electronics

MDPI

Article

Uncertainty-Guided Selective Temporal Refinement for mmWave Radar Human–Background Recognition

Endong Meng [1], Zhiqing Miao [1], Sichen Pu [1], Luxia Ai [2] and Shaoying Shi [1,*]

[1] School of Information and Electronic Engineering, East China Normal University, Shanghai 200241, China
[2] School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China
* Correspondence: syshi@ce.ecnu.edu.cn

Abstract

This study evaluates the incremental benefit of selective temporal refinement beyond Static Random Forest (RF) classification and causal Majority-3 stabilization in mmWave radar human–background recognition. Static RF uses 41-D current-frame features and its intrinsic tree-probability disagreement for routing, without an additional trained router. When disagreement exceeds a development-selected threshold and consecutive three-frame (K3) history is valid, the system invokes Temporal RF with a 54-D input combining 41-D current-frame features and a 13-D temporal extension; otherwise, Static predictions are retained. All paths use the same causal Majority-3. Across eight independent test acquisitions (16,742 frames), Static, Selective, and Always produce 86, 72, and 71 final errors. Relative to Static, Selective reduces Background false positives by 14, decreasing residual errors by 16.28% while retaining two Human misses. Selective invokes Temporal RF 479 times (2.8611% of all frames), versus 15,758 times (94.1226%) for Always, reducing additional temporal calls by 96.96%. In host inference tests, Fused Selective and Optimized Always have mean latencies of 10.642 and 11.681 ms/frame, respectively, an approximately 8.9% reduction. On two independent acquisitions at an additional logistics site, Selective reduces final errors from 66 to 56. These results show that disagreement-guided selective temporal refinement provides incremental improvement beyond static classification and causal output stabilization under the evaluated conditions, with infrequent additional temporal invocation.

Keywords: mmWave radar human–background recognition; selective temporal refinement; ensemble disagreement; uncertainty-guided inference; Random Forest; conditional temporal computation

Check for updates

Academic Editor: Kefeng Ji

Received: 17 September 2026
Revised: 19 October 2026
Accepted: 26 October 2026
Published: 27 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Millimeter-wave (mmWave) radar provides range, angle, radial velocity, and reflectivity for contactless human sensing [1]. Its sparse, nonuniform point clouds fluctuate with posture and motion, while environmental reflections and transient clutter can resemble human returns. Current-frame classification can leave residual Human misses and Background false positives. Historical information can supplement the evidence available in an individual frame: radar activity sensing uses temporal accumulation, inter-frame motion estimation, and recurrent representations to describe changes that are difficult to interpret instantaneously [2–6]. These approaches motivate temporal processing, while leaving open whether every observation benefits from additional learned inference.

Continuous recognition also allows inexpensive use of prediction history. Causal Majority-3 suppresses short-term output reversals without revisiting radar features or executing another classifier. It accounts for the main reduction in Static RF errors in this study, making stabilized Static RF the reference for the additional learned stage. The question is whether, after full Static inference and common causal stabilization, Temporal RF can improve selected history-valid observations enough to justify its execution and resident-model costs. This requires evaluating recognition gains alongside computational allocation, including both corrected and introduced errors.

Efficient temporal representations and input-adaptive inference address complementary aspects of this question. Compact spatio-spectro-temporal, recurrent, and state-space radar models reduce representation or execution costs [7–10], while edge-oriented systems consider constrained implementations [11–15]. Dynamic inference instead adapts the computation assigned to an input; dynamic tree ensembles, for example, can reduce the number of executed learners [16]. This study applies conditional computation between two classifiers. Static RF always executes in full, and its intrinsic tree-level disagreement determines whether a second forest should process short-term feature history.

RF provides individual-tree probabilities during classification. Their dispersion provides ensemble-consistency information beyond the mean prediction and is related to existing work on uncertainty in probabilistic tree ensembles [17]. Reusing this information offers an inexpensive allocation signal without treating it as a new uncertainty estimator. Its usefulness must be demonstrated empirically: low disagreement can accompany an error, and high disagreement does not guarantee that the temporal model can correct it. RF serves as a fixed backbone for studying allocation; neural point-cloud encoders provide external references rather than evidence of universal RF superiority. Both forests remain resident, and fewer Temporal RF calls need not imply a proportional decrease in total runtime or memory.

The study makes three connected contributions. First, it develops a disagreement-guided allocation mechanism for mmWave human–background recognition: after full Static RF inference, Temporal RF is invoked only when disagreement exceeds a development-selected threshold and causal history is valid. Existing ensemble outputs are reused without training a separate router. Second, it combines a 41-D current-frame representation with a compact 13-D temporal extension to form a 54-D input with a three-frame causal window and explicit Static fallback. All three paths use the same causal Majority-3, separating feature-history refinement from output stabilization. Window and feature-group comparisons examine these design choices. Third, controlled comparisons quantify observation selection, incremental recognition gains, and computational costs through equal-call routing comparisons, alternative confidence scores, window and feature ablations, host latency and process RSS measurements, and acquisition-level paired analysis. Additional held-out logistics recordings evaluate the fixed system beyond the original sites. These contributions concern a task-specific mechanism and its empirical evaluation, building on established conditional-inference and stabilization methods.

On the original held-out test, common Majority-3 gives 86, 72, and 71 errors for Static, Selective, and Always Temporal, respectively. Selective reduces residual errors by 14 (16.28%) while retaining two Human misses. Its all-frame temporal invocation ratio is 2.8611%, compared with 94.1226% for Always. Results and Discussion relate this observed increment to host latency, resident memory, routing blind spots, and acquisition-level uncertainty; two supplemental recordings extend the evaluation to one additional site.

Section 2 reviews related work, and Section 3 describes the method. Section 4 presents the experimental setup and evaluation protocol, followed by the results in Section 5. Section 6

discusses the mechanisms, computational trade-offs, and scope of the evidence, and Section 7 concludes the study.

2. Related Work

2.1. MmWave Radar Human Sensing and Point-Cloud Recognition

Radar human sensing spans presence detection, human–background discrimination, identity recognition, and activity analysis. These tasks have different objectives: detecting a person or distinguishing Human from Background does not identify that person or classify an activity. Range–Doppler maps, micro-Doppler spectra, point clouds, and multidomain representations likewise offer different information [1,18–20]. Public activity benchmarks and single-radar HAR studies examine comparison protocols and generalization [21,22]; their reported classification rates cannot be treated as direct performance references for the present binary task. At a finer sensing scale, Hu et al. investigate contactless eye-blink detection with UWB radar [23]. This study provides human-sensing context but uses a different radar modality and target behavior, rather than serving as a direct human–background classification baseline.

Representation approaches range from compact statistical features with SVM or RF to learned point-cloud encoders. RF and LSTM have been compared in radar classification [24]. RadHAR accumulates and voxelizes sparse observations [3], point-cloud recurrent networks combine spatial patterns with shape changes [6], and MiliPoint evaluates PointNet++, DGCNN, and PointTransformer on radar activity data [2]. PointNet++ learns hierarchical geometric representations, whereas PointMLP uses a residual MLP framework [25,26]. These encoders differ from compact feature-based classifiers in their representations and execution requirements; RF and Extra Trees also appear in resource-aware sensing [12,13,27]. PointNet++ and PointMLP are therefore useful external single-frame references here, while the routing experiments fix the RF backbone to isolate allocation from architectural differences. Sparse radar returns and changing point counts motivate both spatial representation and temporal context [2,3].

2.2. Temporal Modeling of Radar Observations

Temporal radar models exploit changes in scattering geometry, Doppler, and reflectivity that a single frame samples only instantaneously. Micro-Doppler methods describe velocity evolution [19]; point-cloud methods use accumulated windows, spatiotemporal tensors, or recurrent spatial features [3,4,6]. Multiview CNN–LSTM combines radar views with recurrent modeling, while MM-HAT and CubeLearn examine Transformer-based sensing and learning from raw signals, respectively [28–30]. milliFlow estimates motion between consecutive point clouds despite sparsity and changing scattering points [5]. These approaches demonstrate different ways of supplying historical context.

Efficiency-oriented work includes lightweight hybrid Vision Transformers [31], adaptive-thresholding HAR [15], and compact spatio-spectro-temporal, mmMotion, mPCT-LSTM, and RadMamba models [7–10]. Temporal windows and online execution policies vary, and the field also includes adaptive inference [32]. The present study asks whether a second learned temporal classifier should execute after the current-frame RF has produced a prediction. Prediction voting, exponential smoothing, and hysteresis use a different source of temporal information, suppressing output jitter without learned reclassification of radar features. Treating them as inexpensive controls allows additional feature-history processing to be assessed beyond output stabilization, rather than attributing all temporal benefit to the learned stage.

2.3. Lightweight Radar Inference and Input-Adaptive Computation

Model-level efficiency reduces the cost of an inference through compact representations, efficient convolution and attention, pruning, compression, or edge implementation [11,13–15,31]. Input-adaptive inference instead changes the computation performed for a particular observation. Dynamic neural networks can adapt structures or parameters at the instance, spatial, or temporal level, with cascaded models and early exits among the strategies for allocating effort [32].

Dynamic RF/GBT inference adjusts how many ensemble members execute according to the input and computational target [16]. In contrast, the present Static RF always runs completely to provide its prediction and intrinsic tree-level disagreement. The gate then decides whether a separate Temporal RF should use causal feature history to refine the current decision. The conditional transition thus occurs between classifiers rather than within the Static ensemble. A complete Static prediction remains available when no temporal call is made or history is invalid. Both forests remain resident; the avoided operation is additional Temporal RF inference, whose invocation ratio must be distinguished from total runtime, memory use, and energy consumption. Hardware-aware classifiers and implementation optimizations remain complementary to this allocation decision [12,27].

Recent ISAC and radio-frequency hardware work addresses other parts of the sensing system. Liu and Fischione use graph neural networks for coordinated beamforming in multi-cell ISAC [33], whereas Zhang et al. develop a compact CMOS SOI power amplifier for portable radar and ISAC front ends [34]. These studies concern transmission design and front-end hardware, respectively. They provide context for resource-aware sensing systems.

2.4. Uncertainty-Based Ensemble Inference and the Position of This Work

RF trees differ in their training samples and feature selections, providing information in the distribution of their probabilities beyond the ensemble mean. Shaker and Hüllermeier study aleatoric and epistemic uncertainty in probabilistic tree ensembles and evaluate uncertainty rankings through abstention [17]. This work builds on ensemble disagreement and confidence-based selection. Here the standard deviation of individual-tree Human probabilities is used as an empirical allocation signal, not as a calibrated error probability or a new uncertainty decomposition. High disagreement may occur on errors that Temporal RF cannot correct, while low disagreement can accompany incorrect Static predictions.

The action following selection distinguishes the mechanism from uncertainty-based rejection. For a history-valid frame whose disagreement exceeds the fixed threshold, the system obtains a Temporal RF prediction from the current features and short-term history; otherwise it retains Static output. It therefore returns a decision for every observation instead of abstaining on selected inputs. Selection triggers reclassification with additional historical information rather than confidence filtering alone. Relative to a generic cascade, the concrete choices are a fully executed Static forest, a gate reusing its tree probabilities without a separate learned routing network, and a second forest receiving a causal temporal extension. The mechanism combines established conditional-inference and ensemble-consistency ideas for this radar task.

The distinction also motivates comparison with inexpensive alternatives: confidence from the mean probability, predictive entropy, probability margin, and hard-vote entropy may allocate the same additional temporal budget differently. The shared-model, equal-call comparisons in Section 5.4 examine these choices rather than assuming tree disagreement is superior. Together with common Majority-3, the Always Temporal reference, and the host-cost and acquisition-level analyses, these controls test whether the selected observations yield an incremental recognition benefit and what practical cost accompanies it. Together, these evaluations relate temporal evidence and input-adaptive execution to recognition gains and measured costs.

3. Method

3.1. Problem Definition and Overall Framework

The task is conditional temporal computation allocation for mmWave radar human–background recognition. Static RF classifies every frame, and a temporal model supplies additional evidence for observations selected by the disagreement-based routing rule. Let the jth independent acquisition be

$$S_j = \left\{\left(P_{j,t}, y_{j,t}\right)\right\}_{t=1}^{T_j}. \tag{1}$$

Here T_j is the sequence length, $P_{j,t}$ is the point cloud, and $y_{j,t} \in \{0, 1\}$ denotes background or human, respectively. Omitting j when the acquisition is clear, write

$$P_t = \{p_{t,i}\}_{i=1}^{n_t}, \quad p_{t,i} = [x_{t,i}, y_{t,i}, z_{t,i}, v_{t,i}, l_{t,i}]^T. \tag{2}$$

Here n_t is valid point count, (x, y, z) gives spatial coordinates, v is Doppler velocity, and l is preprocessed logarithmic reflected power. Temporal state is reinitialized at acquisition changes or nonconsecutive frame IDs; historical evidence never crosses these boundaries.

Figure 1 summarizes the pathway. Static RF uses a 41-D current-frame representation to produce a human probability and tree-level disagreement in the same inference. Observations with disagreement above the routing threshold and valid history receive a 13-D causal extension and enter Temporal RF. Its probability replaces the Static probability only for routed observations. Thresholding and fixed causal Majority-3 then produce stabilized decisions. Feature-based reinterpretation and prediction-level stabilization are separate operations, allowing their effects to be compared independently.

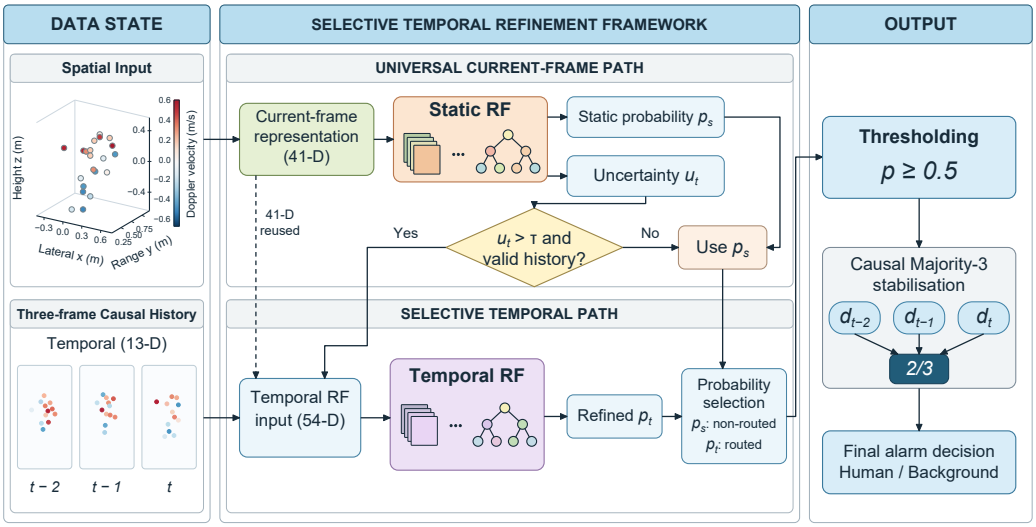

Figure 1. Overall architecture of uncertainty-guided selective temporal refinement.

This structure preserves a direct interpretation of each output: the static probability is the default, temporal reclassification is conditional, and stabilization is applied after class prediction. The separation is useful when comparing mechanisms because a gain from inexpensive voting cannot be mistaken for a gain from learned temporal features, and the latter can be examined with or without the stabilizer.

3.2. Compact Current-Frame Radar Feature Representation

Variable point counts and sparse spatial sampling motivate a compact statistical representation rather than fixed-size voxels. The features describe location, occupancy, geometry, range, radial motion, and reflectivity. They are jointly interpreted by Static RF rather than

subjected to separate hand-designed human-class thresholds. Point count describes the number of returns, while the centroid locates their reflection region:

$$c_t = \frac{1}{n_t} \sum_{i=1}^{n_t} [x_{t,i}, y_{t,i}, z_{t,i}]^T. \tag{3}$$

Coordinate standard deviations and bounding-box extents describe spatial dispersion and scale. Radial statistics are derived from three-dimensional and horizontal range:

$$r_{t,i} = \sqrt{x_{t,i}^2 + y_{t,i}^2 + z_{t,i}^2}, \quad r_{t,i}^{xy} = \sqrt{x_{t,i}^2 + y_{t,i}^2}. \tag{4}$$

The representation includes mean, standard deviation, minimum, and maximum three-dimensional range, and mean and standard deviation of horizontal range. Spatial compactness, covariance eigenvalues $\lambda_{1,t} \geq \lambda_{2,t} \geq \lambda_{3,t} \geq 0$, and shape descriptors summarize point-cloud geometry. These quantities complement location and point count without assuming that human returns always exhibit a fixed geometric form.

Doppler features describe velocity distribution, absolute velocity, and proportions of positive and negative velocities; reflected-power statistics describe central tendency, quartiles, and dispersion. Bounding-box volume and logarithmic point density describe spatial concentration. Table 1 gives the complete inventory, yielding

$$x_t \in \mathbb{R}^{41}. \tag{5}$$

Table 1. Composition of current-frame representation and temporal extension.

Feature Group	Description	Dimension
Point statistics	Number of detected points	1
Centroid position	Mean coordinates (x, y, z)	3
Spatial dispersion	Standard deviation of x, y, z	3
Spatial extent	Bounding-box extent along x, y, z	3
Range statistics	Mean, standard deviation, minimum and maximum of 3-D/XY range	6
Spatial compactness	RMS compactness descriptor	1
Covariance spectrum	Eigenvalues $\lambda_1, \lambda_2, \lambda_3$	3
Shape descriptors	Linearity, planarity, scattering and anisotropy	4
Doppler statistics	Velocity distribution statistics and signed/absolute velocity ratios	9
Reflectivity statistics	Mean, standard deviation, median, quartiles and IQR of reflected power	6
Volume and density	Bounding-box volume and logarithmic point density	2
Current-frame representation	—	41
Centroid displacement	Inter-frame centroid displacement	1
Temporal mean statistics	Three-frame causal means of point count, mean Doppler, mean absolute Doppler, mean log reflected power and RMS compactness	5
First-order differences	Adjacent-frame differences of point count, mean range, RMS compactness, mean Doppler, Doppler standard deviation, mean absolute Doppler and mean log reflected power	7
Temporal extension	—	13
Temporal RF input	Current-frame + temporal extension	54

This 41-D representation is used on every frame. Temporal RF retains all current-frame features and adds historical information. Static–Temporal comparisons therefore retain the same instantaneous evidence while examining short-term features and their allocation.

The nine Doppler descriptors include mean, standard deviation, median, minimum, maximum, mean and maximum absolute velocity, and positive/negative velocity proportions. The six reflectivity descriptors comprise mean, standard deviation, median, lower and upper quartiles, and interquartile range. Together, these descriptors characterize the distributions of velocity and reflected power. Table 1 lists their dimensions.

3.3. Compact Causal Temporal Representation

The temporal representation is causal and compact: it supplements Static RF without encoding a long raw point-cloud sequence. The Native K3 configuration uses the current observation and the previous two observations from the same acquisition. This short window provides recent feature levels together with adjacent-frame changes, while limiting buffering and the history needed before refinement becomes available. All inputs are available on arrival of the current frame, without waiting for future observations. Six current-frame quantities form the base vector

$$g_t = [n_t, \bar{r}_t, c_t^{\mathrm{rms}}, \bar{v}_t, \overline{|v|}_t, \bar{l}_t]^T \in \mathbb{R}^6. \tag{6}$$

Its components are point count, mean range, RMS compactness, mean Doppler, mean absolute Doppler, and mean logarithmic reflected power. Recent levels and adjacent-frame changes describe complementary aspects of short-term variation while avoiding temporal expansion of all 41 features. For three consecutive observations, define

$$\mu_t^{(3)} = \frac{1}{3}\sum_{k=0}^{2} g_{t-k}. \tag{7}$$

The mean is computed elementwise. The 13-D extension includes five causal means for point count, mean Doppler, mean absolute Doppler, mean log reflected power, and RMS compactness. Its seven first-order differences concern point count, mean range, RMS compactness, mean Doppler, Doppler standard deviation, mean absolute Doppler, and mean log reflected power.

Adjacent-frame centroid displacement describes movement of the reflection center:

$$d_t^c = \|c_t - c_{t-1}\|_2. \tag{8}$$

Let $g_{t,j}$ be component j of g_t, $\mu_{t,j}^{(3)}$ its three-frame mean, and $\sigma_{v,t}$ current-frame Doppler standard deviation. The seven first-order differences, one displacement, and five means are concatenated as

$$\begin{aligned}\phi_t = [&g_{t,1} - g_{t-1,1},\ d_t^c,\ g_{t,2} - g_{t-1,2},\ g_{t,3} - g_{t-1,3},\\ &g_{t,4} - g_{t-1,4},\ \sigma_{v,t} - \sigma_{v,t-1},\ g_{t,5} - g_{t-1,5},\ g_{t,6} - g_{t-1,6},\\ &\mu_{t,1}^{(3)},\ \mu_{t,4}^{(3)},\ \mu_{t,5}^{(3)},\ \mu_{t,6}^{(3)},\ \mu_{t,3}^{(3)}]^T \in \mathbb{R}^{13}.\end{aligned} \tag{9}$$

Appending this 13-D extension to the complete current-frame representation gives the 54-D Temporal RF input:

$$z_t = [x_t^T, \phi_t^T]^T \in \mathbb{R}^{54}. \tag{10}$$

The current and required historical frames must belong to the same acquisition and have consecutive original acquisition frame IDs. In Native K3, history validity is determined from this recorded continuity, without joining observations across acquisitions or across frame-ID gaps. At sequence starts, after gaps, or with insufficient history, Temporal RF is not invoked and the Static probability is retained. Neither future observations nor class labels determine history validity, and a label change alone does not interrupt a continuous acquisition. Training likewise uses only samples with valid three-frame history. Figure 2 illustrates consecutive human point clouds and a background example; these qualitative observations do not determine routing. Invocation depends on the disagreement and validity rule defined below.

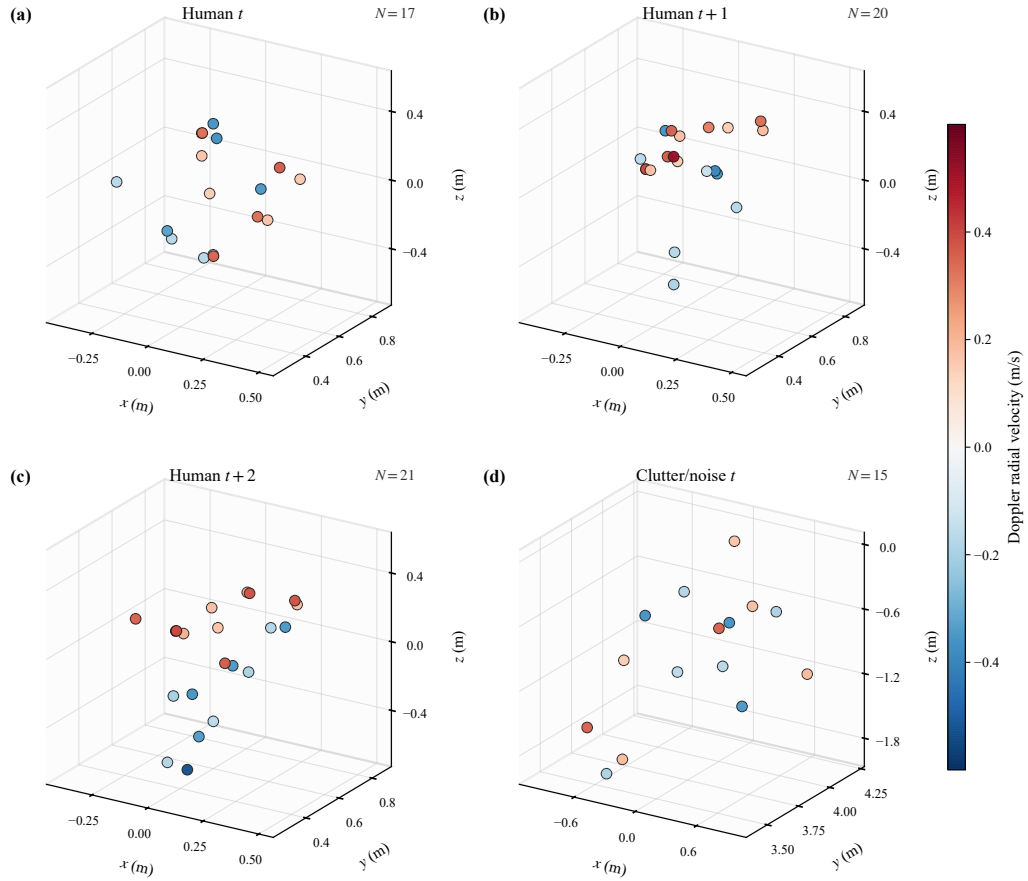

Figure 2. Representative mmWave point clouds from three consecutive human frames and one background frame: (**a**) human at t; (**b**) human at $t + 1$; (**c**) human at $t + 2$; and (**d**) background (clutter/noise) at t.

The selected means describe recent levels of point count, radial motion, reflected power, and spatial compactness. First-order differences capture immediate changes in these quantities, range, and Doppler dispersion, while centroid displacement describes movement of the reflection center. These descriptors supplement the full current-frame input without replicating all 41 statistics across the window. The window and feature-group comparisons evaluate this compact design: K5 produces fewer errors than K3 under the controlled common-K5 protocol, so K3 is a short-history choice rather than an accuracy optimum.

3.4. Static Random Forest and Intrinsic Ensemble Uncertainty

Static RF is the only learned classifier executed on every frame. Its $M_S = 300$ trees produce human-class probabilities $p^S_{t,m}$ from x_t, whose mean is

$$p^S_t = \frac{1}{M_S} \sum_{m=1}^{M_S} p^S_{t,m}. \tag{11}$$

Because trees differ in training samples and feature selection, their prediction dispersion describes internal consistency [17]. The dispersion describes differences among tree probabilities rather than whether the ensemble prediction is correct. Tree-probability disagreement is quantified by their standard deviation, which serves as an empirical routing signal:

$$u_t = \sqrt{\frac{1}{M_S} \sum_{m=1}^{M_S} (p^S_{t,m} - p^S_t)^2}. \tag{12}$$

First- and second-order probability statistics can be accumulated during the same forest traversal, requiring scalar operations over already available outputs rather than

another classification network. This reuses intrinsic ensemble information to determine whether Temporal RF is invoked. The matched-random experiment tests the error reduction associated with routing by this signal.

Static RF uses training-data-based missing-value handling, class-balanced ensemble learning, and a fixed class threshold of 0.5. It provides current-frame recognition and accessible tree-level predictions. The same static model remains the common backbone throughout temporal comparisons. Tree-probability disagreement affects routing through u_t, whereas the mean probability determines the Static class decision. These quantities serve different purposes even though both are derived from the same tree outputs. High-disagreement observations are candidates for additional evidence, whose effect is determined by Temporal RF.

3.5. Uncertainty Gating and Selective Temporal Refinement

Let $v_t^{(3)}$ indicate valid and available three-frame history under the continuity rule in Section 3.3. In the additional history-availability simulation described in Section 4.5, the required historical frames must also remain available to the temporal representation. Routing is deterministic:

$$r_t = v_t^{(3)} \mathbb{I}(u_t > \tau), \quad r_t \in \{0, 1\}. \tag{13}$$

The strict inequality and history condition must both hold. When $r_t = 0$, the Static probability is retained and Temporal RF does not execute. Otherwise, the 300-tree Temporal RF processes the 54-D input and returns p_t^T. The pre-stabilization Selective probability is

$$p_t = (1 - r_t) p_t^S + r_t p_t^T. \tag{14}$$

This is branchwise probability selection, not continuous weighted fusion: routed frames use the Temporal probability directly. An uninvoked model is not evaluated, and an undefined temporal probability is never accessed when history is invalid. The Always Temporal reference instead invokes the same temporal model for every valid-history observation:

$$p_t^A = (1 - v_t^{(3)}) p_t^S + v_t^{(3)} p_t^T. \tag{15}$$

Always retains Static fallback for invalid history. The models, feature representations, and classification threshold are identical between Selective and Always; only invocation sets differ. Always measures the correction obtained by applying the shared Temporal RF throughout the eligible set. The routing threshold is selected using grouped development predictions and remains unchanged during held-out testing, so the resulting test invocation rate is observed rather than prescribed from the test curve. The comparison therefore measures how much of the shared Temporal RF's correction is retained by selecting only a subset of eligible frames, with model capacity and representation held fixed.

3.6. Fixed Causal Majority-3 Stabilizer

Output voting can suppress brief prediction reversals without reclassifying radar features. The same fixed causal Majority-3 is applied after frame-level classification to Static, Selective, and Always pathways. Thresholding gives

$$d_t = \mathbb{I}(p_t \geq 0.5).$$

When three consecutive decisions are available,

$$\hat{y}_t = \mathbb{I}(d_t + d_{t-1} + d_{t-2} \geq 2). \tag{16}$$

Only the current and previous two outputs are used. With insufficient decision history, the original frame-level decision is retained; majority history resets at acquisition changes or

actual continuity breaks. The additional simulation of unavailable temporal-feature history does not remove current observations or reset the decision-history buffer while the recorded stream remains continuous. Majority-3 suppresses isolated prediction reversals through output consistency, whereas Temporal RF uses short-term feature changes to reclassify uncertain observations. Keeping the stabilizer fixed makes it possible to assess learned refinement beyond the same causal output-history baseline. Applying this operation equally to all paths isolates the final increment from learned feature-history refinement, while pre-stabilization comparisons measure its direct effect.

3.7. Temporal Invocation Accounting

Additional Temporal RF execution is measured by its all-frame invocation ratio over all N test frames:

$$\rho_{\text{sel}} = \frac{1}{N}\sum_{t=1}^{N} r_t, \quad \rho_{\text{always}} = \frac{1}{N}\sum_{t=1}^{N} v_t^{(3)}. \tag{17}$$

For $N_{\text{elig}} = \sum_{t=1}^{N} v_t^{(3)}$ history-valid and available frames, the eligible-frame routing rate is $\rho_{\text{sel,elig}} = (\sum_{t=1}^{N} r_t)/N_{\text{elig}}$. This conditional rate describes the fraction of eligible observations that invoke Temporal RF. The denominators in Equation (17), however, are all test frames in both cases. Relative temporal invocation reduction is

$$R_{\text{temp}} = 1 - \frac{\rho_{\text{sel}}}{\rho_{\text{always}}}. \tag{18}$$

Static RF executes on every frame. These ratios quantify the frequency of additional Temporal RF invocation, with both complete forests retained.

4. Experimental Setup and Evaluation Protocol

4.1. Data Acquisition and Sequence-Level Experimental Partitioning

Data were acquired using an Andar ADT6101 FMCW millimeter-wave radar (Hangzhou Andar Technologies Co., Ltd., Hangzhou, China) with an integrated ARM Cortex-M3 processor and a 2-Tx/2-Rx TDM-MIMO configuration. The embedded development environment was Keil μVision 5. The device's nominal 57–64 GHz operating band is distinct from the acquisition configuration used here: the sweep starts at 58 GHz with a configured bandwidth of 2000 MHz. Table 2 reports the actual firmware configuration, including the register-aligned bandwidth and timing. The radar was mounted approximately 1.0 m above the ground, with typical human passages at approximately 3 m. The original and supplemental acquisitions use the same hardware, nominal acquisition settings, and installation conditions. The Cortex-M3 is internal to the radar platform; the dual-RF inference measurements in Section 5.3 were performed on the host, not on this processor.

Table 2. Actual ADT6101 acquisition configuration used in the recordings.

Parameter	Acquisition Setting
Start frequency	58 GHz
Sweep bandwidth	2000 MHz configured; approximately 1998.679 MHz register-aligned
Chirp timing	T_0 = 96 µs (aligned to approximately 95.96 µs), T_1 = 120 µs, T_2 = 10 µs
Aligned chirp period	Approximately 225.96 µs
Effective chirps per frame	64 total; 32 per transmit channel
MIMO scheme	2-Tx/2-Rx TDM-MIMO
Raw ADC sampling rate	Approximately 7.142857 MS/s
FIR decimation	Factor of 2
Post-decimation sampling rate	Approximately 3.571429 MS/s
ADC samples	256 per chirp per receive channel
Frame period and rate	50 ms; 20 Hz

The 256 ADC samples per chirp and receive channel are signal-acquisition samples, distinct from the sparse point-cloud detections produced by signal processing. At the 50-ms frame interval, Native K3 spans approximately 100 ms between its oldest and current frames when sampling is continuous. This is a causal historical span, not the complete end-to-end alarm latency. Figure 3 shows the radar hardware and representative acquisition configurations, including human-target observations and background recordings under repositioned and stationary platform conditions.



Figure 3. Radar platform and acquisition settings. (**a**) Human-target recording in an underground parking environment (radar height approximately 1.0 m); (**b**) ADT6101 FMCW module and acquisition hardware; (**c**) Background recording during radar-platform repositioning; (**d**) Background recording with stationary radar and non-human target; (**e**) Human-target recording in a distinct indoor public area.

The original dataset comprises 40 independent recording sessions from two physical sites: 20 from an underground garage and 20 from a laboratory building. These complete acquisitions retain the fixed 32-sequence training/development and eight-sequence held-out partition. Human sequences contain a person entering or occupying the sensing region, whereas background sequences contain non-human targets and environmental returns without a person. Ground-truth labels were assigned frame by frame with reference to the corresponding video recordings: label 1 denotes a person within the radar sensing region, and label 0 denotes no person in that region.

Table 3 summarizes 40 sequences: 32 training/development sequences contain 66,968 frames, and 8 held-out sequences contain 16,742 frames, including 8250 human and 8492 background frames from four sequences each.

Table 3. Dataset composition and sequence-disjoint evaluation protocol.

Partition	Sequences	Human Frames	Background Frames	Total Frames
Training/development	32	33,251	33,717	66,968
Held-out human test	4	8250	0	8250
Held-out background test	4	0	8492	8492
Held-out total	8	8250	8492	16,742

The development and test sets are

$$S_{\text{dev}} = \{S_j\}_{j=1}^{32}, \quad S_{\text{test}} = \{S_j\}_{j=33}^{40}$$

and satisfy

$$S_{\text{dev}} \cap S_{\text{test}} = \emptyset. \tag{19}$$

Partitioning by complete acquisition prevents neighboring observations from the same recording from appearing in both development and testing. Temporal statistics remain within sequences and restart after discontinuities, as specified in Section 3. Adjacent frames share background layout, installation conditions, and motion trajectories. The 16,742 paired test outcomes therefore represent eight independent acquisitions rather than 16,742 independent experimental conditions.

Two additional held-out acquisitions were recorded at a parcel logistics station outside both original acquisition sites. The location contains dense parcels, metal shelving, and normal personnel and logistics activity; these characteristics describe the setting rather than measured multipath, occlusion, or clutter severity. Each acquisition is a separate continuous recording. In L01, a person enters the sensing region from side A and exits at side B, then enters from B and exits at A; L02 follows the reverse order. The direction labels describe passage order at this single site. The two recordings were acquired independently, without artificial noise or subdivision of one recording. Labels follow actual entry and exit: a background label is assigned only when no person is within the sensing region, including other personnel at the station. Label transitions within a continuous acquisition do not reset temporal history; independent acquisition boundaries and actual recording discontinuities do, as specified in Section 3. Historical eligibility is not determined using future labels, and histories are never joined across acquisitions.

These supplemental acquisitions contribute neither to Static or Temporal RF training nor to imputation fitting, development-threshold selection, or hyperparameter adjustment. Evaluation uses the existing models, preprocessing, fixed thresholds, and causal output rules. They are evaluated separately in Section 5.7 and are not added to the original 40 acquisitions or used to repeat the 32/8 partition in Table 3.

4.2. Model Training and Parameter Selection

Model fitting, imputation statistics, routing-threshold selection, and EMA and Hysteresis parameter selection were performed using only the development acquisitions. Static RF uses the 41-D current-frame representation, whereas Temporal RF uses the 54-D representation for samples with valid causal history. Both classifiers use a fixed decision threshold of 0.5. Four-fold StratifiedGroupKFold was used to generate grouped out-of-fold (OOF) development predictions, with complete acquisition sequences kept within individual folds.

The routing threshold was determined from the grouped OOF predictions under a predefined maximum temporal-invocation constraint. Candidate thresholds were ranked

primarily by balanced accuracy, with Recall, FPR, and lower temporal activation used sequentially to resolve ties. The resulting threshold was

$$\tau = 0.4665. \tag{20}$$

The implemented threshold is $\tau = 0.46647615158762407$; the value in Equation (20) is rounded only for presentation.

4.3. Evaluation Metrics and Error Definitions

Human is treated as the positive class. FN denotes a human miss and FP a background false alarm. Recall and FN are reported alongside aggregate errors to distinguish Human misses from Background false positives. The metrics are

$$\text{Recall} = \frac{TP}{TP + FN}, \quad \text{FPR} = \frac{FP}{FP + TN}, \quad \text{FA/1000} = 1000\text{FPR}. \tag{21}$$

Here FA/1000 measures false alarms per 1000 background observations. Balanced accuracy and total errors are

$$\text{BA} = \frac{1}{2}\left[\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right]. \tag{22}$$

$$E = FN + FP. \tag{23}$$

For history-valid Static errors $\mathcal{E}_{S,\text{elig}}$ and routed set $\mathcal{R}$, define

$$C_{\text{err}} = \frac{|\mathcal{E}_{S,\text{elig}} \cap \mathcal{R}|}{|\mathcal{E}_{S,\text{elig}}|}. \tag{24}$$

This measures error coverage, not the error rate of routed frames. Under uniform random selection, expected capture equals the selected fraction of eligible frames. For all-frame error counts E_S, E_{sel}, and E_A, retained benefit is

$$B_{\text{ret}} = \frac{E_S - E_{\text{sel}}}{E_S - E_A}. \tag{25}$$

The denominator is the observed improvement of the shared Always Temporal reference. Together, capture and retained benefit distinguish selecting erroneous Static predictions from correcting them. All-frame error counts use exactly the same test observations for Static, Selective, Always, and random routing. Their differences therefore quantify net corrections or introduced errors under each strategy. Eligible-frame capture uses a different set restricted by history validity; its denominator must not be confused with total test errors or with the number of temporal invocations.

4.4. Single-Frame Classification Baselines

Standard RF uses the Static feature representation. SVM provides a non-tree statistical-learning reference, while PointNet++ and PointMLP classify raw point clouds independently using their respective representations and training pipelines. Preprocessing, fitting, model selection, and early stopping exclude test data. SVM and the deep baselines receive no proposed temporal module; subsequent mechanism comparisons retain the same RF backbone and vary temporal processing only.

RF/SVM and host inference used Python 3.11.7, NumPy 1.26.4 and scikit-learn 1.2.2; neural baselines used Python 3.9.23, NumPy 2.0.2 and PyTorch 2.5.1 (CUDA 12.1). The recorded neural baselines use 512 points per frame, sampled with replacement when necessary, without pooling points across frames. PointNet++ uses the SSG architecture with XYZ, log-power, Doppler, and range inputs. It uses Adam with learning rate 10^{-3} and

weight decay 10^{-4}, and a fixed 20-epoch training budget with StepLR (step 20, factor 0.7). Its final-epoch checkpoint is evaluated without development-based checkpoint selection or early stopping. Training-only XYZ jitter and Z rotation are applied after sampling; held-out inference uses no augmentation.

A fresh copy of the PointMLP implementation was obtained from the code repository accompanying its paper. Its input and training procedures were adapted to the present mmWave radar human–background recognition task, and the model was retrained. The task-adapted implementation uses the same input representation as PointNet++, with seed 2024, batch size 24, and 100 training epochs using Adam (learning rate 10^{-5}, weight decay 10^{-5}) and cosine scheduling ($T_{\max} = 100$, minimum learning rate 10^{-6}), without early stopping. Within the original 32-acquisition training/development pool, fold 0 of a four-fold stratified grouped split assigns 24 acquisitions to training and eight to development. Checkpoints are selected by full-development BA, then lower development loss, then earlier epoch; held-out test data are excluded from this selection.

4.5. Temporal Controls and Selective Computation Experimental Design

Static uses no history. Majority-3 performs causal voting, EMA smooths successive probabilities exponentially, and Hysteresis uses upper/lower thresholds to suppress state changes. EMA and Hysteresis parameters are selected from development OOF predictions. These output-level controls differ from learned feature-history reclassification. Selective and Always share both forests, the 54-D temporal input, and the class threshold, differing only in which history-valid frames invoke Temporal RF. Final evaluation applies common Majority-3:

$$\text{Static} + \text{Maj3}, \quad \text{Selective} + \text{Maj3}, \quad \text{Always} + \text{Maj3}. \tag{26}$$

Static versus Always measures the available temporal correction; output controls measure inexpensive stabilization; Selective versus Always evaluates computation allocation. Common stabilization then tests whether selective refinement adds benefit beyond prediction-history processing. Holding both forests, their feature representations, and the classification threshold fixed isolates routing from differences in model capacity or decision sensitivity. The design thus separates temporal evidence, allocation of learned processing, and output stabilization.

Window and feature-group comparisons examine the selected causal representation under their respective evaluation protocols. The controlled window comparison uses a common K5-valid evaluation set, and its K3 condition is distinguished from the Native K3 configuration. Feature-group ablations retain the 41-D current-frame representation and vary the temporal extension. These comparisons examine the selected design without redefining the Native K3 operating point.

The additional history-availability experiment uses fixed-seed masks at 0, 5, and 10% unavailability conditions. All current observations remain in the stream, but masked past frames cannot contribute to Temporal RF history. Eligibility is recomputed under each mask, and observations without the required available history retain Static predictions. The trained forests, routing threshold, and classification threshold remain fixed. Actual acquisition boundaries and frame-ID gaps continue to reset temporal and decision history, whereas the simulated restriction on access to historical features does not clear the common Majority-3 buffer during an otherwise continuous stream.

4.6. Matched-Random Routing Control and Uncertainty-Budget Diagnostic

Matched-random routing samples the same number of observations as Selective from the same history-valid set, retaining the models, probability replacement, and class threshold. Repeating selection 1000 times yields distributions of errors, Static-error capture,

and retained benefit. Reported random intervals are 95% empirical intervals rather than full trial ranges.

The diagnostic routes observations in descending tree-probability disagreement order at eligible-frame target budgets of 0, 1, 3, 5, 10, and 20%, recording errors, capture, and retained benefit. Each target budget is defined relative to the history-valid set, rather than all test frames used for the invocation ratios defined in Section 3.7.

4.7. Acquisition-Level Statistical Evaluation

The original eight complete held-out acquisitions are the independent sampling units for statistical evaluation; their 16,742 correlated frames are not independent acquisition conditions. The stratified acquisition-cluster bootstrap uses 10,000 replicates with fixed seed 2026, resampling complete acquisitions with replacement within the Human and Background strata while retaining the original four-acquisition count in each stratum. Every selected acquisition contributes all its frames, preserving within-sequence dependence. Differences in Recall, FPR, FA/1000, BA, Accuracy, error rate, and error counts are recomputed for Static + Majority-3 versus Selective + Majority-3 to obtain 95% cluster-bootstrap intervals. Sequence-wise comparisons describe where the observed final improvement occurs.

4.8. Failure-Case Decomposition Protocol

Paired failure decomposition uses Static and Selective predictions before Majority-3 to separate learned refinement from stabilization:

$$\begin{aligned} C &= \{t : d_t^S \neq y_t, d_t^{\text{sel}} = y_t\}, \\ I &= \{t : d_t^S = y_t, d_t^{\text{sel}} \neq y_t\}, \\ F &= \{t : d_t^S \neq y_t, d_t^{\text{sel}} \neq y_t\}. \end{aligned} \tag{27}$$

Here C contains corrected Static errors, I newly introduced errors, and F shared failures. Net reduction is

$$\Delta E_{\text{net}} = |C| - |I|. \tag{28}$$

The sets are further separated into human misses and background false positives and aggregated by acquisition. This distinguishes background suppression from an increase in human misses. The same paired definitions are also applied to the final decisions $\hat{y}_t$ after the common causal Majority-3, keeping their stage distinct from the pre-stabilization sets above. Pre-Majority-3 correction measures learned feature-history reclassification, whereas Post-Majority-3 correction measures changes in final outputs. Final changes are additionally separated according to whether Temporal RF is invoked on the current frame. Changes at non-routed frames can arise because a refined earlier decision remains in the current three-decision voting window; this is forward propagation through causal output history, not influence from future frames. Representative point clouds and prediction transitions provide qualitative context rather than evidence of specific physical failure causes.

4.9. Host Inference Measurement Protocol

The host measurements concern inference from precomputed current-frame and temporal features, excluding raw point-cloud feature extraction. The runtime diagnostics measure imputation, RF inference, routing where required, and causal Majority-3 using `perf_counter_ns`; warm-up, context restoration, file I/O, and correctness checks are outside the timed interval. RF execution uses `n_jobs=1` and BLAS/OpenMP uses one thread, with no refitting or change to the stored model parameters. Fused Selective obtains the Static mean probability and population standard deviation from one traversal of the individual-tree probabilities; this fusion does not require individual probabilities to be binary. Optimized Always invokes Temporal RF directly on history-valid frames and uses Static only

for fallback on invalid history, avoiding redundant Static inference on eligible frames. Thus, the latency comparison concerns these explicit execution paths rather than a Static-first implementation of Always.

The host benchmark uses five repeated rounds on fixed continuous segments covering all eight acquisitions, with causal voting context restored outside the timed interval. These segments provide a common measurement workload rather than a random sample or full-test timing pass. Fused Selective latency and RSS are obtained under the same measurement procedure. Process RSS is measured after warm-up, and the sampled peak uses 20-ms polling, which can miss instantaneous peaks.

5. Results and Analysis

The comparisons evaluate single-frame classification, temporal refinement, routing, and performance after common stabilization. Total errors are reported together with Human misses and Background false positives.

5.1. Single-Frame Baseline Comparison

Table 4 compares single-frame classifiers without temporal refinement or output stabilization on the same 16,742 held-out frames, comprising 8250 Human and 8492 Background frames. All derived metrics are calculated from the reported FN and FP counts under these common class denominators.

Table 4. Single-frame recognition performance on the held-out acquisitions.

Method	Recall (%)	FPR (%)	FA/1000	BA (%)	Accuracy (%)	FN	FP	Errors
Static RF	99.9152	2.5907	25.9067	98.6622	98.6441	7	220	227
SVM	99.9030	2.2727	22.7273	98.8152	98.7994	8	193	201
PointNet++	99.4061	1.6839	16.8394	98.8611	98.8532	49	143	192
PointMLP	99.1758	2.0137	20.1366	98.5810	98.5725	68	171	239

Static RF has 99.9152% Human Recall, with 7 misses and 220 background false positives. SVM has 201 total errors and 8 misses, while PointNet++ has the fewest total errors among these references (192), but more Human misses than Static RF (49 versus 7). The task-adapted PointMLP yields 68 misses and 171 false positives, totaling 239 errors. These trade-offs motivate reporting FN and FP alongside total errors.

RF's observed Human Recall and accessible ensemble disagreement motivate its use as the fixed backbone for the controlled temporal-routing study. Its residual errors are predominantly background false positives, leaving scope for background-error reduction without replacing the single-frame classifier. The following comparisons keep this backbone fixed to distinguish learned feature-history refinement from inexpensive prediction-history stabilization.

5.2. Incremental Benefit Beyond Causal Output Stabilization

Majority-3 reduces Static errors from 227 to 86, providing the common prediction-history baseline for evaluating learned refinement. EMA and Hysteresis yield 104 and 206 errors, respectively (Table 5). These controls suppress output reversals without reclassifying radar features. EMA and Hysteresis parameters are selected on development data, as specified in Section 4.5, while the same fixed Majority-3 is applied to Static, Selective, and Always in the final comparison.

Table 5. Inexpensive output-history controls applied to the same Static RF predictions.

Output Processing	Recall (%)	FPR (%)	FA/1000	BA (%)	Accuracy (%)	FN	FP	Errors
Static RF	99.9152	2.5907	25.9067	98.6622	98.6441	7	220	227
Majority-3	99.9758	0.9892	9.8917	99.4933	99.4863	2	84	86
EMA	99.9636	1.1894	11.8935	99.3871	99.3788	3	101	104
Hysteresis	99.9636	2.3905	23.9049	98.7866	98.7696	3	203	206

Table 6 groups the three RF paths by evaluation stage; Selective and Always use the same Temporal RF and all paths use the same classification threshold. Before Majority-3, errors decrease from 227 for Static to 169 for Selective and 161 for Always. Selective removes 58 errors (approximately 25.6%), with Human misses falling from 7 to 3; Always has 2 misses and removes eight additional errors relative to Selective. Thus, pre-stabilization refinement reduces Background false positives without reducing Human Recall. The 169 Selective errors precede Majority-3, whereas the 86 Static errors follow it.

Table 6. RF system performance before and after the same causal Majority-3. Invocation percentages use all 16,742 test frames.

System	Recall (%)	FPR (%)	FA/1000	BA (%)	FN	FP	Errors	Temporal Calls	Invocation (%)
Before Majority-3									
Static	99.9152	2.5907	25.9067	98.6622	7	220	227	0	0.0000
Selective	99.9636	1.9548	19.5478	99.0044	3	166	169	479	2.8611
Always	99.9758	1.8724	18.7235	99.0517	2	159	161	15,758	94.1226
After common Majority-3									
Static	99.9758	0.9892	9.8917	99.4933	2	84	86	0	0.0000
Selective	99.9758	0.8243	8.2431	99.5757	2	70	72	479	2.8611
Always	99.9758	0.8125	8.1253	99.5816	2	69	71	15,758	94.1226

After common Majority-3, Static, Selective, and Always produce 86, 72, and 71 errors, respectively. Selective removes 14 errors relative to Static, entirely through fewer background false positives (84 to 70), while Human misses remain at 2. The 16.28% reduction is relative to the Static error count, not a 16.28-percentage-point increase in Accuracy; the corresponding Accuracy increase is approximately 0.0836 percentage points. Always produces one fewer final error than Selective.

Figure 4 places the final outcomes alongside Temporal RF use: Selective invokes the temporal classifier 479 times, whereas Always invokes it on all 15,758 history-valid frames. Their all-frame invocation ratios are 2.8611% and 94.1226%, respectively; the Selective eligible-frame routing rate is separately 3.04%. The matched-random and budget diagnostics in Section 5.4 address routing allocation separately. The main result is fewer residual Background errors beyond common causal stabilization with sparse Temporal RF execution. Its computational cost is examined next.

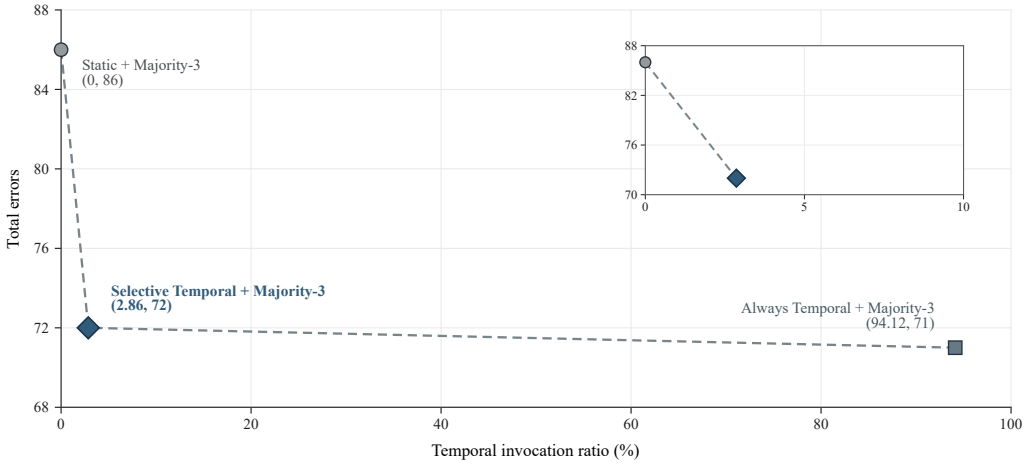

Figure 4. Reliability–computation trade-off.

5.3. Host Latency and Memory Costs

Table 7 reports host latency and process RSS under the protocol in Section 4.9. The latency paths are Static, Fused Selective, and Optimized Always: Selective derives the Static mean and tree disagreement in a shared traversal, whereas Always evaluates Temporal RF directly on eligible frames and Static only where history is invalid. Optimized Always therefore avoids redundant Static inference on eligible frames; the comparison concerns these specific implementations.

Table 7. Host inference latency and process memory. Latency is in ms/frame and memory in MiB.

Metric	Static	Fused Selective	Optimized Always
Mean latency	10.638	10.642	11.681
P95 latency	12.382	15.381	15.918
Warm RSS	183.352	193.508	193.297
Sampled peak RSS	184.500	195.141	194.410

Note. Latency and RSS for Fused Selective use the same measurement procedure. RSS is a process working-set measurement rather than isolated model RAM. Sampled peaks use 20-ms polling.

The host latency measurements give a Selective mean approximately 8.9% below Optimized Always, while its mean is close to Static (10.642 versus 10.638 ms/frame). Its P95 is approximately 24.2% higher than Static, despite being approximately 3.37% below Always.

Selective and Always retain Temporal RF in memory. Both dual-model processes have RSS approximately 10 MiB above Static, with little difference between them. Process RSS includes the runtime and preloaded data as well as the models; it is not isolated model RAM. The 96.96% reduction in Temporal RF calls therefore describes invocation count, not an equivalent reduction in total latency, CPU utilization, or energy. The measurements cover host inference from precomputed features, excluding radar-board deployment and the raw-sensor-to-alarm processing chain.

5.4. Evidence for Disagreement-Based Allocation

Table 8 compares routing signals at the same budget of 479 Temporal RF calls, using the same Static RF, Temporal RF, held-out test set, and history-valid candidate set. All results in this subsection are evaluated before Majority-3, so that they measure allocation for learned reclassification without the additional effect of output stabilization. Tree-probability standard deviation yields 169 errors, compared with 171 for the three mean-confidence scores and 177 for hard-vote entropy.

Table 8. Routing-signal comparison at 479 Temporal RF calls, before Majority-3. Matched-random errors are the mean of 1000 trials.

Routing Signal	Total Errors
RF tree-probability standard deviation	169
Distance of mean probability from 0.5	171
Mean predictive entropy	171
Probability margin	171
Hard-vote entropy	177
Matched random (1000-trial mean)	225.05

The three mean-confidence scores are monotonically related. For a binary mean probability p, predictive entropy decreases monotonically with distance $|p - 0.5|$, while the two-class probability margin equals $2|p - 0.5|$. Prioritizing higher entropy or smaller distance/margin therefore follows the same mean-confidence ordering, apart from the handling of ties. Here mean predictive entropy denotes entropy computed from the mean predictive probability. Tree-probability standard deviation instead uses differences among individual-tree probabilities, whereas hard-vote entropy discretizes those probabilities into tree decisions.

Matched-random routing selects 479 history-valid frames per trial across 1000 repetitions. Models and processing rules are identical to Selective, so the comparison isolates observation selection at equal invocation counts rather than greater temporal model use.

Figure 5 shows 169 errors for Selective versus a random mean of 225.05, close to Static. Selective captures 70.33% of eligible Static errors; random capture averages 3.02%, with a 95% empirical interval of 0.96–5.74%. Retained benefit is 87.88% for Selective versus a random mean of 2.96%. Capture measures whether routing selects eligible Static errors, while retained benefit measures their net reclassification gain relative to the same Always reference. The empirical random-routing distribution conditions on this fixed dataset and varies routing selections; it is distinct from resampling independent acquisitions in the cluster-bootstrap analysis. Together, these comparisons show that selecting frames by disagreement reduces errors relative to matched-random routing on this fixed test set.

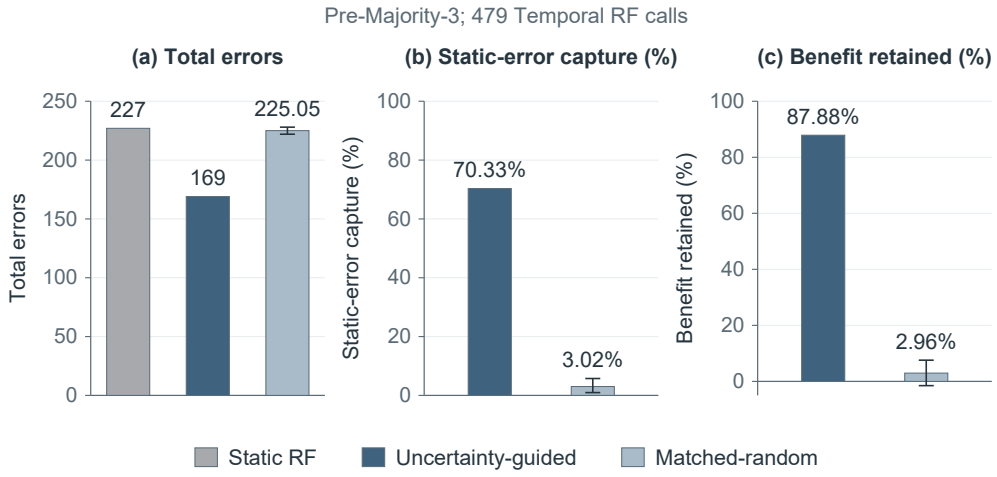

Figure 5. Pre-Majority-3 uncertainty-guided versus matched-random routing at 479 Temporal RF calls. Error bars show the 2.5th and 97.5th percentiles of 1000 random-routing trials on the fixed test set: 95% empirical intervals, not acquisition-bootstrap confidence intervals.

Figure 6 examines how errors, Static-error capture, and retained benefit change as progressively more history-valid observations are routed.

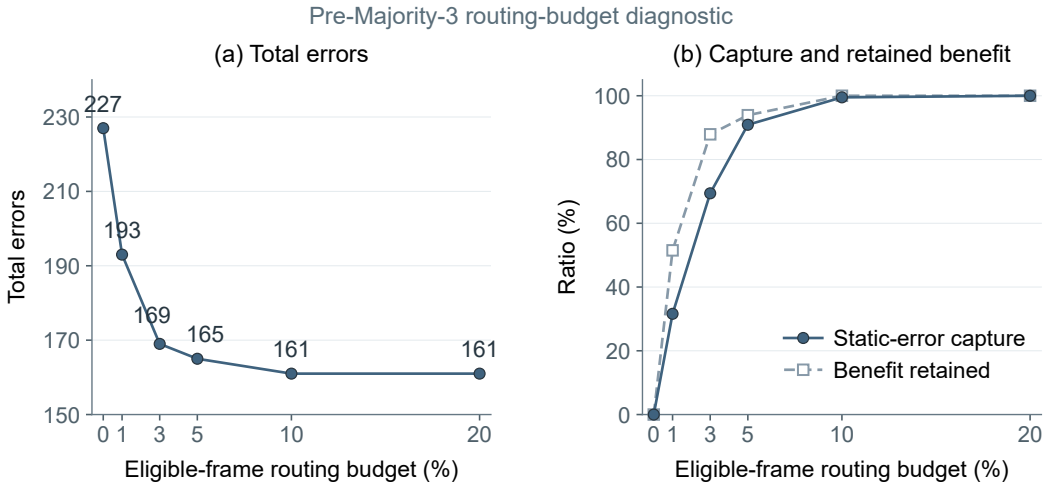

Figure 6. Pre-Majority-3 routing-budget diagnostic. The horizontal axis gives eligible-frame budgets (0, 1, 3, 5, 10, and 20%) relative to 15,758 Native K3 history-valid frames. The development-selected threshold is not reselected from this curve.

The budget diagnostic shows that benefit is concentrated at low eligible-frame budgets. Static-error capture exceeds 90% near 5%, while changes become smaller at budgets of 10% or higher. The horizontal axis uses the 15,758 Native K3 history-valid frames as its denominator. At the fixed development-selected threshold, the 479 Selective calls correspond to an eligible-frame routing rate of approximately 3.04%, but to an all-frame invocation ratio of 2.8611% over all 16,742 test frames. The diagnostic 3% budget is consequently not identical to the observed main-system operating point. The matched-random control evaluates selection at a fixed invocation count, whereas the budget curve varies that count. The next comparison examines the historical representation supplied to selected frames.

5.5. Compact Causal Representation and History Availability

Table 9 separates window length from feature-group composition under their respective evaluation protocols. Panel A uses the Controlled Common-K5 history-eligibility protocol for all windows, with causal Majority-3 applied consistently. K5 has fewer errors than K3 for both Selective (69 versus 77) and Always (65 versus 74), so K3 is not the accuracy-optimal window in this comparison. Native K3, whose final errors are 72 and 71 in Table 6, uses a different history-eligibility protocol; comparing its 72 Selective errors directly with Controlled K3's 77 would not isolate window length. K3 remains the selected shorter-history configuration, requiring two preceding frames rather than four for K5. The shorter history requires less feature buffering.

Panel B examines the compact temporal extension described in Section 3.3. Mean features describe the recent state level, whereas differences and displacement reflect short-term changes. The full 13-D combination produces fewer errors than either group alone under the tested configuration, supporting their complementary contribution under this configuration. The 41-D condition in this ablation is an input to the temporal-classifier path and is not the original Static RF prediction path; its 95 errors must not be equated with the 86 errors of Static + Majority-3. The panels isolate feature composition and window length under their respective protocols.

Table 9. Causal window and feature-group comparisons after common Majority-3. Panels A and B follow different evaluation protocols and must be interpreted within each panel.

Configuration	Selective Errors	Always Errors
A. Controlled Common-K5 window comparison		
K2	106	98
K3	77	74
K5	69	65
B. Temporal-branch feature-group ablation		
41-D current-frame input	95	95
46-D: current frame + means	84	81
49-D: current frame + dynamics	85	83
54-D: current frame + full 13-D extension	72	71

The history-availability sensitivity analysis retains every current frame for evaluation, as defined in Section 4.5. At 0, 5, and 10% simulated unavailability, valid K3 frames number 15,758, 14,244, and 12,786, and Selective calls number 479, 431, and 393, respectively. Selective FN/FP/errors are 2/70/72, 2/70/72, and 2/68/70; the corresponding Always results are 2/69/71, 2/69/71, and 2/67/69, while Static remains at 86 errors in all three conditions. Ineligible temporal histories retain the Static output. The simulated masks do not additionally clear the Majority-3 prediction buffer, unlike actual acquisition boundaries or frame-ID gaps, and therefore do not represent sensor packet loss. The experiment characterizes historical-feature unavailability rather than demonstrating improved recognition from missing history or robustness to physical frame loss.

5.6. Acquisition-Level Gains, Failure Patterns, and Statistical Uncertainty

Table 10 reports final outcomes for each of the eight original held-out acquisitions after common causal Majority-3. Only S07 and S08 show a net reduction, by eight and six errors, respectively, and both are Background acquisitions. They account for 78 of Static's 86 final errors, so the observed aggregate benefit is concentrated where most residual background errors occur. S01 and S03 have no final errors, whereas S02 and S04 each retain one Human miss, giving two Human misses for all three paths. The other six acquisitions have unchanged final error counts, including six shared errors in S06. The data therefore support localized background-false-positive reduction rather than improvement in every acquisition.

Table 10. Final paired outcomes on the original eight held-out acquisitions after common Majority-3. C and I denote final Selective corrections and introduced errors relative to Static; net reduction is $C - I$.

Acquisition	Class	Frames	Static + Maj3 Errors	Selective + Maj3 Errors	Always + Maj3 Errors	C	I	Net Reduction
S01 human1	Human	1710	0	0	0	0	0	0
S02 human10	Human	2038	1	1	1	0	0	0
S03 human13	Human	2415	0	0	0	0	0	0
S04 human19	Human	2087	1	1	1	0	0	0
S05 noise11	Background	1831	0	0	0	0	0	0
S06 noise15	Background	2142	6	6	6	0	0	0
S07 noise3	Background	2415	28	20	20	10	2	8
S08 noise6	Background	2104	50	44	43	9	3	6
Total	—	16,742	86	72	71	19	5	14

The 14-error final reduction consists of 19 corrected errors and five introduced errors, all Background false positives. Of these paired changes, seven corrections and two introduced errors occur at currently routed frames, while 12 corrections and three introduced errors occur at non-routed frames through the causal Majority-3 history. Temporal RF directly replaces the probability only for the selected frame, but that changed decision can affect subsequent outputs while it remains in the voting window. Final changes can therefore extend beyond directly routed frames through forward causal propagation, without future information.

The allocation analysis in Section 5.4 also reveals routing blind spots. Before Majority-3, the 227 Static errors comprise 147 with valid history that are routed, 62 with valid history that are not routed, and 18 without valid history. Thus, the 70.33% capture ratio is 147/209 among the 209 history-valid Static errors at this pre-stabilization stage; it is not a capture rate for the final errors. Low disagreement can accompany a wrong prediction, and some correctable errors are missed by routing. However, Always corrects only eight of the 62 eligible but unselected errors, leaving 54 incorrect under the same Temporal RF. The routing blind spot and the temporal classifier's correction limit are therefore distinct. At this pre-Majority stage, Selective corrects 68 Static errors and introduces 10, giving 58 net corrections, including four corrected Human misses and no newly introduced Human miss. These pre-stabilization counts describe direct learned refinement, whereas the final 19/5 decomposition includes output stabilization. Figure 7 retains the original qualitative examples of corrections, introduced errors, and shared failures at this earlier stage.

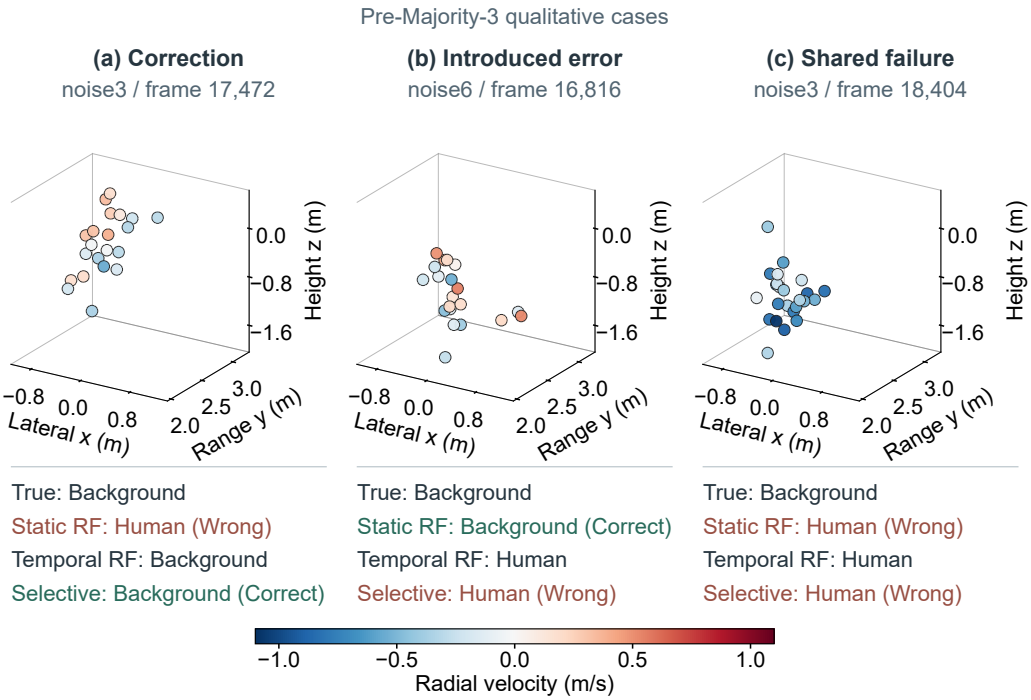

Figure 7. Representative pre-Majority-3 prediction transitions after selective temporal refinement; qualitative point-cloud examples.

Table 11 reports uncertainty for the final Selective-minus-Static comparison using the stratified acquisition-cluster protocol in Section 4.7. The observed FP difference is -14, with a 95% interval of $[-28, 0]$. The BA difference is approximately $+0.0008243$ on the 0–1 scale, equivalent to approximately 0.08243 percentage points, with an interval of $[0, 0.0015490]$. All reported intervals include or touch zero, so the eight-acquisition evidence does not establish a statistically significant improvement at the acquisition level.

Table 11. Post-Majority-3 differences (Selective minus Static) with 95% stratified acquisition-cluster bootstrap intervals.

Metric Difference	Observed Difference	95% Cluster Interval
Recall	0	[0, 0]
FPR	−0.0016486	[−0.0030980, 0]
FA/1000	−1.6486	[−3.0980, 0]
Balanced Accuracy	+0.0008243	[0, +0.0015490]
Accuracy	+0.0008362	[0, +0.0016509]
Error rate	−0.0008362	[−0.0016509, 0]
FP count	−14	[−28, 0]
Total errors	−14	[−28, 0]

Note. Recall, FPR, BA, Accuracy, and error-rate differences use the 0–1 fraction scale. FA/1000 is per 1000 Background frames; FP and total-error differences are counts. These are related summaries of the same paired comparison, not independent significance tests.

5.7. Supplemental Evaluation in an Unseen Logistics Environment

The fixed system is evaluated on the two independent continuous logistics-station recordings described in Section 4.1, from one physical site not used in the original training/development acquisitions. Table 12 reports each recording and their aggregate separately from the original eight-acquisition test. The models, preprocessing, class and routing thresholds, and causal Majority-3 are fixed; neither supplemental recording is used to retune the system. The 2309 frames comprise 991 Human and 1318 Background observations. This provides supplemental evaluation at an additional physical site, separate from the original dataset split.

Table 12. Supplemental held-out logistics recordings after common causal Majority-3. Temporal invocation percentages use all frames within the corresponding row.

Recording	Frames	Static Errors	Selective Errors	Always Errors	Selective Calls	Invocation (%)
L01	1302	42	37	37	175	13.4409
L02	1007	24	19	19	114	11.3208
Total	2309	66	56	56	289	12.5162

Note. Aggregate FN/FP are 9/57 for Static, 9/47 for Selective, and 9/47 for Always. Per-recording FN/FP are not inferred from total errors.

Selective reduces the final error count from 66 to 56, entirely through ten fewer Background false positives (57 to 47), while Human misses remain at nine. Each recording contributes five fewer total errors. Selective and Always have the same aggregate errors and FN/FP counts; this agreement does not establish identical frame-wise predictions, statistical equivalence, or non-inferiority. The observations show additional false-positive reduction by the fixed selective mechanism in these two recordings at the new site.

The all-frame Temporal invocation ratio is 12.5162% (289/2309), compared with 2.8611% (479/16,742) on the original test. The invocation burden therefore depends on the input distribution rather than remaining near 2.86% across scenes.

These recordings provide limited external evidence from one additional site. Their frames remain temporally dependent, and the original eight-acquisition bootstrap intervals do not apply to this set or to a pooled ten-acquisition analysis. The observations extend the measured false-positive reduction to the tested logistics site, while broader generalization and statistical significance require further independent acquisition evidence.

6. Discussion

6.1. Incremental Refinement Beyond Causal Output Stabilization

Majority-3 and learned refinement use different temporal information. Output voting suppresses brief prediction reversals, whereas Temporal RF reclassifies the current observation using recent feature changes. Table 6 separates their effects: Pre-Majority-3 errors are 227, 169, and 161 for Static, Selective, and Always, compared with 86, 72, and 71 after common stabilization. Majority-3 accounts for most of the reduction from the original Static output. The learned component's increment is the decrease from 86 to 72 final errors, with Background false positives falling from 84 to 70 and Human misses unchanged at two. The reduction is 16.28% of residual errors, rather than an equivalent percentage-point increase in Accuracy.

The paired changes in Section 5.6 show how refinement can help or harm the final output. The 14-error net reduction comprises 19 corrections and five introduced errors. Seven corrections and two introduced errors occur at directly routed frames; the remaining 12 corrections and three introduced errors propagate through causal Majority-3. A refined decision can affect subsequent votes while it remains in the prediction history, extending the effect beyond the frame on which Temporal RF executes without using future observations. Thus, feature-history reclassification complements broadly applied output stabilization, but its value depends on the selected observations and on the cost of retaining and executing the additional classifier.

6.2. Allocation Signal, Routing Blind Spots, and Design Choices

The matched-budget comparison shows that the location of temporal calls matters. At 479 calls before Majority-3, tree-probability standard deviation yields 169 errors versus a matched-random mean of 225.05 (Table 8). Its advantage over the mean-confidence alternatives is only two errors: distance from 0.5, mean predictive entropy, and probability margin each yield 171 errors, while hard-vote entropy yields 177. This small observed difference does not establish statistical superiority, and the three binary mean-confidence scores have monotonically related rankings. Tree-probability disagreement instead uses variation among individual-tree probabilities to request additional temporal evidence without a separately trained router.

The error decomposition distinguishes limits of selection from limits of reclassification. Before Majority-3, 147 of the 227 Static errors have valid history and are routed, 62 are history-valid but unselected, and 18 have invalid history. The resulting 147/209 capture rate, approximately 70.33%, concerns eligible Static errors before stabilization, not final corrections. Of the 62 unselected errors, Always Temporal corrects eight, identifying opportunities missed by routing; the other 54 remain incorrect under the same temporal model. Improving selection alone therefore cannot recover all errors with this classifier and representation. History-invalid frames remain outside the candidate set and use Static fallback.

The representation experiments indicate that the information supplied to Temporal RF also affects correction. In the controlled comparison over a common history-valid set (Table 9), Selective produces 106, 77, and 69 final errors for K2, K3, and K5, respectively. These results indicate that a longer temporal window can provide more historical information and reduce recognition errors under the evaluated conditions. Window length also determines the dependence of temporal inference on consecutive valid observations. K3 requires the current frame and the preceding two frames to be continuously valid, whereas K5 requires five consecutive valid observations. In an operating radar system, missing frames or acquisition discontinuities may temporarily invalidate history; a longer window makes the temporal eligibility condition more demanding and requires a longer history to be accumulated again after an interruption. K3 is therefore selected as an engineering trade-off

between recognition performance and temporal availability, reducing dependence on a longer uninterrupted observation history to accommodate variable acquisition conditions.

In the feature-group ablation, Selective errors are 95 with the 41-D temporal-branch input, 84 with means, 85 with dynamics, and 72 with the full 54-D input. Recent levels and short-term changes therefore provide complementary information in the tested 13-D extension. The 41-D condition is a temporal-branch ablation, not Static RF.

6.3. Practical Computational Trade-Offs

Conditional execution reduces the frequency of additional inference rather than its cost per call. Selective invokes Temporal RF 479 times versus 15,758 for Always, a 96.9603% reduction, while their final error counts are 72 and 71 after common Majority-3. The measured mean host latency is 10.642 ms/frame for Fused Selective and 11.681 ms/frame for Optimized Always, approximately 8.9% lower (Table 7). Selective still executes the complete Static RF on every frame, so 96.96% fewer additional calls does not imply the same reduction in total computation, CPU utilization, end-to-end latency, power, or energy.

The comparison with Static exposes the cost of that increment. Static and Selective have similar means (10.638 and 10.642 ms/frame), but Selective's P95 is higher: 15.381 versus 12.382 ms/frame, compared with 15.918 ms/frame for Optimized Always. A near-Static mean therefore coexists with additional tail latency. Process RSS is 193.508 MiB for Selective, 183.352 MiB for Static, and 193.297 MiB for Always. The approximately 10 MiB increase over Static accompanies retention of Temporal RF; skipping a call does not release the second model.

These costs apply to the implementations and measurement scope in Section 4.9. Fused Selective derives the Static mean and tree-probability standard deviation in one traversal, whereas Optimized Always runs Temporal RF directly on eligible frames and Static only for fallback. The measured latency difference therefore compares these execution paths, not instruction-identical implementations differing only in routing. Timing includes inference from precomputed features and causal output processing, excluding raw point-cloud feature extraction and the complete sensor-to-alarm chain. Hardware constraints currently prevent measurement of dual-RF deployment on the ADT6101's Cortex-M3, target-processor latency, complete board memory use, and energy.

6.4. Acquisition Scope, Statistical Uncertainty, and Generalization Limits

The primary test comprises four Human and four Background acquisitions, totaling 16,742 frames: 8250 Human and 8492 Background. Improvement occurs only in S07 and S08, where errors decrease from 28 to 20 and from 50 to 44 (Table 10). These two Background acquisitions account for 78 of Static's 86 final errors. Their contribution therefore reflects the concentration of residual false positives, rather than a uniform gain across acquisitions. The other six have unchanged error counts, including six shared errors in S06. The allocation conclusions apply to the fixed RF backbone and shared Temporal RF; PointNet++ and PointMLP provide external model comparisons under the protocols in Section 4.4.

The stratified acquisition-cluster intervals in Table 11 retain each sampled acquisition as a whole, preserving its frame count and within-recording dependence. The final Selective-minus-Static FP difference is -14 with a 95% interval of $[-28, 0]$. The FPR difference is -0.0016486 with an interval of $[-0.0030980, 0]$, and the Balanced Accuracy difference is $+0.0008243$ with an interval of $[0, +0.0015490]$. The latter is approximately 0.08243 percentage points on a 0–100 scale. All reported intervals include or touch zero, so these eight acquisitions do not establish statistically significant improvement at the acquisition level. The point estimates quantify the observed increment, whereas the cluster intervals account for acquisition-level sampling uncertainty that frame-level paired tests cannot resolve.

The supplemental test adds two independent continuous recordings at one previously unused logistics station to the evidence from the original garage and laboratory-building sites, which supplied 20 acquisitions each. Dense parcels, metal shelving, and personnel activity describe the new setting without quantifying multipath or occlusion. With models, preprocessing, and thresholds fixed before this test, evaluation on the 2309 frames yields 66 errors for Static and 56 for both Selective and Always after Majority-3 (Table 12). Selective reduces Background false positives from 57 to 47 while Human misses remain at nine. The additional site therefore shows the same direction of false-positive reduction, but the 289 Temporal calls give an all-frame invocation ratio of 12.5162%, compared with 2.8611% on the original test.

The two recordings extend the evaluation to one additional site, with 2309 temporally related frames rather than 2309 independent conditions. They provide limited external evidence across acquisition settings, while broader transfer across people and radar devices remains untested. Equal aggregate FN/FP counts for Selective and Always also leave frame-wise agreement, equivalence, and non-inferiority unresolved. Within this scope, the study provides controlled evidence of incremental recognition gains and measured host costs for conditional temporal refinement. Further work should examine more independent sites and participants, varied conditions and actual history interruptions, and target-processor implementations with system-level memory, latency, and energy measurements. History windows and routing strategies also warrant further evaluation.

7. Conclusions

This study evaluates conditional temporal refinement beyond causal output stabilization for mmWave radar human–background recognition. After complete Static RF inference, intrinsic tree-level probability disagreement selects history-valid observations for an additional Temporal RF, without a separately learned gating network. The contribution is this concrete allocation mechanism and its controlled empirical evaluation, connecting existing ensemble-consistency and conditional-inference ideas with short-term radar evidence. Tree-probability disagreement serves as an empirical allocation signal within this task-specific design.

Majority-3 provides the main error reduction, while learned refinement supplies a limited further increment. With the same stabilizer, Static, Selective, and Always produce 86, 72, and 71 errors on the eight original held-out acquisitions. Selective yields 14 fewer Background false positives while Human misses remain at two. Selective makes 479 Temporal RF calls versus 15,758 for Always. This 96.96% reduction differs from the measured host benefit: Fused Selective's mean latency is approximately 8.9% below Optimized Always, with no statistical equivalence established for recognition. Relative to Static, Selective retains the additional temporal model and incurs higher measured P95 and process RSS despite a similar mean. The resulting trade-off is an observed reduction in residual false positives with measured host latency and resident-memory costs.

The increment on the original test is concentrated in S07 and S08, and acquisition-cluster intervals touch zero. On two additional continuous recordings at one previously unseen logistics site, errors decrease from 66 for Static to 56 for Selective, with Human misses unchanged, while the all-frame temporal invocation ratio rises from 2.8611% to 12.5162%. Hardware constraints currently prevent validation of dual-RF deployment on Cortex-M3, board energy, and complete end-to-end alarm latency.

https://doi.org/10.3390/electronics15214980

Author Contributions: Conceptualization, E.M. and S.S.; methodology, E.M.; software, E.M.; validation, E.M.; investigation and data curation, E.M. and S.P.; writing—original draft preparation, E.M.; writing—review and editing, Z.M., L.A. and S.S.; supervision, S.S.; project administration, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study involved non-invasive mmWave radar sensing for algorithm evaluation and did not involve medical intervention, clinical experiments, or the collection of sensitive personal information. The data acquisition procedure followed institutional guidelines for non-invasive sensing research.

Informed Consent Statement: Informed consent was obtained from all participants involved in the data acquisition. The participants agreed to the use of the collected radar sensing data and related acquisition images for academic research and publication purposes.

Data Availability Statement: The data presented in this study are available from the corresponding author upon reasonable request due to privacy considerations related to human-participant data acquisition.

Acknowledgments: The authors thank the participants involved in radar data acquisition.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kong, H.; Huang, C.; Yu, J.; Shen, X. A survey of mmwave radar-based sensing in autonomous vehicles, smart homes and industry. *IEEE Commun. Surv. Tutor.* **2025**, *27*, 463–508. [CrossRef]
2. Cui, H.; Zhong, S.; Wu, J.; Shen, Z.; Dahnoun, N.; Zhao, Y. Milipoint: A point cloud dataset for mmwave radar. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 62713–62726. [CrossRef]
3. Singh, A.D.; Sandha, S.S.; Garcia, L.; Srivastava, M. Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar. In Proceedings of the 3rd ACM Workshop on Millimeter-Wave Networks and Sensing Systems, Los Cabos, Mexico, 25 October 2019; pp. 51–56.
4. Yu, C.; Xu, Z.; Yan, K.; Chien, Y.R.; Fang, S.H.; Wu, H.C. Noninvasive human activity recognition using millimeter-wave radar. *IEEE Syst. J.* **2022**, *16*, 3036–3047. [CrossRef]
5. Ding, F.; Luo, Z.; Zhao, P.; Lu, C.X. milliFlow: Scene Flow Estimation on mmWave Radar Point Cloud for Human Motion Sensing. In *Proceedings of the Computer Vision—ECCV 2024*; Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G., Eds.; Springer : Cham, Switzerland, 2025; pp. 202–221. [CrossRef]
6. Kim, Y.; Alnujaim, I.; Oh, D. Human activity classification based on point clouds measured by millimeter wave MIMO radar with deep recurrent neural networks. *IEEE Sens. J.* **2021**, *21*, 13522–13529. [CrossRef]
7. Dang, V.N.; Hoang, N.C.; Nguyen, Q.C.; Le, M.T. Advancing robust human activity recognition via informative mmWave radar characteristics and a lightweight spatio-spectro-temporal network. *Measurement* **2025**, *256*, 118056. [CrossRef]
8. Zeng, R.; Zhang, J.; Mao, Y.; Yang, Z.; Shi, L. mmMotion: A Real-Time Human Motion Detection Smart Home System Based on mmWave Radar. *IEEE Internet Things J.* **2025**, *12*, 54699–54714. [CrossRef]
9. Wu, Y.; Fioranelli, F.; Gao, C. Radmamba: Efficient human activity recognition through a radar-based micro-Doppler-oriented mamba state-space model. *IEEE Trans. Radar Syst.* **2026**, *4*, 261–272. [CrossRef]
10. Wang, X.; Chen, Y.; Wang, F.; Zheng, K.; Zong, Y. MPCT-LSTM: A lightweight human activity recognition model for 3D point clouds in millimeter-wave radar. *Digit. Signal Process.* **2025**, *164*, 105263. [CrossRef]
11. Luo, F.; Khan, S.; Li, A.; Huang, Y.; Wu, K. EdgeActNet: Edge intelligence-enabled human activity recognition using radar point cloud. *IEEE Trans. Mob. Comput.* **2024**, *23*, 5479–5493. [CrossRef]
12. Singh, Y.P.; Gupta, A.; Chaudhary, D.; Wajid, M.; Srivastava, A.; Mahajan, P. Hardware deployable edge ai solution for posture classification using mmwave radar and low-computational machine learning model. *IEEE Sens. J.* **2024**, *24*, 26836–26844. [CrossRef]
13. Fard, A.S.; Mashhadigholamali, M.; Zolfaghari, S.; Abedi, H.; Chakraborty, M.; Borzì, L.; Daneshtalab, M.; Shaker, G. Lightweight FMCW radar framework for human activity recognition under limited data conditions. *Sci. Rep.* **2026**, *16*, 9650. [CrossRef]
14. Zhu, J.; Lou, X.; Ye, W. Lightweight deep learning model in mobile-edge computing for radar-based human activity recognition. *IEEE Internet Things J.* **2021**, *8*, 12350–12359. [CrossRef]
15. Li, Z.; Le Kernec, J.; Abbasi, Q.; Fioranelli, F.; Yang, S.; Romain, O. Radar-based human activity recognition with adaptive thresholding towards resource constrained platforms. *Sci. Rep.* **2023**, *13*, 3473. [CrossRef]

16. Daghero, F.; Burrello, A.; Macii, E.; Montuschi, P.; Poncino, M.; Jahier Pagliari, D. Dynamic decision tree ensembles for energy-efficient inference on IoT edge nodes. *IEEE Internet Things J.* **2024**, *11*, 742–757. [CrossRef]
17. Shaker, M.H.; Hüllermeier, E. Aleatoric and epistemic uncertainty with random forests. In *Proceedings of the International Symposium on Intelligent Data Analysis, Konstanz, Germany, 27–29 April 2020*; Springer: Cham, Switzerland, 2020; pp. 444–456.
18. Huang, X.; Patel, N.; Tsoi, K.P. Application of mmwave radar sensor for people identification and classification. *Sensors* **2023**, *23*, 3873. [CrossRef]
19. Hanif, A.; Muaz, M.; Hasan, A.; Adeel, M. Micro-Doppler based target recognition with radars: A review. *IEEE Sens. J.* **2022**, *22*, 2948–2961. [CrossRef]
20. Abdu, F.J.; Zhang, Y.; Fu, M.; Li, Y.; Deng, Z. Application of deep learning on millimeter-wave radar signals: A review. *Sensors* **2021**, *21*, 1951. [CrossRef]
21. Gorji, A.; Khalid, H.-U.-R.; Bourdoux, A.; Sahli, H. On the generalization and reliability of single radar-based human activity recognition. *IEEE Access* **2021**, *9*, 85334–85349. [CrossRef]
22. Yang, S.; Le Kernec, J.; Romain, O.; Fioranelli, F.; Cadart, P.; Fix, J.; Ren, C.; Manfredi, G.; Letertre, T.; Sáenz, I.D.H.; et al. The human activity radar challenge: Benchmarking based on the 'radar signatures of human activities' dataset from glasgow university. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 1813–1824. [CrossRef]
23. Hu, J.; Jiang, H.; Liu, D.; Xiao, Z.; Zhang, Q.; Min, G.; Liu, J. Real-Time Contactless Eye Blink Detection Using UWB Radar. *IEEE Trans. Mob. Comput.* **2024**, *23*, 6606–6619. [CrossRef]
24. Schumann, O.; Wöhler, C.; Hahn, M.; Dickmann, J. Comparison of random forest and long short-term memory network performances in classification tasks using radar. In *Proceedings of the 2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6.
25. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5099–5108.
26. Ma, X.; Qin, C.; You, H.; Ran, H.; Fu, Y. Rethinking network design and local geometry in point cloud: A simple residual MLP framework. *arXiv* **2022**, arXiv:2202.07123.
27. Sonny, A.; Kumar, A.; Cenkeramaddi, L.R. Carry object detection utilizing mmWave radar sensors and ensemble-based extra tree classifiers on the edge computing systems. *IEEE Sens. J.* **2023**, *23*, 20137–20149. [CrossRef]
28. Khalid, H.-U.-R.; Gorji, A.; Bourdoux, A.; Pollin, S.; Sahli, H. Multi-view CNN-LSTM architecture for radar-based human activity recognition. *IEEE Access* **2022**, *10*, 24509–24519. [CrossRef]
29. Yan, J.; Zeng, X.; Zhou, A.; Ma, H. MM-HAT: Transformer for millimeter-wave sensing based human activity recognition. In *Proceedings of the GLOBECOM 2022–2022 IEEE Global Communications Conference*; IEEE: Piscataway, NJ, USA, 2022; pp. 547–553.
30. Zhao, P.; Lu, C.X.; Wang, B.; Trigoni, N.; Markham, A. Cubelearn: End-to-end learning for human motion recognition from raw mmwave radar signals. *IEEE Internet Things J.* **2023**, *10*, 10236–10249. [CrossRef]
31. Huan, S.; Wang, Z.; Wang, X.; Wu, L.; Yang, X.; Huang, H.; Dai, G.E. A lightweight hybrid vision transformer network for radar-based human activity recognition. *Sci. Rep.* **2023**, *13*, 17996. [CrossRef]
32. Han, Y.; Huang, G.; Song, S.; Yang, L.; Wang, H.; Wang, Y. Dynamic neural networks: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 7436–7456. [CrossRef]
33. Liu, X.; Fischione, C. Coordinated Beamforming for Multi-Cell ISAC Using Graph Neural Networks. *IEEE Trans. Wirel. Commun.* **2026**, *25*, 5876–5889. [CrossRef]
34. Zhang, Y.; Zhu, N.; Meng, F.; Wang, K.; Ma, K.; Seng Yeo, K. A 6–10.1-GHz Compact CMOS SOI Power Amplifier Using a 6:1 Dual-Mode-Optimized Interleaved Transformer for Portable Radar and ISAC Front Ends. *IEEE J. Solid-State Circuits* **2026**, *early access*. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.