electronics

MDPI

Article

Physics-Guided Domain-Quality Selective Learning for Cross-Domain Wi-Fi CSI Fall Recognition

Xiaobao Li [1], Tianjian Yang [2], Mayila Wumaier [1], Donghao Li [1], Xu Liu [1] and Jinlong Liu [1,*]

[1] School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China; 23b905015@stu.hit.edu.cn (X.L.); 26s105185@stu.hit.edu.cn (D.L.); 25s005091@stu.hit.edu.cn (X.L.)
[2] School of Mathematics and Statistics, Hunan Normal University, Changsha 410081, China; 202430121214@hunnu.edu.cn
* Correspondence: yq20@hit.edu.cn

Abstract

The Physics-Guided Domain-Quality Selective Network (PG-DQSNet) combines CSI motion representations, domain supervision, and class-conditional review scores for cross-domain fall recognition. Packet-centered amplitude dynamics and sanitized relative-phase differences feed a multi-scale temporal encoder. The selector uses capture quality, calibrated probability, dropout disagreement, and prototype distance. We evaluate 2011 recordings with 429 demonstrated falls from three participants and three rooms using acquisition-group-separated fitting, calibration, and test partitions. Matched deterministic inference gives person/room-disjoint macro F2 of 0.608/0.537 against amplitude-only input at 0.526/0.373. The original frozen policy gives recall of 0.941/0.762 with specificity of 0.064/0.319. At a descriptive 10% test false-positive rate, matched PG-AP recall is 0.111/0.189. Learned selection gives room-disjoint AURC of 0.507 against confidence ordering at 0.653; person-disjoint values are 0.686/0.681. Frozen review thresholds defer 19.7%/33.8% of recordings. Class-specific counts, matched-inference controls, perturbation tests, and split ledgers document the measured recognition and review behavior.

Keywords: physics-guided CSI representation; relative-phase motion encoding; multi-scale temporal encoding; domain-adversarial learning; class-conditional selective prediction

1. Introduction

Wi-Fi channel state information (CSI) observes human motion through changes in multipath propagation. Contact-free sensing supports motion monitoring through existing wireless links. The receiver records complex channel measurements. WiFall and RT-Fall established the feasibility of wireless fall recognition [1,2]. Subsequent approaches studied amplitude distributions and learned temporal representations [3,4]. The learning task separates falls from daily activities across people, rooms, and acquisition runs.

CSI contains motion evidence and acquisition-dependent variation. Multipath geometry, receiver gain, packet timing, and the performer affect the observed tensor. Cross-domain wireless representations address related questions in gesture and activity recognition [5,6]. Acquisition-group separation controls the evaluation of closely related recordings [7,8]. An explicit review decision provides a second output for uncertain observations.

We introduce the Physics-Guided Domain-Quality Selective Network (PG-DQSNet). Measurement-derived amplitude and relative-phase channels expose temporal motion structure to a multi-scale encoder. Participant and room adversaries supervise the training representation [9]. A separate class-conditional error-risk selector combines capture quality

Check for updates

Academic Editor: Stefano Scanzio

Received: 16 October 2026
Revised: 14 November 2026
Accepted: 18 November 2026
Published: 20 November 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.

with classifier evidence. Its evidence includes Platt-transformed probability, dropout disagreement, and distance to training-class prototypes [10–13]. The classifier and selector have separate optimization stages.

The corpus contains 2011 audited recordings with 429 demonstrated falls. It covers three participants, three rooms, six dates, and 13 acquisition runs. Eight model configurations share identical outer partitions and three seeds. The evaluation asks how the complete AP configuration changes calibrated-threshold F2 across held-out domains. It also measures risk ordering at matched coverage and review workload under a frozen threshold. The contribution combines an explicit motion representation, a modular selective architecture, and a record-level grouped benchmark.

2. Related Work

2.1. Wireless Fall and Activity Recognition

WiFall detects falls from motion-induced wireless changes [1]. RT-Fall develops contactless detection with commodity Wi-Fi devices [2]. DapFall uses a dynamic amplitude probability-density profile [3]. Recurrent architectures learn activity structure from CSI sequences [4]. Surveys organize sensing by signal processing, recognition, and estimation tasks [14]. Our representation specifies each amplitude and phase transformation before temporal encoding.

2.2. Domain Representations and Selective Decisions

Widar3.0 studies cross-domain gesture recognition [5]. Environment-independent recognition separates activity from environment-dependent features [6]. Gradient reversal trains a task embedding against domain discrimination [9]. The participant and room heads use identities from the fitting partition. CSI and capture quality supply the inference inputs.

Selective classification adds an accept/review rule to the classifier [11]. Probability calibration estimates the relationship between scores and observed outcomes [12]. Monte Carlo dropout supplies an approximate disagreement signal [10]. Distribution-shift studies motivate evaluation on intact held-out domains [13]. Our learned and confidence selectors use identical PG-AP predictions. Matched coverage measures error ordering. A calibration-frozen threshold determines the measured review fraction. SIMAC studies task-oriented multimodal sensing and communication representations [15]. Our task representation uses a single Wi-Fi CSI stream and a separate review score.

2.3. Measurement, Temporal Encoding, and Transfer

The Intel 5300 CSI tool provides a reproducible example of commodity channel measurement [16]. SpotFi studies localization from wireless measurements [17]. CSI behavior-recognition surveys organize motion-sensing applications and processing choices [18]. CSI-power sensing provides a phase-offset-insensitive measurement route [19]. These studies motivate explicit measurement processing. Our recorded tensor has 52 subcarriers and a fixed 3 s event span.

Long short-term memory models represent recurrent temporal dependencies [20]. Residual networks use skip connections to train deep representations [21]. Squeeze-and-excitation blocks reweight feature channels [22]. Group normalization normalizes feature groups within each observation [23]. PG-DQSNet uses group-normalized temporal residual blocks with parallel depthwise branches and channel gating.

Transfer-learning surveys distinguish source and target learning settings [24]. Domain-adaptation theory relates generalization to differences between domains [25]. Deep CORAL aligns feature correlations [26]. Our implementation uses participant and

room adversaries on fitting identities. The reported experiments evaluate complete held-out participants or rooms.

2.4. Probability Estimation and Decision Evaluation

Supervised probability-estimation studies examine the calibration of classifier outputs [27]. The recognition-error/rejection tradeoff supplies a decision-theoretic basis for abstention [28]. Our selective stage estimates errors from detached classifier evidence and capture quality.

Extremely randomized trees and support-vector networks provide established classification families [29,30]. We implement ExtraTrees and an RBF SVM as feature-based comparators. ROC analysis summarizes score ranking across decision thresholds [31]. Precision–recall analysis describes ranking in relation to positive-class frequency [32,33]. Brier score evaluates probabilistic forecasts [34]. Bootstrap resampling summarizes variability in observed units [35]. These references support the comparator and evaluation definitions used below.

3. PG-DQSNet: Representation and Learning

Recording i contains complex CSI $H_i \in \mathbb{C}^{T\times S}$ and acquisition-quality vector $q_i \in \mathbb{R}^{13}$. Its supervision includes binary label $y_i \in \{0,1\}$ and participant/room identities (d_i^p, d_i^r). The implemented tensor has $T = 200$ and $S = 52$. Figure 1 separates inference and training supervision. The physics guidance comprises propagation-motivated signal transformations and same-event observation consistency.

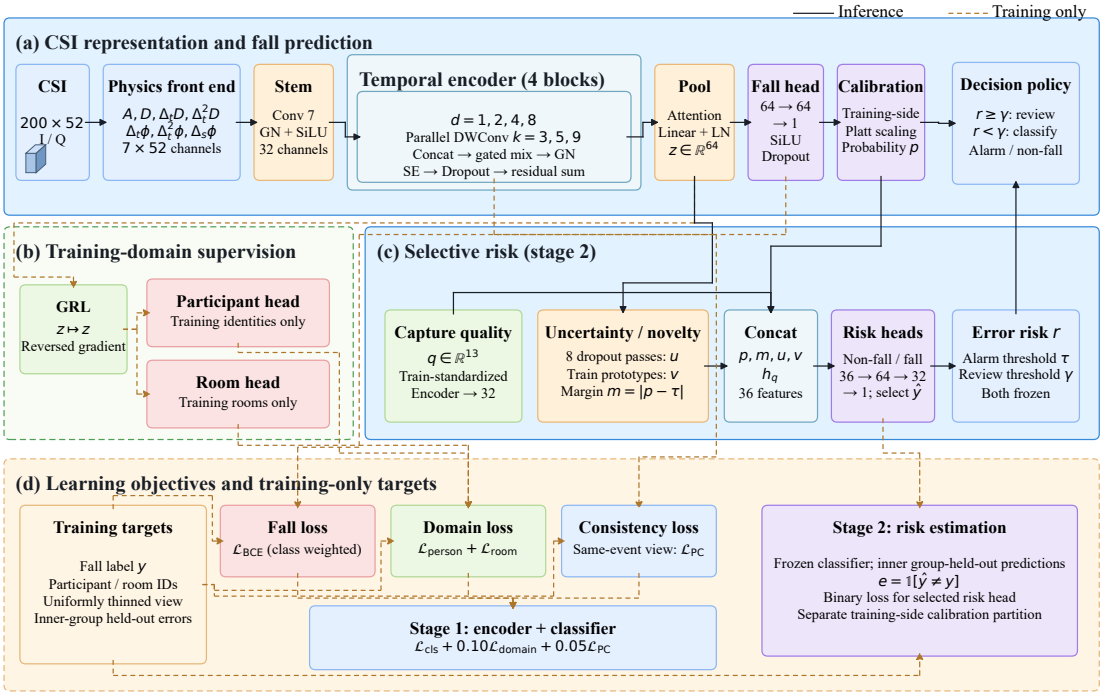

Figure 1. Code-grounded PG-DQSNet architecture. Solid arrows denote inference and dashed arrows denote training supervision. The temporal encoder has four residual blocks with parallel depthwise kernels. Domain labels supervise the representation during training; capture quality enters only the risk selector. Stage 2 uses a frozen classifier and inner acquisition-group-held-out predictions. The tensor icon is schematic.

3.1. Observation Model and Amplitude Channels

A useful OFDM observation model separates motion-sensitive propagation from receiver nuisance:

$$H_{t,s} = g_t e^{j(\alpha_t s + \beta_t)} \sum_{\ell=1}^{L} a_{\ell,t} e^{-j2\pi f_s \tau_{\ell,t}} + \epsilon_{t,s}. \tag{1}$$

Here g_t denotes packet-dependent gain. The affine term $\alpha_t s + \beta_t$ describes receiver phase nuisance. Path ℓ has amplitude $a_{\ell,t}$ and delay $\tau_{\ell,t}$. Equation (1) motivates common-gain reduction, affine-phase sanitization, and relative motion channels. The fitted affine trend can contain receiver offsets and a propagation-delay structure. Removing it discards the absolute delay and phase slope; the retained channels describe circular residual variation.

We first reduce common amplitude variation within each packet and then standardize using the fitting partition only:

$$A_{t,s} = \log(1 + |H_{t,s}|) - \text{median}_k \log(1 + |H_{t,k}|), \qquad X_{t,s} = \frac{A_{t,s} - \mu_s^{\text{fit}}}{\max(\sigma_s^{\text{fit}}, 0.05)}. \tag{2}$$

Single-frame repair and timestamp-based interpolation precede this training-side standardization. A nine-frame moving average with replicated boundary values provides the slow component:

$$B_{t,s} = \frac{1}{9} \sum_{k=-4}^{4} X_{t+k,s}, \quad D_{t,s} = X_{t,s} - B_{t,s}, \quad V_{t,s} = D_{t,s} - D_{t-1,s}, \quad J_{t,s} = V_{t,s} - V_{t-1,s}. \tag{3}$$

The first differences are zero-padded. For gain $c > 0$, $\log(1 + c|H|) - \log(1 + |H|)$ depends on $|H|$. Packet-median centering reduces common gain variation and is not an exact multiplicative-gain invariant. The amplitude branch concatenates $[X, D, V, J]$, making slow change, motion variation, and rapid transitions available to the same encoder.

3.2. Sanitized Relative-Phase Channels

The receiver stores interleaved imaginary and real components. We unwrap phase across subcarriers, fit an affine trend on sufficiently strong subcarriers, and retain the circular residual:

$$(\hat{\alpha}_t, \hat{\beta}_t) = \arg\min_{a,b} \sum_{s \in \mathcal{U}_t} [\text{unwrap}(\angle H_{t,s}) - as - b]^2, \qquad \phi_{t,s} = \text{wrap}(\angle H_{t,s} - \hat{\alpha}_t s - \hat{\beta}_t - c_t). \tag{4}$$

The valid set $\mathcal{U}_t$ uses an amplitude gate, and c_t is the median residual. With fewer than three valid subcarriers, median centering replaces the affine fit. Interpolation is performed through sine and cosine before reconstructing the angle. Defining $\text{wrap}(u) = \text{atan2}(\sin u, \cos u)$ gives

$$P^t_{t,s} = \text{wrap}(\phi_{t,s} - \phi_{t-1,s}), \quad P^{tt}_{t,s} = \text{wrap}(P^t_{t,s} - P^t_{t-1,s}), \quad P^s_{t,s} = \text{wrap}(\phi_{t,s} - \phi_{t,s-1}). \tag{5}$$

The AP representation is $U = [X, D, V, J, P^t, P^{tt}, P^s] \in \mathbb{R}^{364\times200}$. The A variant contains 208 channels. Relative phase supplies temporal and subcarrier motion differences.

3.3. Multi-Scale Gated Temporal Encoding

A kernel-seven convolution maps U to 32 channels, followed by group normalization and SiLU. Four residual blocks use dilation $d \in \{1, 2, 4, 8\}$. Within block b, the three depthwise branches operate in parallel:

$$\begin{aligned} C_b &= \text{Concat}_{k\in\{3,5,9\}} \text{DWConv}_{k,d_b}(Z_b), \\ [M_b, G_b] &= \text{Conv}_{1\times1}(C_b), \qquad R_b = \text{GN}(M_b \odot \sigma(G_b)). \end{aligned} \tag{6}$$

Channel gating pools R_b over time and reweights the residual update:

$$s_b = \sigma(W_{b,2}\, \text{SiLU}(W_{b,1}\, \text{Avg}_t\, R_b)), \qquad Z_{b+1} = Z_b + \text{Dropout}(R_b \odot s_b). \tag{7}$$

Attention pooling learns which temporal positions contribute to the recording embedding:

$$\omega_t = \frac{\exp[w_2^\top \tanh(W_1 Z_t)]}{\sum_v \exp[w_2^\top \tanh(W_1 Z_v)]}, \qquad z = \mathrm{Dropout}\Big(\mathrm{SiLU}\Big(\mathrm{LN}\Big(W_z \sum_t \omega_t Z_t + b_z\Big)\Big)\Big) \in \mathbb{R}^{64}. \tag{8}$$

The fall head is a $64 \to 64 \to 1$ multilayer perceptron with SiLU and dropout, producing logit $\ell = f_\theta(z)$.

3.4. Domain Supervision and Observation Consistency

The binary objective uses the ratio of non-fall to fall counts in the fitting partition as positive weight w_+:

$$\mathcal{L}_{\mathrm{cls}} = -\frac{1}{N} \sum_i [w_+ y_i \log \sigma(\ell_i) + (1 - y_i) \log(1 - \sigma(\ell_i))]. \tag{9}$$

Participant and room heads classify only identities present in the fitting partition. A gradient-reversal operator is an identity in the forward pass and reverses the encoder gradient [9]:

$$\mathcal{R}_\lambda(z) = z, \qquad \frac{\partial \mathcal{R}_\lambda}{\partial z} = -\lambda I, \qquad \mathcal{L}_{\mathrm{dom}} = \mathrm{CE}(h_p(\mathcal{R}_\lambda(z)), d^p) + \mathrm{CE}(h_r(\mathcal{R}_\lambda(z)), d^r). \tag{10}$$

A head is omitted when its fitting partition has one identity. The reversal strength follows $\lambda(e) = 2/[1 + \exp(-10e/E)] - 1$, with $E = 18$ epochs.

The consistency view retains the three-second event. Uniform sampling reduces 200 temporal positions to 100. Interpolation restores 200 positions. A detached teacher supervises the view outputs (ℓ', z'):

$$\begin{aligned} \mathcal{L}_{\mathrm{PC}} &= \mathrm{BCE}(\sigma(\ell'), \mathrm{sg}[\sigma(\ell)]) + 0.25[1 - \cos(\mathrm{sg}[z], z')], \\ \mathcal{L}_1 &= \mathcal{L}_{\mathrm{cls}} + 0.10\mathcal{L}_{\mathrm{dom}} + 0.05\mathcal{L}_{\mathrm{PC}}. \end{aligned} \tag{11}$$

The operator sg stops gradients. The no-PC configuration removes this term. The DQSNet reference configuration also removes the deterministic physics front end. Uniform temporal thinning preserves event duration and retained-observation order. It assumes the class remains stable after coarser sampling. Interpolation smooths transitions faster than the reduced sampling interval; it does not reconstruct discarded high-frequency motion.

3.5. Calibrated Probability, Uncertainty, and Novelty

Eight dropout passes generate probabilities $\tilde{p}_m$ and embeddings z_m [10]. Their mean and population standard deviation define the classifier score and disagreement. Platt calibration is fitted on an outer training-side calibration partition:

$$\bar{p} = \frac{1}{8} \sum_{m=1}^{8} \tilde{p}_m, \qquad u = \sqrt{\frac{1}{8} \sum_{m=1}^{8} (\tilde{p}_m - \bar{p})^2}, \qquad p = \sigma(a\,\mathrm{logit}(\bar{p}) + b). \tag{12}$$

The calibration coefficients (a, b) are learned without outer test access [12]. With unit-normalized class prototypes fitted from training embeddings,

$$c_k = \frac{\text{mean}_{i:y_i=k}\, z_i}{\|\text{mean}_{i:y_i=k}\, z_i\|_2}, \qquad v = 1 - \max_{k\in\{0,1\}} \cos(\bar{z}, c_k), \qquad \bar{z} = \frac{1}{8}\sum_{m=1}^{8} z_m. \tag{13}$$

Novelty measures cosine distance to fitting-class prototypes. The embedding dimension and prototype construction remain fixed across folds.

3.6. Class-Conditional Risk and Decision Policy

The 13 acquisition features are frame count, duration, median interval, maximum gap, interval variation, sequence gaps, duplicate frames, saturation, repairs, PHY fraction, 95th-percentile gap, long gaps, and excluded-source count. Fitting-partition standardization precedes a quality encoder with output $h_q \in \mathbb{R}^{32}$. These features describe capture integrity. A 36-dimensional selector input combines quality with classifier evidence:

$$s = [p, |p-\tau|, u, v, h_q], \qquad \hat{y} = 1[p \geq \tau], \qquad r = \sigma\big(g_{\hat{y}}(s)\big). \tag{14}$$

Each risk head is $36 \to 64 \to 32 \to 1$. Probability, disagreement, and novelty are detached from the classifier. The risk stage updates the quality encoder and both risk heads.

To fit the risk targets, three-fold inner acquisition-group splits generate predictions on unseen inner groups. For each such prediction, $e_i = 1[\hat{y}_i \neq y_i]$ is an error target and

$$\mathcal{L}_2 = -\frac{1}{N_{\text{OOF}}}\sum_i [e_i \log r_i + (1-e_i)\log(1-r_i)]. \tag{15}$$

Inner probabilities use the outer calibration map and alarm threshold. Each inner model derives novelty from its own fitting-class prototypes. The alarm threshold maximizes calibration F_2 over $\{0.10, 0.11, \ldots, 0.90\}$. The review threshold is the calibration 90th percentile, corresponding to a nominal 10% review target. Both are frozen before evaluating the outer domain:

$$a(H,q) = \begin{cases} \text{review}, & r \geq \gamma, \\ \text{alarm}, & r < \gamma \text{ and } p \geq \tau, \\ \text{non-fall}, & r < \gamma \text{ and } p < \tau. \end{cases} \tag{16}$$

3.7. Record Allocation and Fitting Sequence

Let F, C, and T denote outer fitting, calibration, and test records. Their acquisition groups are disjoint. Every fitted component follows this sequence:

1. Fit signal and quality normalization on F. Fit the outer encoder on F for 18 epochs. Derive class prototypes from deterministic embeddings of F.
2. Compute outer scores on C. Fit the Platt map g_C on (s_C, y_C) and select τ_C on C using the fixed F2 grid.
3. Divide F into three inner acquisition-group folds. Fit each inner encoder and normalization on $F \setminus V_k$. Derive its prototypes from $F \setminus V_k$ and score V_k with eight dropout passes.
4. Apply g_C and τ_C to the inner scores. Construct error targets $e_i = 1[1[g_C(s_i) \geq \tau_C] \neq y_i]$ for $i \in F$ and collect disagreement and inner-prototype distance.
5. Freeze the outer classifier. Fit the quality encoder and risk heads on pooled inner-held-out records in F for 250 updates.
6. Score C using the outer classifier and fitted selector. Set γ_C to its higher-order 90th risk percentile. Apply all frozen components to T.

Risk-head fitting records and review-threshold calibration records are disjoint. Calibration records are reused for the probability map, alarm threshold, and review threshold. Inner error targets use the outer-model map; Supplementary Section S5 reports the score-transfer diagnostic.

4. Experimental Design

4.1. Acquisition Hardware and Recording Procedure

Figure 2 documents the laboratory equipment. The receiver is an Espressif ESP32-S3-DevKitC-1 (Espressif Systems, Shanghai, China) with an on-board PCB inverted-F antenna; the H3C Magic NX30 Pro router (New H3C Technologies Co., Ltd., Hangzhou, China) has five external rod antennas. The corpus also includes TP-Link TL-XDR3010 AX3000 (TP-LINK Technologies Co., Ltd., Shenzhen, China) and MERCURY X306G AX3000 (Shenzhen Meikexing Communication Technology Co., Ltd., Shenzhen, China) routers under metadata labels `tplink` and `mercury`. The author-supplied room sketch gives the transmitter offsets as 2.3 m from the west wall and 1.4 m from the south wall, and the receiver offsets as 5.4 m and 4.4 m, respectively (Supplementary Figure S2). The receiver height is 0.75 m. The H3C router height is 1.30 m. The sketch documents one arrangement and is not linked to a named room in the acquisition sidecars.



Figure 2. H3C configuration example of the laboratory CSI acquisition arrangement and hardware. (**a**) The unmarked scene photograph is annotated at the author-identified device positions. Tx denotes the Wi-Fi router, and Rx denotes the CSI receiver. (**b**) ESP32-S3-DevKitC-1 receiver with an on-board PCB inverted-F antenna. (**c**) H3C Magic NX30 Pro router with five external antennas.

All records use router-ping acquisition with a nominal 10 ms packet interval and 3000 ms capture duration. After the action prompt, the participant has a 10 s preparation interval before the immediate 3 s capture timer begins. The instructed action is completed within that 3 s capture window, and records are not cropped around an event timestamp. Primary channels 11 and 6 occur in 1708 and 303 records, corresponding to 2.462 and 2.437 GHz. The receiver reports 104 interleaved I/Q values for 52 selected subcarriers. In the ESP-IDF v5.5.5 receiver structure, `cwb=0` denotes 20 MHz and `cwb=1` denotes 40 MHz; the corpus contains 49 and 1962 records with these codes. Sidecars record link-distance settings of 3, 4, and 6 m for 279, 1181, and 551 records. The collector version is `console-csi-v7-research`. Supplementary Table S1 provides the retained hardware and wireless metadata.

Demonstrated falls use a foam mat for protection.

4.2. Corpus and Deterministic Preparation

The input package contains 2713 sidecar-linked recording occurrences. Hash and metadata checks precede deterministic integrity and shape screening. Shape screening requires 240–361 frames, a 2.4–3.6 s timestamp span, a maximum gap of at most 100 ms, no sequence gaps, and at most two duplicate frames. The audit excludes 645 shape-quarantined occurrences (124 falls and 521 non-falls) and 35 marked retakes. Supplementary Tables S2–S4 report participant–room–run counts, active heads, and exclusions by action and domain. It also excludes 14 long calibration captures, five calibration retakes, and three duplicated calibration captures. The retained corpus contains 2011 unique recordings with 429 falls and 1582 non-falls. The data provider confirmed three distinct people and three distinct rooms. Table 1 reports the included class counts for each domain.

Table 1. Included recordings by held-out identity and room. Each group is evaluated in full. Falls are demonstrated forward or backward falls.

Grouping	Domain	Falls	Non-Falls	Total
Participant	P001	94	185	279
Participant	ycw	62	188	250
Participant	yxd	273	1209	1482
Room	R01	218	612	830
Room	R02	20	57	77
Room	csR01	191	913	1104

Forward and backward falls are positive labels. Standing, sitting, walking, bending, lying down, static/empty-room recordings, and the recorded pick-up action are negative labels. The usable recordings span 22 September–1 October 2026, contain 13 acquisition runs, and are standardized to 200 × 52 tensors over a 3 s window. Canonical signal hashes confirm no duplicate CSI sequence among the included records. All included records satisfy the relative-phase metadata gate: complex CSI enabled, interleaved I/Q order, 52 subcarriers, channel filtering disabled, and zero manual scale shift. Figure 3 shows the action counts and actual signal examples.

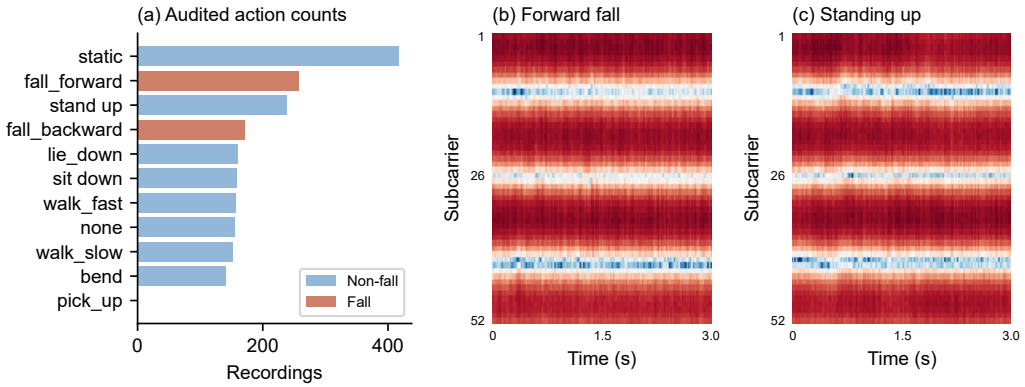

Figure 3. Audited corpus and real amplitude examples. (**a**) Included action counts. (**b**,**c**) Three-second, 52-subcarrier tensors from forward-fall and standing-up recordings, selected by median within-action temporal-difference energy. Colors encode per-frame-median-centered log amplitude.

4.3. Partitioning, Comparators, and Optimization

The primary protocols leave one complete participant or one complete room out. Each protocol has three outer groups and seeds 17, 29, and 43. A three-fold stratified group split of the remaining pool supplies fitting and calibration partitions. Acquisition groups combine raw-stream parent directory, date, and room. Fitting, calibration, and testing use

disjoint acquisition groups. Person-disjoint evaluation permits room overlap. Room-disjoint evaluation permits participant overlap. The two protocols measure each named exclusion separately. P001 and ycw occur in R01; yxd occurs in all three rooms. Holding out R02 or csR01 measures room transfer for yxd. Holding out yxd leaves a fitting partition from R01. Router identity also varies with acquisition groups.

All eight models share these partitions. PG-AP is the complete amplitude–phase configuration. PG-A uses amplitude channels. The no-DA and no-PC configurations omit domain supervision and observation consistency. DQSNet is the internal reference architecture with the physics front end and consistency removed. The convolutional baseline uses kernel-seven and kernel-five layers, group normalization, global pooling, and a binary head. ExtraTrees uses 250 trees and minimum leaf size two. The prespecified RBF SVM uses $C = 2$. An additional six-candidate SVM diagnostic selects C within each outer fitting partition and gives person/room-disjoint test AUROC/F2 of 0.422/0.607 and 0.344/0.166, compared with 0.419/0.608 and 0.578/0.384 for fixed $C = 2$ (Supplementary Table S11). Tree and SVM inputs combine 13 motion descriptors, 52 subcarrier means, and 52 standard deviations. Feature scaling uses the fitting partition.

Neural classifiers train for 18 epochs with AdamW. Learning rate, weight decay, batch size, and dropout are 10^{-3}, 10^{-4}, 128, and 0.15. Gradient-norm clipping uses a bound of five. The width-32 encoder produces 64-dimensional embeddings. PG-AP uses eight dropout passes at inference. The remaining neural configurations use deterministic inference. The matched-inference analysis in Section 5.1 additionally evaluates the PG ablations under deterministic inference and eight-pass inference separately. Each mode refits calibration on identical records and uses the same 81 alarm thresholds. Architecture and optimizer settings are fixed across PG ablations. The configuration selection budget is one fixed training setting per model and seed, with no outer-test tuning. Published Wi-Fi methods are discussed by input and task; their dataset-specific scores are not transferred into this corpus. PG-AP adds three inner group-held-out fits and 250 full-batch risk updates. The risk optimizer uses learning rate 10^{-3} and weight decay 10^{-3}. The experiment contains 144 outer classifier fits and 54 inner PG-AP fits. Checkpoints, partitions, and test predictions are retained.

4.4. Metrics and Aggregation

The primary score weights recall more strongly than precision:

$$F_2 = \frac{5\,\text{precision}\,\text{recall}}{4\,\text{precision} + \text{recall}}, \qquad \text{recall} = \frac{TP}{TP + FN}, \qquad \text{specificity} = \frac{TN}{TN + FP}. \tag{17}$$

Scores are averaged over seeds within each held-out domain. The three domain means receive equal weight. Figure intervals use 10,000 resamples of the three observed means and summarize their descriptive spread. The independent domain count remains three. The unit of descriptive comparison is the held-out domain. Supplementary Table S5 gives paired F2 contrasts for every held-out domain. Repeated seeds are not additional participants or rooms. AUROC and average precision (AP) evaluate test-probability ranking across thresholds. Brier score measures mean squared probability error. Expected calibration error uses ten equal-width probability bins.

For selection, recordings are ordered by increasing estimated risk. Coverage c retains the lowest-risk fraction and measures its error:

$$R(c) = \frac{\sum_{i \in \mathcal{A}(c)} 1[\hat{y}_i \neq y_i]}{|\mathcal{A}(c)|}, \qquad \text{AURC} = \frac{1}{n}\sum_{k=1}^{n} R(k/n), \qquad \rho_{\text{review}} = \frac{1}{n}\sum_{i} 1[r_i \geq \gamma]. \tag{18}$$

Matched-coverage curves compare selectors at identical realized automation. The frozen-threshold review fraction reports the separate operating workload. The confidence comparator assigns $r_{\text{conf}} = 1 - |p - \tau|$. Both selectors use the same probability p and alarm threshold τ. The review threshold uses the higher-order calibration quantile.

5. Results

5.1. Grouped Classification and Matched Inference

Table 2 preserves the original frozen-policy results. PG-AP gives person/room-disjoint F2 of 0.609/0.522 and recall of 0.941/0.762. Specificity is 0.064/0.319, corresponding to FPR of 0.936/0.681. The always-alarm reference gives F2 of 0.623/0.596, recall one, and specificity zero. Its larger F2 reflects recall weighting and class prevalence. Table 3 reports precision and balanced accuracy at these frozen policies.

Table 2. Domain-macro classification results. Each model score averages three seeds per domain. All three domains receive equal weight. P: person-disjoint; R: room-disjoint. The last row assigns an alarm to every recording.

Model	P			R		
	$F_2 \uparrow$	Recall ↑	Specificity ↑	$F_2 \uparrow$	Recall ↑	Specificity ↑
PG-DQSNet-AP	0.609	0.941	0.064	0.522	0.762	0.319
PG-DQSNet-A	0.526	0.770	0.271	0.373	0.546	0.477
PG-DQSNet w/o DA	0.578	0.860	0.151	0.463	0.676	0.328
PG-DQSNet w/o PC	0.558	0.815	0.171	0.484	0.716	0.347
DQSNet	0.550	0.792	0.251	0.480	0.723	0.315
CNN	0.500	0.675	0.332	0.405	0.608	0.362
ExtraTrees	0.585	0.834	0.237	0.285	0.382	0.784
RBF SVM	0.608	0.953	0.032	0.384	0.638	0.392
Always-alarm reference	0.623	1.000	0.000	0.596	1.000	0.000

Table 3. Precision (Pr) and balanced accuracy (BA) at the original frozen policies. BA is the mean of recall and specificity.

Model	P Pr	P BA	R Pr	R BA
PG-AP	0.257	0.502	0.293	0.541
PG-A	0.270	0.521	0.247	0.512
no-DA	0.259	0.506	0.247	0.502
no-PC	0.252	0.493	0.267	0.531
CNN	0.258	0.504	0.210	0.485
ExtraTrees	0.276	0.535	0.351	0.583
SVM	0.254	0.492	0.149	0.515

PG-AP precision is 0.257/0.293 and balanced accuracy is 0.502/0.541 for person/room exclusion. Table 4 reports the matched-inference ablations.

With the deterministic inference held constant, PG-AP gives F2 of 0.608/0.537 against PG-A at 0.526/0.373, no-DA at 0.578/0.463, and no-PC at 0.558/0.484. Changing PG-AP alone to eight dropout passes gives 0.608/0.531 in the controlled rerun. Supplementary Table S5 provides paired domain contrasts. These comparisons separate inference averaging from input and training components.

Table 4. Matched-inference ablations. D: deterministic; M8: eight dropout passes. Each mode refits Platt calibration and selects the alarm threshold on identical calibration records using the same 81-point grid. Values are domain-macro F_2.

Model	P D	P M8	R D	R M8
PG-AP	0.608	0.608	0.537	0.531
PG-A	0.526	0.538	0.373	0.368
no-DA	0.578	0.581	0.463	0.461
no-PC	0.558	0.573	0.484	0.474
DQSNet	0.550	0.553	0.480	0.506

5.2. Ranking and Recall at Controlled False-Positive Rates

Table 5 reports the threshold-independent ranking metrics. Original PG-AP AUROC is 0.501/0.524. ExtraTrees reaches person-disjoint AUROC 0.641; SVM reaches room-disjoint AUROC 0.578. PG-AP Brier scores are 0.223/0.214 and ten-bin calibration errors are 0.150/0.175.

Table 5. Threshold-independent ranking of the saved test probabilities. Domain-macro averaging follows Table 2. AUROC and average precision (AP) increase with better ranking.

Model	Person-Disjoint		Room-Disjoint	
	AUROC ↑	AP ↑	AUROC ↑	AP ↑
PG-DQSNet-AP	0.501	0.279	0.524	0.279
PG-DQSNet-A	0.519	0.299	0.546	0.275
PG-DQSNet w/o DA	0.476	0.257	0.513	0.256
PG-DQSNet w/o PC	0.490	0.264	0.518	0.260
DQSNet	0.513	0.282	0.526	0.277
CNN	0.485	0.248	0.531	0.245
ExtraTrees	0.641	0.392	0.468	0.288
RBF SVM	0.419	0.217	0.578	0.325

Table 6 reports recall at controlled test false-positive rates under matched deterministic inference.

Table 6. Descriptive test ROC envelopes: maximum recall at FPR at most 5% or 10%. Neural models use matched deterministic inference. Diagnostic thresholds use test labels; these are not frozen deployment policies.

Model	P 5%	P 10%	R 5%	R 10%
PG-AP	0.067	0.111	0.055	0.189
PG-A	0.086	0.137	0.037	0.129
no-DA	0.044	0.078	0.029	0.078
no-PC	0.047	0.084	0.035	0.077
CNN	0.047	0.071	0.024	0.068
ExtraTrees	0.169	0.244	0.105	0.133
SVM	0.026	0.047	0.022	0.222

At a descriptive test FPR of at most 10%, deterministic PG-AP recalls 11.1%/18.9% of falls. This point provides at least 90% specificity at the listed sensitivity. A separate policy chooses thresholds using calibration negatives alone: a nominal 5% calibration FPR yields test recall/FPR of 0.186/0.183 and 0.243/0.222. The 10% calibration target yields 0.279/0.285 and 0.313/0.291. These frozen outcomes quantify threshold transfer across held-out domains. Supplementary Table S6 includes all four targets and both inference modes.

5.3. Class-Specific Selective Decisions

Learned and confidence selectors use identical original PG-AP outputs. Room-disjoint AURC is 0.507 versus 0.653; person-disjoint AURC is 0.686 versus 0.681. The nominal 10% calibration threshold produces realized review fractions of 19.7%/33.8%. Table 7 reports the class-specific frozen-policy counts. Figure 4 shows risk ordering and the resulting review workload.

Table 7. Frozen review counts combine three seeds: 1287 fall and 4746 non-fall prediction instances per protocol, from 429 and 1582 unique recordings. Deferred cases are not counted as corrected predictions.

Protocol	True Class	Alarm	Non-Fall	Deferred
Person	Fall	852	129	306
Person	Non-fall	2935	599	1212
Room	Fall	505	331	451
Room	Non-fall	2429	1091	1226

Domain-macro retained recall/specificity is 0.902/0.108 for person exclusion and 0.596/0.421 for room exclusion. Automatic alarms cover 74.6%/41.6% of all falls. Deferred fall fractions are 19.5%/37.6% under domain-macro averaging. Deferred falls remain unresolved in this evaluation.

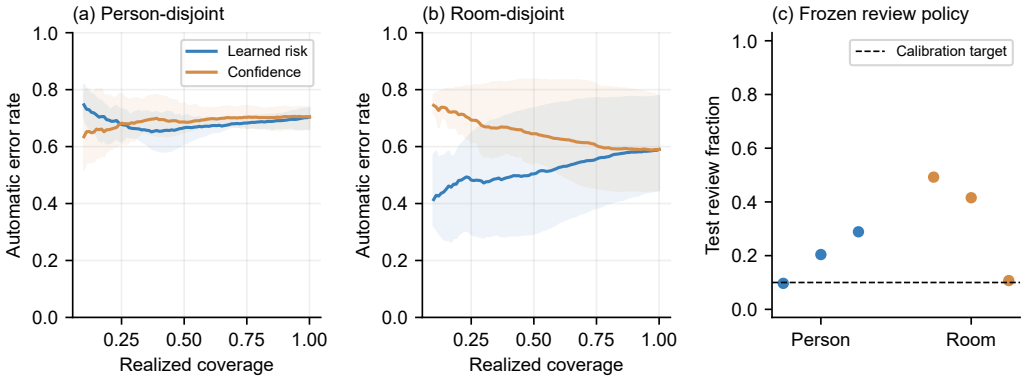

Figure 4. Risk ordering and frozen review workload. Both selectors use identical classifier outputs. Curves report equal retained coverage. Review points apply calibration-frozen thresholds. In panel (**c**), blue dots represent the three held-out participants and orange dots represent the three held-out rooms; each dot averages three seeds. The dashed line marks the nominal 10% calibration review target.

A review decision creates an offline flag. A prospective interface can provide a designated operator with the three-second CSI trace, probability, quality indicators, and risk score. No human adjudication or response-delay study was conducted. On an RTX 2050 at batch size one, eight-pass classification takes 25.49–27.97 ms and the selector takes 0.50–0.74 ms. Timing excludes acquisition, preprocessing, communication, and human review.

5.4. Signal Transformations and Selector Controls

Supplementary Figure S1 shows the same real recording before and after processing. On 42 predefined action–participant–room strata, amplitude gains of 0.5 and 2 give mean centered-log RMS changes of 0.074 and 0.055. Controlled time-varying affine-phase slopes of 0.002, 0.010, and 0.030 rad/subcarrier give circular sanitized-phase RMS changes of 0.288, 0.293, and 0.320 rad. Unwrapping, amplitude gating, and circular centering affect the residual response. Supplementary Section S4 specifies the perturbations. Under the same deterministic CNN, raw amplitude, power, and sanitized-phase inputs give person/room-disjoint F2 of 0.364/0.427, 0.331/0.404, and 0.586/0.419. These controls hold the CNN and training schedule fixed. PG-A versus PG-AP in Table 4 holds the multi-scale

encoder fixed. Quality-only, evidence-only, and combined selectors use identical original inner-held-out targets and 250-update budgets. Their person/room-disjoint AURC is 0.676/0.507, 0.657/0.527, and 0.685/0.508. Quality-only routing retains predicted-class head assignment. Supplementary Table S7 gives the review fractions and fold results.

6. Conclusions

PG-DQSNet specifies a CSI representation and selective-learning pipeline for 2011 recordings. Matched deterministic inference gives person/room-disjoint macro F2 of 0.608/0.537 against amplitude-only input at 0.526/0.373. The original frozen policy yields recall/specificity of 0.941/0.064 and 0.762/0.319. At a descriptive 10% test FPR, matched PG-AP recall is 0.111/0.189. Learned risk ordering gives room-disjoint AURC of 0.507 against confidence ordering at 0.653; corresponding person-disjoint values are 0.686/0.681. Class-specific deferral counts, acquisition-group ledgers, signal perturbations, and selector controls make the measured recognition and review decisions traceable.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/electronics15225415/s1, Supplementary material containing acquisition metadata, acquisition-group allocations, screening counts, paired domain contrasts, signal and selector controls, calibration diagnostics, reproduction details, additional SVM and covariance-alignment comparisons, and the prepared blinded review protocol.

Author Contributions: Conceptualization, X.L. (Xiaobao Li); Methodology, X.L. (Xiaobao Li); Software, X.L. (Xiaobao Li); Validation, X.L. (Xiaobao Li) and T.Y.; Formal analysis, T.Y., M.W. and D.L.; Investigation, T.Y., D.L. and X.L. (Xu Liu); Resources, X.L. (Xu Liu); Data curation, M.W. and X.L. (Xu Liu); Writing—original draft, X.L. (Xiaobao Li); Writing—review and editing, X.L. (Xiaobao Li) and J.L.; Visualization, D.L. and X.L. (Xu Liu); Supervision, J.L.; Project administration, J.L.; Funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review for this study was handled at the school level by the School of Electronics and Information Engineering, Harbin Institute of Technology, which issued a signed and stamped statement confirming that the study meets ethical review requirements. Harbin Institute of Technology has no separate institutional ethics committee for this type of study, and no separate committee approval number was assigned.

Informed Consent Statement: Informed consent was obtained from all participants involved in the study.

Data Availability Statement: The reviewer supplement and reviewer experiment package contain audit counts, split ledgers, analysis scripts, training configurations, and prediction outputs used in this revision. Source recordings and model checkpoints remain in the original experiment archive.

Acknowledgments: The authors thank their academic mentors for guidance during the study.

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

References

1. Han, C.; Wu, K.; Wang, Y.; Ni, L.M. WiFall: Device-free fall detection by wireless networks. In *IEEE INFOCOM 2014—IEEE Conference on Computer Communications*; IEEE: Piscataway, NJ, USA, 2014; pp. 271–279. [CrossRef]
2. Wang, H.; Zhang, D.; Wang, Y.; Ma, J.; Wang, Y.; Li, S. RT-Fall: A Real-Time and Contactless Fall Detection System with Commodity WiFi Devices. *IEEE Trans. Mob. Comput.* **2017**, *16*, 511–526. [CrossRef]
3. Chen, Z.; Zhao, Y.; Chen, G.; Yang, Y.; Huang, W.; Xu, C.-Z. DapFall: Dynamic Amplitude Probability Density Profile-Based Wi-Fi CSI Sensing for Fall Detection. *IEEE Sens. J.* **2025**, *25*, 14252–14265. [CrossRef]
4. Ding, J.; Wang, Y. WiFi CSI-Based Human Activity Recognition Using Deep Recurrent Neural Network. *IEEE Access* **2019**, *7*, 174257–174269. [CrossRef]

5. Zheng, Y.; Zhang, Y.; Qian, K.; Zhang, G.; Liu, Y.; Wu, C.; Yang, Z. Zero-Effort Cross-Domain Gesture Recognition with Wi-Fi. In Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services, Seoul, Republic of Korea, 17–21 June 2019; pp. 313–325. [CrossRef]
6. Jiang, W.; Miao, C.; Ma, F.; Yao, S.; Wang, Y.; Yuan, Y.; Xue, H.; Song, C.; Ma, X.; Koutsonikolas, D.; et al. Towards Environment Independent Device Free Human Activity Recognition. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*; ACM: New York, NY, USA, 2018; pp. 289–304. [CrossRef]
7. Kaufman, S.; Rosset, S.; Perlich, C.; Stitelman, O. Leakage in data mining. *ACM Trans. Knowl. Discov. From Data* **2012**, *6*, 1–21. [CrossRef]
8. Varga, D. Exposing Data Leakage in Wi-Fi CSI-Based Human Action Recognition: A Critical Analysis. *Inventions* **2024**, *9*, 90. [CrossRef]
9. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.* **2016**, *17*, 1–35. Available online: https://jmlr.org/papers/v17/15-239.html (accessed on 12 October 2026).
10. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *Int. Conf. Mach. Learn.* **2016**, *48*, 1050–1059. Available online: https://proceedings.mlr.press/v48/gal16.html (accessed on 12 October 2026).
11. Geifman, Y.; El-Yaniv, R. Selective Classification for Deep Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Volume 30. Available online: https://papers.nips.cc/paper/2017/hash/4a8423d5e91fda00bb7e46540e2b0cf1-Abstract.html (accessed on 12 October 2026).
12. Guo, C.; Pleiss, G.; Sun, Y.; Weinberger, K.Q. On Calibration of Modern Neural Networks. *Int. Conf. Mach. Learn.* **2017**, *70*, 1321–1330. Available online: https://proceedings.mlr.press/v70/guo17a.html (accessed on 12 October 2026).
13. Ovadia, Y.; Fertig, E.; Ren, J.; Nado, Z.; Sculley, D.; Nowozin, S.; Dillon, J.; Lakshminarayanan, B.; Snoek, J. Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. In Proceedings of the Advances in Neural Information Processing Systems 32 (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019; Volume 32. Available online: https://papers.nips.cc/paper/2019/hash/8558cb408c1d76621371888657d2eb1d-Abstract.html (accessed on 12 October 2026).
14. Ma, Y.; Zhou, G.; Wang, S. WiFi Sensing with Channel State Information. *ACM Comput. Surv.* **2019**, *52*, 1–36. [CrossRef]
15. Peng, Y.; Xiang, L.; Yang, K.; Jiang, F.; Wang, K.; Wu, D.O. SIMAC: A Semantic-Driven Integrated Multimodal Sensing and Communication Framework. *IEEE J. Sel. Areas Commun.* **2026**, *44*, 673–688. [CrossRef]
16. Halperin, D.; Hu, W.; Sheth, A.; Wetherall, D. Tool release. *ACM SIGCOMM Comput. Commun. Rev.* **2011**, *41*, 53. [CrossRef]
17. Kotaru, M.; Joshi, K.; Bharadia, D.; Katti, S. SpotFi. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*; ACM: New York, NY, USA, 2015; pp. 269–282. [CrossRef]
18. Yousefi, S.; Narui, H.; Dayal, S.; Ermon, S.; Valaee, S. A Survey on Behavior Recognition Using WiFi Channel State Information. *IEEE Commun. Mag.* **2017**, *55*, 98–104. [CrossRef]
19. Wang, Z.; Zhang, J.A.; Wu, K.; Chen, K.; Xu, M.; Guo, Y.J. Bistatic Passive Sensing via CSI Power. *IEEE Trans. Commun.* **2026**, *74*, 13222–13236. [CrossRef]
20. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778. [CrossRef]
22. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2018; pp. 7132–7141. [CrossRef]
23. Wu, Y.; He, K. Group Normalization. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–19. [CrossRef]
24. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
25. Ben-David, S.; Blitzer, J.; Crammer, K.; Kulesza, A.; Pereira, F.; Vaughan, J.W. A theory of learning from different domains. *Mach. Learn.* **2010**, *79*, 151–175. [CrossRef]
26. Sun, B.; Saenko, K. Deep CORAL: Correlation Alignment for Deep Domain Adaptation. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 443–450. [CrossRef]
27. Niculescu-Mizil, A.; Caruana, R. Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning—ICML '05*; ACM: New York, NY, USA, 2005; pp. 625–632. [CrossRef]
28. Chow, C. On optimum recognition error and reject tradeoff. *IEEE Trans. Inf. Theory* **1970**, *16*, 41–46. [CrossRef]
29. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
30. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
31. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]
32. Saito, T.; Rehmsmeier, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE* **2015**, *10*, e0118432. [CrossRef] [PubMed]

33. Davis, J.; Goadrich, M. The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd International Conference on Machine Learning—ICML '06, Pittsburgh, PA, USA, 25–29 June 2006; pp. 233–240. [CrossRef]
34. Brier, G.W. Verification of Forecasts Expressed in Terms of Probability. *Mon. Weather Rev.* **1950**, *78*, 1–3. [CrossRef]
35. Efron, B. Bootstrap Methods: Another Look at the Jackknife. *Ann. Stat.* **1979**, *7*, 1–26. [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.