S. Bahk is with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea (e-mail: sbahk@etri.re.kr). S. Bahk and J. Paek are the corresponding authors.

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grants funded by the Korea government (MSIT) (IITP-2026-RS-2024-00398157; IITP-2026-RS-2024-00429088), and also by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00359450).

3) **Non-Intrusive Sensing:** Abstract signal analysis preserves privacy, excluding optical or anatomical imaging.

However, the overhead geometry imposes unique physical challenges distinct from conventional planar sensing. First, the system confronts *subject-dependent geometric variability*, where inconsistent hyperbolic trajectories complicate robust feature extraction. Second, the top-down perspective induces *non-stationary Doppler signatures* compounded by *spectral aliasing*, where rapid frequency sweeping and limb velocities exceeding the Nyquist limit render deterministic motion analysis unfeasible.

To address these complexities, we propose *SAFE*, a spatially-aligned feature extraction framework for metal detection via top-view UWB radar. First, *Kinematic Parameter Estimation* compensates for subject-dependent geometric distortions induced by individual motion characteristics. This enables the compensation of hyperbolic range migration, transforming distorted time-domain signals into a normalized person-centric coordinate system. Building on this alignment, *Dual-Domain High-Resolution Maps* constructs a spatial magnitude map (M) characterizing static scattering morphology and a micro-Doppler spectrogram (S) capturing limb articulation dynamics. Finally, we perform *multi-view feature fusion* over the extracted representations. By synthesizing flattened map vectors with supplementary statistical descriptors to restore global structural context, we employ an eXtreme Gradient Boosting (XGBoost) [12] classifier to robustly identify metal-positive pass-through events across variations in walking speed and orientation.

Our main contributions are summarized as follows:

- We identify fundamental challenges in ceiling-mounted UWB sensing, including subject-dependent geometric distortion, kinematic clutter, and signal aliasing, which hinder the detection of weak metallic reflections.
- We propose *SAFE*, a physics-aware detection framework that resolves these challenges through trajectory-based spatial alignment and a dual-stream ensemble learning architecture, ensuring robust performance in data-constrained environments.
- We implement *SAFE* and validate its performance via extensive experiments, demonstrating that it achieves an Area under the Curve of 0.921 with over 3,800$\times$ computational efficiency compared to deep learning models, confirming its feasibility for real-time edge deployment.

The remainder of this paper is organized as follows: We provide background on UWB radar sensing in §II and outline the key challenges of pass-through sensing in §III. We then present our proposed approach in §IV, followed by a detailed evaluation in §V. We review related work in §VI, provide a discussion on practical considerations and limitations in §VII, and conclude the paper in §VIII.

II. Background

This section provides background on UWB radar sensing, focusing on the structure of the channel impulse response (CIR) and the electromagnetic differences between human tissue and metallic objects.

A. UWB Radar and CIR Fundamentals

Impulse Radio UWB (IR-UWB) radars transmit sub-nanosecond pulses over gigahertz-scale bandwidths, providing fine range resolution and the ability to distinguish closely spaced scattering paths. Each pulse transmitted at time instant t produces a one-dimensional (1-D) CIR, where the fast-time delay maps directly to the range r. Following a commonly used interpretation of the CIR as a superposition of multipath components [13], the received signal $\gamma(r;t)$ at a specific range r given time t is expressed in the following form:

$$\gamma(r;t) = \gamma_{\text{bg}}(r) + \gamma_{\text{tgt}}^{\text{LOS}}(r;t) + \gamma_{\text{tgt}}^{\text{NLOS}}(r;t) + n(r;t). \quad (1)$$

The background term $\gamma_{\text{bg}}(r)$ captures static reflections from dominant environmental structures such as floors and nearby walls, while $n(r;t)$ denotes additive measurement noise. The signal reflected by the target comprises a line-of-sight (LOS) component $\gamma_{\text{tgt}}^{\text{LOS}}(r;t)$ and non-line-of-sight (NLOS) components $\gamma_{\text{tgt}}^{\text{NLOS}}(r;t)$ resulting from additional scattering paths created by the target and its interaction with the environment. In this fast-time view, the CIR provides a high-resolution range profile for each individual radar pulse.

By stacking these sequential CIR frames, we construct a two-dimensional radargram. The temporal axis of this accumulation corresponds to the elapsed time t, constituting the slow-time dimension. Transforming the two-dimensional radargram along this dimension into the Doppler domain effectively isolates the moving target from static environmental clutter. Within the Doppler spectrum, a concealed metallic object manifests as a *high-intensity signature* due to its significant radar cross-section (RCS), while strictly adhering to the torso's bulk motion. This distinction between the complex micro-Doppler kinematics of limbs and the rigid motion of carried objects provides a fundamental basis for feature discrimination using time-frequency analysis.

B. Electromagnetic Characteristics of Metallic Objects

Metallic objects exhibit fundamental scattering properties that render them highly distinguishable in UWB radar systems. From an electromagnetic perspective, metals act as near-Perfect Electric Conductors (PECs) [14], presenting a significant impedance mismatch with the surrounding air. This results in a reflection coefficient of $\Gamma \approx -1$, causing the incident UWB pulse to undergo near-total reflection. Unlike biological tissues, which behave as lossy dielectric volumes and induce diffuse scattering, metallic surfaces generate strong and coherent reflections. This physical contrast–between the strong surface reflection of metal and the weaker, scattered reflection of the body–establishes the theoretical baseline for detection.

These electromagnetic attributes create distinctive signatures in both the time and Doppler domains. In the time-domain CIR, metallic targets manifest as sharp, high-amplitude impulses with minimal spreading, whereas the human body produces a smeared response with energy distributed over a longer duration due to complex tissue interactions [15]. This distinction extends to the Doppler domain. While the human body generates a distributed micro-Doppler spread

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2026.3731584

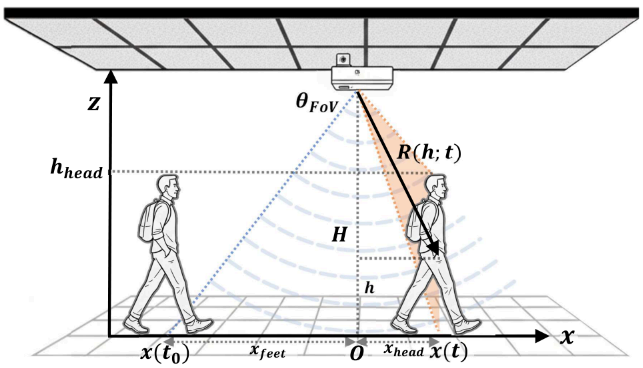


Fig. 2. Pass-through sensing geometry of an overhead UWB radar. At time t, reflections from height h map to range $R(h;t)$, yielding the orange-highlighted interval $[R_{\min}, R_{\max}]$ in the CIR.

due to non-rigid movements, a rigid metallic object reflects a narrow, high-intensity peak. Ultimately, this dual-domain energy concentration–seen as a sharp impulse in time and a strong peak in Doppler–provides a definitive signature for separating metallic threats from the human body.

C. XGBoost

XGBoost is a scalable and highly efficient implementation of gradient boosting decision trees [12]. Unlike parallel ensemble methods such as Random Forest which construct independent trees simultaneously [16], XGBoost builds the model in a sequential manner. The core principle involves iteratively adding new trees to predict the "residuals" (errors) from previous steps. This approach efficiently transforms a collection of weak learners into a single strong predictor.

Two distinguishing mathematical properties of XGBoost make it superior to traditional gradient boosting machines. First, XGBoost incorporates a regularized learning objective that penalizes model complexity (e.g., number of leaves), which is critical for preventing overfitting on finite radar datasets. Second, it employs a second-order approximation of the loss function, utilizing both the gradient and Hessian to determine the optimal tree structure. These features collectively enable faster convergence and robust generalization performance, allowing the model to effectively handle complex, high-dimensional features.

III. Problem and Challenges

This section establishes the system model, analyzes the signal behavior specific to the overhead UWB geometry, and outlines the resulting challenges that complicate the detection of metal-induced perturbations.

A. System Model

As illustrated in Fig. 2, we project the 3D sensing geometry onto a 2D Cartesian plane defined by the along-track (x) and vertical (z) axes, assuming the pedestrian passes directly beneath the radar (i.e., cross-range $y \approx 0$) without loss of generality. This assumption holds because any constant cross-range offset y is mathematically equivalent to simply substituting the vertical height z with $z_{\text{eff}} = \sqrt{z^2 + y^2}$, rendering the formulation identical to the zero-cross-range case. Let the origin $O(0, 0)$ denote the radar's ground projection, placing the sensor at $S(0, H)$. Within the x-z plane, we model the pedestrian as point scatterers translating along x at varying heights h ($0 \le h < h_{\text{head}}$). Assuming a constant velocity v, the horizontal displacement relative to the nadir becomes $x(t) = v(t - t_c)$, where t_c denotes the origin crossing time. Consequently, the distance $R(h;t)$ from the radar to a scatterer at height h follows a hyperbolic trajectory:

$$R(h;t) = \sqrt{x(t)^2 + (H-h)^2} = \sqrt{v^2(t-t_c)^2 + (H-h)^2}. \quad (2)$$

Eq. (2) indicates that the trajectory in the range-time domain is subject-dependent; the walking speed v governs the curvature, while the scatterer's height h determines the vertex depth.

The radar's conical FoV restricts observability to scatterers satisfying $|x(t)| \le (H - h)\tan\theta_{\text{FoV}}$. Under this constraint, we define the valid range extent $[R_{\min}(t), R_{\max}(t)]$ where the target is detectable. This extent is governed by the horizontal entry thresholds for the feet ($h = 0$) and head ($h = h_{\text{head}}$):

$$x_{\text{feet}} = H\tan\theta_{\text{FoV}}, \quad x_{\text{head}} = (H - h_{\text{head}})\tan\theta_{\text{FoV}}. \quad (3)$$

Here, x_{feet} sets the maximum observation window, corresponding to the entry time t_0 ($|x(t_0)| = x_{\text{feet}}$) in Fig. 2.

Based on these spatial thresholds, the maximum boundary $R_{\max}(t)$ tracks the feet:

$$R_{\max}(t) = \sqrt{x(t)^2 + H^2}, \quad \text{for } |x(t)| \le x_{\text{feet}}. \quad (4)$$

Conversely, the minimum boundary $R_{\min}(t)$ is piecewise, determined by the head trajectory within the FoV and the cone boundary otherwise:

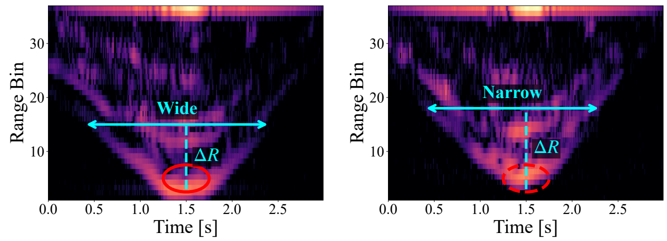
$$R_{\min}(t) = \begin{cases} \sqrt{x(t)^2 + (H - h_{\text{head}})^2}, & \text{if } |x(t)| \le x_{\text{head}} \\ \frac{|x(t)|}{\sin\theta_{\text{FoV}}}, & \text{if } x_{\text{head}} < |x(t)| \le x_{\text{feet}}. \end{cases} \quad (5)$$

Consequently, the dominant LoS reflections are confined within the region bounded horizontally by $|x(t)| \le x_{\text{feet}}$ and vertically by the range interval $[R_{\min}(t), R_{\max}(t)]$, effectively rendering NLoS components ($\gamma_{\text{tgt}}^{\text{NLOS}}$) negligible.

B. Challenges

The overhead pass-through geometry introduces fundamental physical and operational challenges that preclude the use of standard radar processing pipelines. These difficulties stem from the unique sensing geometry and the complex scattering behavior of the human body in motion.

Subject-Dependent Geometric Distortion. In contrast to frontal radar imaging characterized by linear target approaches, the overhead geometry induces a nonlinear hyperbolic range migration. As derived in §III-A, both the signal trajectory (Eq. (2)) and the valid signal space (Eqs. (3) to (5)) are explicitly parameterized by the kinematic state vector $\mathbf{\Theta} = [h, v, t_c]^\top$. This parametric dependency implies that the resulting spatial signature is strictly subject-dependent (i.e. humans with different heights and sizes). Consequently, direct analysis of the raw range-time map is infeasible, as the geometric distortions induced by motion variability preclude the extraction of consistent quantitative features.



(a) Subject A (Slow / Tall) (b) Subject B (Fast / Short)

Fig. 3. Visual demonstration of geometric distortion caused by motion variance. The red circles indicate the vertical shift caused by subject height (h_{head}), while the blue arrows represent the temporal spreading caused by walking speed (v).

Fig. 3 visualizes this geometric variance using preprocessed CIR images from two distinct subjects. In each plot, the red circle indicates the trajectory's vertex (nadir), determined by the subject's height (h_{head}) and arrival time (t_c). The cyan arrow represents the temporal width measured at a fixed vertical depth (ΔR) from the vertex, connected by the dashed line. For the tall (1.82 m) and slow-walking Subject A (Fig. 3(a)), the low speed leads to an extended dwell time within the radar beam, resulting in a significantly "Wide" temporal spread (cyan arrow). In contrast, for the short (1.65 m) and fast-walking Subject B (Fig. 3(b)), the rapid traversal generates a steep trajectory, exhibiting a "Narrow" width at the identical relative depth. This visual evidence confirms that the target's spatial signature is intrinsically coupled with kinematic attributes, thereby necessitating a comprehensive spatial normalization scheme that accounts for this multi-parametric dependency.

Kinematic Clutter and Signal Aliasing. While the proposed spatial alignment effectively rectifies geometric distortions, relying solely on CIR is often insufficient due to the lack of kinematic information and the risk of location-dependent overfitting. To compensate for these limitations, conventional radar systems typically incorporate micro-Doppler signatures as a complementary feature set to capture robust motion dynamics.

In the overhead configuration, however, such kinematic analysis is severely hindered by the rapid, non-linear variation of the Doppler frequency. Based on the hyperbolic geometry defined in §III-A, the instantaneous radial velocity $v_r(t)$ and Doppler shift $f_D(t)$ are strictly governed by $\mathbf{\Theta}$:

$$v_r(t;\mathbf{\Theta}) = \frac{v^2(t-t_c)}{\sqrt{(H-h)^2+(v(t-t_c))^2}}, \quad f_D(t;\mathbf{\Theta}) = \frac{2v_r(t;\mathbf{\Theta})}{\lambda}, \quad (6)$$

where $\lambda = c/f_c$ denotes the wavelength at the radar center frequency f_c, and c is the speed of light. This formulation indicates that the Doppler signature is highly subject-dependent; the walking speed v scales the frequency magnitude, while the height h dictates the rate of the zero-crossing transition. Consequently, $f_D(t;\mathbf{\Theta})$ becomes highly non-stationary, rapidly sweeping across the full bandwidth from $-\frac{2v}{\lambda}$ to $+\frac{2v}{\lambda}$. Such drastic spectral evolution violates the stationary assumption of standard fast Fourier transform (FFT) processing, rendering frequency analysis insufficient for precise feature extraction.

Furthermore, accurately analyzing such non-stationary Doppler signatures is obstructed by spectral aliasing stemming from the mismatch between the radar's limited sampling rate and rapid body motion. Typical IR-UWB radar configurations ($f_c \approx 7.29\,\text{GHz}$) generally operate at a frame rate on the order of tens of Hertz [17], [18]. Even with the higher frame rate of 100 frames per second (FPS) employed in our experimental setup, the maximum unambiguous radial velocity is strictly constrained by the Nyquist limit [19]:

$$v_{\text{nyq}} = \frac{\lambda \cdot \text{FPS}}{4} \approx \frac{0.041 \cdot 100}{4} \approx 1.02\,\text{m/s}. \quad (7)$$

In contrast, biomechanical studies indicate that the relative velocity of swinging limbs frequently reaches $2.5{\sim}3.0\,\text{m/s}$ [20], far exceeding v_{nyq}. This inevitably causes the high-velocity Doppler components to wrap around and overlap within the baseband spectrum, rendering quantitative analysis via deterministic models practically infeasible.

IV. Proposed Framework: *SAFE*

SAFE is a framework for detecting concealed metallic objects using overhead UWB radar. The system transforms raw, time-variant signals into robust spatial representations and classifies metal-positive pass-through events through three sequential stages (Fig. 4): (1) *Kinematic Parameter Estimation*, which recovers pedestrian motion to compensate subject-dependent hyperbolic range migration; (2) *Dual-Domain Signal Representation*, which constructs spatial and micro-Doppler maps to mitigate geometric distortion and non-stationary Doppler effects; and (3) *Multi-View Feature Fusion*, which combines spatial scattering and motion cues for metal-presence classification.

A. Kinematic Parameter Estimation

To robustly estimate $\mathbf{\Theta}$ from the noisy radar signal, we employ the multi-step procedure summarized in Algorithm 1, while Fig. 5 visualizes its intermediate outputs. The primary objective of this process is to isolate the valid LoS component $\gamma_{\text{tgt}}^{\text{LOS}}$ from the signal superposition defined in Eq. (1). However, as visualized in Fig. 5(a), the raw CIR is dominated by the static background γ_{bg}, where the floor reflection forms the maximum intensity peak at range H. We exploit the shadowing effect on this peak to pinpoint the pedestrian's timing. Using this temporal anchor, we establish a coarse spatiotemporal region of interest (ROI) designed to encompass the theoretical valid signal space derived in §III-A. Temporally, we apply a window of $\pm t_{\text{window}}$ sufficient to cover the maximum beam traversal time; spatially, we strictly constrain the search to range bins below the floor level ($R < H$). This spatial constraint effectively isolates the interval $[R_{\min}, R_{\max}]$ dominated by $\gamma_{\text{tgt}}^{\text{LOS}}$ while simultaneously suppressing strong interference from floor reflections. Within this region, a High-Pass Filter (HPF) with cutoff f_{cut} removes the residual γ_{bg}, yielding the clutter-suppressed range-time intensity map $s(t,r)$ as shown in Fig. 5(b). Here, the spatial coordinate r corresponds to the geometric range $R(h;t)$ derived in §III-A.

While the HPF effectively mitigates static clutter, the remaining dynamic signatures exhibit a significant range spread

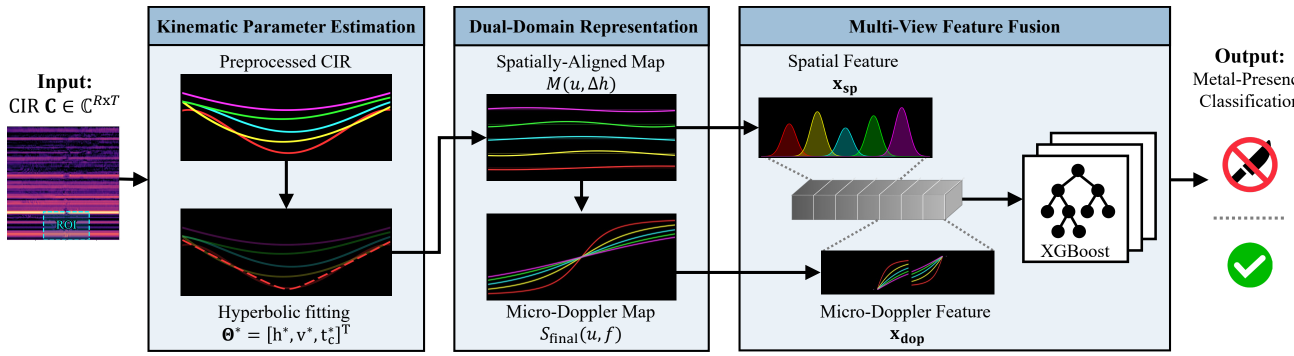


Fig. 4. Conceptual overview of the proposed *SAFE* framework. The pipeline is structured into three key phases: (1) **Kinematic Parameter Estimation** for robust trajectory recovery, (2) **Dual-Domain Signal Representation** for generating high-resolution spatial and micro-Doppler maps, and (3) **Multi-View Feature Fusion** for metal-presence classification using early fusion.

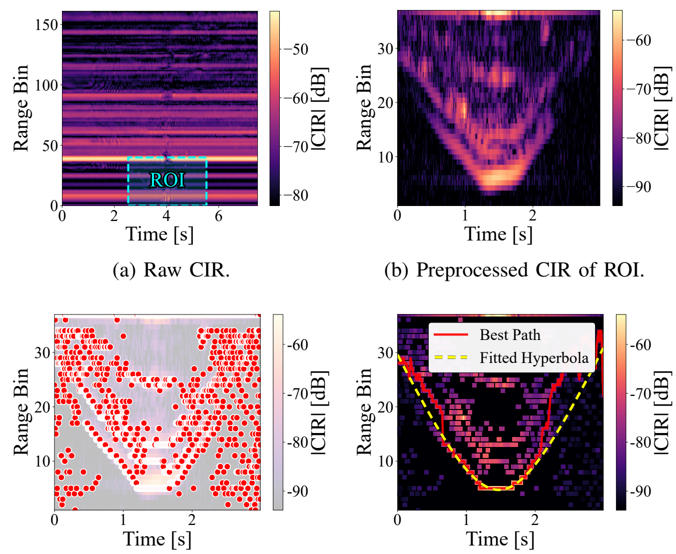
Algorithm 1 *Kinematic Parameter Estimation*

Require: CIR $\mathbf{C} \in \mathbb{C}^{R\times T}$

// Phase 1: ROI Localization
Identify the floor-reflection bin r_f
Detect the pedestrian interval from attenuation at r_f and extract the corresponding ROI
// Phase 2: CIR Preprocessing
Apply high-pass filtering to obtain the clutter-suppressed map $s(t, r)$
// Phase 3: Peak Candidate Extraction
for each frame t **do**
  $\mathcal{P}_t \leftarrow \mathrm{NMS}(s(t,:))$
end for
// Phase 4: Trajectory Optimization and Fitting
for $t = 2, \ldots, T$ **do**
  Update the minimum-cost trajectories over $\mathcal{P}_{t-1} \rightarrow \mathcal{P}_t$ using signal fidelity and kinematic regularization
end for
Select the minimum-cost trajectory $\mathcal{P} = \{b_1, \ldots, b_T\}$
Fit $\boldsymbol{\Theta} = [h, v, t_c]^\top$ using $R(\boldsymbol{\Theta}; t)$ from Eq. (2):
$\boldsymbol{\Theta}^* \leftarrow \arg\min_{\boldsymbol{\Theta}} \sum_{t=1}^{T} [b_t - R(\boldsymbol{\Theta}; t)]^2$
return $\boldsymbol{\Theta}^* = [h^*, v^*, t_c^*]^\top$

across the interval $[R_{\min}, R_{\max}]$, driven by the combined effects of the body's vertical extent and non-rigid articulation. To address this spatial ambiguity, we resolve the problem by targeting the most distinct feature: the minimum range boundary $R_{\min}$ corresponding to the pedestrian's head. To robustly extract this boundary, we employ a Non-Maximum Suppression (NMS) algorithm. NMS is a standard preprocessing technique widely utilized to extract distinct signal peaks from continuous range profiles, forming a sparse set of candidate nodes prior to trajectory optimization [21]. In our context, this operation sharpens the range-profile peaks by suppressing diffuse echoes from the limbs while preserving prominent local intensity maxima as candidate nodes for the head trajectory. This process effectively filters out the distributed body reflections, yielding a refined search space for robust $R_{\min}$ estimation (Fig. 5(c)).

Building upon this discrete search space, a Dynamic Programming (DP) algorithm is employed to synthesize a globally consistent trajectory $\mathcal{P} = \{b_1, \ldots, b_T\}$ that is robust against



(a) Raw CIR.
(b) Preprocessed CIR of ROI.
(c) CIR after NMS.
(d) Hyperbolic Fitting based on DP.

Fig. 5. Visual overview of the kinematic parameter estimation.

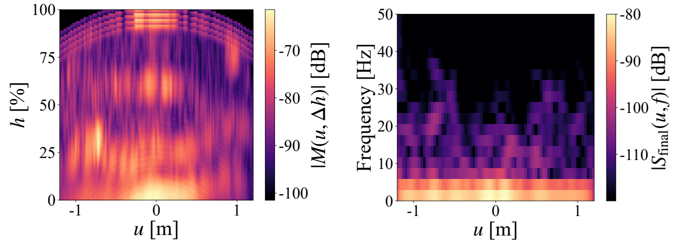
signal dropouts. Unlike greedy peak-picking methods, our DP approach optimizes the entire path by minimizing a composite cost function that balances measurement evidence with physical motion constraints:

$$C(\mathcal{P}) = \sum_t \Big(\underbrace{-\lambda_{\text{emit}} s(t, b_t)}_{\text{Signal Fidelity}} + \underbrace{\lambda_{\text{vel}} |b_t - b_{t-1}| + \lambda_{\text{acc}} |b_t - 2b_{t-1} + b_{t-2}|}_{\text{Kinematic Regularization}} \Big), \quad (8)$$

where $s(t, b_t)$ denotes the signal intensity at the candidate range b_t. The regularization terms, weighted by λ_{vel} and λ_{acc}, penalize erratic velocity changes and acceleration, thereby enforcing Newtonian inertia and ensuring the recovered path follows a smooth hyperbolic curvature. The resulting optimal path (red trajectory in Fig. 5(d)) is fitted to the theoretical model defined in Eq. (2) via non-linear least squares. This optimization yields the best-fit hyperbola (visualized as yellow curve) and determines the precise kinematic parameters $\boldsymbol{\Theta}^*$ essential for the subsequent spatial alignment.

B. Dual-Domain Signal Representation

Leveraging the precisely estimated kinematic parameters $\boldsymbol{\Theta}^*$, the second stage transforms the raw and distorted radar



(a) Spatially aligned u-h map (M). (b) Micro-Doppler Spectrogram (S).

Fig. 6. Visual overview of the motion-compensated feature extraction. (a) Person-centric mapping yields a stationary body-aligned profile. (b) Micro-Doppler reveals rhythmic limb motions about the body centroid.

signals into two complementary high-resolution feature maps. This transformation is critical because raw radar data is heavily dependent on the sensing geometry and the subject's speed. By actively compensating for the dependency, we construct a *person-centric* coordinate system that standardizes the target representation, yielding both a spatial magnitude map for morphological analysis and a micro-Doppler spectrogram for kinematic profiling.

To standardize the geometric representation independent of the pedestrian's walking speed, we first construct a person-centric spatial grid of fixed dimensions $N_u \times N_h$. In this coordinate system, the vertical axis Δh represents the relative height from the radar, while the horizontal axis u follows the linear mapping $u(t) = v^*(t - t_c^*)$, representing spatial displacement. We employ a Gaussian-weighted Backward Projection (BP) algorithm to populate this grid. For each pixel $(u_j, \Delta h_i)$, the theoretical sensor-to-target distance is calculated as $R_{i,j} = \sqrt{(h^* - \Delta h_i)^2 + u_j^2}$. The magnitude $M(u_j, \Delta h_i)$ is then computed by aggregating the radar signal magnitudes of $s(t, r)$ around this trajectory:

$$M(u_j, \Delta h_i) = \textstyle\sum_k W(r_k - R_{i,j}) \cdot |s(t_j, r_k)|, \tag{9}$$

where r_k denotes the discrete physical range of the k-th sample bin, and $W(\cdot)$ is a Gaussian interpolation kernel with standard deviation σ_W. This rectification process effectively eliminates the hyperbolic range migration caused by the overhead geometry, resulting in a stationary target profile where anatomical features are spatially aligned to consistent pixel locations (Fig. 6(a)).

To capture fine-grained kinematic details, we generate a Micro-Doppler spectrogram focused on the subject's periodic motion. Leveraging the rectified spatial map $M(u, \Delta h)$, we analyze the periodic amplitude modulations along the subject's trajectory. First, a Short-Time Fourier Transform (STFT) is applied to the spatial magnitude along the horizontal displacement axis u:

$$\mathcal{S}(u, f, \Delta h) = \left| \textstyle\sum_n M(u + n, \Delta h) \cdot w(n) e^{-j2\pi f n} \right|^2, \tag{10}$$

where $w(n)$ is an analysis window of length N_w applied to suppress spectral leakage. This transformation decomposes the spatial profile along u into its frequency components f for each height offset Δh.

Subsequently, to synthesize a unified kinematic signature, we perform a Gaussian-weighted spectral aggregation along the height axis Δh:

$$S_{\text{final}}(u, f) = \textstyle\sum_{\Delta h} \mathcal{S}(u, f, \Delta h) \cdot G(\Delta h; \sigma). \tag{11}$$

$G(\Delta h; \sigma)$ is a normalized Gaussian weighting function over the normalized height axis, defined as

$$G(\Delta h; \sigma) = \frac{\exp\left(-\frac{(\Delta h - \mu_h)^2}{2\sigma^2}\right)}{\sum_{\Delta h'} \exp\left(-\frac{(\Delta h' - \mu_h)^2}{2\sigma^2}\right)}, \tag{12}$$

with $\mu_h = 50$ denoting the center of the normalized height axis. This aggregation effectively attenuates background clutter from peripheral heights while emphasizing the distinct Doppler modulations of the torso and limbs. Fig. 6(b) visualizes this spectrogram $S_{\text{final}}(u, f)$, demonstrating the distinct Doppler modulations induced by limb articulations. By combining this dynamic signature with the static spatial map, our framework creates a holistic representation that remains robust against both structural variations and motion artifacts.

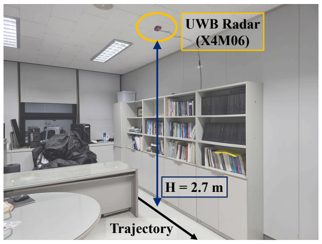
C. Multi-View Feature Fusion

Since tree-based models require vectorized input, the generated spatial map M and the micro-Doppler representation S_{final} are transformed into 1-D feature vectors, $\mathbf{x}_{\text{sp}}$ and $\mathbf{x}_{\text{dop}}$, respectively. For the Doppler stream, $S_{\text{final}} \in \mathbb{R}^{17\times 32}$ is log-compressed and flattened into $\mathbf{x}_{\text{dop}} \in \mathbb{R}^{544}$ to preserve fine-grained displacement-frequency dynamics. For the spatial stream, directly flattening $M(u, \Delta h)$ would induce excessive dimensionality and redundancy. We therefore construct a compact centered spatial descriptor from the power map $|M(u, \Delta h)|^2$. Specifically, the map is aggregated along the displacement axis u to obtain four Δh-wise profiles: mean power, standard deviation, maximum power, and kurtosis. Each profile is interpolated to 32 bins. The dominant peak of the mean-power profile is then shifted to a fixed center bin, and the same shift is applied to all four profiles to reduce residual spatial offsets across subjects. We further append two normalized profiles: the peak-normalized mean profile and the coefficient-of-variation profile. The resulting spatial feature vector $\mathbf{x}_{\text{sp}}$ consists of six 32-bin profiles, yielding $\mathbf{x}_{\text{sp}} \in \mathbb{R}^{192}$. These complementary vectors are then concatenated into a unified representation:

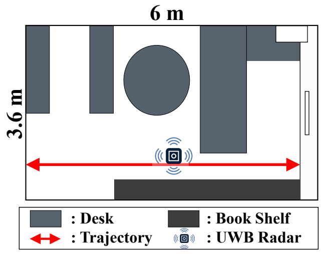
$$\mathbf{x}_{\text{unified}} = \text{Concat}\left(\mathbf{x}_{\text{sp}}, \mathbf{x}_{\text{dop}}\right) \in \mathbb{R}^D. \tag{13}$$

This early fusion strategy enables the classifier to simultaneously learn from fine-grained micro-Doppler dynamics and compact spatial scattering structure.

Finally, an XGBoost classifier processes this unified vector to resolve non-linear decision boundaries for metal-presence classification. To improve robustness under limited radar samples and complex clutter conditions, we employ lightweight regularization strategies during training. Specifically, we apply data augmentation by randomly perturbing feature magnitudes through scale-based jittering, which expands the effective training distribution while preserving class semantics. The optimized model yields a probability score $P(y = 1|\mathbf{x}_{\text{unified}})$, where $y = 1$ denotes a metal-positive pass-through event.

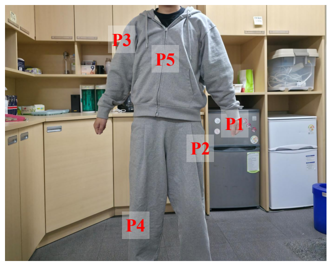


(a) Indoor testbed setup.



(b) Testbed schematic (top view).

(c) Metallic targets.



(d) Target locations on the human body.

Fig. 7. Experimental setup.

TABLE I
KEY IMPLEMENTATION PARAMETERS RELATED TO *SAFE*.

Symbol	Description	Value
Part 1: Hardware & Geometric Configuration		
f_c	Center Frequency	7.29 GHz
B	Bandwidth	1.5 GHz
PRF	Pulse Repetition Frequency	15.2 MHz
FPS	Frame Rate	100 Hz
H_{UWB}	Mounting Height (Ceiling)	2.7 m
θ_{FoV}	Field of View	$\pm 32.5°$
Part 2: Implementation Details		
A. Preprocessing & Trajectory Estimation		
f_{cut}	HPF Cutoff Frequency	0.3 Hz
t_{window}	ROI Temporal Window	± 1.5 s
ΔR_{DP}	DP Search Grid (Range)	0.05 m
λ_{emit}	Signal Fidelity Weight	2.0
λ_{vel}	Velocity Penalty Weight	1.5
λ_{acc}	Acceleration Penalty Weight	0.8
B. Dual-Domain Map Generation		
U_{lim}	Horizontal Spatial Limit	± 1.2 m
$N_u \times N_h$	Spatial Map Dimensions	300×100
N_{w}	STFT Window Size (Hamming)	32
Overlap	STFT Overlap Ratio	90%
σ_W	BP Gaussian Kernel Std. Dev.	2.0
σ	Spectral Gaussian Std. Dev.	2.0
C. XGBoost Classifier Settings		
N_{est}	Number of Estimators	3,000
η	Learning Rate	0.01
$d_{\max}$	Max Tree Depth	6
r_{sub}	Subsample Ratio	0.8
γ_{reg}	Column Sample by Tree	0.7

V. EVALUATION

In this section, we evaluate *SAFE* and verify the impact of key components via an ablation study. We validate the robustness of *SAFE* across various body locations, and benchmark it against learning-based baselines to demonstrate its superior accuracy and computational efficiency.

A. Evaluation Setup

We first describe the hardware, testbed, and the data acquisition environment. Subsequently, we provide the implementation details including algorithmic parameters and the cross-validation protocol.

Testbed and Data Acquisition. To validate the proposed framework under realistic pass-through scenarios, we construct a testbed within an indoor environment utilizing a commercial UWB radar (Novelda X4M06) [17], as comprehensively depicted in Figs. 7(a) and 7(b). The X4M06 acquires CIR frames at 100 FPS, with consecutive CIR frames forming the slow-time sequence used by *SAFE*. The testbed is deployed in a 6 m $\times$ 3.6 m $\times$ 2.7 m furnished office space to represent a typical cluttered indoor setting containing static obstacles such as desks and bookshelves. Within the testbed, participants follow the nominal path labeled as 'trajectory' in the figures, without strictly adhering to the radar's nadir. This setup captures natural spatial deviations and walking behaviors, where geometric variables such as lateral offset and walking speed are not restricted. Detailed specifications regarding the radar configuration composition are summarized in Table I.

The dataset consists of pass-through events collected from 12 participants (11 males and 1 female) with the informed consent of all individuals, to evaluate the fundamental metal-presence detection capability. The participants' body profiles spanned heights of 1.59–1.82 m and weights of 50–95 kg. The *Metal-Negative* class consists of scenarios where participants walked without carrying any metallic objects. In contrast, the *Metal-Positive* class is constructed based on the metallic objects and body locations illustrated in Figs. 7(c) and 7(d). The primary scenarios involve a metallic spray canister (25 cm $\times$ 6.7 cm) selected to emulate concealed weapons and carried at the most common positions: P1 (hand) and P2 (pocket). This object serves as a proxy target because its physical dimensions and metallic properties are comparable to typical concealed weapons.

Beyond the primary dataset, we conduct exploratory trials on a subset of 3 participants to assess the system's robustness under harsher operational conditions. These trials introduce two key variations to evaluate generalization capabilities. First, regarding *target diversity*, we test a range of metallic objects with varying shapes and RCS: scissors, knives, and box cutters. As detailed in Fig. 7(c), all of these targets have significantly smaller physical dimensions compared to the primary canister. Second, to examine *placement diversity* and robustness against severe occlusion, we conceal targets in challenging body locations: the upper arm (P3), lower leg (P4), and back (P5), as visualized in Fig. 7(d). Overall, the dataset comprises 4,276 pass-through events, consisting of 2,708 samples from the primary scenarios utilized for both training and testing, and an additional 1,568 samples for testing in §V-D.

Implementation Details. We implement *SAFE* in Python and conduct trace-driven evaluations on a standard PC (Intel Core i7, 16GB RAM) using the collected dataset. All processing-time measurements are performed using CPU-only execution on a workstation equipped with an Intel Core i7-14700KF CPU and 16 GB of RAM [22]. Key algorithmic parameters are empirically optimized to balance accuracy and efficiency, as detailed in Table I.

To evaluate the system's robustness against subject-dependent variations, we adopt a *Leave-One-Subject-Out (LOSO) Cross-Validation* protocol. Unlike random partitioning, which risks data leakage by allowing subject-specific features to permeate both training and testing sets, LOSO ensures strict separation. In each fold, one participant is completely withheld for testing ($N_{\text{test}} = 1$), while the remaining subjects are used for training. This process iterates through all participants, guaranteeing that the reported accuracy reflects the system's generalization capability to *unseen users* [23]. Crucially, we maintain this rigorous subject-independent protocol across all exploratory scenarios—involving diverse objects and placements—to verify robustness without bias.

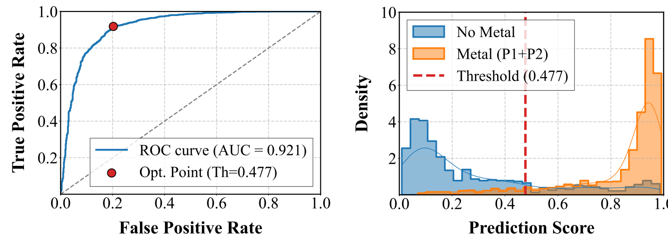
B. End-to-End Overall Performance

To evaluate the discrimination performance and determine the optimal operating point, we analyze the Receiver Operating Characteristic (ROC) curve shown in Fig. 8(a). The proposed method achieves AUC of 0.921, indicating robust separability. We utilize Youden's J statistic [24] to identify the optimal decision threshold that maximizes the difference between the True Positive Rate (TPR) and False Positive Rate (FPR), resulting in a value of 0.477. At this threshold, *SAFE* achieves a TPR of 91.8%, an FPR of 20.3%, and a false-negative rate of 8.2%. This decision boundary is visually corroborated by the prediction score distributions in Fig. 8(b), which illustrate the probability densities for the 'No Metal' and 'Metal (P1+P2)' classes. As indicated by the red dashed line, the chosen threshold effectively separates the two distributions, minimizing the overlap region responsible for misclassification. The operating threshold can be adjusted according to deployment-specific requirements to balance false alarms and missed detections.

Applying this global threshold, Fig. 9 presents the classification accuracy for each subject. The system achieves a global average accuracy of 87.8% (indicated by the red dashed line) alongside a high F1-score of 0.91, demonstrating both high overall precision and a balanced detection capability. Interestingly, the accuracy for the P1 scenario (95.5%) is over 10% higher than those recorded for the 'No Metal' (79.7%) and P2 (88.2%) cases. This improvement is attributed to the rhythmic arm swing in P1 which amplifies the micro-Doppler signature of the carried metal. Conversely, Subjects S4 and S9 yield the lowest accuracies of 70.7% and 73.1%, respectively. This performance drop is primarily attributed to the subjects' smaller physical stature, which likely resulted in radar signatures that are under-represented in the training data. Nevertheless, the system maintains high accuracy ($> 84\%$) for 10 out of 12 subjects, confirming the generalization power of our framework despite individual variations.

C. Ablation Study

To validate the contribution of each module, Table II presents an ablation study comparing the proposed framework against five modified configurations.



(a) ROC curve. (b) Prediction score distribution.

Fig. 8. Overall classification performance: (a) ROC curve of *SAFE* (solid) with the chance-level no-discrimination reference line (dashed; AUC = 0.5), and (b) prediction score distributions showing the selected decision threshold.

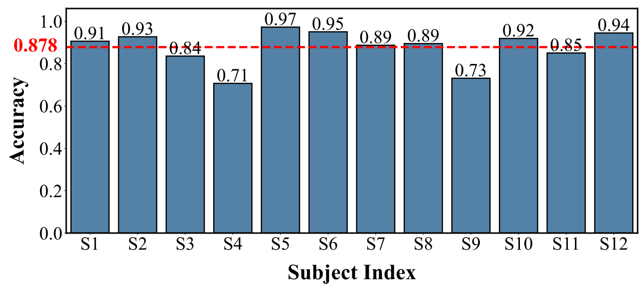


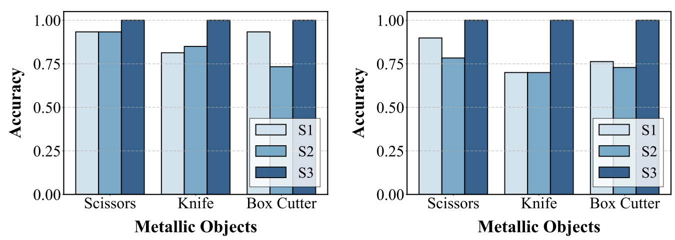
Fig. 9. Classification accuracy for each subject. The red dashed line indicates the global average accuracy.

Impact of Kinematic Parameter Estimation. Configurations (1) and (2) examine the role of our physics-aware ROI extraction by evaluating partial preprocessing steps. In Configuration (1), the raw radar data is segmented with t_{window} and processed with a HPF (f_{cut}) to remove static clutter, yet the full range axis remains preserved. Consequently, while static clutter is mitigated, the classifier is still exposed to dynamic NLoS multipath reflections ($\gamma_{\text{tgt}}^{\text{NLOS}}$) and noise in non-target range bins, limiting the AUC to 0.745 and accuracy to 70.0%. Configuration (2) attempts to address this by applying naive range-bin cropping. However, without kinematic parameter estimation, this fixed cropping fails to compensate for *subject-dependent geometric distortions* in §III-B, resulting in spatial misalignment of the target features and yielding only a marginal improvement to 0.788 in AUC and 72.1% in accuracy. These results confirm that naive data reduction is insufficient; precise alignment via kinematic estimation is essential for isolating reliable target signatures.

Importance of Multi-View Feature Fusion. Configurations (3) through (5) validate our fusion strategy against single-domain and raw-input baselines. First, using either micro-Doppler features alone (Configuration 3) or spatial features alone (Configuration 4) fails to reach the peak performance of the full framework. While both configurations maintain a relatively high AUC (≈ 0.89) compared to the ROI-absent baselines, they exhibit a notable drop in accuracy (84.6% and 80.1%, respectively) and fail to match the proposed method's AUC of 0.921. This confirms that kinematic dynamics and spatial structure provide complementary information; relying on a single view leaves the model vulnerable to ambiguity, whereas fusing them enhances robustness. Furthermore, Configuration (5) investigates the direct use of $M(u, \Delta h)$ instead of spatial feature vectors. Contrary to the assumption that using

TABLE II
ABLATION STUDY RESULTS: CONTRIBUTION OF KEY MODULES

Configuration	Description	AUC (Δ)	Accuracy (Δ)
(*) Proposed	**Full Framework (Physics-aware ROI + Hybrid)**	**0.921 (-)**	**87.8% (-)**
(1) w/o ROI	Full Range (HPF Only No Cropping)	0.745 (↓0.176)	70.0% (↓17.8%)
(2) w/o Kinematics	Naive ROI Cropping (No Kinematic Estimation)	0.788 (↓0.133)	72.1% (↓15.7%)
(3) w/o Spatial	Doppler Features Only (No Spatial Range Profile)	0.895 (↓0.026)	84.6% (↓3.2%)
(4) w/o Doppler	Spatial Features Only (No Micro-Doppler STFT)	0.890 (↓0.031)	80.1% (↓7.7%)
(5) w/ Raw Map	Using Full Temporal Map	0.892 (↓0.029)	81.2% (↓6.6%)



(a) Scenario P1

(b) Scenario P2

Fig. 10. Detection accuracy comparison for P1 and P2 scenarios across subjects S1, S2, and S3.

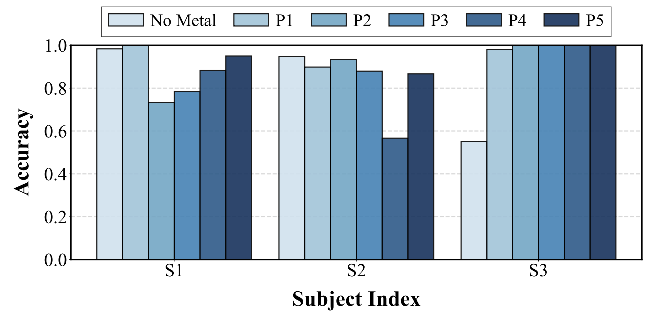


Fig. 11. Comparison of classification accuracy for each subject (S1, S2, S3) under different scenarios (No metal and P1–P5).

all available data yields better results, the accuracy degrades to 81.2%, despite a comparable AUC of 0.892. This indicates that the raw map contains significant redundancy and noise, which hampers effective learning on limited data. Consequently, our explicit feature extraction strategy proves to be more robust, as it effectively isolates discriminative target signatures while suppressing non-informative clutter.

D. Generalization and Robustness Analysis

To verify the system's physical robustness against both metallic target diversity and severe occlusion scenarios, we conduct a comprehensive analysis on the exploratory dataset described in §V-A. To rigorously assess generalization to unseen users, we adhere to a strict *subject-independent protocol*, where the test set consists exclusively of subjects disjoint from the training data. Crucially, all reported results rely on a fixed decision threshold of 0.477, calibrated solely on the primary training distribution, to demonstrate operational stability without the aid of test-time parameter tuning.

Fig. 10 plots the detection accuracy of *SAFE* across subjects for **diverse metallic objects**, including scissors, knives, and box cutters. Among the tested objects, scissors exhibit the highest detection accuracy up to 91.4%, which can be attributed to their larger exposed metallic surface area, producing stronger and more persistent radar reflections. In contrast, more compact objects such as knives (83.6%) and box cutters (79.6%) yield slightly lower accuracies, reflecting their reduced metallic extent and less distinctive scattering signatures. Despite object-specific differences, the limited performance gap across categories indicates that *SAFE* captures generic metallic scattering characteristics, achieving a mean accuracy of 84.9%.

Subsequently, we evaluate the robustness of *SAFE* against **severe occlusion** by testing metallic targets concealed at challenging body locations defined in Fig. 7(d): upper arm (P3), lower leg (P4), and back (P5). These placements introduce substantially different spatial configurations and micro-Doppler characteristics compared to the standard training positions (P1 and P2), and are therefore not explicitly observed during training. As shown in Fig. 11, *SAFE* maintains high detection accuracy across these unseen placements, achieving average accuracy of 88.10%, which indicates a strong degree of spatial generalization. Notably, the back-carried configuration (P5) achieves the highest accuracy (93.89%), surprisingly outperforming even the standard front-pocket training case (P2). The upper-arm configuration (P3) follows with a robust accuracy of 88.75%. The lower-leg placement (P4) exhibits a relatively lower accuracy of 81.67%; however, detailed analysis reveals that this drop is largely driven by a specific outlier (S2 at 56.67%), whereas other subjects maintained high accuracy (e.g., S3 at 100%). The observation that *SAFE* exhibits only minimal performance degradation on these strictly excluded configurations serves as strong evidence that the model captures intrinsic target features rather than relying on spatial priors. These results highlight the practical robustness of the proposed framework and its applicability to real-world scenarios involving diverse and unpredictable object placements.

E. Comparison with Learning Baselines

To evaluate the trade-off between classification performance and computational efficiency, we conduct a comparative study against state-of-the-art Deep Learning (DL) models. We select MobileNetV2 [25] and Vision Transformer (ViT)-Base [26] as representative lightweight convolutional neural network (CNN) and Transformer architectures, respectively, and train both models using transfer learning from ImageNet [27]. For DL-based baselines, dual-stream inputs comprising the

TABLE III
COMPARISON OF CLASSIFICATION ACCURACY AND COMPUTATIONAL COMPLEXITY. (G DENOTES GIGA (10^9) OPERATIONS)

Category	Model Architecture	Accuracy	FLOPs
Deep Learning	MobileNetV2 (Pre-trained)	74.4%	0.3 G
	MobileNetV2 (Scratch)	57.4%	
	ViT-Base (Pre-trained)	76.5%	17.6 G
	ViT-Base (Scratch)	76.4%	
Machine Learning	Logistic Regression	77.5%	4,434
	Random Forest	80.7%	28,617
	XGBoost (*SAFE*)	**87.8%**	**77,634**

CIR derived spatial map $s(t, r)$ and the corresponding micro-Doppler spectrogram are resized to meet the input requirements of each architecture. For classical machine learning baselines, Logistic Regression and Random Forest are evaluated using the proposed physics-aware features.

Accuracy Analysis. As summarized in Table III, DL models fail to utilize the radar data effectively due to the limited dataset size. MobileNetV2 exhibits severe overfitting, with accuracy dropping to 57.4% without pre-training, while ViT-Base saturates around 76.5% regardless of initialization. Remarkably, even the simplest Logistic Regression (77.5%) outperforms both DL architectures. This empirical evidence proves that our explicit physics-aware feature engineering extracts significantly more robust discriminative patterns than implicit features learned by data-hungry deep neural networks. Among all methods, the proposed *SAFE* (XGBoost) outperforms all baselines with a peak accuracy of 87.8%, establishing its distinct advantage in identifying low-RCS threats.

Computational Efficiency. Beyond accuracy, *SAFE* offers distinct advantages in computational cost. As shown in Table III, MobileNetV2 and ViT-Base require approximately 0.3 billion and 17.6 billion floating-point operations (FLOPs) per inference, respectively. In contrast, our method operates with only 77,634 FLOPs. This represents a massive computational reduction of $\approx$3,800$\times$ compared to MobileNetV2, which is widely adopted for real-time mobile tasks, and over 200,000$\times$ against ViT-Base. Although XGBoost incurs a higher computational load than Logistic Regression (4,434 FLOPs) and Random Forest (28,617 FLOPs), this margin is negligible on modern microcontrollers given the substantial accuracy gain ($>$ 6%) over other baselines. To complement the FLOPs analysis, we measure the CPU-only processing latency of *SAFE* on an Intel Core i7-14700KF CPU. Feature generation and classification require an average of 82.1 ms per pass-through event, while classifier inference alone requires 54.4 ms. Based on the 3.26$\times$ Geekbench 6.7.1 single-core performance difference between the evaluation CPU and Raspberry Pi 5, the corresponding latency on the Raspberry Pi 5 is estimated to be approximately 267.7 ms [28], [29]. Although this value is a benchmark-based projection rather than an on-device empirical measurement, the results support the feasibility of low-latency edge processing without GPU acceleration.

VI. RELATED WORK

This section reviews existing sensing modalities for concealed weapon detection.

Magnetic-based methods are widely adopted for security screening due to their cost-effectiveness and low-power operation. These approaches are generally categorized into two distinct mechanisms: *passive magnetometry*, which detects distortions in the ambient geomagnetic field [30], and *active electromagnetic induction (EMI)*, which measures perturbations in a locally transmitted field caused by conductive targets [31]. Although inherently privacy-preserving, both face distinct hurdles: passive sensors are blind to non-ferrous metals [32], while active systems are prone to false alarms from personal items [33] and pose safety risks for individuals with medical implants [6].

mmWave radars are widely explored for high-resolution screening, yet face significant deployment hurdles. Pei et al. [34] introduced a multi-view fusion framework for metallic environments, while Gao et al. [35] proposed a multi-shot decision mechanism for general indoor scenarios. However, insufficient resolution restricts these linear systems to coarse classification. To address resolution limitations, Lee et al. [36] employed nonlinear harmonic radars with fully digital phased arrays. Nevertheless, their vulnerability to environmental noise renders them unsuitable for ceiling-mounted configurations where clutter from orthogonal surfaces obscures target signatures. Although complex-valued federated learning [37] attempts to address this, its excessive computational cost precludes real-time edge deployment.

Conversely, efforts to miniaturize these systems for portable or privacy-preserving use often introduce operational rigidities or resolution trade-offs. For instance, 'mmSense' [38] achieves real-time capability via lightweight model but suffers from poor lateral resolution due to sparse antenna arrays. Alternatively, Lu et al. [39] enforce rigid protocols, requiring subjects to rotate 360 degrees for sufficient spatial coverage. Even sensor fusion methods [40] integrate depth cameras yet fail to generalize due to static background reliance.

UWB radars have emerged as a compelling modality for security screening, offering superior penetration through clothing compared to mmWave sensors and higher resolution than magnetic detectors. Leveraging these capabilities, imaging-based researchers have sought to reconstruct the spatial distribution of concealed items. Ariyoshi et al. [41] reported approximately 90% detection performance but required two opposing radar panels, each comprising 198 transmit and 198 receive antenna elements. Khan et al. [42] reported an overall $\mathrm{AUC}_{\mathrm{eff}}$ of 88.8, but their method relies on reconstructed 3D radar images and computationally demanding image-based inference. Carrer and Yarovoy [43] applied computer-vision-based automatic target recognition to 3D UWB radar images but did not report a directly comparable system-level AUC or event-level walk-through accuracy. Furthermore, these reconstruction-based approaches require explicit motion compensation to mitigate non-rigid body motion and self-occlusion, while the computational costs of back-projection, 3D reconstruction, and high-dimensional feature extraction limit their suitability for lightweight edge deployment. In contrast, *SAFE* achieves an AUC of 92.1% without explicit 3D reconstruction, using a single ceiling-mounted UWB radar and a lightweight feature-based processing pipeline.

Conversely, feature-based methods prioritize signal signatures over spatial imaging to bypass computational bottlenecks. The Late-Time Response (LTR) technique [44] theoretically isolates the unique resonant frequencies of metallic objects independent of aspect angle. Yet, in practice, this utility is nullified by the body coupling effect: the human body acts as a lossy medium that absorbs electromagnetic energy and damps the resonance signal when weapons are worn close to the skin [45]. Meanwhile, although ceiling-mounted UWB systems have been successfully deployed for human activity monitoring, such as posture classification [46] and fall detection [47], they have not been adapted for concealed weapon detection. To address the unresolved challenge of detecting low-RCS threats on moving subjects in overhead sensing, we propose a robust framework that overcomes motion blur and body clutter to enable reliable detection where conventional methods fail.

VII. Discussion and Limitations

This section reviews the practical perspectives and design boundaries of *SAFE* as a pre-screening system.

Threat Classification vs. Metal-Presence Pre-Screening. As identified in our evaluation, *SAFE* excels at isolating significant metallic signatures from human body clutter. However, the current single-sensor setup responds to strong metallic signatures rather than determining the exact shape or semantic category of the corresponding object. Consequently, benign high-RCS metallic items, such as tablets or other large portable electronic devices, may also produce metal-positive alarms. Therefore, *SAFE* is best positioned as a rapid and unobtrusive “pre-screening” layer with potential false positives. An alarm from *SAFE* should be interpreted as a metal-presence cue requiring secondary screening, not as an automatic threat decision. In this workflow, pedestrians without a metal-positive alarm can continue through the screening path, while only alarm-triggering cases are routed to quick secondary inspection, reducing the operational burden compared with universal manual screening. Future work will explore integrating multi-static UWB radar networks or multiple-input multiple-output (MIMO) arrays to extract richer spatial information and support finer differentiation between benign and threatening metallic objects.

Privacy-Preserving Operation. The privacy advantage of *SAFE* stems from the non-imaging nature of UWB radar. Unlike optical cameras or body-imaging scanners, *SAFE* does not capture facial appearance, clothing texture, or anatomical body images. The estimated kinematic parameters are short-window auxiliary variables used for geometric alignment, not for identity recognition or long-term behavioral tracking. In practical deployment, raw CIR frames can be processed locally and discarded after feature extraction, retaining only the alarm score or short-lived metadata required for system auditing. This local and non-imaging processing preserves the privacy benefits of UWB sensing while supporting real-time pass-through screening.

Multi-Pedestrian Flow and Spatial Constraints. The top-view geometry is inherently suited for high-traffic scenarios. *SAFE* leverages this by requiring only a short 2.4 m screening zone ($U_{\text{lim}} = \pm 1.2\,\text{m}$), which supports non-stop passage and provides operational potential for high-throughput screening. This compact window enables rapid motion estimation and localized processing, providing a natural path toward efficient multi-pedestrian operation. However, the current evaluation considers isolated single-pedestrian traversals and does not quantify the maximum sustainable pedestrian arrival rate, the minimum inter-pedestrian spacing, nor simultaneous multi-pedestrian performance. Future work will empirically characterize these capabilities by varying the arrival interval and temporal overlap between consecutive pedestrians, as well as the number of pedestrians simultaneously occupying the screening zone.

Evaluation Coverage and Generalizability. Current evaluation is limited in terms of participant diversity and environmental coverage. The participant cohort includes 11 male and one female participant and is therefore not gender-balanced. Nevertheless, under the strict LOSO protocol, the female participant was evaluated as an unseen subject using a model trained exclusively on the male participants. The maintained performance in this fold suggests that the learned representation is not restricted solely to gait patterns observed from the male participants. However, a single female participant is insufficient to establish generalization across genders.

Furthermore, all experiments were conducted in a single furnished indoor environment and therefore do not establish cross-environment generalizability. Performance across broader demographic groups and under different building structures, clutter configurations, and dynamic environmental interference remains uncharacterized. Evaluation with a larger and more gender-balanced participant cohort across diverse public venues constitutes important future work.

VIII. Conclusion

This work demonstrates that a ceiling-mounted UWB radar enables a new class of unobtrusive pass-through screening systems for high-traffic public spaces. Overhead sensing faces critical physical constraints, primarily geometric distortions and spectral aliasing. We address these challenges by developing a physics-aware pipeline, *SAFE*, that synergizes kinematic parameter estimation with a dual-domain signal representation, enabling the normalization of diverse walking patterns. Extensive experiments on 4,276 single-pedestrian walking sequences confirm that *SAFE* achieves 87.8% metal-presence detection accuracy with a 20.3% FPR during natural walking without requiring a stop-and-scan procedure, demonstrating the feasibility of overhead UWB as a lightweight metal-presence pre-screening filter. Our results validate the system’s robustness across diverse conditions, including variations in subjects, threat locations, and metallic object types. We anticipate that further scaling the subject pool will continue to elevate detection performance by capturing an even broader spectrum of human gait patterns.

References

[1] Institute for Economics & Peace, "Global terrorism index 2025: Measuring the impact of terrorism," Sydney, Australia, March 2025, http://visionofhumanity.org/resources, Accessed: 2026-01-11.

[2] Europol, "European Union Terrorism Situation and Trend Report," Publications Office of the European Union, Luxembourg, 2024.

[3] M. G. Stewart and J. Mueller, "Cost-benefit analysis of airport security: Are airports too safe?" *Journal of Air Transport Management*, vol. 35, pp. 19–28, 2014.

[4] Cybersecurity and Infrastructure Security Agency, "Public Venue Security Screening Guide: Touchless Screening Annex," U.S. Department of Homeland Security, 2021.

[5] National Urban Security Technology Laboratory, "Walk-Through Metal Detectors Market Survey Report," U.S. Department of Homeland Security, 2014.

[6] W. Boivin, J. Coletta, and L. Kerr, "Characterization of the magnetic fields around walk-through and hand-held metal detectors," *Health physics*, vol. 84, no. 5, pp. 582–593, 2003.

[7] S. Demirci, H. Cetinkaya, E. Yigit, C. Ozdemir, and A. Vertiy, "A study on millimeter-wave imaging of concealed objects: Application using back-projection algorithm," *Progress In Electromagnetics Research*, vol. 128, 2012.

[8] Persistence Market Research, "Millimeter Wave Body Scanner Market Forecast, 2033 (Market Summary)," 2024.

[9] D. Sheen, D. McMakin, and T. Hall, "Three-dimensional millimeter-wave imaging for concealed weapon detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 9, pp. 1581–1592, 2001.

[10] K. Wu, "Wi-metal: Detecting metal by using wireless networks," in *IEEE International Conference on Communications (ICC)*, 2016, pp. 1–6.

[11] M. Cheraghinia, A. Shahid, S. Luchie, G.-J. Gordebeke, O. Caytan, J. Fontaine, B. V. Herbruggen, S. Lemey, and E. D. Poorter, "A Comprehensive Overview on UWB Radar: Applications, Standards, Signal Processing Techniques, Datasets, Radio Chips, Trends and Future Research Directions," *IEEE Communications Surveys & Tutorials*, vol. 27, no. 4, pp. 2283–2324, 2025.

[12] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[13] H. S. Lee, "Device-Free Localization Using Commercial UWB Transceivers," https://arxiv.org/abs/2507.13938, 2025.

[14] M. Gashinova, M. Cherniakov, and A. Vasalos, "UWB signature analysis for detection of bodyworn weapons," in *CIE International Conference on Radar*. IEEE, 2006, pp. 1–4.

[15] S. Singh, Q. Liang, D. Chen, and L. Sheng, "Sense through wall human detection using UWB radar," *EURASIP Journal on Wireless Communications and Networking*, no. 1, p. 20, 2011.

[16] L. Breiman, "Random Forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[17] Novelda AS, *X4M06 Datasheet X4 Radar Development Kit.*, https://intopion.com/sdesigner/product/creaca_RES/X4M06/datasheet.pdf, Accessed: 2026-01-11.

[18] ——, *XeThru X4M300 UWB Radar Module Datasheet*, https://device.report/m/ae1c2aab87d461b251588d028b2bf58222a76fafe26f73f4b7beec1e4d6918c0.pdf, Accessed: 2026-01-11.

[19] G. Li, G. He, X. Zou, and P. S. Ray, "A Velocity Dealiasing Scheme for C-band Weather Radar Systems," *Advances in Atmospheric Sciences*, vol. 31, no. 1, pp. 17–26, 2014.

[20] M. P. Murray, A. B. Drought, and R. C. Kory, "Walking Patterns of Normal Men," *The Journal of Bone and Joint Surgery*, vol. 46, no. 2, pp. 335–360, 1964.

[21] Y. Wang, Z. Jiang, X. Gao, J.-N. Hwang, G. Xing, and H. Liu, "RODNet: Radar Object Detection Using Cross-Modal Supervision," in *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 504–513.

[22] Intel Corporation, "Intel Core i7-14700KF Processor Specifications," https://www.intel.com/content/www/us/en/products/sku/236789/intel-core-i7-processor-14700kf-33m-cache-up-to-5-60-ghz/specifications.html, 2023, Accessed: 2026-08-04.

[23] S. U. Rehman, A. Ali, A. M. Khan, and C. Okpala, "Human Activity Recognition: A Comparative Study of Validation Methods and Impact of Feature Extraction in Wearable Sensors," *Algorithms*, vol. 17, no. 12, p. 556, 2024.

[24] M. D. Ruopp, N. J. Perkins, B. W. Whitcomb, and E. F. Schisterman, "Youden index and optimal cut-point estimated from observations affected by a lower limit of detection," *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, vol. 50, no. 3, pp. 419–430, 2008.

[25] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE conference on computer vision and pattern recognition (CVPR)*, 2018, pp. 4510–4520.

[26] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *International Conference on Learning Representations (ICLR)*, 2021.

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 248–255.

[28] Primate Labs, "Geekbench 6 CPU Benchmark Result: Intel Core i7-14700KF," https://browser.geekbench.com/v6/cpu/17871799, 2026, Geekbench 6.7.1 for Windows AVX2, Accessed: 2026-08-04.

[29] ——, "Geekbench 6 CPU Benchmark Result: Raspberry Pi 5 Model B," https://browser.geekbench.com/v6/cpu/18389676, 2026, Geekbench 6.7.1 Preview for Linux AArch64, Accessed: 2026-08-04.

[30] W. Lv, P. Huang, Y. Yang, Q. Luo, S. Xie, and C. Fu, "A Novel Method of Magnetic Sources Edge Detection Based on Gradient Tensor," *Minerals*, vol. 14, no. 7, p. 657, 2024.

[31] R. Thomas, B. Salmon, D. Holloway, and J. Olivier, "Machine learning classification of metallic objects using pulse induction electromagnetic data," *Measurement Science and Technology*, vol. 35, no. 6, 2024.

[32] M. Roskosz, P. Mazurek, J. Kwaśniewski, and J. Wu, "Use of Different Types of Magnetic Field Sensors in Diagnosing the State of Ferromagnetic Elements Based on Residual Magnetic Field Measurements," *Sensors*, vol. 23, no. 14, 2023.

[33] J. Makkonen, L. A. Marsh, J. Vihonen, A. Järvi, D. W. Armitage, A. Visa, and J. Peyton, "KNN classification of metallic targets using the magnetic polarizability tensor," *Measurement Science and Technology*, vol. 25, no. 5, p. 055105, 2014.

[34] D. Pei, D. Gong, K. Liu, X. Zeng, S. Zhang, M. Chen, and K. Zheng, "mmCTD: Concealed Threat Detection for Cruise Ships via mmWave Radar," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 12, pp. 18 434–18 451, 2024.

[35] X. Gao, H. Liu, S. Roy, G. Xing, A. Alansari, and Y. Luo, "Learning to Detect Open Carry and Concealed Object with 77GHz Radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 4, pp. 791–803, 2022.

[36] W. Lee, T. Jeong, D. Lee, K. Woo, K.-Y. Lee, C.-R. Yun, C. Seo, J. Oh, K. C. Hwang, and S. K. Hong, "Fully Digital Phased Array Harmonic Radar for Detecting Concealed Electronic Devices," *IEEE Journal of Microwaves*, vol. 5, no. 4, pp. 868–881, 2025.

[37] H. Mahdipour, J. Laviada, F. Las-Heras Andrés, and M. Sookhak, "Leveraging Complex-Valued Federated Learning for Accurate and Privacy-Respectful Threat Detection Based on Millimeter-Wave Imaging," *IEEE Access*, vol. 13, pp. 91 581–91 594, 2025.

[38] K. Mitchell, K. Kassem, C. Kaul, V. Kapitany, P. Binner, A. Ramsay, D. Faccio, and R. Murray-Smith, "mmSense: Detecting concealed weapons with a miniature radar sensor," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[39] Y. Lu, W. Chen, H. Liu, and A. Zhou, "Study on feasibility of remote metal detection using millimeter wave radar for convenient and efficient security check," *CCF Transactions on Pervasive Computing and Interaction*, vol. 3, no. 3, pp. 284–299, 2021.

[40] C. Kaul, K. J. Mitchell, K. Kassem, A. Tragakis, V. Kapitany, I. Starshynov, F. Villa, R. Murray-Smith, and D. Faccio, "AI-Enabled Sensor Fusion of Time-of-Flight Imaging and mmWave for Concealed Metal Detection," *Sensors*, vol. 24, no. 18, p. 5865, 2024.

[41] M. Ariyoshi, K. Ogura, T. Sumiya, N. S. Khan, S. Yamanouchi, and T. Nomura, "Practical Implementation of Motion-Robust Radar Imaging and Whole-Body Weapon Detection for Walk-Through Security Screening," *IEICE Transactions on Communications*, vol. E106-B, no. 11, pp. 1244–1255, 2023.

[42] N. S. Khan, K. Ogura, T. Sumiya, and M. Ariyoshi, "Real-Time Concealed Weapon Detection on 3D Radar Images for Walk-Through Screening System," in *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023.

[43] L. Carrer and A. G. Yarovoy, "Concealed Weapon Detection Using UWB 3-D Radar Imaging and Automatic Target Recognition," in *Proc. European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2014, pp. 2786–2790.

[44] M. Gashinova, M. Cherniakov, and L. Daniel, "Late time response signature analysis in UWB sensors," *IEICE Proceedings Series*, vol. 13, no. 4B1-5, 2007.

[45] A. Vasalos, "Late time response analysis in UWB radar for concealed weapon detection: feasibility study," Ph.D. dissertation, University of Birmingham, 2012.

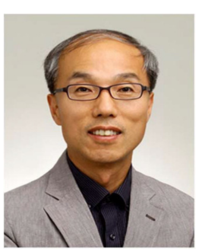
[46] W. Lu, C. Bird, M. Sandhu, and D. Silvera-Tawil, "Office Posture Detection Using Ceiling-Mounted Ultra-Wideband Radar and Attention-Based Modality Fusion," *Sensors*, vol. 25, no. 16, 2025.

[47] W. Lu, S. Kumar, M. Sandhu, and Q. Zhang, "An Unobtrusive Fall Detection System Using Ceiling-mounted Ultra-wideband Radar," in *International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2023, pp. 1–5.

Dokyun Ryoo is currently a Ph.D. candidate at the Department of Electrical and Computer Engineering, Seoul National University, Seoul, South Korea, where he is a member of the Ubiquitous Network Laboratory (NETLAB). He received the B.S. degree in electrical and computer engineering from Seoul National University in 2020. His research interests include ultra-wideband (UWB) systems, mobile cyber-physical systems (CPS), and robust wireless communications, with a particular focus on application-aware sensing systems.

Jihwan Suh received the B.S. degree in Informatics and Mathematical Science from Kyoto University, Kyoto, Japan, in 2013, and the Ph.D. degree from the School of Electrical and Computer Engineering, Seoul National University, Seoul, South Korea, in 2026. His research interests include ultra-reliable low-latency communication (URLLC), reinforcement learning for communication optimization, AI-driven communication systems, and wireless communication security.

Jeongyeup Paek is a professor at the Department of Computer Science and Engineering, Chung-Ang University, Seoul, Republic of Korea. He received his B.S. degree from Seoul National University in 2003 and his M.S. degree from University of Southern California (USC) in 2005, both in Electrical Engineering. He then received his Ph.D. degree in Computer Science from USC in 2010. He was a research intern at Deutsche Telekom R&D Labs USA in 2010, and then joined Cisco Systems in 2011 where he was a Technical Leader in the Internet of Things Group. In 2014, he was with the Hongik University, Department of Computer Information Communication as an assistant professor. He is an IEEE senior member.



Saewoong Bahk received the B.S. and M.S. degrees in electrical engineering from Seoul National University (SNU), in 1984 and 1986, respectively, and the Ph.D. degree from the University of Pennsylvania, in 1991. He was with AT&T Bell Laboratories as a Member of Technical Staff from 1991 to 1994. He was a Professor at SNU from 1994 through 2026. He has led many industrial projects on 3G through 6G and IoT connectivity supported by the Korean industry. He has published more than 300 technical articles and holds more than 100 patents. He received the KICS Haedong Scholar Award in 2012. He has been President of the Electronics and Telecommunications Research Institute (ETRI) in Korea since May 2026. He received a national medal of honor for his contributions to education in 2025 and is a Fellow of the National Academy of Engineering of Korea (NAEK).