UNIVERSITY OF WEST ATTICA

eISSN:3136-7658

Applied Manufacturing Intelligence Journal

Vol. 5, No. 3, July, 2026



Multi-Rate Mamba2-UNet for Factory Wearable Safety Event Detection

Indrek Jõhvi[1,*], Daan Eelmaa[1], Gunnar Hõbe[1]

[1] *Institute of Mobile IT, Võru Vocational College, Võru, 65602, Estonia*

Received 13 April 2026; Revised 23 July 2026; Accepted 27 July 2026; Published 30 July 2026

Abstract

Factory wearable systems collect inertial, physiological, and proximity signals at different sampling rates, while many event detectors require a shared temporal grid. This study proposes a multi-rate Mamba2-UNet model for safety event detection in industrial wearable networks, integrating rapid inertial changes, gradual posture transitions, and intermittent proximity observations. The architecture combines rate-specific Mamba2 encoders, a cross-rate alignment gate, and a temporal UNet decoder to reconstruct frame-level event boundaries while retaining long-range context. The evaluation dataset comprises five channel groups operating at three acquisition rates and covers slips, near collisions, fatigue posture, and tool-zone intrusion. Against LSTM, temporal convolutional network, Transformer-UNet, and single-rate Mamba baselines, the proposed model achieves 94.8% event F1, 96.1% safety-critical recall, 31 ms median model inference latency, and 0.038 s boundary error on the main test split. Removing cross-rate alignment reduces F1 by 3.4 percentage points, while replacing Mamba2 blocks with convolutional encoders increases latency by 17 ms. Class-specific evaluation connects these differences to abrupt transitions, sustained posture changes, and sparse environmental observations. The architectural contribution is the integration of native-rate state retention with frame-level reconstruction. The reported comparisons describe the stated test split, and the computational timing concerns inference on an available input window. These results support further evaluation of rate-aware state space modeling for accurate and timely wearable safety monitoring.

Keywords: Multi-Rate Wearables, Mamba2-UNet, Safety Event Detection, Industrial Sensing, Temporal State Space Modeling

1. Introduction

Nowadays, many modern factories use wearable sensors to detect unsafe behavior and fatigue-related behavior in workers in advance, as well as proximity hazards [1]. Inertial units in the helmet, wrist and belt can obtain high-frequency data of slip, impact and sudden rotation; physiological sensors and zone beacons generally provide lower-frequency but contextually rich evidence of effort, posture and location [2]. The extended sensing environment can be applied to safety supervision, and even in cases of visual occlusion or absent personal protective equipment, dynamic work cell changes are difficult to monitor by conventional cameras and machine logs [3]. Wearable safety events have a temporal structure of sudden changes in motion and prolonged physiological or postural deviations [4]. Hand-vibration patterns preceding tool misuse typically occur within a subsecond interval, whereas fatigue-related posture may persist for several minutes. Proximity alarms are usually sparse and asynchronously time-stamped packets. Resampling all streams to a shared clock can introduce artificial correlations, blur event boundaries, and increase computational cost [5]. These temporal differences motivate an architecture that explicitly represents heterogeneous sampling schedules and the physical processes associated with each stream. Rapid body motion requires sufficient inertial sampling

Corresponding author: indrec.jo@vvk.ee

https://doi.org/10.68406/amij.2026.vol5iss3np3:36-51

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

resolution to preserve impact signatures, while physiological context develops over longer intervals. Proximity observation is based on packet arrival and the geometry of the working area. Event evidence becomes available at different times, and contextual observations may arrive after the initial motion transition. The detector therefore requires temporal integration with sufficient resolution for timely warning.

Previous wearable safety studies have developed classifiers for falls, ergonomic risk, and near misses, commonly using a shared sampling frequency or compact feature window [6]. These representations provide limited temporal detail for events with heterogeneous sensing schedules. Deep recurrent models support sequential learning, while sustained factory operation places demands on memory access and state updates under intermittent observations [7]. Temporal convolutional networks support parallel computation, with fixed receptive fields determining the temporal scales represented by each layer [8]. Attention-based models capture long-range dependencies, and their quadratic token-interaction cost becomes substantial when high-frequency inertial signals are combined with slower physiological and proximity observations [9]. State space models offer linear-time sequence processing, while many activity-recognition pipelines continue to address irregular sampling during preprocessing [10]. Factory safety monitoring requires classification accuracy, boundary precision, timely response, and stable operation across heterogeneous sensors. Similar local movements can correspond to different hazards: routine tool relocation may resemble restricted-zone entry, and post-lift recovery may resemble the onset of sustained fatigue. Task context and environmental conditions therefore contribute to event interpretation. A suitable temporal representation must integrate brief motion transitions, sustained body states, and intermittent environmental observations within a shared inference pathway. Variation in movement intensity and sensor orientation across workers further motivates robust temporal encoding.

This paper develops a Multi-Rate Mamba2-UNet model for factory wearable safety event detection. Its contribution is the integration of native-rate Mamba2 branches, timestamp-aware cross-rate alignment, and boundary-sensitive temporal reconstruction. Each sensing rate is processed by a dedicated branch, and aligned branch states are decoded into event probabilities by a temporal UNet. The five wearable channel groups comprise tri-axial acceleration, tri-axial angular velocity, heart-rate variability, posture inclination, and ultra-wideband proximity. Native-rate processing preserves fast transitions and retains slower context for segmentation. The engineering evaluation examines accuracy, safety-critical recall, latency, boundary error, calibration, and robustness to missing packets. The intended setting is continuous monitoring in assembly and material-handling areas where workers alternate between stationary tasks, walking, and equipment interaction. A segmentation output associates each alert with a bounded interval and describes its probability trajectory. Timestamp-aware fusion connects this output to heterogeneous sensor states. Controlled changes in packet availability, sensor orientation, and window duration examine sensitivity to acquisition and operating conditions. The experimental protocol specifies the signal rates, event categories, data split, and operating settings used to interpret the reported performance. Acquisition timing is central because the distribution of samples across branches determines the amount and timing of evidence available to the decoder. The integration is evaluated within the stated experimental scope.

The paper follows the progression from heterogeneous wearable acquisition to temporally resolved safety detection, connecting the sensing assumptions with the architecture and its experimental evaluation. Section 2 reviews wearable safety detection and multi-rate temporal modeling, including synchronous representations and state space memory. Section 3 defines the signal groups, event labels, and alignment objectives for retaining native acquisition schedules while producing predictions on a shared evaluation grid. Section 4 describes the rate-specific Mamba2 encoder, content-aware alignment gate, temporal UNet decoder, and training objective for class recognition and boundary localization. Section 5 presents the evaluation protocol and compares detection performance, module ablations, sensor perturbations, and computational requirements. The tables document signal organization and principal performance measures. The figures examine class-specific detection, parameter sensitivity, and the relationship between throughput and model size. The accompanying analysis relates aggregate performance to event characteristics, acquisition conditions, and model configuration. This organization separates the definition of the sensing problem from the mechanisms used to encode and align its observations, then examines how those mechanisms affect recognition and execution. Section 6 concludes the study, identifies limitations of the evaluation scope, and outlines field validation and adaptation research relevant to industrial deployment. The methodological and experimental sections jointly establish the conditions under which the reported comparisons should be interpreted.

2. Related Work

2.1. Wearable Safety Event Detection

Wearable safety monitoring has developed from rule-based alarms toward data-driven recognition of hazardous movement and workload patterns. Wearable stress monitoring illustrates the use of physiological measurements and learned features to characterize demanding conditions [11]. Such approaches offer interpretable indicators, while their behavior depends on sensor placement, task cadence, and the measured population. RFID and wearable activity-recognition methods combine acquisition, preprocessing, and machine-learning representations to identify human

behavior [12]. Window-level aggregation can weaken the onset and offset information required for timely warning. A near-collision event can contain a brief proximity observation, simultaneous trunk rotation, and delayed recovery. Combining these components into an averaged feature vector can obscure their temporal sequence and weaken the connection between classification and alarm timing.

Deep learning represents wearable sequences through local and longer-range temporal features. ActivityMamba combines convolutional processing with Mamba-based modeling for efficient human activity recognition [13]. This example motivates the joint use of local movement texture and broader sequence context, while differing from the proposed native-rate architecture and temporal decoder. Factory wearable streams have heterogeneous schedules: inertial measurements may run at 100 Hz, physiological observations at 1-10 Hz, and proximity packets at irregular intervals. Synchronization and packet availability therefore influence the evidence seen by a sequence model. Multiresolution wavelet-attention provides another approach to extracting informative features from sensor-based activity sequences [14]. These directions support investigation of temporal scale and representation quality, while native-rate alignment and dense event-boundary reconstruction remain the focus of the present method.

2.2. Multi Rate Temporal Modeling and State Space Networks

Multi-rate temporal modeling addresses measurements acquired by devices with different bandwidths. Interpolation, downsampling, and filtering provide established ways to organize heterogeneous observations, but their assumptions can constrain nonlinear event representation. Energy management and imputation in wearable devices address incomplete sensing under resource constraints [15]. A slip onset is a short nonstationary burst, whereas fatigue posture develops gradually; a uniform representation can attenuate the burst or repeat slower observations excessively. Multimodal fusion with transformer techniques offers a learned approach to combining activity-recognition evidence [16]. Such fusion can capture complementary information, while its computational requirements and acquisition assumptions must be considered for high-frequency wearable input.

State space sequence models maintain temporal context through compact propagated memory. TF-Mamba provides a lightweight state-space approach to Wi-Fi-based human activity recognition [17]. Its sensing setting differs from body-worn inertial and physiological acquisition, but it illustrates the use of efficient state-space sequence processing for activity-related signals. A direct single-rate detector requires a shared input clock, and preservation of slower context depends on the representation used when high-rate observations dominate. Timestamp-supervised action segmentation investigates temporal labeling and boundary estimation under sparse supervision [18]. This supports the importance of interval localization as a separate modeling objective. The UNet-style decoder in the present architecture is used for dense temporal reconstruction. The design challenge is to integrate rate-aware encoding, asynchronous alignment, and boundary-sensitive decoding within one wearable safety framework.

3. Multi Rate Wearable Event Formulation

3.1. Signal Groups and Event Labels

The studied system contains wearable nodes placed at the helmet, wrist, and waist of a worker operating in a mixed assembly and material-handling area. Each node contributes a different view of safety behavior. The wrist unit records acceleration and angular velocity associated with tool use and sudden hand motion. The waist unit reports trunk inclination and low-frequency posture drift. The helmet unit carries proximity packets and head-motion cues relevant to near collision events. The sensing system preserves separate acquisition schedules: inertial channels are treated as high-rate streams, posture channels as medium-rate streams, and heart-rate or proximity variables as low-rate streams. This design reflects practical wearable networks, where power budget, communication protocol, and sensor physics impose different sampling schedules.

Let $x_r(t)$ denote the observation sequence from rate group r, where r identifies the high-, medium-, or low-rate stream. Each group has its own timestamp set and channel dimension. The target $y(t)$ is a frame-level safety label defined on an evaluation grid that supports event boundary scoring. The model estimates the posterior probability of safety event class c at evaluation frame t while processing each rate group at its native resolution. Figure 1 shows the workflow from heterogeneous wearable acquisition to frame-level event labeling.

$$\mathcal{X} = \{x_{r,k} \mid r \in \{H, M, L\}, k = 1, \dots, N_r\}, y_t \in \{0,1\}^C, t = 1, \dots, T \quad (1)$$

The following equation defines the observation collection and the frame-level label space:

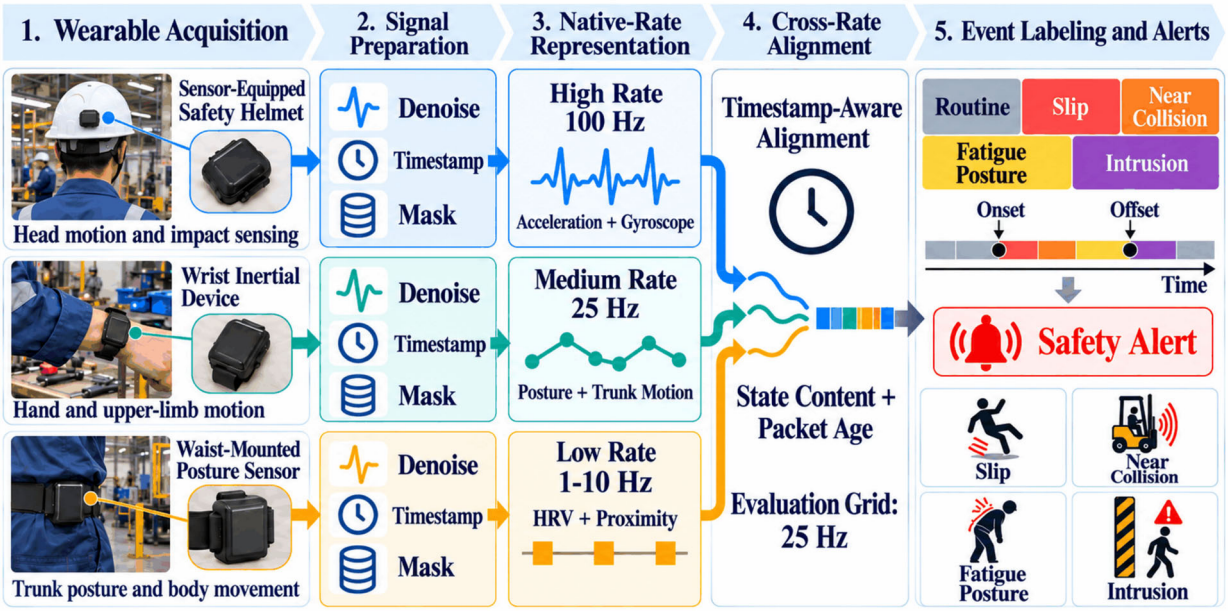


Figure 1. Multi-rate wearable safety event detection workflow from heterogeneous factory sensors to frame-level safety labels

The high-rate group comprises tri-axial acceleration and angular velocity at 100 Hz. The medium-rate group contains posture inclination and a derived, smoothed trunk-acceleration feature at 25 Hz. The low-rate group contains heart-rate variability and proximity observations at 1-10 Hz, depending on packet availability. These are channel groups associated with the wearable nodes described above; derived features are distinct from additional physical sensing devices. Event categories comprise routine operation, slip or stumble, near collision, fatigue posture, and tool-zone intrusion. They cover abrupt and gradual hazards that require complementary information across sampling rates. Acceleration can reveal a slip before postural recovery, sustained inclination can indicate fatigue posture, and proximity, wrist motion, and task phase jointly inform intrusion detection.

For rate group r with sampling interval Δ_r, the raw stream consists of observation vectors $x_{r,k}$. The evaluation interval Δ_e is specified independently of the native sensing intervals. Each branch retains its native sequence and exposes hidden states to the fusion module operating on the evaluation grid. The event target y_t is encoded as a one-hot vector, with C denoting the number of classes and T the number of evaluation frames. A boundary tolerance of 150 ms is used during evaluation to distinguish delayed alarms from acceptable temporal shifts. The observation equation preserves the separate sequence length N_r of each rate group while assigning the prediction target to a shared evaluation grid.

Label construction follows an event-segmentation formulation with explicitly bounded intervals. Each annotated event has an onset, an active interval, and an offset. The model produces a probability trajectory that rises within the warning deadline and falls after the event ends. This requirement demands greater temporal specificity than assigning one label to an entire window. If a 12 s window contains a 0.8 s slip and 11.2 s of normal recovery, onset-aware labels identify the short hazardous interval within the longer sequence. The evaluation grid therefore preserves dense labels, while the loss function assigns additional weight to boundary frames and safety-critical classes. Routine frames are retained because false alarms during ordinary work can reduce confidence in the monitoring system and increase unnecessary intervention.

3.2. Cross Rate Alignment Objective

The alignment problem is both temporal and semantic. A low-rate proximity packet can be highly informative when it arrives near a burst of wrist acceleration, while the same packet may have limited relevance during routine walking. The model uses learned alignment weights conditioned on timestamp and signal content. Each branch emits a hidden state $h_{r,k}$, and a cross-rate gate maps nearby states onto the evaluation grid. The gate is trained jointly with the event loss so that alignment emphasizes states associated with accurate detection. Its fusion weights depend on state content, timestamp offset, and event context.

Branch states contribute to the evaluation-frame representation through the alignment equation:

$$u_t = \sum_{r\in\{H,M,L\}} \sum_{k} \alpha_{t,r,k} W_r h_{r,k},$$
$$\alpha_{t,r,k} = softmax_{\,r,k}\left(\phi\left(h_{r,k}, \tau_t - \tau_{r,k}\right)\right) \quad (2)$$

Here, u_t is the fused representation at evaluation frame t, W_r projects the branch state into a shared feature space, and $\alpha_{t,r,k}$ is the normalized contribution of a candidate state. The timestamps τ_t and $\tau_{r,k}$ identify the evaluation frame and native observation, respectively. The softmax is evaluated over the candidate states admitted to the alignment neighborhood. Within the scoring function ϕ, the timestamp offset is normalized by the native sampling interval of the corresponding branch to account for differences in state density.

The content term is computed from compact projections of the branch hidden state, the preceding fused state, and a learned event-query vector. Acceleration evidence receives a content-dependent contribution when its state resembles a safety-relevant transition. Proximity evidence is weighted using the surrounding motion and posture context associated with near collision or intrusion. The alignment equation consequently provides a shared representation whose composition changes with the event evidence available at each frame. Native timestamps, packet availability, and branch-state content remain accessible to the gate, supporting asynchronous observations and irregular low-rate updates. Candidate-state selection defines the observations admitted to alignment; softmax weighting determines their relative contributions within that set. Timestamp scaling inside the score accounts for the native interval of each branch.

4. Proposed Mamba2 UNet Architecture

4.1. Rate Specific Mamba2 Encoding

The proposed architecture is organized around three design principles. First, each rate group is encoded at its natural temporal resolution to preserve high-frequency transients and the native density of low-frequency observations. Second, fusion determines when a slow contextual cue is relevant to a fast event boundary. Third, a hierarchical temporal decoder reconstructs dense event probabilities because factory hazards commonly contain a short onset, an active high-risk interval, and a recovery tail. Figure 2 presents the native-rate encoder, alignment module, and temporal decoding pathway.

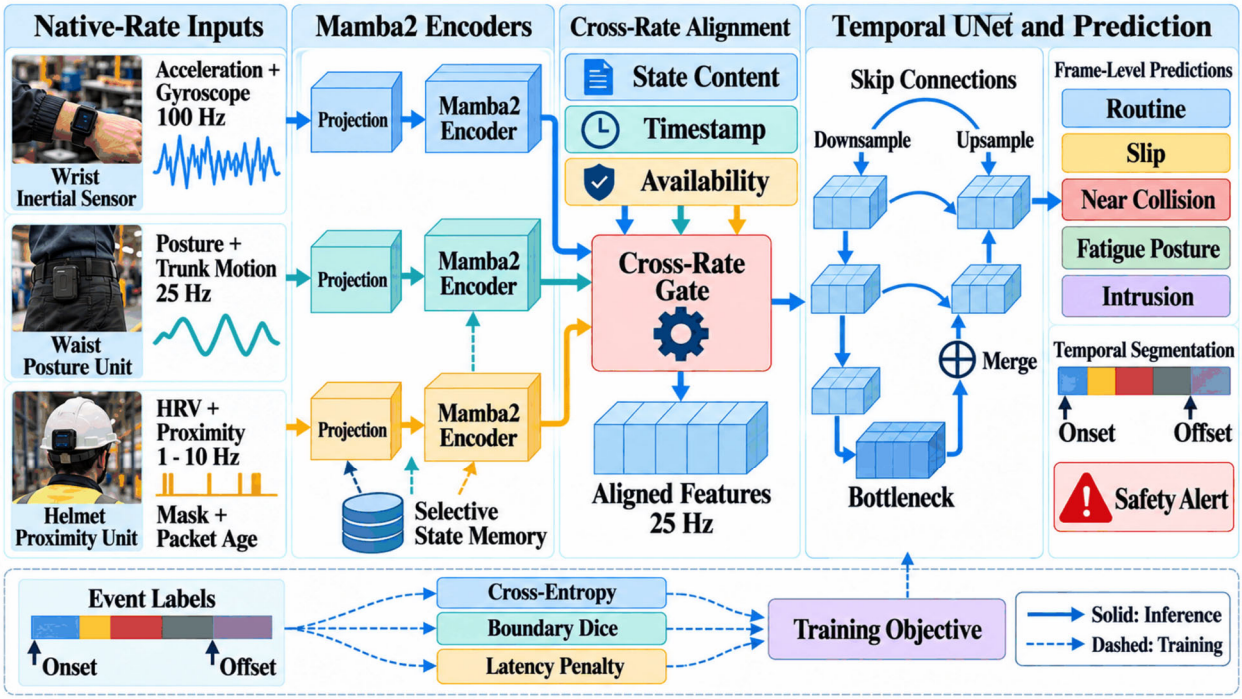


Figure 2. Architecture of the proposed Multi Rate Mamba2 UNet with native-rate branches, cross-rate alignment gate, and temporal decoder

Each rate branch begins with a shallow projection layer that maps the sensor channels into a common hidden dimension. Stacked Mamba2 blocks then process the projected sequence. The high-rate branch uses a limited block depth to control latency, while the medium- and low-rate branches use larger state dimensions to retain longer behavioral context. Selective state updates accommodate the temporal imbalance between frequent routine motion and sparse hazardous transitions. For branch r at native step k, the state-transition equation describes memory propagation and

output formation: The displayed recurrence is a compact selective-state abstraction. Reproducing the actual Mamba2 encoder additionally requires verified branch depths, state dimensions, discretization settings, and block implementation details.

$$s_{r,k} = A_r(x_{r,k})s_{r,k-1} + B_r(x_{r,k})x_{r,k},$$
$$z_{r,k} = C_r(x_{r,k})s_{r,k} \tag{3}$$

The state-transition equation expresses the branch dynamics in a compact state space form. The internal memory $s_{r,k}$ combines the preceding state with the current observation, while $z_{r,k}$ exposes the encoded information to subsequent projection and alignment. The observation-dependent operators A_r, B_r , and C_r regulate state retention, input contribution, and output extraction, respectively.

Selective parameterization allows the state transition and input contribution to vary with the current observation. During routine wrist motion, the branch can preserve stable context. A sudden rotation accompanied by an acceleration burst can increase the contribution of new evidence. This behavior is particularly relevant to slip and tool-zone intrusion events, where the earliest discriminative frames determine alarm timeliness. The branch output $z_{r,k}$ is projected into the fusion feature space and passed to the cross-rate gate. This projection preserves the temporal index associated with each native-rate observation.

The encoder applies residual normalization after each Mamba2 block and a compact channel-mixing layer before the branch output projection. The high-rate branch uses shorter temporal pooling to preserve transient timing, whereas the low-rate branch uses wider state memory and native packet timing. Missing low-rate observations are represented by a mask channel and a time-since-update feature. These auxiliary variables distinguish a stable measurement from an unavailable observation. Training and inference use the same masking convention, allowing the encoding and fusion pathways to learn how packet availability affects the reliability of temporal evidence.

4.2. Multi Rate Fusion and UNet Decoding

The fusion module receives hidden states from all rate branches and maps them onto the evaluation grid. It computes a content-aware weight for each branch and nearby timestamp candidate, forming the aligned representation u_t. A temperature parameter regulates how strongly the gate concentrates its contribution on a dominant branch or distributes it across multiple rates. A moderate temperature supports events whose evidence is distributed across sensing modalities. A near collision, for example, can combine a low-rate proximity observation with a high-rate braking movement and a medium-rate torso turn.

The aligned representation is decoded by a temporal UNet. Downsampling stages aggregate event context over progressively larger windows, while upsampling stages restore frame-level resolution through skip connections from corresponding encoder features. This arrangement connects local onset texture with the surrounding behavior needed for event interpretation. The decoder produces the class probability sequence $p_{t,c}$ on the evaluation grid. Recognition accuracy, boundary localization, and timely response are coupled through the training-loss equation:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda_b \mathcal{L}_{Dice}^{boundary} + \lambda_\ell max(0, d_t - d_{max}) \tag{4}$$

In this equation, $\mathcal{L}_{\mathrm{CE}}$ denotes weighted cross-entropy, and $\mathcal{L}_{\mathrm{Dice}}^{\mathrm{boundary}}$ measures agreement around annotated event boundaries. The coefficients λ_b and λ_ℓ control the contributions of boundary reconstruction and delayed recognition. The quantity d_t represents detection delay for a safety-critical event, while $d_{\max}$ specifies the permitted delay.

The latency term assigns greater importance to timely recognition of slip, near collision, and intrusion than to the more gradual development of fatigue posture. The boundary Dice component penalizes predictions with correct class identity but imprecise temporal extent. Their joint contribution encourages early recognition with bounded event intervals. Skip projections preserve local transition features during reconstruction, helping the decoder recover onset and offset information after contextual aggregation. Increasing the boundary coefficient emphasizes interval localization, while increasing the delay coefficient strengthens the penalty beyond the permitted deadline. These training coefficients are distinct from the probability threshold used by the alarm rule.

Each downsampling stage halves the evaluation-frame resolution and increases the channel dimension, allowing the bottleneck to represent several seconds of event context. Each upsampling stage restores temporal resolution and concatenates the corresponding encoder-side feature map. These skip connections retain the short onset patterns needed to distinguish a slip from routine tool movement within a longer contextual window. The final layer is a one-dimensional convolution followed by class-wise sigmoid probabilities, supporting multi-label use while retaining the mutually exclusive labels adopted in the reported evaluation. The main results apply the decision threshold directly to the probability sequence, so latency and boundary error reflect the temporal output of the model. The reported evaluation concerns mutually exclusive labels. Multilabel compatibility is an architectural property whose empirical performance and conflict-resolution behavior require separate assessment.

5. Experimental Data and Analysis

5.1. Evaluation Protocol and Comparative Detection Performance

The evaluation uses a factory wearable safety dataset containing mixed-rate sensing streams from assembly and logistics tasks. It includes 42.6 h of annotated recordings, 18 worker-role profiles, and 11 task templates. The five channel groups comprise acceleration, angular velocity, posture inclination, heart-rate variability, and proximity. Annotations cover 1,840 routine windows and 1,126 safety events: 312 slip or stumble events, 264 near-collision events, 286 fatigue-posture events, and 264 tool-zone intrusions. The reported 12:3:3 split allocates the 18 worker-role profiles to training, validation, and testing. The test set contains held-out profiles and task combinations. The model uses 12 s windows with 50% overlap and evaluates predictions on a 25 Hz frame grid. Participant variability and transfer across activity-recognition settings motivate careful separation of evaluation populations [19]. A worker-disjoint design requires all windows from one profile to remain within its assigned partition. Window overlap describes temporal sampling within recordings and is separate from the identity criterion used to define the split.

The compared methods comprise a feature-based random forest, a two-layer LSTM, a temporal convolutional network, Transformer-UNet, single-rate Mamba, and the proposed Multi-Rate Mamba2-UNet. All neural baselines use the same data splits and labels, with comparable hidden dimensions. Single-rate models receive streams resampled to 25 Hz, while the proposed model retains native rates and aligns hidden states through its cross-rate gate. Hyperparameters are selected on the validation set, and final results are computed once on the test set. Metrics include event F1, safety-critical recall, false alarms per hour, boundary error, expected calibration error, median inference latency, and throughput. Temporal representation learning provides relevant context for sensor-based activity recognition [20]. Event F1, class-specific precision and recall, and the conditional confusion matrix are separate summaries whose denominators depend on interval matching, aggregation, and the admitted sample population. Boundary error and inference latency separately describe temporal localization and computational response.

Table 1 summarizes the dataset organization and signal-rate groups. The high-rate inertial group contains the largest number of samples per event but provides onset evidence complemented by the slower context required for fatigue posture and intrusion classification. The low-rate group has fewer observations, but its proximity packets reduce ambiguity in near collision and restricted-zone scenarios. At 100 Hz, 4.26 million sample vectors correspond to approximately 11.83 h of one continuously observed stream. Reconciling this count with 42.6 h requires the valid-duration and channel-counting conventions used for Table 1.

Table 1. Dataset organization and wearable signal-rate groups

Group	Channels	Rate	Role in Detection	Samples or Events
High rate	Acceleration, angular velocity	100 Hz	Slip onset, hand motion, impact texture	4.26 million samples
Medium rate	Posture inclination, trunk motion	25 Hz	Fatigue posture, recovery state, body orientation	1.06 million samples
Low rate	Heart-rate variability, proximity	1-10 Hz	Exertion context, collision and zone evidence	0.18 million packets
Labels	Routine, slip, near collision, fatigue, intrusion	25 Hz grid	Frame-level event scoring and boundary analysis	1,126 events

Table 2 reports the principal detection results on the test split. The proposed model reaches 94.8% event F1, 96.1% safety critical recall, and 0.42 false alarms per hour. Compared with Transformer-UNet, it improves F1 by 2.7 percentage points and reduces median latency from 49 ms to 31 ms. Compared with single-rate Mamba, it improves boundary error from 0.052 to 0.038, indicating an improvement in event-edge localization alongside class recognition. The displayed recalls for slip, near collision, and intrusion average to 96.1%; verification of the implemented aggregation convention requires the evaluation code.

Table 2. Main detection performance on the worker-disjoint test split

Method	Event F1 (%)	Safety Recall (%)	False Alarms/h	Boundary Error (*s*)	Latency (*ms*)
Random Forest	81.6	82.7	1.84	0.118	18
LSTM	86.9	88.1	1.21	0.083	44
TCN	89.4	90.2	0.96	0.071	37
Transformer-UNet	92.1	93.4	0.61	0.055	49
Single-rate Mamba	91.7	92.8	0.57	0.052	45
Multi Rate Mamba2 UNet	94.8	96.1	0.42	0.038	31

The dataset also includes controlled nuisance conditions that are common in wearable deployments. Packet loss was injected into the low-rate channel at 10%, 20%, 30%, and 40% levels during robustness testing, but the main model was selected on the clean validation split to avoid overfitting to a single corruption pattern. Wearable placement sensitivity was assessed through controlled rotation of the sensor-coordinate orientation of wrist, waist, and helmet streams, which approximates loose straps, different body shapes, and partial device movement during work. Class imbalance was handled by inverse-frequency weighting in the cross-entropy term, and the reported test set retained the evaluation class distribution. This choice keeps the false-alarm metric meaningful because routine operation remains the dominant condition. All reported values are averages over the fixed test split, and threshold sensitivity is examined across a specified operating range. These point values describe the fixed test split. Variability across repeated training runs, workers, and factories requires a separate uncertainty assessment.

Figure 3(a) reports the six methods: event F1 is 81.6% for random forest, 86.9% for LSTM, 89.4% for the temporal convolutional network, 92.1% for Transformer-UNet, 91.7% for single-rate Mamba, and 94.8% for the proposed model. The 2.7 percentage-point margin over Transformer-UNet is smaller than the 13.2-point margin over random forest, indicating that temporal representation already accounts for a substantial part of the separation between methods. The additional gain is compatible with the joint use of native-rate encoding and boundary reconstruction, with module-specific contributions examined in the ablation analysis. Figure 3(b) resolves recall for slip, near collision, and intrusion, giving 95.6%, 96.8%, and 95.9% for the proposed detector. The corresponding single-rate values are 92.5%, 93.7%, and 92.1%. The larger gain for intrusion, at 3.8 percentage points, is consistent with the need to combine wrist motion with sparse zone evidence. Slip recognition relies more directly on high-rate motion texture, which offers a plausible explanation for its smaller improvement. These values describe class-specific detection behavior under the stated evaluation split. The relative gains should also be read against the six-method ordering. LSTM improves on the feature classifier by 5.3 points, and temporal convolution adds a further 2.5 points. Those changes indicate that the evaluated streams contain sequential information accessible through progressively richer temporal representations. The remaining differences concern how that information is aligned and reconstructed. Energy-efficient wearable communication and monitoring motivate explicit consideration of operational resource constraints [21].

Figure 3(c) shows routine-operation false alarms of 1.84, 1.21, 0.96, 0.61, 0.57, and 0.42 per hour in the same method order. Relative to Transformer-UNet, the proposed model reduces the alarm rate by approximately 31.1%. Because critical recall also increases, this change is compatible with better separation between hazardous and routine segments at the selected operating point. Figure 3(d) gives boundary errors of 0.031 s for slip, 0.036 s for near collision, 0.047 s for fatigue posture, and 0.039 s for intrusion. The differences across events reflect their temporal characteristics: fatigue onset is gradual, so a small shift in an annotated transition can affect boundary scoring more than for an abrupt slip. Boundary errors are interval-matching summaries on the evaluation grid and can be fractional averages even when individual predictions use discrete frames. Figure 3(e) presents confidence-bin calibration, with expected calibration error of 0.024 for the proposed detector and 0.041 for Transformer-UNet. Calibration assessment relates predicted confidence to empirical outcome frequency across bins [22]. In the evaluation, reduced calibration error suggests that a probability threshold can be interpreted more consistently across the central confidence range, while rare high-risk cases still warrant class-specific inspection. A reduction of 0.19 alarms per hour corresponds to approximately one fewer nuisance alert over a five-hour monitoring interval under a stationary event mixture. Such a conversion illustrates burden at the reported operating point; actual shift-level counts would depend on task changes, correlated errors, and the duration of uninterrupted routine observation. The four displayed boundary errors average to 0.03825 s, which rounds to 0.038 s at the precision used in Table 2; the underlying weighting convention requires confirmation from the evaluation code.

Figure 3(f) separates median inference latency from its 90th percentile. The proposed model records 31 ms and 58 ms, compared with 49 ms and 91 ms for Transformer-UNet and 45 ms and 82 ms for single-rate Mamba. The median improvement over Transformer-UNet is 36.7%, while the reduction at the upper percentile is 36.3%. Similar relative changes at both percentiles suggest that the measured efficiency advantage extends beyond the typical window. Model inference latency measures computation after input preparation; event onset delay also depends on the arrival of discriminative observations and the alert decision rule. The two quantities therefore represent different parts of the warning process. Native-rate processing limits the repetition of slow observations, and state space encoding gives the architecture a computational explanation for the observed pattern. Hardware identity, processor utilization, warm-up, and batching affect the absolute values. The reported measurements compare methods under a shared execution convention; processor-specific gateway benchmarking determines the service budget for each deployment. Upper-percentile latency is especially relevant when several windows compete for execution resources. An average value can conceal occasional slow inference, and those windows may coincide with the arrival of multiple sensor updates. Recording latency distributions separately for routine and hazardous windows would help determine whether computation becomes slower precisely when an alert is required. Inference timing concerns computation on an available input window. Startup buffering, the 12 s context requirement, sensor communication, and the alert rule contribute separately to complete warning latency.

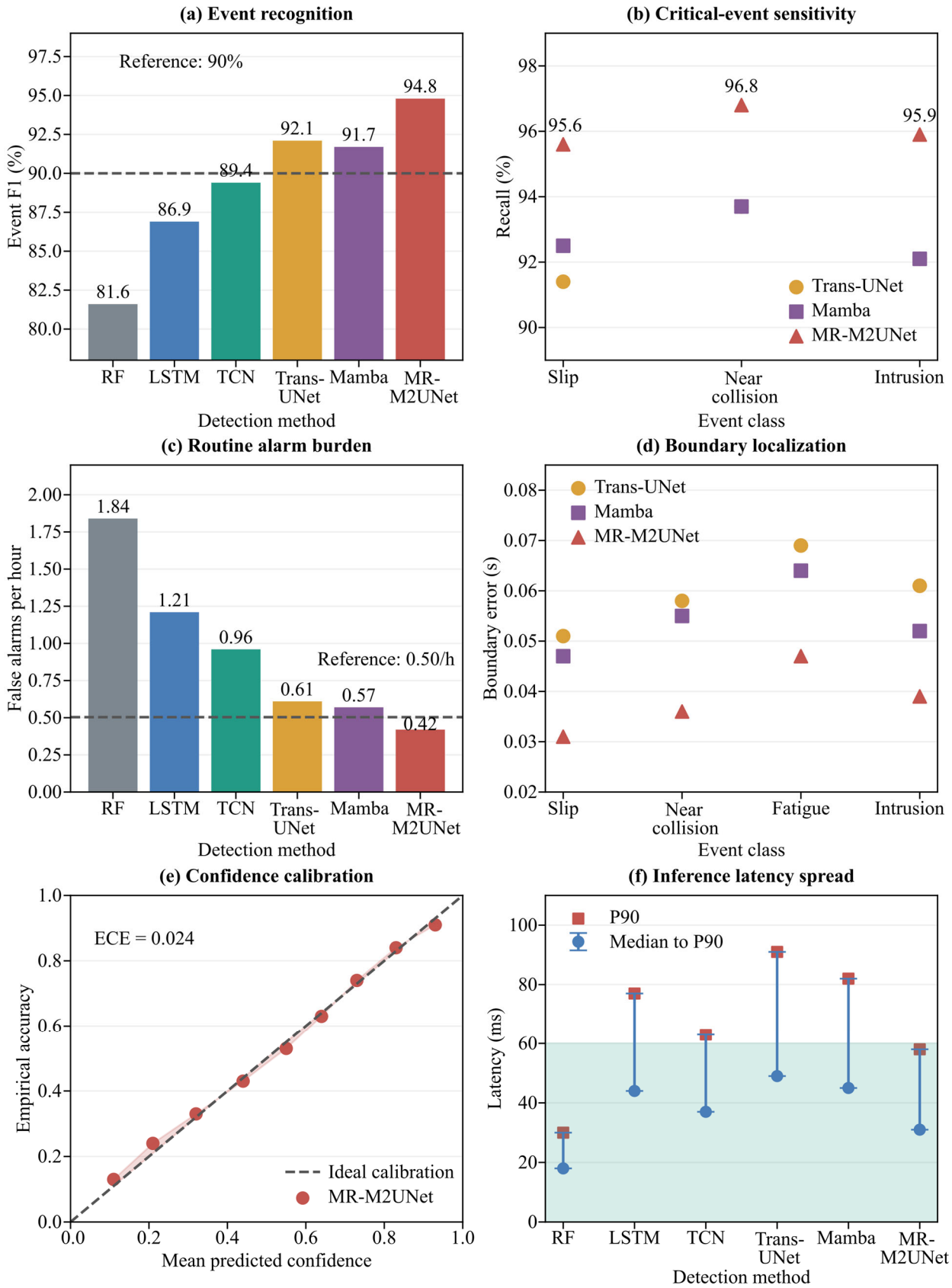


Figure 3. Comparative detection behavior: (a) event F1 across methods; (b) safety critical recall by event type; (c) false alarms per hour; (d) boundary error by event type; (e) calibration by confidence bin; (f) latency distribution

The comparative analysis depends on the weighting of event categories. Routine segments occupy more time than brief hazards, and long fatigue intervals contribute many evaluation frames. A frame-dominated score can favor accurate long-interval predictions even when short-event onsets are poorly localized. Event-level F1 evaluates recognized intervals, while the boundary metric exposes temporal displacement. Precision and recall should also be compared at a

shared operating threshold before class-specific alarm routing is introduced. For slip events, median onset delay is 29 ms for the proposed model, 44 ms for single-rate Mamba, and 52 ms for Transformer-UNet. These summaries measure the interval between annotated onset and the alert decision. Authorized and unauthorized handling can produce similar wrist movement, making zone context relevant to interpreting a high movement score. For near collision, packet age affects whether proximity evidence remains informative as relative position changes. These mechanisms create class-specific errors requiring inspection of sensor timing and event context alongside aggregate throughput and F1. Stratification by packet age and motion intensity would help distinguish stale-observation errors from ambiguity in event labels and show whether recall gains persist under incomplete context.

5.2. Ablation and Event Diagnostics

The ablation study isolates the contributions of native-rate encoding, alignment, state space blocks, and dense decoding. Figure 4(a) reports median component latencies of 4 ms for preprocessing, 13 ms for branch encoding, 6 ms for alignment, and 8 ms for decoding. Encoding accounts for approximately 41.9% of the 31 ms total, while alignment accounts for 19.4%; optimization focused exclusively on the gate would therefore leave most computation unchanged. Figure 4(b) compares five variants. Removing alignment lowers F1 from 94.8% to 91.4% and increases boundary error from 0.038 s to 0.057 s. A flat decoder yields 90.7% F1 and 0.061 s error, while a convolutional encoder yields 91.9% and 0.054 s. Single-rate Mamba gives 91.7% and 0.052 s. The flat decoder has the largest accuracy and boundary penalties, suggesting that temporal reconstruction contributes directly to interval localization. The gate removal also affects both measures, which is consistent with its role in selecting asynchronous context. The ablation tests change one module at a time and provide local comparisons; interactions between module capacity, training dynamics, and the loss weights require a broader factorial evaluation. The latency decomposition establishes a practical optimization order. Encoding and decoding jointly occupy 21 ms, so changes to their sequence lengths or operator implementation have a larger potential effect than small preprocessing reductions. Any optimization should retain the same boundary-matching protocol because a faster decoder can produce less precise temporal output if its resolution is reduced.

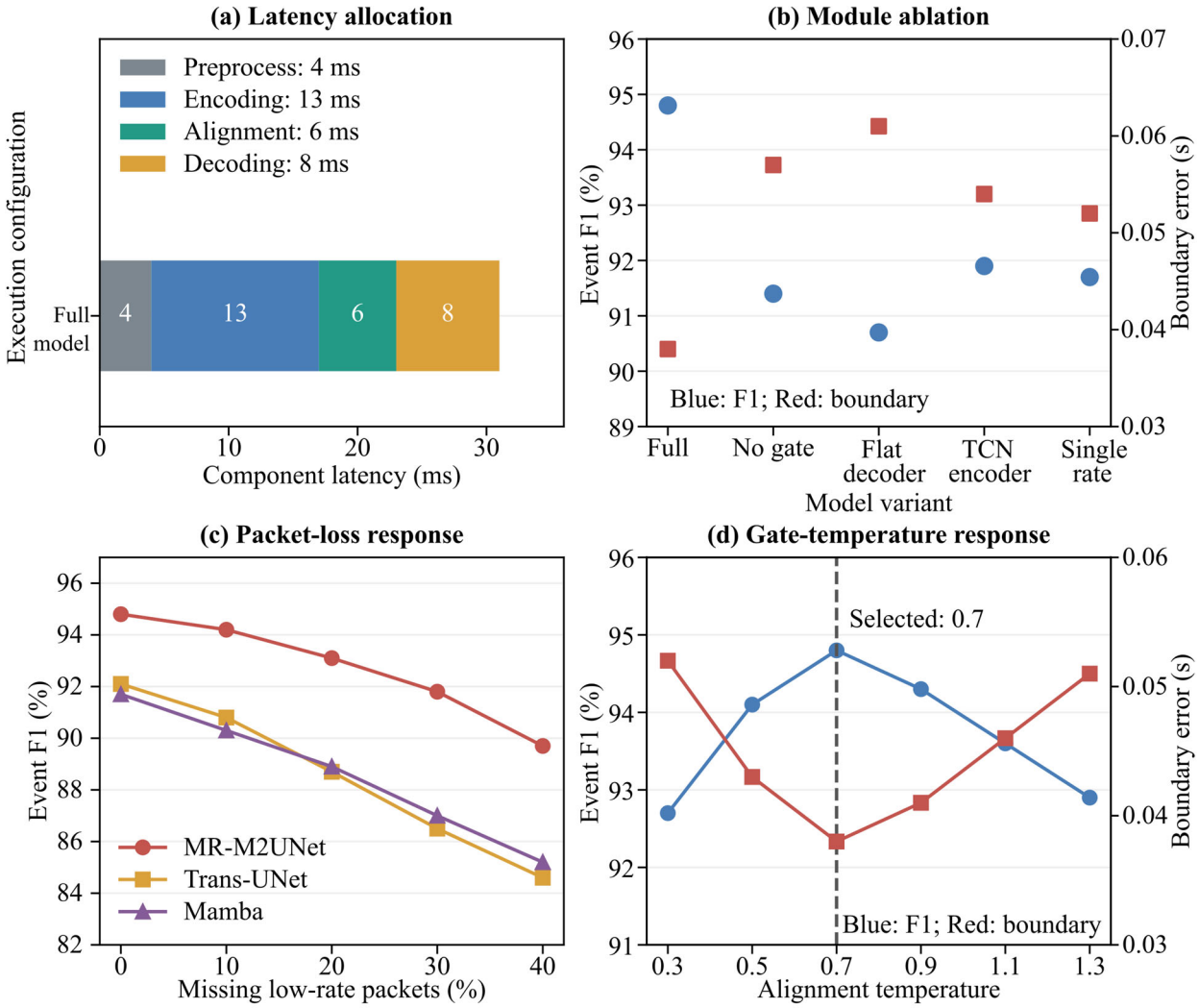


Figure 4. Ablation and sensitivity: (a) latency allocation; (b) accuracy and boundary response; (c) missing-packet robustness; (d) alignment-temperature sensitivity

Robust multi-sensor estimation uses observation availability and reliability to regulate the contribution of individual channels [23]. Figure 4(c) examines random loss of low-rate packets at 0%, 10%, 20%, 30%, and 40%. Proposed-model F1 changes from 94.8% to 94.2%, 93.1%, 91.8%, and 89.7%, respectively, compared with 92.1% to 84.6% for Transformer-UNet and 91.7% to 85.2% for single-rate Mamba across the same range. The advantage over Transformer-UNet expands from 2.7 to 5.1 percentage points, although the absolute decline of 5.1 points remains operationally relevant. Figure 4(d) examines alignment temperatures of 0.3, 0.5, 0.7, 0.9, 1.1, and 1.3. F1 values are 92.7%, 94.1%, 94.8%, 94.3%, 93.6%, and 92.9%, with boundary errors of 0.052, 0.043, 0.038, 0.041, 0.046, and 0.051 s. The intermediate setting produces the strongest joint response. Sharper weighting can discard complementary evidence, while diffuse weighting can dilute a discriminative stream. This interpretation follows the specified gate behavior and constitutes a mechanism to test through recorded branch-state trajectories. The packet-loss sweep assumes independent removal of observations. Consecutive missing packets could produce a longer interval of stale proximity information even at the same overall loss percentage. A burst-loss experiment would therefore provide a different robustness test, particularly for near collision, where the relative distance between a worker and moving equipment can change rapidly.

Figure 5(a) presents precision and recall of 96.2% and 95.6% for slip, 94.9% and 96.8% for near collision, 92.8% and 93.4% for fatigue posture, and 95.1% and 95.9% for intrusion. Fatigue has the weakest pair of scores, consistent with overlap between sustained unsafe inclination and ordinary recovery.

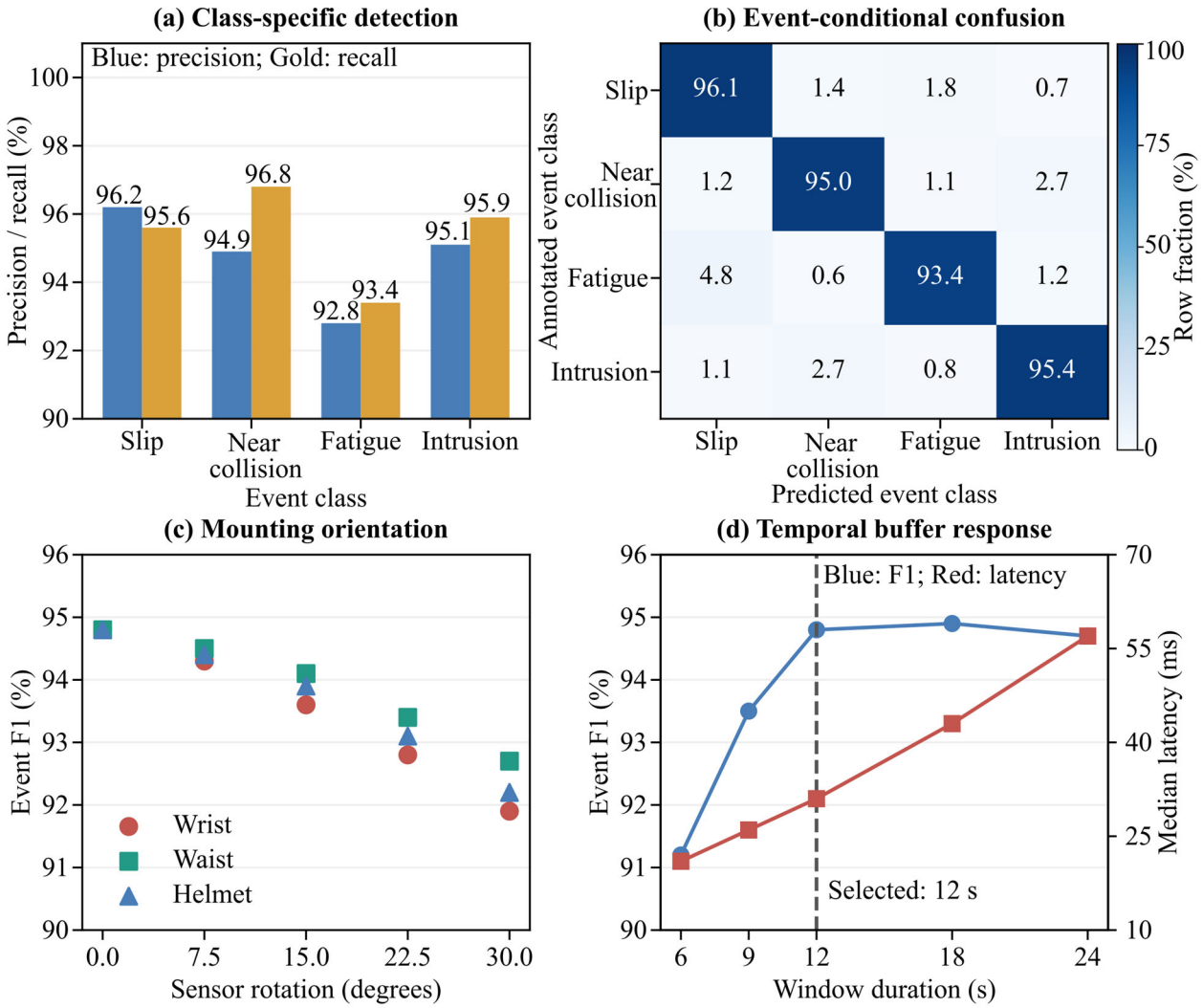


Figure 5. Event diagnostics: (a) class-wise precision and recall; (b) hazardous-event conditional confusion matrix; (c) sensor rotation sensitivity; (d) window duration response

Figure 5(b) displays a row-normalized, event-conditional confusion matrix covering the four hazardous classes. Its rows sum to 100%; diagonal values are 96.1%, 95.0%, 93.4%, and 95.4%. Fatigue-to-slip confusion is 4.8%, while near-collision-to-intrusion confusion is 2.7%. This matrix conditions on hazardous event samples and addresses a different denominator from the full routine-inclusive recall metric. The difference between its diagonal and the recall values therefore follows the evaluation populations. Near collision and intrusion share proximity cues, whereas fatigue and slip can share posture recovery patterns. The displayed errors identify which pairs need additional contextual separation and which label transitions deserve closer annotation review. Row normalization makes the conditional error distribution

visible but suppresses the number of observations supporting each row. The class event counts in the protocol provide complementary prevalence information. A deployment evaluation should additionally retain the raw confusion counts and routine-inclusive errors so that rare classes and repeatedly observed long events can be interpreted with their actual denominators. Class-conditioned error analysis examines the distribution of incorrect predictions across labels [24]. Each matrix row is normalized over the admitted hazardous-event population for that row. The diagonal expresses conditional agreement, whereas a broader recall calculation can use a different evaluation denominator.

Wang et al. [25] investigated compensation strategies for sensor displacement in wrist-worn human activity recognition. Figure 5(c) uses rotation levels of 0, 7.5, 15, 22.5, and 30 degrees, with proposed-model F1 of 94.8%, 94.3%, 93.6%, 92.8%, and 91.9% for wrist rotation. At 30 degrees, waist and helmet rotation give 92.7% and 92.2%, respectively. These controlled rotations preserve the angular unit across devices and isolate orientation disturbance from translational displacement. Figure 5(d) compares window durations of 6, 9, 12, 18, and 24 s. F1 is 91.2%, 93.5%, 94.8%, 94.9%, and 94.7%, while median latency is 21, 26, 31, 43, and 57 ms. Moving from 12 to 18 s adds 12 ms for a 0.1-point F1 change; moving to 24 s adds further cost with slightly lower F1. The selected 12 s buffer preserves the context needed by the evaluated fatigue pattern while limiting computational load. Its suitability depends on the persistence of hazards and the cadence of available observations. Longer-duration fatigue processes would require persistent state beyond this buffer and a separate evaluation of history retention. Orientation disturbance changes the direction of acceleration and angular-velocity components while preserving their temporal organization. This creates a useful controlled test of representation sensitivity. Strap looseness would add time-varying movement between sensor and body, and should be assessed through a separate perturbation model whose frequency content is specified before its detection impact is interpreted.

5.3. Computational Efficiency and Operational Sensitivity

Figure 6 compares parameter count against throughput for all six methods, with marker area encoding event F1. Random forest, LSTM, temporal convolution, Transformer-UNet, single-rate Mamba, and the proposed model occupy 1.9, 3.2, 2.8, 5.6, 3.4, and 3.8 million parameters and process 820, 1,040, 1,320, 1,140, 1,650, and 1,820 frames per second, respectively. The proposed model has approximately 59.6% higher throughput and 32.1% fewer parameters than Transformer-UNet. Compared with single-rate Mamba, it adds 0.4 million parameters while increasing throughput by approximately 10.3%. Deployment-oriented benchmarking relates model size and execution capacity to the number of concurrent sensing streams [26]. Parameter count describes storage and model capacity, whereas throughput also reflects sequence organization, operator implementation, and parallel execution. The ordering of methods consequently reflects both parameter size and execution characteristics. The parameter budget is consistent with a compact alignment and reconstruction path, with operator-level profiling needed to attribute execution differences to particular kernels. The 1,500-frame-per-second reference line marks a service-capacity reference target, allowing methods to be compared against an explicit capacity criterion alongside accuracy. At a 25 Hz evaluation grid, the stated capacity would nominally accommodate multiple simultaneous prediction streams. A capacity estimate obtained by dividing aggregate throughput by the frame rate depends on consistent batching and execution assumptions. Input buffering, scheduling, and repeated overlapping windows must be included before that estimate is translated into a monitored-worker limit.

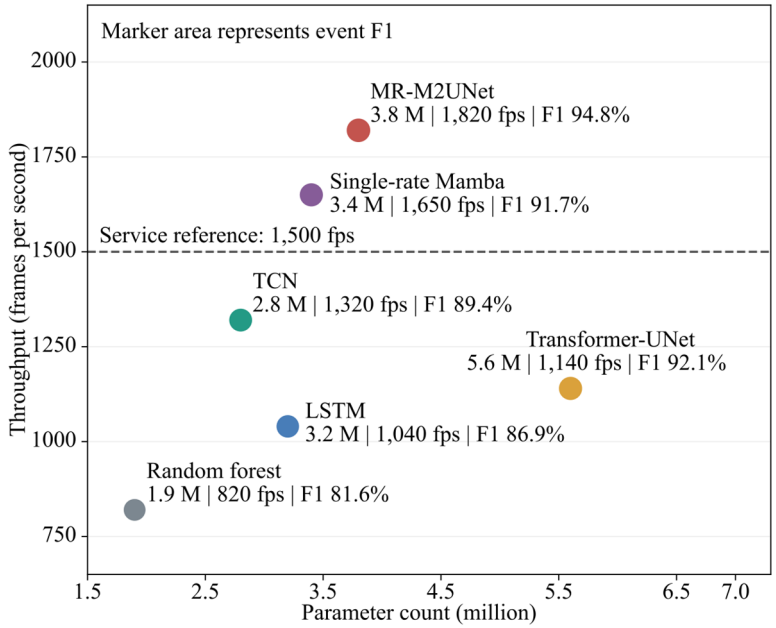


Figure 6. Computational trade-offs: throughput versus parameter count, with marker area representing event F1 and a 1,500-frame-per-second service-capacity reference.

The memory profile adds a separate constraint on concurrent processing. At batch sizes of 1, 4, 8, 16, and 32, the proposed model uses 0.31, 0.52, 0.83, 1.42, and 2.66 GB. Transformer-UNet uses 0.46, 0.81, 1.32, 2.18, and 4.05 GB, while single-rate Mamba uses 0.28, 0.49, 0.78, 1.31, and 2.44 GB. At batch size 16, the proposed model requires 34.9% less memory than Transformer-UNet and 8.4% more than single-rate Mamba. The differences describe the cost of retaining rate-specific representations and decoder features relative to an attention-based synchronized sequence. Gate inspection can make the relation between sensor evidence and an alert traceable during post-event examination [27]. In the inspected event windows, routine gate weights for high-, medium-, and low-rate groups are 0.52, 0.31, and 0.17. Near collision shifts them to 0.41, 0.16, and 0.43, while fatigue gives 0.31, 0.48, and 0.21. These normalized weights describe event-conditioned allocation; their interpretation concerns the allocation of model weights. The gate values sum to unity within each evaluated condition, permitting direct comparison of contributions across rate groups. Their normalization also couples the groups: an increase in one weight necessarily changes the available share for others. Evidence attribution would therefore benefit from complementary channel-removal tests and temporal inspection of the associated state features.

Probability calibration supports the interpretation of scores used in threshold-based alarm routing [28]. The threshold scenario spans 0.35 to 0.75 in 0.05 increments and yields F1 values of 92.7%, 93.8%, 94.5%, 94.8%, 94.6%, 94.0%, 93.2%, 92.1%, and 90.7%. The displayed coarse sweep peaks at 0.50, whereas the principal comparison uses the validation-selected operating threshold of 0.52. These values refer to different selection resolutions. At 0.35, false alarms rise to 0.73 per hour, compared with 0.42 at the selected setting. At 0.75, false alarms decrease to 0.25 per hour and intrusion recall falls by 5.2 percentage points. An alarm policy can evaluate these competing changes against the hazard distribution and the cost of interruption. A setting for a material-handling area may assign more weight to near-collision recall than one for a controlled assembly station. Probability quality and class-specific sensitivity should consequently be examined alongside the aggregate threshold curve before a site-level policy is specified. The coarse threshold spacing is suitable for displaying the broad operating response, while validation selection can use a finer grid. Distinguishing the coarse sweep from validation selection preserves the interpretation of the displayed values and the selected operating threshold. Site-specific evaluation should record both the adopted threshold and the event mixture against which its recall and nuisance-alarm burden were assessed.

Inspection of 126 misdetected segments identifies the contextual distinctions that remain difficult for wearable-only inference. Fatigue windows account for 34.1% of the false-negative subset, near-collision windows with incomplete proximity observations for 27.0%, and intrusion windows with handling-like wrist motion for 19.8%. The denominator for these percentages is the false-negative subset of the inspected segments. Routine recovery after lifting contributes 31.7% of false alarms, and tool repositioning contributes 24.6%. Slowly changing trunk orientation makes the annotated fatigue boundary sensitive to the duration used to define unsafe posture. Sparse proximity updates create ambiguity when relative movement changes rapidly between packets. Authorized and unauthorized handling can produce similar kinematics while differing in zone permission. Each error group therefore points to a specific missing distinction: sustained posture criteria, packet freshness, or task authorization. Task schedules and machine-state signals could help distinguish these cases. Their utility requires evaluation using synchronized records and explicit access permissions. A subsequent error audit should retain the same event-matching convention and report class counts so that changes in annotation or event prevalence can be separated from changes in model behavior. An audit record should preserve sensor timestamps, packet availability, label boundaries, and the decision threshold for each inspected segment. These fields make it possible to reconstruct whether the failure arose before fusion, during event reconstruction, or at the alarm rule. Separating those stages supports targeted correction when a particular class remains difficult after changes in deployment conditions.

6. Conclusion

This paper presented a Multi-Rate Mamba2-UNet model for factory wearable safety event detection. The framework combines native-rate encoding, content-aware alignment, and temporal reconstruction to represent abrupt motion hazards, sustained posture changes, and sparse proximity evidence. The study developed the signal formulation, architecture, training objective, and evaluation needed to examine recognition, boundary localization, and execution requirements. Its contribution concerns the integration of these functions under the stated evaluation conditions. Broader operational effectiveness remains a subject for deployment validation.

The evaluation scope is limited by the range of plants, equipment configurations, and communication conditions represented in the study. Fatigue posture creates annotation uncertainty because its onset can develop gradually. The model assumes known wearable identity, sensor type, and approximate placement during inference. Mounting errors and equipment changes may require adaptation or sensor-quality assessment. Reproducibility also depends on verified acquisition counts, scoring conventions, and implementation settings. Broader evaluation should examine these sources of variation over sustained operating periods.

Future work will investigate online personalization, uncertainty-aware alarm routing, and embedded implementation on low-power gateways. Larger field studies could examine worker and supervisor responses to calibrated outputs over consecutive working weeks. Further development could incorporate self-supervised pretraining on unlabeled shift data, continual learning for new task patterns, and privacy-preserving federation across factories. These directions would support adaptation and evaluation under a wider range of industrial conditions.

7. Declarations

7.1 Author Contributions

Indrek Jõhvi contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Daan Eelmaa contributes to validation, analysis, data collection, draft preparation. Gunnar Hõbe contributes to conceptualization, validation, analysis. All authors have read and agreed with the manuscript before its submission and publication.

7.2 Funding and Acknowledgments

This research received no specific financial support from any funding agency.

7.3 Institutional Review Board Statement

Not applicable.

7.4 Informed Consent Statement

Not applicable.

7.5 Date Availability

Not applicable.

7.6 Declaration of Competing Interest

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

7.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

8. References

[1] Scholz, C., Cao, H.-L., Imrith, E., Roshandel, N., Firouzipouyaei, H., Burkiewicz, A., Amighi, M., Menet, S., Sisavath, D. W., Paolillo, A., Rottenberg, X., Gerets, P., Cheyns, D., Dahlem, M., Ocket, I., Genoe, J., Philips, K., Stoffelen, B., Van den Bosch, J., … Vanderborght, B. (2025). Sensor-enabled safety systems for human–robot collaboration: A review. IEEE Sensors Journal, 25(1), 65–88. https://doi.org/10.1109/JSEN.2024.3496905

[2] Taylor, W., Hill, D., Adam, R., Cooper, J., Abbasi, Q. H., & Imran, M. A. (2024). Hybrid sensing for fatigue detection using wearables and RF. IEEE Sensors Journal, 24(14), 22764–22772. https://doi.org/10.1109/JSEN.2024.3404637

[3] Cunico, F., Aldegheri, S., Avogaro, A., Boldo, M., Bombieri, N., Capogrosso, L., Caputo, A., Carra, D., Centomo, S., Cheng, D. S., Cinquetti, E., Cristani, M., De Marchi, M., Demrozi, F., Emporio, M., Fummi, F., Geretti, L., Germiniani, S., Giachetti, A., … Quaglia, D. (2024). Enhancing safety and privacy in Industry 4.0: The ICE laboratory case study. IEEE Access, 12, 154570–154599. https://doi.org/10.1109/ACCESS.2024.3479411

[4] Chen, S., Zhu, C., Chen, X., & Yi, J. (2025). Machine learning-based real-time walking activity and posture estimation in construction with a single wearable inertial measurement unit. IEEE Transactions on Automation Science and Engineering, 22, 16144–16156. https://doi.org/10.1109/TASE.2025.3573966

[5] Alruwaili, O., Yousef, A., & Armghan, A. (2024). Monitoring the transmission of data from wearable sensors using probabilistic transfer learning. IEEE Access, 12, 97460–97475. https://doi.org/10.1109/ACCESS.2024.3428444

[6] Jiang, Z., Al-Qaness, M. A. A., Al-Alimi, D., Ewees, A. A., Abd Elaziz, M., Dahou, A., & Helmi, A. M. (2024). Fall detection

systems for Internet of Medical Things based on wearable sensors: A review. IEEE Internet of Things Journal, 11(21), 34797–34810. https://doi.org/10.1109/JIOT.2024.3421336

[7] Ghorrati, Z., Esmaeili, A., & Matson, E. T. (2024). A multi-section hierarchical deep neural network model for time series classification: Applied to wearable sensor-based human activity recognition. IEEE Access, 12, 137851–137869. https://doi.org/10.1109/ACCESS.2024.3462515

[8] Yussif, S. B., Xie, N., Yang, Y., Huang, Y., Wang, G., & Du, Z. (2024). EGTCN: An efficient graph and temporal convolution network for sensor-based human activity recognition in federated learning. IEEE Sensors Journal, 24(21), 34892–34906. https://doi.org/10.1109/JSEN.2024.3446290

[9] Zhou, X., Yuan, J., Fan, L., Niu, X., Zha, K., & Liu, X. (2025). MSMFT: Multi-stream multimodal factorized transformer for human activity recognition. IEEE Sensors Journal, 25(6), 10402–10416. https://doi.org/10.1109/JSEN.2025.3529889

[10] Li, S., Zhu, T., Duan, F., Chen, L., Ning, H., Nugent, C., & Wan, Y. (2025). HARMamba: Efficient and lightweight wearable sensor human activity recognition based on bidirectional Mamba. IEEE Internet of Things Journal, 12(3), 2373–2384. https://doi.org/10.1109/JIOT.2024.3463405

[11] Taskasaplidis, G., Fotiadis, D. A., & Bamidis, P. D. (2024). Review of stress detection methods using wearable sensors. IEEE Access, 12, 38219–38246. https://doi.org/10.1109/ACCESS.2024.3373010

[12] Kanjilal, R., Kucuk, M. F., & Uysal, I. (2025). Human activity recognition: A review of RFID and wearable sensor technologies powered by AI. IEEE Journal of Radio Frequency Identification, 9, 180–199. https://doi.org/10.1109/JRFID.2025.3561345

[13] Luo, F., Li, A., Jiang, B., Khan, S., Wu, K., & Wang, L. (2025). ActivityMamba: A CNN-Mamba hybrid neural network for efficient human activity recognition. IEEE Transactions on Mobile Computing, 24(8), 6797–6811. https://doi.org/10.1109/TMC.2025.3544573

[14] Wieland, C., & Pankratius, V. (2025). Enhancing sensor-based human activity recognition with multiresolution wavelet-attention. IEEE Sensors Journal, 25(14), 26920–26930. https://doi.org/10.1109/JSEN.2025.3571572

[15] Hussein, D., Yamin, N., & Bhat, G. (2024). EMI: Energy management meets imputation in wearable IoT devices. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 43(11), 3792–3803. https://doi.org/10.1109/TCAD.2024.3448379

[16] Aidarova, S., Nurmakhan, T., Myrzakhan, R., Fazli, S., & Yazici, A. (2025). Advancing activity recognition with multimodal fusion and transformer techniques. IEEE Sensors Journal, 25(11), 19632–19649. https://doi.org/10.1109/JSEN.2025.3555663

[17] Liu, J., Huang, Y., Shi, X., Ren, X., Mi, T., & Qiu, R. C. (2025). TF-Mamba: A lightweight state-space model for Wi-Fi-based human activity recognition. IEEE Sensors Journal, 25(13), 23184–23194. https://doi.org/10.1109/JSEN.2024.3520857

[18] Patsch, C., Wu, Y., Salihu, D., Zakour, M., & Steinbach, E. (2024). TSCL: Timestamp supervised contrastive learning for action segmentation. IEEE Robotics and Automation Letters, 9(9), 7485–7492. https://doi.org/10.1109/LRA.2024.3427553

[19] Ashry, S., Das, S., Rafiei, M., Baumbach, J., & Baumbach, L. (2025). Transfer learning of human activities based on IMU sensors: A review. IEEE Sensors Journal, 25(3), 4115–4126. https://doi.org/10.1109/JSEN.2024.3510097

[20] Chen, X., Zhou, X., Sun, M., & Wang, H. (2025). Temporal contrastive learning for sensor-based human activity recognition: A self-supervised approach. IEEE Sensors Journal, 25(1), 1839–1850. https://doi.org/10.1109/JSEN.2024.3491933

[21] Obiri, N. M., & Van Laerhoven, K. (2024). A survey of LoRaWAN-integrated wearable sensor networks for human activity recognition: Applications, challenges and possible solutions. IEEE Open Journal of the Communications Society, 5, 6713–6735. https://doi.org/10.1109/OJCOMS.2024.3484002

[22] Dong, J., Jiang, Z., Pan, D., Chen, Z., Guan, Q., Zhang, H., Gui, G., & Gui, W. (2025). A survey on confidence calibration of deep learning-based classification models under class imbalance data. IEEE Transactions on Neural Networks and Learning Systems, 36(9), 15664–15684. https://doi.org/10.1109/TNNLS.2025.3565159

[23] Shi, J., Zhu, Y., He, J., Xu, Z., Chu, L., Braun, R., & Shi, Q. (2025). Human activity recognition based on feature fusion of millimeter wave radar and inertial navigation. IEEE Journal of Microwaves, 5(2), 409–424. https://doi.org/10.1109/JMW.2025.3539957

[24] Ankalaki, S. (2024). Simple to complex, single to concurrent sensor-based human activity recognition: Perception and open challenges. IEEE Access, 12, 93450–93486. https://doi.org/10.1109/ACCESS.2024.3422831

[25] Wang, H., Wang, X., Lu, C., Yuan, M., Wang, Y., Yu, H., & Li, H. (2024). Enhancing human activity recognition in wrist-worn sensor data through compensation strategies for sensor displacement. IEEE Access, 12, 95058–95070. https://doi.org/10.1109/ACCESS.2024.3422256

[26] Yu, J., Zhang, L., Cheng, D., Huang, W., Wu, H., & Song, A. (2024). Improving human activity recognition with wearable

sensors through BEE: Leveraging early exit and gradient boosting. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 32, 3452–3464. https://doi.org/10.1109/TNSRE.2024.3457830

[27] El Marhraoui, Y., Dahaoui, I., Bouchard, B., & Gaboury, S. (2026). FoCLAP-Net: Toward an interpretable model for body posture recognition via multimodal temporal data from wearable sensors. IEEE Internet of Things Journal, 13(8), 17209–17222. https://doi.org/10.1109/JIOT.2026.3661422

[28] Dong, J., Jiang, Z., Pan, D., & Yu, H. (2025). Combining priors with experience: Confidence calibration based on binomial process modeling. Proceedings of the AAAI Conference on Artificial Intelligence, 39(15), 16317–16326. https://doi.org/10.1609/aaai.v39i15.33792