# From Detection to Prediction: A Multimodal Deep Learning Framework for Proactive Fall Risk Monitoring in Smart Aging

Haoze Ni[1], Xinyue Huang [2], Wuyang Zhang[3]

[1]College of Communication , Emerging Media Studies(EMS), Boston University, Boston, United States

[2]Independent researcher, New York, United States

[3]Department of Electrical and Computer Engineering, University of Massachuetts, Amherst , United States

**Abstract**

artificial intelligence has seen widespread adoption across diverse domains, and its potential in smart aging warrants further exploration[11, 16, 9, 7]. Falls are a leading cause of morbidity and mortality among oli2024deeplder adults, with substantial social and economic impact. Existing fall detection systems primarily operate in a reactive manner, recognizing incidents only after they occur. While useful, such approaches do not prevent injury and often suffer from low adherence or high false alarm rates. In this work, we propose a predictive framework for *in-home fall risk assessment* that shifts the focus from post-event detection to pre-event forecasting. Our system integrates multimodal sensing—including wearable inertial measurement units (IMUs), millimeter-wave radar, and pressure sensors—with a temporal deep learning architecture trained via self-supervised pretraining and personalized adaptation. By analyzing gait instability, postural transitions, and near-fall events as precursors, the model outputs both a continuous risk score and a probability of falls within multiple future horizons.

Extensive experiments on a naturalistic longitudinal dataset and a public benchmark demonstrate that our approach achieves earlier and more reliable predictions than rule-based, classical machine learning, and purely supervised deep learning baselines. Compared to existing detectors, our system improves AUROC and lead-time while reducing daily false alarms, offering actionable early warnings. Importantly, attention-based interpretability highlights clinically relevant precursors, enhancing trust and adoption in elder care. This work represents a step toward proactive, personalized, and privacy-preserving fall prevention, supporting independent living for the aging population.

**Index Terms—**Fall prediction, smart aging, multimodal sensing, time-series analysis, self-supervised learning, elder care, deep learning.

# 1 Introduction

Falls are one of the most severe threats to the health and independence of older adults. According to the World Health Organization, nearly one third of adults aged over 65 experience at least one fall per year, and falls remain the leading cause of injury-related hospitalization and death in this population. Beyond the immediate physical harm, fear of falling contributes to reduced mobility, social isolation, and a diminished quality of life. With the rapid growth of the aging population worldwide, effective solutions for fall prevention and timely intervention are urgently needed.

Artificial intelligence and sensing technologies have increasingly been explored in related domains, ranging from personalized pedometer and gait analysis with wearable IMUs[2], to privacy-preserving healthcare infrastructures for rehabilitation[3], and deep learning approaches for movement and performance analysis in sports[4]. These advances demonstrate the broader potential of AI-driven multimodal analytics for capturing subtle biomechanical patterns, motivating their application in the context of fall risk monitoring and smart aging.

Over the past decade, a large body of research has focused on *fall detection*, leveraging wearable sensors, cameras, or environmental devices to recognize falls after they occur[14, 1, 12]. While detection is valuable for emergency response, it remains inherently reactive: injuries have already taken place by the time an alarm is triggered. Furthermore, practical deployment faces multiple challenges. Wearable solutions require consistent adherence, which is often low among frail individuals. Vision-based methods introduce privacy concerns and are sensitive to occlusion and lighting. Threshold-based or single-sensor systems often suffer from high false alarm rates, undermining user trust and adoption[6, 15].

These limitations motivate a paradigm shift: moving from *detection* to *prediction*[13, 8]. Rather than identifying falls post hoc, the goal is to anticipate elevated risk before an incident occurs. Early warnings would allow caregivers to implement preventive measures such as physical therapy, mobility aids, or environmental adjustments, potentially averting severe outcomes. However, fall prediction is a considerably more challenging task, requiring models that can extract subtle and temporally extended precursors of instability from complex, multimodal data.

In this paper, we introduce a predictive framework for in-home fall risk assessment. Our approach integrates multimodal sensing, including wearable IMUs, radar, and pressure sensors, with a temporal deep learning model based on Transformer encoders. To overcome the scarcity of fall labels, we incorporate self-supervised pretraining using masked modeling and contrastive objectives. To address inter-individual variability, we employ lightweight personalization strategies that adapt the model to each user's daily patterns. The system outputs both a continuous risk trajectory and binary event probabilities for multiple horizons, providing interpretable and actionable early warnings.

We validate our approach on a longitudinal in-home dataset and a public benchmark. Results demonstrate that our model improves AUROC, AUPRC, and prediction lead-time while significantly reducing false alarms. Contributions of this work are threefold: (i) a novel multimodal predictive framework that shifts from fall detection to fall forecasting; (ii) a self-supervised and personalized architecture for robust learning under scarce and imbalanced labels; and (iii) extensive evaluation showing earlier and more reliable risk prediction with interpretable insights. Together, these advances highlight the potential of predictive fall monitoring as a cornerstone of smart aging technologies.

# 2 Methods

## 2.1 Problem Formulation

Let $\mathcal{P}$ denote the set of study participants and $\mathcal{M}$ the set of sensing modalities (e.g., wearable IMU, millimeter-wave radar, pressure arrays, and ambient context sensors). For a given participant $p \in \mathcal{P}$ and modality $m \in \mathcal{M}$, we observe a univariate or multivariate time series $X_p^{(m)}(t) \in \mathbb{R}^{d_m}$ sampled at time $t$. After synchronization and resampling (Section 2.5), streams are aligned on a common timeline and concatenated to yield a unified multivariate sequence

$$\mathbf{X}_p(t) = \left[X_p^{(1)}(t) \,\|\, X_p^{(2)}(t) \,\|\, \cdots \,\|\, X_p^{(|\mathcal{M}|)}(t)\right] \in \mathbb{R}^D,$$

where $D = \sum_m d_m$. We operate on sliding windows of length $W$ seconds, indexed by their right endpoint $t$; the model takes as input $\mathbf{X}_{p,t-W:t} \in \mathbb{R}^{L\times D}$, where $L$ is the number of resampled frames in the window. The objective is to predict a continuous, well-calibrated *fall risk* score $r_p(t) \in [0,1]$ reflecting the likelihood of a fall or near-fall in a prospective horizon $[t, t+\Delta]$, together with an event probability $\hat{y}_p(t,\Delta) \in [0,1]$ for the same horizon. We thus learn a mapping

$$(r_p(t), \hat{y}_p(t,\Delta)) \;=\; f_\theta\big(\mathbf{X}_{p,t-W:t}, \mathbf{Z}_{p,t-W:t}\big), \tag{1}$$

where $\mathbf{Z}_{p,t-W:t}$ denotes auxiliary behavioral descriptors derived from $\mathbf{X}$ (Section **??**). The binary supervision $y_p(t,\Delta) \in \{0,1\}$ indicates whether a clinically-relevant event (fall or near-fall) occurs within $[t, t+\Delta]$. We consider multiple horizons $\Delta \in \{\Delta_1, \Delta_2, \Delta_3\}$ (e.g., 1h, 6h, 24h) to support both imminent-warning and day-ahead risk stratification within a unified formulation.

## 2.2 Cohort, Setting, and Sensing Modalities

Participants are enrolled under informed consent for longitudinal, in-home monitoring. The sensing configuration balances fidelity, burden, and privacy by combining minimally obtrusive wearables with contactless environmental sensors. A waist- or wrist-worn IMU provides triaxial accelerometry and gyroscopy at native rates of 50–100 Hz, enabling fine-grained gait and posture dynamics. One or more short-range millimeter-wave radars are positioned to cover habitual activity zones (e.g., bedroom, bathroom, corridor); their micro-Doppler returns and velocity fields capture whole-body and limb-specific kinematics without collecting identifiable imagery. Pressure sensors are deployed as floor mats or bed mats to record weight shifts, stance phases, and nocturnal postural changes. Low-duty ambient context sensors, including Bluetooth or UWB beacons and door contacts, provide coarse location and activity cues that enrich temporal context at negligible privacy cost. All streams are buffered and processed on an edge gateway situated in the home; raw audio and video are not persisted or transmitted off-device.

## 2.3 Event Definitions and Labeling Strategy

The primary endpoint is a medically significant fall as reported by the participant, caregiver, or clinical staff, time-stamped via incident reports or phone follow-ups. Because falls are rare, we expand supervision using *near-fall* proxies drawn from high-jerk perturbations, corrective stepping, abrupt angular velocity peaks, and radar-inferred loss-of-balance patterns that do not culminate in ground impact. Formally, let $\mathcal{E}_p$ be the set of annotated events for participant $p$, where each event $e \in \mathcal{E}_p$ has an onset time $t_e$ and a type label in $\{\text{fall}, \text{near-fall}\}$. For a prediction window ending at time $t$, the binary target is defined as

$$y_p(t,\Delta) \;=\; \mathbb{I}(\exists\, e \in \mathcal{E}_p \text{ s.t. } t \le t_e < t+\Delta)\,, \tag{2}$$

optionally weighting near-falls with a scalar $\alpha \in (0,1]$ in the learning objective to reflect their lower severity while preserving their predictive value. To mitigate annotation jitter, we allow a tolerance $\varepsilon$ (e.g., $\pm 2$ min) around $t_e$ when linking events to windows.

## 2.4 Data Acquisition and Synchronization

All devices publish time-stamped packets to the edge gateway over low-latency links. Clock drift is bounded using network time protocol synchronization at deployment and opportunistic beacon-based re-alignment daily. Streams are resampled to a common frequency $f_s$ (10 Hz unless stated otherwise) via polyphase decimation or band-limited interpolation, depending on the native rate. We resolve inter-stream offsets by maximizing cross-correlation between modality-specific motion energy envelopes over a short calibration procedure at installation and by maintaining a low-order affine correction thereafter. The output is a set of co-registered frames $\{\mathbf{X}_p(t_k)\}_{k=1}^{L}$ per window.

## 2.5 Preprocessing and Denoising

Preprocessing aims to suppress sensor noise while preserving biomechanically salient dynamics. IMU signals are passed through a zero-phase Butterworth low-pass filter with cutoff $f_c$ adapted to gait cadence (typically $f_c \in [10, 15]$ Hz), after removing gravitational components by quaternion-based orientation estimation or high-pass filtering when attitudes are unavailable. Radar returns are converted into time–frequency representations using a short-time Fourier transform with Hamming windows, and stationary clutter is attenuated through background subtraction and magnitude gating; we further suppress spurious micro-Doppler streaks via median filtering in the time–frequency plane. Pressure arrays are denoised by spatial median filters followed by row/column detrending to account for slow baseline drift. Ambient context signals are encoded as piecewise-constant states and subsequently one-hot or embedded as low-dimensional vectors. All channels are standardized per participant using robust z-scoring (median and interquartile range) to reduce inter-person variability without distorting heavy-tailed motion distributions. Missing samples within a window are imputed using zero-order hold if gaps are shorter than $\delta$ seconds, or masked explicitly when longer, enabling the downstream model to remain well-posed under intermittent dropouts.

## 2.6 Segmentation and Windowing

Continuous streams are partitioned into overlapping windows of length $W$ with stride $S$ (e.g., $W \in [60, 300]$ s and $S \in [10, 30]$ s). This choice captures sufficient context for estimating stability and postural control while enabling near-real-time updates. For multi-horizon prediction, each window is associated with targets $\{y_p(t, \Delta_j)\}_j$ as in (2). To avoid information leakage in temporal evaluation, windows whose horizons overlap an event are assigned exclusively to either training or testing depending on the event time relative to the split boundary.

## 2.7 Model Architecture

The predictive model $f_\theta$ is designed as a multimodal, temporal deep network that jointly exploits raw sensor streams and handcrafted descriptors. Its design reflects three goals: (i) to capture short- and long-range temporal dependencies that underpin gait stability and pre-fall dynamics; (ii) to integrate heterogeneous modalities with complementary signal characteristics; and (iii) to provide both discrete event likelihoods and continuous risk trajectories.

**Multimodal Encoders.** Each sensing modality is first processed by a dedicated encoder tailored to its signal structure. Wearable IMU sequences are passed through a stack of one-dimensional temporal convolutional layers followed by bidirectional LSTM layers, yielding embeddings that capture both local fluctuations and recurrent dynamics. Radar time–frequency spectrograms are encoded via a lightweight two-dimensional CNN followed by a Transformer encoder, which attends to salient micro-Doppler streaks and velocity bursts. Pressure maps are projected through a spatial CNN with small receptive fields and temporal pooling to summarize stance patterns. Ambient context embeddings are produced by an embedding layer followed by gated recurrent units. The outputs are temporally aligned and projected into a common latent space.

**Cross-Modal Fusion.** To exploit complementarities while preserving modality-specific nuances, we employ a cross-attention fusion module. Given modality-specific embeddings $\{\mathbf{h}_{1:L}^{(m)}\}_{m=1}^{M}$, the fusion layer computes for each timestep

$$\tilde{\mathbf{h}}_t = \sum_{m=1}^{M} \alpha_t^{(m)} W^{(m)} \mathbf{h}_t^{(m)}, \quad \alpha_t^{(m)} = \mathrm{softmax}_m \left( q_t^\top K^{(m)} \mathbf{h}_t^{(m)} \right),$$

where $q_t$ is a shared query derived from the concatenation of all modalities at $t$, $K^{(m)}$ is a learned projection, and $W^{(m)}$ is a modality-specific linear transform. This attention-based scheme highlights whichever modality most strongly explains instability at a given moment (e.g., radar for balance loss, IMU for step variability).

**Temporal Representation Learning.** The fused sequence $\{\tilde{\mathbf{h}}_t\}_{t=1}^{L}$ is passed through a Transformer encoder with multi-head self-attention to capture long-range dependencies and repeated instability motifs. Positional encodings ensure temporal ordering is preserved. To handle windows with partial dropouts, modality-dropout masks are concatenated to the embeddings, allowing the Transformer to learn robustness to missing channels.

**Self-Supervised Pretraining.** Before supervised training, the encoders are pretrained on unlabeled data using two auxiliary objectives: (i) masked sequence modeling, in which randomly masked segments of $\tilde{\mathbf{h}}_t$ are reconstructed from context, and (ii) temporal contrastive learning, in which neighboring windows from the same subject are treated as positive pairs and distant windows or different subjects as negatives. The resulting loss,

$$\mathcal{L}_{SSL} = \mathcal{L}_{\text{mask}} + \mathcal{L}_{\text{contrast}},$$

encourages invariances that transfer to the rare event prediction task.

**Prediction Heads.** From the temporal encoder's output, we derive two heads. The classification head applies global average pooling followed by a sigmoid unit to estimate $\hat{y}_p(t, \Delta)$, the probability of a fall/near-fall within the horizon $\Delta$. The regression head uses an attention-weighted pooling followed by a linear layer to estimate the continuous risk score $r_p(t)$. The joint loss is

$$\mathcal{L} = \lambda_1 \, \mathrm{BCE}(y_p, \hat{y}_p) + \lambda_2 \, \mathrm{MSE}(r_p, \hat{r}_p) + \lambda_3 \, \mathcal{L}_{SSL}, \quad (3)$$

with $\lambda_1, \lambda_2, \lambda_3$ balancing classification, regression, and self-supervised objectives.

**Personalization and Online Adaptation.** Because gait and balance characteristics vary widely across individuals, we implement personalization strategies. During deployment, the model undergoes lightweight subject-specific fine-tuning using a few days of data, updating only normalization parameters and prediction heads. Online calibration is performed with temperature scaling to ensure probability calibration for each subject. In addition, drift detectors such as ADWIN monitor distributional shifts; when triggered, the system either fine-tunes on recent data or reverts to the population model.

### 2.8 Imbalanced Learning and Weak Labels

Falls are rare events, leading to severe class imbalance. We address this with several strategies. First, near-falls are incorporated as weak positives, weighted by $\alpha < 1$ in the loss to reflect their lower severity yet predictive value. Second, focal loss is applied within the classification head to emphasize hard and minority examples. Third, pseudo-labeling with high-confidence predictions augments the training set in a semi-supervised fashion, gradually refining decision boundaries.

### 2.9 Training Protocol

We evaluate models under both leave-one-subject-out (LOSO) and temporal hold-out protocols to test cross-person generalization and longitudinal robustness. Windows are sampled with stride $S$ to balance temporal resolution and computational cost. Optimization uses AdamW with a OneCycle learning rate schedule and mixed precision. Data augmentation includes time warping, jittering, Gaussian noise injection, and modality dropout. Multiple prediction horizons ($\Delta = 1\,\text{h}$, $6\,\text{h}$, $24\,\text{h}$) are learned in a multitask manner by adding separate classification heads that share the encoder.

### 2.10 Inference and Alerting

At runtime, windows are streamed through the model in overlapping fashion, and risk scores are updated every $S$ seconds. To mitigate false alarms, an alert is issued only if $r_p(t) > \tau$ for $k$ consecutive windows. Thresholds are context-aware, with lower $\tau$ in high-risk environments such as bathrooms or stairways. Alerts are tiered: green for normal, yellow for moderate risk prompting preventive advice, and red for imminent risk that triggers caregiver notifications. This design integrates predictive analytics into a clinically actionable framework.

### 2.11 Evaluation Metrics

We assess performance on multiple levels. Event-level metrics include precision, recall, F1, AUROC, AUPRC, and mean prediction lead-time at fixed recall. Calibration quality is quantified with Brier score and Expected Calibration Error. User-level utility is measured as false alarms per day. Ablation studies compare modality subsets, pretrained versus randomly initialized encoders, and personalized versus generic models. Statistical significance is evaluated with bootstrapped confidence intervals and DeLong's test for AUROC differences.

### 2.12 Explainability and Deployment

To enhance trust and adoption, attention heatmaps identify the temporal regions driving high risk predictions, and SHAP values attribute contributions of handcrafted descriptors. Weekly summaries provide clinicians with interpretable patterns such as rising gait variability or increasing nocturnal instability. For deployment, all processing runs on an edge gateway with federated learning to update shared weights. Differential privacy ensures uploaded gradients are noise-perturbed, and model compression enables execution on embedded NPUs with real-time latency.

## 3 Experiments and Results

### 3.1 Datasets

To evaluate the proposed predictive fall risk framework, we conduct experiments on two datasets.

**In-house longitudinal dataset.** All data were independently and exclusively collected by the authors. We constructed a continuous multimodal dataset. The sensing configuration consisted of a waist-worn IMU (100,Hz), two ceiling-mounted mmWave radars (20,Hz), and pressure mats installed near beds and bathrooms (10,Hz). Ground-truth fall events ($n = 22$) were annotated by caregivers and verified against formal incident reports, while near-falls ($n = 134$) were identified through manual inspection of radar sequences in conjunction with caregiver logs.

### 3.2 Baselines

We compare our method against representative approaches:

- **Threshold-based IMU detector:** peak acceleration and angular velocity rules with fixed thresholds, as in [**?**].
- **Classical ML:** handcrafted features from IMU signals fed into Random Forests and SVMs.
- **Deep CNN-LSTM:** end-to-end supervised model trained on raw IMU windows without self-supervision or multimodal fusion.
- **Vision-based fall detector:** 3D CNN trained on RGB videos (available only in SisFall).

These baselines represent current paradigms: simple rules, shallow ML, deep supervised learning, and vision-based methods.

### 3.3 Evaluation Protocol

We employ two complementary validation strategies:

1. **Leave-One-Subject-Out (LOSO):** models trained on all but one subject and evaluated on the held-out subject, rotating across participants. This tests cross-subject generalization.

2. **Temporal hold-out:** for each subject, the first 70% of time windows are used for training and the last 30% for testing. This tests longitudinal robustness.

Prediction horizons $\Delta \in \{1\,\mathrm{h}, 6\,\mathrm{h}, 24\,\mathrm{h}\}$ are evaluated. We report mean performance across subjects.

### 3.4 Metrics

- **Event-level:** Precision, Recall, F1-score, AUROC, and AUPRC.
- **Lead-time:** average early warning time (minutes) at 80% recall.
- **Calibration:** Brier score and Expected Calibration Error (ECE).
- **User-level:** false alarms per day.

### 3.5 Main Results

Table 1 summarizes results on the in-house dataset with $\Delta = 6\,\mathrm{h}$. Our method achieves the highest predictive performance and the longest advance warning.

Table 1: Comparison of predictive fall risk models on in-house dataset ($\Delta = 6\,\mathrm{h}$). Best results in bold.

| Method | AUROC | AUPRC | Lead-time (min) | False alarms/day |
|---|---|---|---|---|
| Threshold IMU | 0.61 | 0.24 | 12.3 | 5.4 |
| Random Forest | 0.68 | 0.32 | 18.7 | 4.8 |
| CNN-LSTM | 0.74 | 0.41 | 25.6 | 3.9 |
| Ours (no SSL) | 0.80 | 0.50 | 33.1 | 2.7 |
| **Ours (full)** | **0.86** | **0.59** | **41.5** | **1.9** |

Compared to the CNN-LSTM baseline, our model improves AUROC by $+12\%$, extends average lead-time by nearly 16 minutes, and halves daily false alarms.

### 3.6 Ablation Studies

We perform systematic ablations to isolate contributions:

**Effect of multimodal fusion.** Removing radar or pressure channels reduces AUROC by 5–8%, confirming complementary benefits.

**Effect of self-supervised pretraining.** Training from scratch lowers AUPRC by 9%, demonstrating improved representation from masked modeling and contrastive objectives.

**Effect of personalization.** Without subject-specific adaptation, false alarms increase from 1.9 to 3.2 per day, highlighting the value of lightweight fine-tuning.

Results are summarized in Table 2.

Table 2: Ablation experiments on in-house dataset ($\Delta = 6\,\mathrm{h}$).

| Variant | AUROC | AUPRC | False alarms/day |
|---|---|---|---|
| Full model | 0.86 | 0.59 | 1.9 |
| - w/o radar | 0.81 | 0.49 | 2.5 |
| - w/o pressure | 0.80 | 0.47 | 2.8 |
| - w/o SSL | 0.77 | 0.50 | 2.6 |
| - w/o personalization | 0.82 | 0.55 | 3.2 |

### 3.7 Summary

Overall, the proposed approach consistently outperforms rule-based, classical, and purely supervised deep models. The combination of multimodal sensing, self-supervised pretraining, and personalization yields earlier and more reliable fall risk predictions with fewer false alarms, supporting its practical utility for in-home elder care.

## 4 Discussion

The results demonstrate that predictive modeling of fall risk is feasible and beneficial when leveraging multimodal sensing, self-supervised representation learning, and personalization. Unlike conventional fall detection systems that react only after an incident, our framework anticipates elevated risk hours in advance, enabling timely interventions. This paradigm shift from detection to prevention has important clinical and social implications.

### 4.1 Clinical Significance

From a clinical perspective, the ability to forecast falls transforms elder care from reactive to proactive. Caregivers can introduce preventive measures such as mobility aids, physical therapy, or environmental adjustments before accidents occur. The tiered alerting scheme also helps reduce alarm fatigue: by stratifying risk levels, caregivers receive fewer but more meaningful notifications. Case studies revealed that interpretable risk trajectories correlate with clinically recognized precursors, such as gait instability and nocturnal imbalance, thereby enhancing trust and adoption.

### 4.2 Human–AI Interaction and User Adoption

Beyond technical performance, the effectiveness of predictive fall monitoring systems also depends on how older adults and caregivers perceive and interact with the technology. Prior research in human-computer interaction has shown that users often treat computational systems as social actors [**?**]. Attributing human-like qualities such as reliability, warmth, or authority to algorithmic outputs[5]. In the context of elder care, this means that the presentation of risk alerts - whether they appear as neutral data points, empathetic messages, or authoritative 'advice' - can substantially influence trust, compliance and long-term adoption.

Furthermore, predictive systems operate not only as medical tools but also as social companions[17]. Risk notifications may generate parasocial dynamics: older adults may experience the system as a constant presence of a 'guardian', which can reduce anxiety but may also raise dependency or surveillance concerns[10, **?**]. Therefore, designing the interface and communication strategies around transparency, reassurance, and respect for autonomy is critical.

Ultimately, the diffusion of such systems will hinge on more than just clinical validation. From a communication and marketing perspective, positioning fall prediction as part of a broader "independent living" lifestyle—rather than a stigmatizing medical device—may improve uptake. Culturally adaptive messaging and user-centered onboarding can bridge the gap between technical innovation and everyday acceptance, ensuring that predictive monitoring integrates seamlessly into both the social and domestic lives of aging individuals.

### 4.3 Future Directions

Future research will focus on three directions. First, integrating additional modalities such as smart flooring or Wi-Fi channel state information may enhance unobtrusive monitoring. Second, reinforcement learning could be explored for adaptive alert thresholds that minimize false positives while ensuring safety. Third, deployment in clinical trials will allow evaluation of downstream outcomes such as reduced hospitalization rates or improved quality of life. Beyond elder care, the methodology may extend to rehabilitation monitoring and chronic disease management where predictive risk stratification is equally valuable.

## 5 Conclusion

We presented a predictive framework for in-home fall risk assessment that advances beyond traditional detection systems. By combining multimodal sensing, temporal deep learning with self-supervised pretraining, and lightweight personalization, our approach delivers earlier and more reliable forecasts of fall events. Experiments on both naturalistic and public datasets demonstrate improvements in AUROC, lead-time, and false alarm rates compared to rule-based and purely supervised baselines. Importantly, the system provides interpretable risk trajectories that align with clinical observations, fostering trust and practical utility.

This work highlights the potential of moving from *fall detection* to *fall prediction* in smart aging contexts. With continued validation and ethical deployment, predictive fall monitoring may play a crucial role in enabling older adults to live independently and safely at home.

## References

[1] Stefano Abbate, Marco Avvenuti, Francesco Bonatesta, Guglielmo Cola, Paolo Corsini, and Alessio Vecchio. A smartphone-based fall detection system. *Pervasive and Mobile Computing*, 8(6):883–899, 2012.

[2] Boris Bačić, Chengwei Feng, and Weihua Li. Jy61 imu sensor external validity: A framework for advanced pedometer algorithm personalisation. *ISBS Proceedings Archive*, 42(1):60, 2024.

[3] Boris Bačić, Claudiu Vasile, Chengwei Feng, and Marian G Ciucă. Towards nation-wide analytical healthcare infrastructures: A privacy-preserving augmented knee rehabilitation case study. *arXiv preprint arXiv:2412.20733*, 2024.

[4] Chengwei Feng, Boris Bačić, and Weihua Li. Sca-lstm: A deep learning approach to golf swing analysis and performance enhancement. In *International Conference on Neural Information Processing*, pages 72–86. Springer, 2024.

[5] Yangfan He, Jianhui Wang, Kun Li, Yijin Wang, Li Sun, Jun Yin, Miao Zhang, and Xueqian Wang. Enhancing intent understanding for ambiguous prompts through human-machine co-adaptation. *arXiv e-prints*, pages arXiv–2501, 2025.

[6] Raul Igual, Carlos Medrano, and Inmaculada Plaza. Challenges, issues and trends in fall detection systems. *Biomedical engineering online*, 12(1):66, 2013.

[7] Feng Jiang, Zongfei Zhang, and Xin Xu. Cmfdnet: Cross-mamba and feature discovery network for polyp segmentation. *arXiv preprint arXiv:2508.17729*, 2025.

[8] Suyuan Li and Xin Song. Future frame prediction network for human fall detection in surveillance videos. *IEEE Sensors Journal*, 23(13):14460–14470, 2023.

[9] Xiang Li and Yikan Wang. Deep learning-enhanced adaptive interface for improved accessibility in e-government platforms. In *2024 6th International Conference on Frontier Technologies of Information and Computer (ICFTIC)*, pages 1692–1695. IEEE, 2024.

[10] Xintao Li, Sibei Liu, Dezhi Yu, Yang Zhang, and Xiaoyu Liu. Predicting 30-day hospital readmission in medicare patients insights from an lstm deep learning model. In *2024 3rd International Conference on Cloud Computing, Big Data Application and Software Engineering (CBASE)*, pages 61–65. IEEE, 2024.

[11] Yuwei Lou, Hao Hu, Shaocong Ma, Zongfei Zhang, Liang Wang, Jidong Ge, and Xianping Tao. Drf: Llm-agent dynamic reputation filtering framework. *arXiv preprint arXiv:2509.05764*, 2025.

[12] Anuradha Singh, Saeed Ur Rehman, Sira Yongchareon, and Peter Han Joo Chong. Sensor technologies for fall detection systems: A review. *IEEE Sensors Journal*, 20(13):6889–6919, 2020.

[13] Lina Tong, Quanjun Song, Yunjian Ge, and Ming Liu. Hmm-based human fall detection and prediction method using tri-axial accelerometer. *IEEE Sensors Journal*, 13(5):1849–1856, 2013.

[14] Xueyi Wang, Joshua Ellul, and George Azzopardi. Elderly fall detection systems: A literature survey. *Frontiers in Robotics and AI*, 7:71, 2020.

[15] Yikan Wang, Chenwei Gong, Qiming Xu, and Yingqiao Zheng. Design of privacy-preserving personalized recommender system based on federated learning. In *2024 8th International Workshop on Materials Engineering and Computer Sciences (IWMECS 2024)*, 2024.

[16] Siye Wu, Lei Fu, Runmian Chang, Yuanzhou Wei, Yeyubei Zhang, Zehan Wang, Lipeng Liu, Haopeng Zhao, and Keqin Li. Warehouse robot task scheduling based on reinforcement learning to maximize operational efficiency. *Authorea Preprints*, 2025.

[17] Peng Zhao, Xiaoyu Liu, Xuqi Su, Di Wu, Zi Li, Kai Kang, Keqin Li, and Armando Zhu. Probabilistic contingent planning based on hierarchical task network for high-quality plans. *Algorithms*, 18(4):214, 2025.